

Информатика и её применения

Том 17 Выпуск 4 Год 2023

СОДЕРЖАНИЕ

Нелинейная регуляризация обращения линейных однородных операторов с помощью метода блочной пороговой обработки О. В. Шестаков, Е. П. Степанов	2
Рынок с марковской скачкообразной волатильностью III: алгоритм мониторинга цены риска по дискретным наблюдениям цен активов А. В. Борисов	9
Процедура построения множества Парето для дифференцируемых критериальных функций Я. И. Рабинович	17
Непараметрический алгоритм автоматической классификации данных дистанционного зондирования В. П. Тубольцев, А. В. Лапко, В. А. Лапко	23
Технология многофакторной классификации математического контента электронной системы обучения А. В. Босов, А. В. Иванов	32
Расширяемый подход к слиянию данных в распределенных вычислительных средах В. В. Сазонтьев, С. А. Ступников, В. Н. Захаров	42
Решение задачи оптимального управления запасом непрерывного продукта в стохастической модели регенерации со случайными стоимостными характеристиками П. В. Шнурков	48
Границы незавершенной работы в системе с повторными вызовами разных классов и показательным временем обслуживания И. В. Пешкова	57
Модели совместного обслуживания трафика eMBB и URLLC на основе приоритетов в промышленных развертываниях 5G NR Д. В. Иванова, Е. В. Маркова, С. Я. Шоргин, Ю. В. Гайдамака	64
Модели учета влияния статистических характеристик трафика вычислительных сетей на эффективность прогнозирования средствами машинного обучения С. Л. Френкель, В. Н. Захаров	71
Аннотирование параллельных корпусов: подходы и направления развития А. А. Гончаров	81
Степень семантической близости дискурсивных отношений: методы и инструменты расчета О. Ю. Инькова, М. Г. Кружков	88
Научная парадигма информатики: классификация объектов предметной области И. М. Зацман	96
Об авторах	104
Авторский указатель за 2023 г.	106
2023 Author Index	110
Правила подготовки рукописей	114
Requirements for manuscripts	117

НЕЛИНЕЙНАЯ РЕГУЛЯРИЗАЦИЯ ОБРАЩЕНИЯ ЛИНЕЙНЫХ ОДНОРОДНЫХ ОПЕРАТОРОВ С ПОМОЩЬЮ МЕТОДА БЛОЧНОЙ ПОРОГОВОЙ ОБРАБОТКИ*

О. В. Шестаков¹, Е. П. Степанов²

Аннотация: Методы пороговой обработки коэффициентов вейвлет-разложений стали популярным инструментом для регуляризации обратных статистических задач благодаря своей простоте, вычислительной эффективности и возможности адаптации как к виду обрабатываемых операторов, так и к особенностям исследуемой функции. Наиболее плодотворным данный подход оказался при обращении линейных однородных операторов, возникающих в некоторых задачах обработки сигналов и изображений. В работе рассматривается метод блочной пороговой обработки, в котором коэффициенты разложения обрабатываются группами, что позволяет учитывать информацию о соседних коэффициентах. В модели данных с аддитивным гауссовским шумом проводится анализ несмещенной оценки среднеквадратичного риска и показывается, что при определенных условиях данная оценка становится сильно состоятельной и асимптотически нормальной. Данные свойства позволяют строить асимптотические доверительные интервалы для теоретического среднеквадратичного риска рассматриваемого метода.

Ключевые слова: линейный однородный оператор; вейвлеты; блочная пороговая обработка; несмещенная оценка риска; асимптотическая нормальность; сильная состоятельность

DOI: 10.14357/19922264230401

EDN: PGKKYE

1 Введение

При решении обратных статистических задач возникает проблема обращения некоторого оператора, и если в наблюдаемых данных содержится шум, то необходимо применять методы регуляризации. Примером может служить восстановление сигнала в беспроводных каналах, зашумленного за счет гео- и гелиомагнитных условий, состояния ионосферы Земли. Во многих ситуациях рассматриваемый оператор оказывается линейным и однородным. В таких случаях часто применяются методы подавления шума с помощью вейвлет-разложения и процедур пороговой обработки. Они приобрели свою популярность благодаря адаптивности, вычислительной эффективности и асимптотической оптимальности.

Чаще всего процедуры пороговой обработки применяются отдельно к каждому коэффициенту вейвлет-разложения. Коэффициент сравнивается с пороговым значением, и если его абсолютная величина оказывается меньше этого значения, то он обнуляется. Среди таких процедур наиболее распространены процедуры жесткой и мягкой пороговой обработки. Однако они имеют свои недостат-

ки и часто не достигают оптимальных результатов. В работе [1] рассмотрен метод блочной пороговой обработки, при котором коэффициенты обрабатываются не отдельно, а группами. Цель такого подхода заключается в использовании информации о соседних коэффициентах. Получаемые оценки имеют оптимальный (в минимаксном смысле) порядок среднеквадратичного риска для различных классов функций [2].

Для практического анализа погрешности данного метода можно использовать несмещенную оценку среднеквадратичного риска [3], которая дает возможность оценивать качество обработанного сигнала на основе только наблюдаемых данных. В данной работе исследуются статистические свойства этой оценки и показывается, что для довольно широкого класса пространств Бесова она является асимптотически нормальной и сильно состоятельной. В работе [4] подобные исследования проведены для задачи непараметрического оценивания функции сигнала.

Для методов пороговой обработки отдельных коэффициентов свойства оценки среднеквадратичного риска при обращении линейных однородных операторов исследованы в работах [5–8] в предпо-

* Работа выполнена при поддержке программы развития Московского университета, проект № 23-Ш03-03.

¹ Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Московский центр фундаментальной и прикладной математики; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

² Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, estepanov@lvk.cs.msu.ru

ложении о принадлежности исследуемой функции сигнала классу равномерно регулярных по Липшицу функций.

2 Обращение линейных однородных операторов с помощью вейглет-вейвлет-разложения

Линейный оператор K называется однородным, если

$$K[f(a(x - x_0))] = a^{-\alpha}(Kf)[a(x - x_0)]$$

для любого $x_0 \in \mathbb{R}$ и любого $a > 0$. Параметр α называется показателем однородности. Примерами однородных линейных операторов служат операторы дифференцирования и интегрирования, преобразование Гильберта, преобразование Абеля и некоторые операторы свертки. Во многих прикладных задачах наблюдаются значения функции Kf , а интерес представляет функция f , т. е. необходимо решать задачу обращения оператора K . Для этого воспользуемся методом вейглет-вейвлет-разложения.

Вейвлет-разложение Kf представляет собой ряд

$$Kf = \sum_{j,k \in \mathbb{Z}} \langle Kf, \psi_{j,k} \rangle \psi_{j,k}, \quad (1)$$

где ψ — некоторая материнская вейвлет-функция, а $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$ (семейство $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ образует ортонормированный базис в $L^2(\mathbb{R})$). Индекс j в (1) называется масштабом, а индекс k — сдвигом.

Поскольку оператор K линеен и однороден, существуют такие функции $u_{j,k}$, что $\langle f, u_{j,k} \rangle = \langle Kf, \psi_{j,k} \rangle$. При этом функция f представляется в виде ряда

$$f = \sum_{j,k \in \mathbb{Z}} \beta_j \langle Kf, \psi_{j,k} \rangle u_{j,k}, \quad (2)$$

где $u_{j,k} = K^{-1} \psi_{j,k} / \beta_{j,k}$, $\beta_j = 2^{\alpha j} \beta_0$, а $\beta_0 = \|K^{-1} \psi\|$ (функции $u_{j,k}$ называются вейглетами). Формула (2) и есть основа метода вейглет-вейвлет-разложения [9]. При соответствующем выборе вейвлет-функции ψ последовательность $\{u_{j,k}\}$ образует устойчивый базис [10]. Также, если ψ имеет r непрерывных производных и r нулевых моментов, определим при $0 < \gamma < r$ и $1 \leq p, q \leq \infty$ полунорму последовательности вейвлет-коэффициентов выражением

$$|Kf|_{B_{p,q}^\gamma} = \left(\sum_{j=0}^{\infty} \left(2^{sj} \left(\sum_k |\langle Kf, \psi_{j,k} \rangle|^p \right)^{1/p} \right)^q \right)^{1/q},$$

где $s = \gamma + 1/2 - 1/p$. Далее будем считать, что ψ удовлетворяет всем необходимым требованиям, а Kf задана на конечном отрезке и принадлежит пространству Бесова $B_{p,q}^\gamma(A)$ ($A > 0$), т. е. $|Kf|_{B_{p,q}^\gamma} \leq A$.

3 Модель данных

При практических измерениях значения функции Kf регистрируются в дискретных отсчетах. К тому же эти значения, как правило, содержат шум. В данной работе рассматривается следующая модель данных:

$$X_i = (Kf)_i + \epsilon_i, \quad i = \overline{1, 2^J},$$

где 2^J — число отсчетов функции сигнала; $(Kf)_i$ — незашумленные значения функции Kf ; ϵ_i — независимые нормально распределенные случайные величины с нулевым средним и дисперсией σ^2 . После применения дискретного вейвлет-преобразования получается следующая модель зашумленных вейвлет-коэффициентов:

$$Y_{j,k} = \mu_{j,k} + \epsilon_{j,k}^W, \quad j = \overline{0, J-1}, \quad k = \overline{0, 2^j - 1},$$

где $\epsilon_{j,k}^W$ независимы и распределены так же, как и ϵ_i , а $\mu_{j,k} = 2^{J/2} \langle Kf, \psi_{j,k} \rangle$ [11].

4 Блочная пороговая обработка

Для подавления шума коэффициенты $Y_{j,k}$ подвергаются некоторой обработке. Наибольшее распространение получили методы жесткой и мягкой пороговой обработки и их модификации [12–20]. При использовании этих методов происходит сравнение абсолютной величины каждого коэффициента с некоторым порогом, и если это значение оказывается меньше порога, то коэффициент считается шумом и обнуляется. Такие методы обрабатывают каждый коэффициент отдельно, не используя информацию о других коэффициентах. Блочная пороговая обработка применяется к группам соседних коэффициентов, т. е. решение об обнулении принимается сразу ко всем коэффициентам из группы.

Пусть $B_{j,1}, \dots, B_{j,M_j}$ — разбиение множества индексов $\{0, \dots, 2^j - 1\}$ на блоки одинаковой длины L (для удобства предположим, что 2^j делится на L). Пусть $S_{j,m}^2 = \sum_{k \in B_{j,m}} Y_{j,k}^2$. Оценки коэффициентов $\mu_{j,k}$ вычисляются по правилу:

$$\hat{\mu}_{j,k} = \left(1 - \frac{TL\sigma^2}{S_{j,m}^2} \right)_+ Y_{j,k}, \quad j = \overline{0, J-1}, \quad k \in B_m,$$

т. е. если величина $\sum_{k \in B_{j,m}} Y_{j,k}^2$ меньше порога $TL\sigma^2$, то все коэффициенты в рассматриваемом блоке обнуляются.

На качество оценок, получаемых с помощью блочной пороговой обработки, естественно, влияют размер блока L и значение порога T . В работе [2] показано, что при $L = \log 2^j$ достигается баланс между локальной и глобальной адаптивностью метода блочной пороговой обработки, и если при этом $T^* \approx 4,50524$ (T^* — это корень уравнения $T - \log T - 3 = 0$), то среднеквадратичный риск оказывается (почти) оптимальным (в минимаксном смысле). В данной работе рассматриваются именно такие значения L и T .

Одним из самых распространенных критериев качества описанного метода служит среднеквадратичный риск, который определяется по формуле:

$$R_J(T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 \mathbb{E}(\hat{\mu}_{j,k} - \mu_{j,k})^2.$$

Вычислить его значение на практике нельзя, поскольку $R_J(T)$ зависит от ненаблюдаемых «чистых» коэффициентов $\mu_{j,k}$. Однако его можно оценить, используя наблюдаемые значения $Y_{j,k}$. Далее будут исследованы свойства такой оценки.

5 Асимптотические свойства несмещенной оценки риска

В [2] показано, что

$$\sum_{k=0}^{2^j-1} \mathbb{E}(\hat{\mu}_{j,k} - \mu_{j,k})^2 = \sum_{m=1}^{M_j} \mathbb{E}F_m[T, L],$$

где

$$F_m(T, L) = \left[L\sigma^2 + \frac{T^2 L^2 \sigma^4 - 2TL\sigma^4(L-2)}{S_{j,m}^2} \cdot \mathbf{1}(S_{j,m}^2 > TL\sigma^2) + (S_{j,m}^2 - 2L\sigma^2) \cdot \mathbf{1}(S_{j,m}^2 \leq TL\sigma^2) \right].$$

Таким образом, величина

$$\hat{R}_J(T) = \sum_{j=0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} F_m(T, L) \quad (3)$$

служит несмещенной оценкой $R_J(T)$, не зависящей от $\mu_{j,k}$.

Покажем, что оценка (3) асимптотически нормальна.

Теорема 1. Пусть K — линейный однородный оператор с показателем однородности $\alpha > 0$, а $Kf \in B_{p,q}^\gamma(A)$ и задана на конечном отрезке. Если $1 \leq p, q \leq \infty$ и $r > \gamma > \max(0, 1/p - 1/2)$, то при $J \rightarrow \infty$

$$\mathbb{P} \left(\frac{\hat{R}_J(T^*) - R_J(T^*)}{D_J} < x \right) \rightarrow \Phi(x), \quad (4)$$

где

$$D_J^2 = \frac{2\sigma^4 \beta_0^4}{2^{4\alpha+1} - 1} \cdot 2^{(4\alpha+1)J},$$

а $\Phi(x)$ — функция распределения стандартного нормального закона.

Доказательство: Выберем $0 < d < 1$ и запишем:

$$\begin{aligned} \hat{R}_J(T^*) - R_J(T^*) &= \\ &= \sum_{j=0}^{[dJ]} \beta_j^2 \sum_{m=1}^{M_j} [F_m(T^*, L) - \mathbb{E}F_m(T^*, L)] + \\ &+ \sum_{j=[dJ]+1}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} [F_m(T^*, L) - \mathbb{E}F_m(T^*, L)]. \end{aligned} \quad (5)$$

Число слагаемых в первой сумме не превосходит $2^{[dJ]+1}$. Кроме того, существует такая константа $C_F > 0$, что

$$|F_m(T^*, L) - \mathbb{E}F_m(T^*, L)| \leq C_F T^* L \text{ п. в.} \quad (6)$$

Применяя неравенство Хефдинга, получаем, что для любого $\delta > 0$ найдется константа $C_\delta > 0$ такая, что

$$\begin{aligned} \mathbb{P} \left(\frac{1}{D_J} \left| \sum_{j=0}^{[dJ]} \beta_j^2 \sum_{m=1}^{M_j} [F_m(T^*, L) - \mathbb{E}F_m(T^*, L)] \right| > \delta \right) \leq \\ \leq \sum_{j=0}^{[dJ]} \mathbb{P} \left(\frac{\beta_j^2}{D_J} \sum_{m=1}^{M_j} [F_m(T^*, L) - \mathbb{E}F_m(T^*, L)] > J^{-1} \delta \right) \leq J \exp \left\{ -C_\delta \frac{2^{J-[dJ]}}{J^2} \right\}, \end{aligned} \quad (7)$$

т. е.

$$\frac{\sum_{j=0}^{[dJ]} \beta_j^2 \sum_{m=1}^{M_j} [F_m(T^*, L) - \mathbb{E}F_m(T^*, L)]}{D_J} \xrightarrow{P} 0 \quad \text{при } J \rightarrow \infty. \quad (8)$$

При $p_1 \leq p_2$ справедливы неравенства [2]

$$\left(\sum_{k=0}^{2^j-1} |\mu_{j,k}|^{p_2} \right)^{1/p_2} \leq \left(\sum_{k=0}^{2^j-1} |\mu_{j,k}|^{p_1} \right)^{1/p_1} \leq 2^{j(1/p_1-1/p_2)} \left(\sum_{k=0}^{2^j-1} |\mu_{j,k}|^{p_2} \right)^{1/p_2}. \quad (9)$$

Так как $f \in B_{p,q}^\gamma(A)$, то

$$2^{js} \left(\sum_{k=0}^{2^j-1} |\mu_{j,k}|^p \right)^{1/p} \leq A 2^{J/2}$$

и из неравенств (9) следует, что при $p \geq 2$

$$\sum_{k=0}^{2^j-1} \mu_{j,k}^2 = \sum_{m=1}^{M_j} \sum_{k \in B_{j,m}} \mu_{j,k}^2 \leq A^2 2^{-2\gamma j+J}. \quad (10)$$

Для произвольного $\varepsilon > 0$ при всех $j > dJ$ существуют не более $A^2 2^{J-2\gamma j+\varepsilon j}$ слагаемых в (10) таких, что

$$\sum_{k \in B_{j,m}} \mu_{j,k}^2 > 2^{-\varepsilon j}.$$

Выделяя эти слагаемые из второй суммы в (5) в отдельную сумму S' и применяя к S' неравенство Хеффдинга, аналогичное (7), получаем, что

$$\frac{S'}{\sigma^2 \sqrt{2^{J+1}}} \xrightarrow{P} 0$$

при $J \rightarrow \infty$. Таким образом, без ограничения общности можно считать, что во второй сумме в (5)

$$\sum_{k \in B_{j,m}} \mu_{j,k}^2 \rightarrow 0$$

при $J \rightarrow \infty$ для всех $m = 1, \dots, M_j$ и $j > dJ$.

Если $p < 2$ и $\gamma > 1/p - 1/2$, то $s > 0$ и из неравенств (9) следует, что

$$\sum_{k=0}^{2^j-1} \mu_{j,k}^2 \leq A^2 \cdot 2^{-2sj+J}. \quad (11)$$

Рассуждая аналогично случаю $p \geq 2$, заключаем, что без ограничения общности можно считать, что во второй сумме в (5) также $\sum_{k \in B_{j,m}} \mu_{j,k}^2 \rightarrow 0$ при $J \rightarrow \infty$ для всех $m = 1, \dots, M_j$ и $j > dJ$.

Рассмотрим дисперсии слагаемых во второй сумме в (5). Имеем $DF_m(T^*, L) = D[S_{j,m}^2 - L\sigma^2 + U_{j,m}(T^*, L)]$, где

$$U_{j,m}(T^*, L) = \left[\frac{(T^*L)^2 \sigma^4 - 2T^*L\sigma^4(L-2)}{S_{j,m}^2} - S_{j,m}^2 + 2L\sigma^2 \right] \cdot \mathbf{1}(S_{j,m}^2 > T^*L\sigma^2).$$

Пусть X_L — случайная величина, имеющая распределение χ_L^2 с L степенями свободы, и C_1 и C_2 — некоторые положительные константы. Тогда в силу (10) или (11) для некоторой константы $1 < C_\chi < 1 + \delta_\chi$ ($0 < \delta_\chi < 1$)

$$\begin{aligned} D[S_{j,m}^2 \cdot \mathbf{1}(S_{j,m}^2 > T^*L\sigma^2)] &\leq \\ &\leq E[(S_{j,m}^2)^2 \cdot \mathbf{1}(S_{j,m}^2 > T^*L\sigma^2)] \leq \\ &\leq C_\chi^2 E[X_L^2 \cdot \mathbf{1}(C_\chi X_L > T^*L)] \leq \\ &\leq C_1(T^*L)^2 P(C_\chi X_L > T^*L) \end{aligned}$$

и аналогично

$$\begin{aligned} D\left(\left[\frac{(T^*L)^2 \sigma^4 - 2T^*L\sigma^4(L-2)}{S_{j,m}^2} + 2L\sigma^2 \right] \cdot \mathbf{1}(S_{j,m}^2 > T^*L\sigma^2) \right) &\leq \\ &\leq C_2(T^*L)^2 P(C_\chi X_L > T^*L). \end{aligned}$$

В силу леммы 2 из работы [2] и вида T^* и L

$$(T^*L)^2 P(C_\chi X_L > T^*L) \rightarrow 0$$

при $J \rightarrow \infty$.

Таким образом, учитывая очевидное соотношение для дисперсии разности случайных величин $X - Y$:

$$(\sqrt{DX} - \sqrt{DY})^2 \leq D(X - Y) \leq (\sqrt{DX} + \sqrt{DY})^2,$$

получаем

$$\lim_{J \rightarrow \infty} \frac{D \sum_{j=[dJ]+1}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} F_m(T^*, L)}{D \sum_{j=[dJ]+1}^{J-1} \sum_{k=0}^{2^j-1} \beta_0^2 2^{2\alpha j} Y_{j,k}^2} = 1. \quad (12)$$

Кроме того, поскольку $Y_{j,k}$ независимы и $DY_{j,k}^2 = 2\sigma^4 + 4\sigma^2 \mu_{j,k}^2$, получаем:

$$\lim_{J \rightarrow \infty} \frac{D \sum_{j=[dJ]+1}^{J-1} \sum_{k=0}^{2^j-1} \beta_0^2 2^{2\alpha j} Y_{j,k}^2}{D^2 J} = 1. \quad (13)$$

Наконец, выполнено условие Линдеберга: для любого $\varepsilon > 0$ при $J \rightarrow \infty$

$$\begin{aligned} \frac{1}{D^2 J} \sum_{j=[dJ]+1}^{J-1} \sum_{m=1}^{M_j} E[\beta_j^4 (F_m(T^*, L) - \\ - EF_m(T^*, L))^2 \cdot \mathbf{1}(\beta_j^2 |F_m(T^*, L) - \\ - EF_m(T^*, L)| > \varepsilon D_J)] \rightarrow 0. \quad (14) \end{aligned}$$

Действительно, в силу (6), (12) и (13) начиная с некоторого J все индикаторы в (14) обращаются в нуль. Объединяя (8), (13) и (14), получаем (4). Теорема доказана.

Докажем теперь свойство сильной состоятельности оценки (3), справедливое при более слабых ограничениях.

Теорема 2. Пусть K — линейный однородный оператор с показателем однородности $\alpha > 0$, а $Kf \in L^2(\mathbb{R})$ и задана на конечном отрезке. Тогда при любом $\theta > 1/2 + 2\alpha$ имеет место сходимость

$$\frac{\widehat{R}_J(T^*) - R_J(T^*)}{2^{\theta J}} \rightarrow 0 \text{ п. в. при } J \rightarrow \infty. \quad (15)$$

Доказательство: Используя неравенство Хеффдинга, с учетом (6) получаем, что для любого $\delta > 0$ найдется константа $C_\delta > 0$ такая, что

$$p_J = \mathbb{P} \left(\frac{|\widehat{R}_J(T^*) - R_J(T^*)|}{2^{\theta J}} > \delta \right) \leq \leq J \exp \left\{ -C_\delta \frac{2^{(2\theta-1-4\alpha)J}}{J^2} \right\},$$

и, поскольку

$$\sum_{J=1}^{\infty} p_J < \infty,$$

в силу леммы Бореля—Кантелли выполнено (15). Теорема доказана.

Теоремы 1 и 2 дают теоретическое обоснование использования значения $\widehat{R}_J(T^*)$ в качестве оценки неизвестной величины риска (погрешности) $R_J(T^*)$, а также позволяют строить асимптотические доверительные интервалы для $R_J(T^*)$.

Литература

- Hall P., Kerkycharian G., Picard D. On the minimax optimality of block thresholded wavelet estimators // Stat. Sinica, 1999. Vol. 9. P. 33–50.
- Cai T. Adaptive wavelet estimation: A block thresholding and oracle inequality approach // Ann. Stat., 1999. Vol. 28. No. 3. P. 898–924. doi: 10.1214/aos/1018031262.
- Stein C. Estimation of the mean of a multivariate normal distribution // Ann. Stat., 1981. Vol. 9. No. 6. P. 1135–1151. doi: 10.1214/aos/1176345632.
- Шестаков О. В. Анализ несмещенной оценки среднеквадратичного риска метода блочной пороговой обработки // Информатика и её применения, 2021. Т. 15. Вып. 2. С. 30–35. doi: 10.14357/19922264210205. EDN: DSQQUA.
- Кудрявцев А. А., Шестаков О. В. Асимптотика оценки риска при вейвлет-вейвлет разложении наблюдаемого сигнала // Т-Comm — Телекоммуникации и транспорт, 2011. № 2. С. 54–57.
- Шестаков О. В. О свойствах оценки среднеквадратичного риска при регуляризации обращения линейного

однородного оператора с помощью адаптивной пороговой обработки коэффициентов вейвлет-вейвлет разложения // Вестн. Тверск. ун-та. Сер. Прикладная математика, 2012. № 1. С. 117–130. EDN: PEZOJH.

- Шестаков О. В. Статистические свойства метода подавления шума, основанного на стабилизированной жесткой пороговой обработке // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 65–69. doi: 10.14357/19922264160207. EDN: WCBWWP.
- Shestakov O. V. Nonlinear regularization of inverse problems for linear homogeneous transforms by stabilized hard thresholding // J. Math. Sci., 2018. Vol. 234. No. 6. P. 780–785. doi: 10.1007/s10958-018-4044-1.
- Abramovich F., Silverman B. W. Wavelet decomposition approaches to statistical inverse problems // Biometrika, 1998. Vol. 85. No. 1. P. 115–129.
- Lee N. Wavelet-vaguelette decompositions and homogeneous equations: PhD Thesis. — West Lafayette, IN, USA: Purdue University, 1997. 93 p.
- Mallat S. A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.
- Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // Biometrika, 1994. Vol. 81. No. 3. P. 425–455. doi: 10.2307/2337118.
- Donoho D., Johnstone I. M. Adapting to unknown smoothness via wavelet shrinkage // J. Am. Stat. Assoc., 1995. Vol. 90. P. 1200–1224. doi: 10.2307/2291512.
- Donoho D., Johnstone I. M. Minimax estimation via wavelet shrinkage // Ann. Stat., 1998. Vol. 26. No. 3. P. 879–921.
- Gao H.-Y. Wavelet shrinkage denoising using the non-negative garrote // J. Comput. Graph. Stat., 1998. Vol. 7. No. 4. P. 469–488.
- Poornachandra S., Kumaravel N., Saravanan T. K., Somaskandan R. WaveShrink using modified hyper-shrinkage function // 27th Annual Conference (International) of the IEEE Engineering in Medicine and Biology Society Proceedings. — Piscataway, NJ, USA: IEEE, 2005. P. 30–32.
- Lin Y., Cai J. A new threshold function for signal denoising based on wavelet transform // Conference (International) on Measuring Technology and Mechatronics Automation Proceedings. — Piscataway, NJ, USA: IEEE, 2010. P. 200–203.
- Huang H.-C., Lee T. C. M. Stabilized thresholding with generalized sure for image denoising // 17th Conference (International) on Image Processing Proceedings. — Piscataway, NJ, USA: IEEE, 2010. P. 1881–1884.
- He C., Xing J., Li J., Yang Q., Wang R. A new wavelet thresholding function based on hyperbolic tangent function // Math. Probl. Eng., 2015. Vol. 2015. Art. 528656. 10 p.
- Zhao R.-M., Cui H.-M. Improved threshold denoising method based on wavelet transform // 7th Conference (International) on Modelling, Identification and Control Proceedings. — Piscataway, NJ, USA: IEEE, 2015. Art. 7409352. 4 p. doi: 10.1109/ICMIC.2015.7409352.

Поступила в редакцию 07.09.23

NONLINEAR REGULARIZATION OF THE INVERSION OF LINEAR HOMOGENEOUS OPERATORS USING THE BLOCK THRESHOLDING METHOD

O. V. Shestakov^{1,2,3} and E. P. Stepanov¹

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

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

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

Abstract: The methods of thresholding the coefficients of wavelet expansions have become a popular tool for regularization of inverse statistical problems due to their simplicity, computational efficiency, and the ability to adapt both to the type of operators and to the features of the function under study. This approach proved to be the most fruitful for inversion of linear homogeneous operators arising in some signal and image processing problems. The paper considers the block thresholding method in which the decomposition coefficients are processed in groups that allows taking into account information about neighboring coefficients. In a data model with an additive Gaussian noise, an unbiased estimate of the mean-square risk is analyzed and it is shown that under certain conditions, this estimate is strongly consistent and asymptotically normal. These properties allow constructing asymptotic confidence intervals for the theoretical mean-square risk of the method under consideration.

Keywords: linear homogeneous operator; wavelets; block thresholding; unbiased risk estimate; asymptotic normality; strong consistency

DOI: 10.14357/19922264230401

EDN: PGKKYE

Acknowledgments

The work was done with the support of MSU Program of Development, Project No. 23-SCH03-03.

References

- Hall, P., G. Kerkycharian, and D. Picard. 1999. On the minimax optimality of block thresholded wavelet estimators. *Stat. Sinica* 9:33–50.
- Cai, T. 1999. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Stat.* 28(3):898–924. doi: 10.1214/aos/1018031262.
- Stein, C. 1981. Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* 9(6):1135–1151. doi: 10.1214/aos/1176345632.
- Shestakov, O. V. 2021. Analiz nesmeshchennoy otsenki srednekvadrachnogo riska metoda blochnoy porogovoy obrabotki [Analysis of the unbiased mean-square risk estimate of the block thresholding method]. *Informatika i ee Primeneniya — Inform. Appl.* 15(2):30–35. doi: 10.14357/19922264210205. EDN: DSQQUAU.
- Kudryavtsev, A. A., and O. V. Shestakov. 2011. Asimptotika otsenki riska pri veyglet-veyvlet razlozhenii nablyudаемого сигнала [The average risk assessment of the wavelet decomposition of the signal]. *T-Comm — Telekommunikatsii i transport* [T-Comm — Telecommunications and Transport] 2:54–57.
- Shestakov, O. V. 2012. O svoystvakh otsenki srednekvadrachnogo riska pri regularizatsii obrashcheniya lineynogo odnorodnogo operatora s pomoshch'yu adaptivnoy porogovoy obrabotki koeffitsientov veyglet-veyvlet razlozheniya [On the properties of estimating the mean-square risk in the regularization of the inversion of a linear homogeneous operator using adaptive thresholding of the coefficients of the vaguelette-wavelet decomposition]. *Vestn. Tversk. un-ta. Ser. Prikladnaya matematika* [Herald of Tver State University. Series: Applied Mathematics] 1:117–130. EDN: PEZOJH.
- Shestakov, O. V. 2016. Statisticheskie svoystva metoda podavleniya schuma, osnovannogo na stabilizirovannoy zhestkoy porogovoy obrabotke [Statistical properties of the denoising method based on the stabilized hard thresholding]. *Informatika i ee Primeneniya — Inform. Appl.* 10(2):65–69. doi: 10.14357/19922264160207. EDN: WCBWWP.
- Shestakov, O. V. 2018. Nonlinear regularization of inverse problems for linear homogeneous transforms by stabilized hard thresholding. *J. Math. Sci.* 234(6):780–785. doi: 10.1007/s10958-018-4044-1.
- Abramovich, F., and B. W. Silverman. 1998. Wavelet decomposition approaches to statistical inverse problems. *Biometrika* 85(1):115–129.
- Lee, N. 1997. Wavelet-vaguelette decompositions and homogeneous equations. West Lafayette, IN: Purdue University. PhD Thesis. 93 p.

11. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
12. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455. doi: 10.2307/2337118.
13. Donoho, D., and I. M. Johnstone. 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* 90:1200–1224. doi: 10.2307/2291512.
14. Donoho, D., and I. M. Johnstone. 1998. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 26(3):879–921.
15. Gao, H.-Y. 1998. Wavelet shrinkage denoising using the non-negative garrote. *J. Comput. Graph. Stat.* 7(4):469–488.
16. Poornachandra, S., N. Kumaravel, T.K. Saravanan, and R. Somaskandan. 2005. WaveShrink using modified hyper-shrinkage function. *27th Annual Conference (International) of the IEEE Engineering in Medicine and Biology Society Proceedings*. Piscataway, NJ: IEEE. 30–32.
17. Lin, Y, and J. Cai. 2010. A new threshold function for signal denoising based on wavelet transform. *Conference (International) on Measuring Technology and Mechatronics Automation Proceedings*. Piscataway, NJ: IEEE. 200–203.
18. Huang, H.-C., and T. C. M. Lee. 2010. Stabilized thresholding with generalized sure for image denoising. *17th Conference (International) on Image Processing Proceedings*. Piscataway, NJ: IEEE. 1881–1884.
19. He, C., J. Xing, J. Li, Q. Yang, and R. Wang. 2015. A new wavelet thresholding function based on hyperbolic tangent function. *Math. Probl. Eng.* 2015:528656. 10 p.
20. Zhao, R.-M., and H.-M. Cui. 2015. Improved threshold denoising method based on wavelet transform. *7th Conference (International) on Modelling, Identification and Control Proceedings*. Piscataway, NJ: IEEE. Art. ID. 7409352. 4 p. doi: 10.1109/ICMIC.2015.7409352.

Received September 7, 2023

Contributors

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

Stepanov Evgeniy P. (b. 1994) — Candidate of Science (PhD) in physics and mathematics, junior scientist, Department of Computing Systems and Automation, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; estepanov@lvk.cs.msu.ru

РЫНОК С МАРКОВСКОЙ СКАЧКООБРАЗНОЙ ВОЛАТИЛЬНОСТЬЮ III: АЛГОРИТМ МОНИТОРИНГА ЦЕНЫ РИСКА ПО ДИСКРЕТНЫМ НАБЛЮДЕНИЯМ ЦЕН АКТИВОВ*

А. В. Борисов¹

Аннотация: Третья часть цикла посвящена задаче оценивания в реальном масштабе рыночной цены риска в финансовой системе, состоящей из банковского депозита, базовых рисков бумаг и их деривативов. Базовые активы описываются моделью со стохастической волатильностью, представленной скрытым марковским скачкообразным процессом (МСП). Исследуемый рынок предполагается безарбитражным, и рыночная цена риска для него является функцией текущего значения МСП. Таким образом, мониторинг цены риска сводится к задаче фильтрации состояния МСП. Статистическая информация поступает в неслучайные дискретные моменты времени и содержит прямые наблюдения цен базовых активов и косвенные зашумленные наблюдения деривативов. Статья содержит решение задачи оптимальной фильтрации состояния МСП, а также соответствующий численный алгоритм. Представлены результаты вычислительного примера, демонстрирующего качество мониторинга в зависимости от состава и вида доступных наблюдений.

Ключевые слова: рыночная цена риска; безарбитражный рынок; марковский скачкообразный процесс; задача оптимальной фильтрации; численный алгоритм

DOI: 10.14357/19922264230402

EDN: OFYELT

1 Введение

Заметка продолжает цикл работ [1, 2], посвященных исследованию модели безарбитражного финансового рынка со стохастической волатильностью, описываемой внешним МСП. Отсутствие арбитража означает неединственность мартингальной меры, и выбор преобладающей меры из множества — актуальная задача, обеспечивающая возможность оптимизации портфельных инвестиций и построения хеджирующих стратегий. Описание этой меры возможно в терминах рыночной цены риска (MPR, market price of risk). Цель предлагаемой работы заключается в решении задачи оценивания MPR в реальном масштабе времени по наблюдениям, поступающим в дискретные неслучайные моменты времени, с привлечением математического аппарата оптимальной фильтрации.

Статья организована следующим образом. В разд. 2 представлена исследуемая модель рынка: эволюция цен базовых активов описывается решением некоторой стохастической дифференциальной системы (СДС), в то время как цены деривативов определяются решениями системы уравнений в частных производных. В разделе предложена модель доступных наблюдений и представлена аргументация в ее пользу. Задачу мониторинга MPR удается свести к задаче оптимальной фильтрации

состояний МСП, определяющего стохастическую волатильность. В разд. 3 представлены утверждение о виде оценки оптимальной фильтрации и алгоритм ее численной реализации. Раздел 4 содержит численный пример, иллюстрирующий зависимость качества оценивания состояний МСП от состава и вида обрабатываемых наблюдений. Заключительные замечания по результатам, представленным в данной части цикла, даны в разд. 5.

2 Модель рынка и постановка задачи мониторинга рыночной цены риска

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

- банковского вклада с детерминированной ставкой r_t ;
- N базовых финансовых инструментов $S_t = \text{col}(S_t^1, \dots, S_t^N)$;
- M деривативов $F_t = \text{col}(F_t^1, \dots, F_t^M)$ с единым сроком погашения T , определяемых платежным требованием $H(S_T) = \text{col}(H^1(S_T), \dots, H^M(S_T))$.

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

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

Цена S_t представляет собой решение СДС

$$dS_t = \text{diag } S_t a(t, Z_t) dt + \text{diag } S_t \sigma(t, Z_t) dw_t, \quad t \in (0, T], \quad S_0 \sim \pi_0(s), \quad (1)$$

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

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

с набором известных детерминированных функций $\{a^\ell(t)\}_{\ell=\overline{1, L}}$ и $\{\sigma^\ell(t)\}_{\ell=\overline{1, L}}$. Распределение начальной цены S_0 таково, что все ее компоненты положительны с вероятностью 1.

Стохастическая волатильность $\sigma(t, Z_t)$ определяется внешним скрытым МСП $Z_t \triangleq \text{col}(Z_t^1, \dots, Z_t^L) \in \{e_1, \dots, e_L\}$ с матрицей интенсивностей переходов $\Lambda(\cdot)$ и начальным распределением π_0^Z . Марковский скачкообразный процесс Z_t представляет собой решение СДС с мартингалом M_t в правой части:

$$dZ_t = \Lambda^\top(t) Z_t dt + dM_t, \quad t \in (0, T], \quad Z_0 \sim \pi_0^Z. \quad (2)$$

В [1] показано, что в условиях безарбитражности рынка справедливая цена F^m m -го дериватива ($m = \overline{1, M}$) определяется формулой:

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

где детерминированные функции $\{F^{m\ell}(t, s)\}_{\ell=\overline{1, L}}$ являются решением задачи Коши системы дифференциальных уравнений в частных производных

$$F_t^{m\ell} = r F^{m\ell} - \sum_{j=1}^L \Lambda^{lj} F^{mj} - \sum_{n=1}^N F_{s^n}^{m\ell} s^n (r - a_n^\ell) - \frac{1}{2} \sum_{i,j=1}^N s^i s^j F_{s^i, s^j}^{m\ell} B_{ij}^\ell, \quad \ell = \overline{1, M},$$

$$m = \overline{1, M}, \quad t \in [0, T], \quad F^{m\ell}(T, s) = H^m(s); \quad (3)$$

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

Согласно [1], МРР θ_t в данной модели определяется формулой

$$\theta_t = \sum_{\ell=1}^L Z_t^\ell (\sigma^\ell(t))^+ (a^\ell(t) - r_t \mathbf{1})$$

(здесь $\mathbf{1}$ — вектор-столбец подходящей размерности, составленный из единиц) и однозначно определяется значением МСП Z_t .

Цены S_t и F_t , а также фактор-процесс Z_t недоступны прямому наблюдению, однако в дискретные моменты $t_i \triangleq ih$, $i \in \mathbb{N}$ (T делится на h нацело и $\mathcal{I} \triangleq T/h$), имеется следующая статистическая информация: безошибочные наблюдения текущей стоимости базовых активов $S_i = S_{t_i}$ и косвенные наблюдения текущей стоимости деривативов $G_i \in \mathbb{R}^{M'}$ (в общем случае $M \neq M'$).

Введем следующие обозначения:

- $\mathcal{O}_i \triangleq \sigma\{S_j, G_j : 0 \leq j \leq i\}$ — семейство σ -алгебр, порожденных всеми наблюдениями, полученными до момента t_i включительно;
- $\mathcal{G}_i \triangleq \sigma\{G_j, 0 \leq j \leq i\}$ — семейство σ -алгебр, порожденных наблюдениями $\{G_j\}$;
- $\mathcal{H}_i \triangleq \sigma\{S_j, Z_{t_j} : 0 \leq j \leq i\}$ — семейство σ -алгебр, порожденных ценами базового актива и МСП на дискретной временной сетке до момента t_i включительно.

Относительно наблюдаемой последовательности $\{G_i\}$ предполагается, что она обладает марковским свойством при известной паре (S, Z) , т. е. для любого $B \in \mathcal{B}(\mathbb{R}^{M'})$ и $i = \overline{1, \mathcal{I}}$ \mathcal{P} -п. н. выполняется цепочка равенств:

$$\begin{aligned} \mathbb{E}\{\mathbf{I}_B(G_i) | \mathcal{H}_i \vee \mathcal{G}_{i-1}\} &= \mathbb{E}\{\mathbf{I}_B(G_i) | S_i, Z_{t_i}, G_{i-1}\} = \\ &= \sum_{m=1}^M Z_{t_i}^m \int_B \Phi_i^m(dq | S_i, G_{i-1}), \quad (4) \end{aligned}$$

где $\{\Phi_i^m(\cdot | s, g)\}_{i=\overline{1, \mathcal{I}}, m=\overline{1, M}}$ — набор известных мер, определяющих соответствующее условное распределение. Также предполагается, что для каждого $i = \overline{1, \mathcal{I}}$ выбрана такая мера μ_i , что $\Phi_i^m \ll \mu_i$ для всех $m = \overline{1, M}$ и $\phi_i^m(\cdot | s, g) \triangleq d\Phi_i^m(\cdot | s, g) / d\mu_i$ — соответствующие производные Радона–Никодима. Для начального наблюдения G_0 также заданы такие мера μ_0 и плотности $\phi_0^m(\cdot | s)$, что для любого $B \in \mathcal{B}(\mathbb{R}^{M'})$ \mathcal{P} -п. н. выполняется равенство

$$\mathbb{E}\{\mathbf{I}_B(G_0) | S_0, Z_0\} = \sum_{m=1}^M Z_0^m \int_B \phi_0^m(q | S_0) \mu_0(dq).$$

Задача мониторинга МРР сводится к вычислению

$$\widehat{Z}_i \triangleq \mathbb{E}\{Z_{t_i} | \mathcal{O}_i\}, \quad i = \overline{1, \mathcal{I}}.$$

Предложенная модель наблюдений нуждается в объяснении.

Во-первых, на практике цены базовых активов и деривативов недоступны наблюдению в непрерывном времени, иначе при выполнении дополнительного условия идентифицируемости [3] значение скрытого МСП можно было бы восстановить безошибочно по $\{S_t\}$, сделав тем самым рынок,

состоящий только из депозита и базовых активов, полным [4, 5].

Во-вторых, в предложенной модели рынка справедливые цены деривативов не могут быть известны безошибочно даже в дискретные моменты времени, иначе структура цены $F^m(t, S_t, Z_t)$ позволила бы восстановить значение Z в момент t_i .

В-третьих, модель рынка известна его участникам, а значит, они обладают знанием набора возможных справедливых цен деривативов $\{F^{m\ell}(t, s)\}$ — решений (3). Единственное, чего они не знают точно, — правильный выбор одного из этих вариантов, определяемый МСП Z_t . Трейдеры в соответствии со своими соображениями выставляют заявки на покупку (bid) и продажу (ask) деривативов по цене, не совпадающей со справедливой, внося в заявки непредумышленные или намеренные ошибки. Именно последовательности заявок bid, ask и реализованных сделок (trade) и служат доступной статистической информацией.

Стохастические модели тройки «bid–ask–trade» достаточно разнообразны (см., например, [6–10]). Здесь предполагается доступность только цен сделок. Предложенная модель наблюдений цен деривативов (4) является достаточно общей и описывает большое число схем, имеющих практическое значение. Ниже на примере случая одного базового инструмента и одного дериватива рассматриваются следующие схемы (возможности модели этими схемами не исчерпываются):

- (1) наблюдения G_i представляют собой истинную справедливую цену, искаженную независимым мультипликативным белым шумом;
- (2) наблюдения G_i цены дериватива — марковская цепь с множеством значений $\{F^\ell(t_i, S_i)\}_\ell$ и матрицей переходных вероятностей $\Gamma(i, S_i, Z_{t_i})$, зависящей от текущего времени, цены базового актива и текущего состояния МСП.

3 Аналитическое решение задачи фильтрации и численный алгоритм

Рассматриваемая модель рынка гарантирует корректное определение вспомогательного процесса $V_t = \text{col}(V_t^1, \dots, V_t^N)$, $V_t^n \triangleq \ln S_t^n / S_0^n$, $n = \overline{1, N}$, который, согласно правилу Ито, представляет собой единственное сильное решение СДС

$$dV_t = \sum_{\ell=1}^L Z_t^\ell \left(a^\ell(t) - \frac{1}{2} b^\ell(t) \right) dt + \sum_{\ell=1}^L Z_t^\ell \sigma^\ell(t) dw_t, \quad t \in (0, T], \quad V_0 \equiv \mathbf{1},$$

где

$$b^\ell(t) \triangleq \text{col} (B_{11}^\ell(t), \dots, B_{NN}^\ell(t)).$$

Если $V_0 \triangleq S_0$ и $V_i \triangleq V_{t_i} - V_{t_{i-1}}$ для любых $i = \overline{1, T}$, то легко проверить, что наблюдения $\{V_i\}$ взаимно однозначны наблюдениям $\{S_i\}$ и порожденные ими наборы σ -алгебр совпадают. Наблюдения V_i имеют плотность распределения при $i \geq 1$, и совместное распределение пары (Z_{t_i}, V_i) относительно $Z_{t_{i-1}}$ определяется плотностями $\{\xi_i^{jk}(v)\}_{i=\overline{1, T}, j, k=\overline{1, L}}$:

$$P \{Z_{t_i} = e_k, V_i \in A | Z_{t_{i-1}} = e_j\} = \int_A \xi_i^{jk}(v) dv.$$

Теорема 1. Оценка оптимальной фильтрации \widehat{Z}_i определяется следующей рекуррентной процедурой:

$$\widehat{Z}_i^\ell = \frac{\widetilde{Z}_i^\ell}{\sum_{j=1}^L \widetilde{Z}_i^j}, \quad \ell = \overline{1, L}, \quad i > 1; \quad (5)$$

$$\widetilde{Z}_i^\ell = \sum_{k=1}^L \phi_i^\ell(G_i | S_i, G_{i-1}) \xi_i^{k\ell}(V_i) \widehat{Z}_{i-1}^k, \quad \ell = \overline{1, L},$$

$$V_i = \text{col} \left(\ln \frac{S_i^1}{S_{i-1}^1}, \dots, \ln \frac{S_i^N}{S_{i-1}^N} \right), \quad (6)$$

с начальным условием

$$\widehat{Z}_0^\ell = \frac{\pi_0^{Z\ell} \phi_0^\ell(G_0 | S_0)}{\sum_{j=1}^L \pi_0^{Zj} \phi_0^j(G_0 | S_0)}, \quad \ell = \overline{1, L}. \quad (7)$$

Доказательство теоремы 1 приведено в приложении.

Основная сложность численной реализации рекурсии (5), (6) связана с проблемой вычисления плотностей $\xi_i^{jk}(v)$. Дело в том, что они представляют собой сдвиг-масштабные смеси гауссиан, причем смешивающее распределение определяется поведением МСП Z на отрезках $[t_{i-1}, t_i]$ при фиксированных начальном и конечном значениях. Для частного случая матрицы интенсивностей переходов Λ , коэффициентов $\{a^\ell\}$ и $\{b^\ell\}$, не зависящих от времени t , в [11] предложена составная схема «средних» прямоугольников, которую предлагается применить и в данной задаче. Введем следующие обозначения:

- $\mathcal{N}(v, m, K)$ — плотность распределения гауссовского случайного вектора со средним m и невырожденной ковариационной матрицей K ;
- $U_i \triangleq (t_{i-1} + t_i)/2$ — середина отрезка $[t_{i-1}, t_i]$;
- $u_{i,m} \triangleq t_{i-1} + h^{1+\alpha}(m - 1/2)$ — середины отрезков более мелкого разбиения с шагом $h^{1+\alpha}$, $m = \overline{1, [h^{-\alpha}]}$, $0 < \alpha \leq 1$;

– $Q^{k\ell}(v, u) \triangleq \exp[(\Lambda_{kk}(u) - \Lambda_{\ell\ell}(u))] \times \mathcal{N}(v, u(a^k(u) + (1/2)b^k(u)) + (h - u)(a^\ell(u) + (1/2)b^\ell(u)), uB^k(u) + (h - u)B^\ell(u))$ — вспомогательная функция.

Для вычисления $\xi_i^{jk}(\cdot)$ предлагается воспользоваться аппроксимацией

$$\xi_i^{jk}(V_i) \approx \delta_{k\ell} e^{\Lambda_{\ell\ell}(U_i)h} \times \mathcal{N}\left(V_i, h\left(a^\ell(U_i) + \frac{1}{2}b^\ell(U_i)\right), hB^\ell(U_i)\right) + (1 - \delta_{k\ell})\Lambda_{k\ell}(U_i)h^{1+\alpha} \sum_{m=1}^{[h^{-\alpha}]} Q^{k\ell}(V_i, u_{i,m}).$$

4 Численный пример

Предлагаемый пример иллюстрирует влияние наблюдений цен дериватива в дополнение к имеющимся наблюдениям цены базового актива на качество оценивания МСП в зависимости от вида этих наблюдений.

На отрезке времени $[0, 1]$ (1 год, 250 торговых дней по 8 часов каждый) моделируется поведение одного базового актива, стохастическая волатильность которого представляет собой МСП с четырьмя возможными состояниями «рост – состояние, предшествующее панике, – паника – рецессия». Ставка по депозиту $r = 2\%$ годовых, на рынке присутствует один дериватив – call-опцион с ценой исполнения 1. Для моделирования (1), (2) были выбраны следующие параметры: $S_0 \equiv 1$; $a = (0,04; 0,035; 0,015; 0,02)$; $\sigma = (0,1; 0,12; 0,25; 0,15)$;

$$\Lambda = \begin{bmatrix} -12,5 & 12,5 & 0 & 0 \\ 100 & -500 & 400 & 0 \\ 0 & 0 & -125 & 125 \\ 40 & 0 & 10 & -50 \end{bmatrix}; \quad \pi_0^Z = \begin{bmatrix} 0,7273 \\ 0,0182 \\ 0,0727 \\ 0,1818 \end{bmatrix}.$$

Интегрирование СДС выполняется методом Эйлера–Маруямы, адаптированным к скачкам [12], с основным шагом $\tau = 10^{-6}$, интегрирование системы (3) — методом, предложенным в [2], с шагом по времени $H = 0,001$ (шаг соответствует 2 ч торгового времени) и шагом по координате $\delta = 0,002$, что соответствует 0,2 п. п. начальной цены базового актива.

Все наблюдения проводятся с шагом $h = 0,0001$, что соответствует 12 мин торгового времени. Сравнивается качество оценок МСП $\{Z_{t_i}\}$, построенных по двум комплексам доступных наблюдений:

- (1) имеются прямые наблюдения цен базового актива $\{S_i\}$, а также цена дериватива с мультипликативным шумом:

$$G_i^{LN} = F_{t_i}\varepsilon_i,$$

где $\{\varepsilon_i\}$ — последовательность независимых одинаково распределенных случайных вели-

чин, имеющих логнормальное распределение с параметрами $a^{LN}(h) = 0,04h$ и $\sigma^{LN}(h) = 0,28h$;

- (2) имеются прямые наблюдения $\{S_i\}$ и косвенные наблюдения цен дериватива $\{G_i^{MC}\}$. Последние представляют собой цепь со значениями $\{F^\ell(t_i, S_i)\}$ — вариантами возможных цен опциона, обладающую марковским свойством при известной траектории $\{Z_t\}$. Условные матрицы переходов этой цепи с компонентами

$$P\{G_i^{MC} = F^k(t_i, S_i) | G_{i-1}^{MC} = F^j(t_{i-1}, S_{i-1}), Z_{t_i} = e_m\} = \Gamma_{jkm},$$

где $\Gamma_m = \|\Gamma_{jkm}\|_{j,k}$ ($m = \overline{1, L}$) являются матричными экспонентами

$$\Gamma_m(h) = \exp(h\Upsilon_m),$$

построенными по «матрицам интенсивностей переходов»

$$\Upsilon_1 = \begin{bmatrix} -30 & 10 & 10 & 10 \\ 100 & -300 & 100 & 100 \\ 100 & 100 & -300 & 100 \\ 100 & 100 & 100 & -300 \end{bmatrix};$$

$$\Upsilon_2 = \begin{bmatrix} -3000 & 1000 & 1000 & 1000 \\ 1 & -3 & 1 & 1 \\ 1000 & 1000 & -3000 & 1000 \\ 1000 & 1000 & 1000 & -3000 \end{bmatrix};$$

$$\Upsilon_3 = \begin{bmatrix} -300 & 100 & 100 & 100 \\ 100 & -300 & 100 & 100 \\ 10 & 10 & -30 & 10 \\ 100 & 100 & 100 & -300 \end{bmatrix};$$

$$\Upsilon_4 = \begin{bmatrix} -300 & 100 & 100 & 100 \\ 100 & -300 & 100 & 100 \\ 100 & 100 & -300 & 100 \\ 10 & 10 & 10 & -30 \end{bmatrix}.$$

Следует отметить, что параметры распределений наблюдений цен деривативов зависят от временного шага h между поступающими наблюдениями.

На рис. 1 представлены ошибки наблюдений $\Delta_i^{LN} = G_i^{LN} - F_{t_i}$ и $\Delta_i^{MC} = G_i^{MC} - F_{t_i}$. Результаты оценивания МСП по различной статистической информации на интервале $[0; 0,2]$ приведены на рис. 2. На нем для всех компонент МСП изображены:

- истинное значение компоненты X_t^ℓ ;
- оценка фильтрации \widehat{X}_t^S , вычисленная только по наблюдениям S_i ;
- оценка фильтрации \widehat{X}_t^{LN} , вычисленная по наблюдениям S_i и G_i^{LN} ;
- оценка фильтрации \widehat{X}_t^{MC} , вычисленная по наблюдениям S_i и G_i^{MC} .

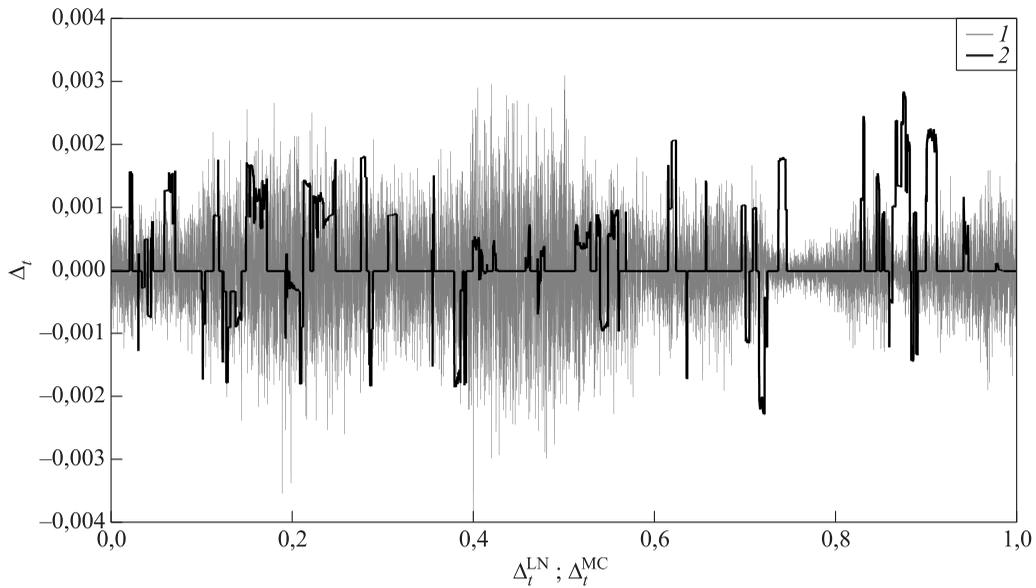


Рис. 1 Ошибки наблюдений цен деривативов: 1 — Δ_t^{LN} ; 2 — Δ_t^{MC}

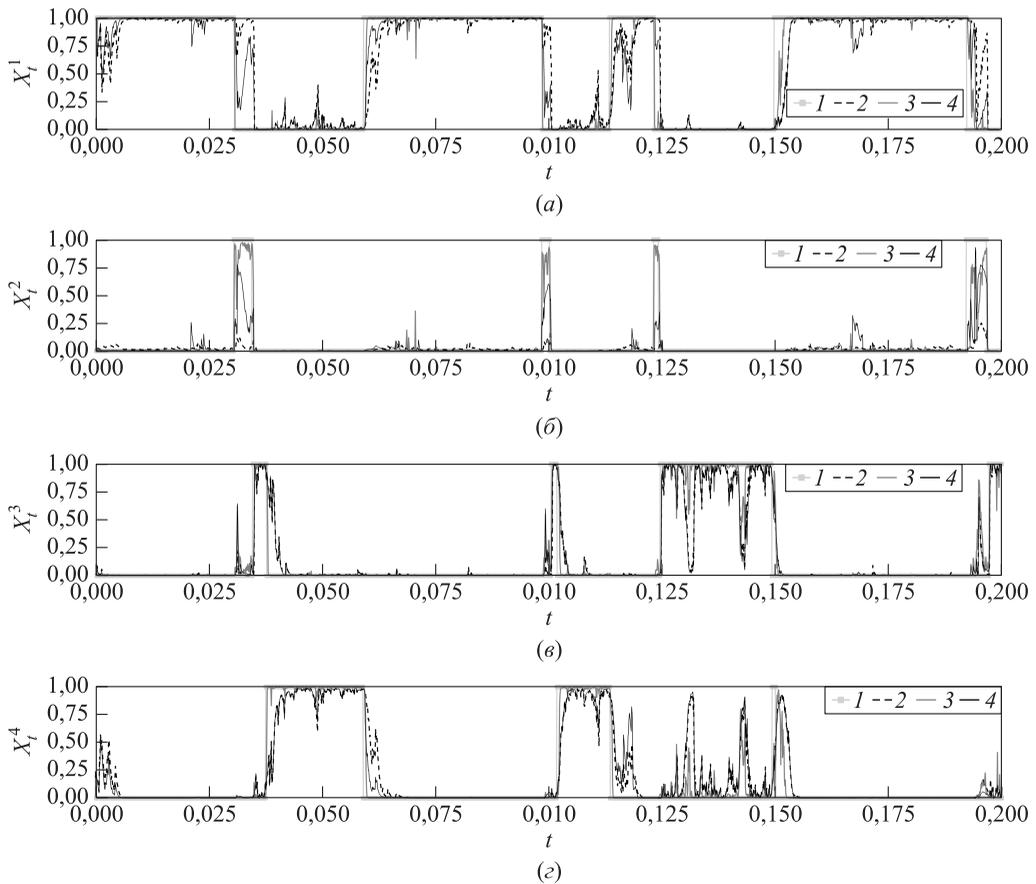


Рис. 2 Оценки компонентов МСП по разным комплексам наблюдений: 1 — X_t^I ; 2 — \hat{X}_t^S ; 3 — \hat{X}_t^{LN} ; 4 — \hat{X}_t^{MC}

По результатам расчетов можно сделать следующие заключения.

1. Параметры в примере выбраны таким образом, что рынок пребывает во 2-м состоянии, предшествующем панике, достаточно короткое время: в среднем 2 ч. За это время можно получить только 20 наблюдений цен базового актива. В то же время значения мгновенной процентной ставки и волатильности во 2-м состоянии близки к соответствующим значениям состояния 1 (рост): 0,035 против 0,04 и 0,12 против 0,1 соответственно. Поэтому вполне ожидаемо, что качество мониторинга состояния 2, выполненного только по наблюдениям цен базового актива, низкое (см. рис. 2, б). Очевидно, что любая дополнительная статистическая информация актуальна для улучшения качества оценивания 2-го состояния.
2. Траектории возможных значений цен опциона $\{F^\ell\}$ визуально малоотличимы друг от друга, а также от наблюдений G_i^{LN} и G_i^{MC} , поэтому на рис. 1 представлены не сами наблюдения, а их ошибки. Видна принципиальная разница характера траекторий этих ошибок: Δ_i^{LN} характерна для белого шума, в то время как в Δ_i^{MC} очевидна временная зависимость.
3. Рисунок 2 позволяет сделать вывод о том, что включение в процедуру фильтрации наблюдений обоих деривативов позволяет повысить качество оценивания. Выборочные средние квадраты значений ошибок наблюдений (см. рис. 1), вычисленные по данным траекториям, для обоих видов наблюдений практически совпадают. Тем не менее бóльший вклад в повышение качества оценивания вносят наблюдения с логнормальным шумом. Этот факт можно объяснить тем, что различия в $\{F^\ell\}$ хотя и малые, но есть. Из-за мультипликативного характера зашумления дисперсии ошибок для разных вариантов F^ℓ различны, что для случая непрерывного времени означает выполнение условий точной идентифицируемости МСП [3]. В рассматриваемом допредельном случае наличие двух наблюдений (S_i, G_i^{LN}) обеспечивает весьма высокое качество оценивания всех компонент МСП Z .

5 Промежуточные выводы

Мониторинг MPR по наблюдениям, полученным в дискретные моменты времени, представляет собой задачу оптимальной фильтрации скрытого МСП. Ее теоретическое решение совместно с численным алгоритмом и иллюстративным примером

приведено в данной части цикла. Однако использованная в данной статье модель наблюдений — это идеализация и упрощение: реальные статистические данные представляют собой потоки значений цен заявок и сделок, фиксируемых в случайные моменты времени, — так называемые мультивариантные точечные процессы (МТП). Обычно в системах электронных торгов они оказываются высокочастотными. Решение задач оптимальной фильтрации состояний МСП по наблюдениям МТП нетривиально как в теоретическом плане, так и с точки зрения практической реализации. Один из вариантов создания эффективных субоптимальных алгоритмов фильтрации основан на использовании диффузионной аппроксимации получаемых МТП. Использованию этого подхода и численному анализу полученных оценок и будет посвящена заключительная часть цикла.

Приложение

Доказательство теоремы 1. Используем метод математической индукции. Пусть на $(i - 1)$ -м шаге известна оптимальная оценка

$$\hat{Z}_{i-1} = \mathbb{E} \{ Z_{t_{i-1}} | \mathcal{O}_{i-1} \}.$$

Обозначим

$$\mathcal{E}(v) \triangleq \text{col}(\exp(v^1), \dots, \exp(v^L)).$$

Свойства условных математических ожиданий гарантируют, что следующая цепочка равенств выполняется \mathcal{P} -п. н. для любых множеств $A \in \mathcal{B}(\mathbb{R}^N)$ и $B \in \mathcal{B}(\mathbb{R}^{M'})$:

$$\begin{aligned} & \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \mathbf{I}_B(\mathbf{G}_i) | \mathcal{O}_{i-1} \right\} = \\ & = \mathbb{E} \left\{ \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \mathbf{I}_B(\mathbf{G}_i) | \mathcal{O}_{i-1} \vee \mathcal{H}_i \right\} | \mathcal{O}_{i-1} \right\} = \\ & = \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \mathbb{E} \left\{ \mathbf{I}_B(\mathbf{G}_i) | \mathcal{O}_{i-1} \vee \mathcal{H}_i \right\} | \mathcal{O}_{i-1} \right\} = \\ & = \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \int_B \phi_i^\ell(q | S_i, \mathbf{G}_{i-1}) \mu_i(dq) | \mathcal{O}_{i-1} \right\} = \\ & = \mathbb{E} \left\{ \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \times \right. \right. \\ & \times \left. \int_B \phi_i^\ell(q | S_i, \mathbf{G}_{i-1}) \mu_i(dq) | \mathcal{O}_{i-1} \vee Z_{t_{i-1}} \right\} | \mathcal{O}_{i-1} \right\} = \\ & = \mathbb{E} \left\{ \mathbb{E} \left\{ Z_i^\ell \mathbf{I}_A(\mathbf{V}_i) \times \right. \right. \\ & \times \left. \int_B \phi_i^\ell(q | \text{diag}(S_{i-1}) \mathcal{E}(\mathbf{V}_i), \mathbf{G}_{i-1}) \mu_i(dq) | \mathcal{O}_{i-1} \vee \right. \\ & \left. \left. \vee Z_{t_{i-1}} \right\} | \mathcal{O}_{i-1} \right\} = \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E} \left\{ \int_A \int_B \phi_i^\ell(q | \text{diag}(\mathbf{S}_{i-1})\mathcal{E}(v), \mathbf{G}_{i-1})\mu_i(dq) \times \right. \\
 &\quad \left. \times \sum_{k=1}^L Z_{t_{i-1}}^k \xi_i^{k\ell}(v) dv \middle| \mathcal{O}_{i-1} \right\} = \\
 &= \int_A \int_B \left[\sum_{k=1}^L \widehat{Z}_{i-1}^k \phi_i^\ell(q | \text{diag}(\mathbf{S}_{i-1})\mathcal{E}(v), \mathbf{G}_{i-1})\xi_i^{k\ell}(v) \right] \times \\
 &\quad \times \mu_i(dq) dv.
 \end{aligned}$$

Выражение в квадратных скобках является производной Радона–Никодима условного распределения тройки (Z_{t_i}, S_i, G_i) относительно \mathcal{O}_{i-1} по мере $\mu_i \times \lambda \times \nu$, где λ — мера Лебега на $\mathbb{R}^{M'}$; ν — считающая мера, сосредоточенная на множестве $\{e_1, \dots, e_L\}$. Тогда по формуле Байеса

$$\begin{aligned}
 \widehat{Z}_i^\ell &= \\
 &= \frac{\sum_{k=1}^L \widehat{Z}_{i-1}^k \phi_i^\ell(G_i | \text{diag}(\mathbf{S}_{i-1})\mathcal{E}(V_i), \mathbf{G}_{i-1})\xi_i^{k\ell}(V_i)}{\sum_{m,j=1}^L \widehat{Z}_{i-1}^m \phi_i^j(G_i | \text{diag}(\mathbf{S}_{i-1})\mathcal{E}(V_i), \mathbf{G}_{i-1})\xi_i^{mj}(V_i)} = \\
 &= \frac{\sum_{k=1}^L \widehat{Z}_{i-1}^k \phi_i^\ell(G_i | S_i, \mathbf{G}_{i-1})\xi_i^{k\ell}(V_i)}{\sum_{m,j=1}^L \widehat{Z}_{i-1}^m \phi_i^j(G_i | S_i, \mathbf{G}_{i-1})\xi_i^{mj}(V_i)},
 \end{aligned}$$

что доказывает истинность формул (5) и (6). Истинность формулы (7) вычисления оценки при $t = 0$ также следует из формулы Байеса. Теорема 1 доказана.

Литература

1. *Борисов А.* Рынок с марковской скачкообразной волатильностью I: мониторинг цены риска как задача оптимальной фильтрации // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 27–33. doi: 10.14357/19922264230204. EDN: GAXCHQ.
2. *Борисов А.* Рынок с марковской скачкообразной волатильностью II: алгоритм вычисления справедливой цены деривативов // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 18–24. doi: 10.14357/19922264230303. EDN: DNJXJB.
3. *Borisov A., Sokolov I.* Optimal filtering of Markov jump processes given observations with state-dependent noises: Exact solution and stable numerical schemes // Mathematics, 2020. Vol. 8. No. 4. Art. 506. 22 p. doi: 10.3390/math8040506.
4. *Shiryayev A.* Essentials of stochastic finance: Facts, models, theory. — Hackensack, NJ, USA: World Scientific, 1999. 834 p.
5. *Karatzas I., Shreve S.* Methods of mathematical finance. — Berlin: Springer, 2016. 416 p.
6. *Bollen N., Smith T., Whaley R.* Modeling the bid/ask spread: Measuring the inventory-holding premium // J. Financ. Econ., 2004. Vol. 72. No. 1. P. 97–141. doi: 10.1016/S0304-405X(03)00169-7.
7. *Lim T., Vath V., Sahut J., Scotti S.* Bid-ask spread modelling, a perturbation approach // Progress in probability / Eds. R. Dalang, M. Dozzi, F. Russo. — Basel: Birkhäuser, 2013. Vol. 67. P. 411–434. doi: 10.1007/978-3-0348-0545-2_21.
8. *Cont R., Kukanov A., Stoikov S.* The price impact of order book events // J. Financ. Economet., 2014. Vol. 12. No. 1. P. 47–88. doi: 10.2139/ssrn.1712822.
9. *Kühn C., Riedel M.* Price setting of market makers: A filtering problem with endogenous filtration // Math. Financ., 2017. Vol. 27. No. 1. P. 251–275. doi: 10.1111/mafi.12079.
10. *Cont R., Müller M.* A stochastic partial differential equation model for limit order book dynamics // SIAM J. Financ. Math., 2021. Vol. 12. No. 2. P. 744–787. doi: 10.1137/19M1254489.
11. *Борисов А.* Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям III: случай мультипликативных шумов // Информатика и её применения, 2020. Т. 14. Вып. 2. С. 10–18. doi: 10.14357/19922264200202. EDN: OVICVK.
12. *Platen E., Bruti-Liberati N.* Numerical solution of stochastic differential equations with jumps in finance. — Berlin: Springer, 2010. 884 p.

Поступила в редакцию 27.01.23

MARKET WITH MARKOV JUMP VOLATILITY III: PRICE OF RISK MONITORING ALGORITHM GIVEN DISCRETE-TIME OBSERVATIONS OF ASSET PRICES

A. V. Borisov

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

Abstract: The third part of the series is devoted to the online estimation of the market price of risk. The market model includes a deposit, underlying, and derivative assets. The model of the underlying asset prices contains stochastic volatility represented by a Markov jump process (MJP). There is no arbitrage in the considered market; so,

the market price of risk is a function of the MJP's current value. So, the MJP monitoring problem transforms into the MJP state filtering one. The statistical data are available at discrete moments and contain the direct observations of the underlying assets and indirect observations of the derivative ones. The paper presents the solution to the optimal filtering problem and the corresponding algorithm of its numerical realization. The paper also contains a numerical example demonstrating the performance of the MJP state estimates in dependence on the type and structure of the available observations.

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

DOI: 10.14357/19922264230402

EDN: OFYELT

Acknowledgments

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

References

1. Borisov, A. 2023. Rynok s markovskoy skachkoobraznoy volatil'nost'yu I: monitoring tseny riska kak zadacha optimal'noy fil'tratsii [Market with Markov jump volatility I: Price of risk monitoring as an optimal filtering problem]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2): 27–33. doi: 10.14357/19922264230204. EDN: GAXCHQ.
2. Borisov, A. 2023. Rynok s markovskoy skachkoobraznoy volatil'nost'yu II: algoritm vychisleniya spravedlivoy tseny derivativov [Market with Markov jump volatility II: Algorithm of derivative fair price calculation]. *Informatika i ee primeneniya — Inform. Appl.* 17(3):18–24. doi: 10.14357/19922264230303. EDN: DNXJGB.
3. Borisov, A., and I. Sokolov. 2020. Optimal filtering of Markov jump processes given observations with state-dependent noises: Exact solution and stable numerical schemes. *Mathematics* 8(4):506. 22 p. doi: 10.3390/math8040506.
4. Shirayev, A. 1999. *Essentials of stochastic finance: Facts, models, theory*. Hackensack, NJ: World Scientific. 834 p.
5. Karatzas, I., and S. Shreve. 2016. *Methods of mathematical finance*. Berlin: Springer. 416 p.
6. Bollen, N., T. Smith, and R. Whaley. 2004. Modeling the bid/ask spread: Measuring the inventory-holding premium. *J. Financ. Econ.* 72(1):97–141. doi: 10.1016/S0304-405X(03)00169-7.
7. Lim, T., V. Vath, J. Sahut, and S. Scotti. 2013. Bid-ask spread modelling, a perturbation approach. *Progress in probability*. Eds. R. Dalang, M. Dozzi, and F. Russo. Basel: Birkhäuser. 67:411–434. doi: 10.1007/978-3-0348-0545-2_21.
8. Cont, R., A. Kukanov, and S. Stoikov. 2014. The price impact of order book events. *J. Financ. Economet.* 12(1):47–88. doi: 10.2139/ssrn.1712822.
9. Kühn, C., and M. Riedel. 2017. Price setting of market makers: A filtering problem with endogenous filtration. *Math. Financ.* 27(1):251–275. doi: 10.1111/mafi.12079.
10. Cont, R., and M. Müller. 2021. A stochastic partial differential equation model for limit order book dynamics. *SIAM J. Financ. Math.* 12(2):744–787. doi: 10.1137/19M1254489.
11. Borisov, A. 2020. Chislennye skhemy fil'tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya III: sluchay mul'tiplikativnykh shumov [Numerical schemes of Markov jump process filtering given discretized observations III: Multiplicative noises case] *Informatika i ee primeneniya — Inform. Appl.* 14(2):10–18. doi: 10.14357/19922264200202. EDN: OVICVK.
12. Platen, E., and N. Bruti-Liberati. 2010. *Numerical solution of stochastic differential equations with jumps in finance*. Berlin: Springer. 884 p.

Received January 27, 2023

Contributor

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

ПРОЦЕДУРА ПОСТРОЕНИЯ МНОЖЕСТВА ПАРЕТО ДЛЯ ДИФФЕРЕНЦИРУЕМЫХ КРИТЕРИАЛЬНЫХ ФУНКЦИЙ

Я. И. Рабинович¹

Аннотация: Универсальная вычислительная процедура многокритериальной оптимизации позволяет аппроксимировать множество Парето при предъявлении различных требований к вектору частных критериев эффективности и множеству допустимых решений. В настоящей работе предполагается, что частные критерии эффективности псевдодогнуты в открытой окрестности компактного выпуклого множества допустимых решений, которое может быть задано дифференцируемыми функциональными ограничениями. Для построения на основе универсальной процедуры конкретных численных методов аппроксимации множества Парето предлагается правило выбора начального приближения и правило перехода от текущего опорного решения к последующему.

Ключевые слова: многокритериальная оптимизация; множество Парето; численные методы аппроксимации; универсальная процедура

DOI: 10.14357/19922264230403

EDN: NEZRGD

1 Введение

Задача построения множества эффективных векторных оценок (множества Парето) чрезвычайно актуальна в самых разнообразных отраслях деятельности, таких как разработка сложных систем технического назначения, перспективное планирование или социология. В настоящей работе в предположении дифференцируемости критериальных функций на основе универсальной процедуры [1] предложен конкретный численный метод аппроксимации множества Парето.

2 Постановка задачи

Пусть в s -мерном евклидовом пространстве \mathbb{R}^s задана m -мерная непрерывная вектор-функция

$$w(x) \in \mathbb{R}^m, \quad (1)$$

образующая вектор частных критериев эффективности, принимающий на непустом компактном множестве допустимых решений

$$X \subset \mathbb{R}^s \quad (2)$$

положительные значения $w(x) > 0$, удовлетворяя соотношениям:

$$\left. \begin{aligned} w(x) \in w(X) \subset \text{int } \mathbb{R}_+^m; \\ w(X) = \{u \in \mathbb{R}^m \mid u = w(x), x \in X\}; \\ \mathbb{R}_+^m = \{u \in \mathbb{R}^m \mid u \geq 0\}, \end{aligned} \right\} \quad (3)$$

так что множество достижимых векторных оценок $w(X)$ из (3) принадлежит внутренности $\text{int } \mathbb{R}_+^m$ неотрицательного органта \mathbb{R}_+^m . Каждый критерий набора $\{w_k(x)\}_{k \in I}$, $I = \{k \mid 1 \leq k \leq m\}$, желательно увеличивать на множестве допустимых решений (2).

Определение 1. Векторная оценка $w \in w(X)$ *эффективна (слабо эффективна)*, если для всякой векторной оценки $u \in w(X)$ система неравенств $u \geq w$ несовместна при условии, что хотя бы одно неравенство строгое (все неравенства строгие).

Векторная оценка $w \in w(X)$ *доминируема* или *определенно неэффективна*, если существует векторная оценка $u \in w(X)$, $u > w$.

Всякое допустимое решение $x \in X$, доставляющее эффективное (слабо эффективное, доминируемое) значение вектора $w(x)$, называется эффективным (слабо эффективным, доминируемым) решением.

В согласии с определением 1 множества эффективных (X_e), слабо эффективных (X_0) и доминируемых (X_d) решений из множества допустимых решений (X) подчиняются соотношениям:

$$X_e \subset X_0 \subset X; X_0 \cap X_d = \emptyset; X = X_0 \cup X_d,$$

где \emptyset — пустое множество; $\cup(\cap)$ — символ объединения (пересечения) множеств. Соответственно множества эффективных $w(X_e)$, слабо эффективных $w(X_0)$, доминируемых $w(X_d)$ и достижимых $w(X)$ векторных оценок удовлетворяют соотношениям:

$$\begin{aligned} w(X_e) \subset w(X_0) \subset w(X); w(X_0) \cap w(X_d) = \emptyset; \\ w(X) = w(X_0) \cup w(X_d). \end{aligned}$$

В настоящей работе рассматриваются методы аппроксимации множества Парето на основе универсальной вычислительной процедуры [1]. Введем

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук; jacrabin@rambler.ru

Определение 2. Если $\|v\|$ — норма вектора $v \in \mathbb{R}^m$, то величина

$$D(W, U) = \sup_{w \in W} \inf_{u \in U} \|w - u\|, \quad \emptyset \neq W, U \subset \mathbb{R}^m,$$

называется *отклонением* множества W от множества U , а величина

$$\Delta(W, U) = \max \{D(W, U), D(U, W)\}, \quad \emptyset \neq W, U \subset \mathbb{R}^m,$$

называется *расстоянием по Хаусдорфу* между множествами W и U .

Определение 3. Дифференцируемая на открытом множестве $A \subset \mathbb{R}^s$ функция f называется *псевдовогнутой* на $A \neq \emptyset$, если для любых $x, y \in A$ неравенство $\langle \nabla f(x), y - x \rangle \leq 0$ влечет неравенство $f(x) \geq f(y)$, где ∇f — градиент функции f ; $\langle \bullet, \bullet \rangle$ — скалярное произведение векторов.

Функция f называется *псевдовыпуклой*, если функция $-f$ псевдовогнута.

Следствие 1.

1. Функция $\sum_{k=1}^m f_k$ псевдовогнута, если ее слагаемые псевдовогнуты.
2. Точка локального максимума псевдовогнутой на выпуклом множестве $A \subset \mathbb{R}^s$ функции f является точкой ее глобального максимума.
3. Пусть функция f псевдовогнута на выпуклом множестве $A \subset \mathbb{R}^s$ и положительно определена на A , так что $f(x) > 0, x \in A$. Тогда
 - (а) функция f^n псевдовогнута на A , если целая степень $n \geq 1$;
 - (б) функция f^n псевдовыпукла на A , если целая степень $n \leq -1$.

Определение 4. Если непустое множество $X \subset \mathbb{R}^s$ задано функциональными ограничениями

$$X = \{x \in \mathbb{R}^s | v_k(x) \geq 0, k \in K\}, \quad K = \{k | 1 \leq k \leq n\},$$

левые части которых дифференцируемы, то в точке $x \in X$ множество

$$H(x) = \{h \in \mathbb{R}^s | \langle \nabla v_k(x), h \rangle > 0, k \in K(x)\},$$

называется *конусом внутренних направлений*, где $K(x) = \{k \in K | v_k(x) = 0\}$ — множество индексов «активных» в точке $x \in X$ функциональных ограничений.

Следствие 2. Если $X \subset \mathbb{R}^s$ — непустой выпуклый компакт, то выполняется включение $X \subset \{x\} + \overline{H}(x)$, где многогранный конус $\overline{H}(x)$ — замыкание конуса внутренних направлений; $A + B$ — векторная сумма множеств.

Укажем также известное полезное свойство псевдовогнутых функций [2].

Лемма 1. Псевдовогнутая на выпуклом многограннике $A = \text{conv}\{a^q\}_{q=1}^r$ функция f достигает минимума в крайней точке:

$$\min_{x \in A} f(x) = \min_{1 \leq q \leq r} f(a^q).$$

3 Универсальная процедура

Дадим в согласии с [1] строгое описание универсальной процедуры аппроксимация множества Парето.

На непустом компактном множестве допустимых решений $X \subset \mathbb{R}^s$ строится последовательность множеств $\{X_t\}_{t=1}^\infty \subset X$. Если $X_1 \subset X$ — произвольное начальное приближение и известно множество $X_t, t \geq 1$, то следующее за ним множество X_{t+1} подчиняется соотношениям:

$$X_{t+1} = \bigcup_{x \in X_t} X_{t+1}(x), \quad X_{t+1}(x) = \{x\} + \bigcup_{J \in M_t(x)} \{h(x, J)\} \subset X. \quad (4)$$

В согласии с (4) всякая опорная точка $x \in X_t$ порождает на следующем $(t + 1)$ -м уровне непустую векторную сумму множеств $X_{t+1}(x)$, причем направление $h(x, J)$ перехода из опорной точки $x \in X_t$ в следующую точку $y = x + h(x, J) \in X_{t+1}(x), J \in M_t(x)$, удовлетворяет условиям:

$$h(x, \emptyset) = 0, \quad h(x, J) \neq 0, \quad \emptyset \neq J \subset I = \{k | 1 \leq k \leq m\}, \quad (5)$$

где ненулевые направления $h(x, J) \neq 0$ подчиняются соотношениям

$$\left. \begin{aligned} & y = x + h(x, J) \in X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon), \\ & \emptyset \neq J \subset I = \{k | 1 \leq k \leq m\}, \\ & Y_J(x, \varepsilon) = \left\{ y \in \mathbb{R}^s \left| \frac{w_k(y)}{w_k(x)} \geq 1 + \sigma \varepsilon^2, \right. \right. \\ & \quad \left. \left. k \in J, \frac{w_k(y)}{w_k(x)} \geq 1 - \sigma \varepsilon, k \in I \setminus J \right\}; \right. \\ & Y(x, \varepsilon) = \left\{ y \in \mathbb{R}^s \left| \sum_{k \in I} \frac{w_k(y)^2}{w_k(x)^2} \leq (1 + \sigma \varepsilon)^2 \right. \right\}, \\ & \quad \varepsilon = \varepsilon_t(x) \in (0, 1), \end{aligned} \right\} \quad (6)$$

$$\sigma = \left[2 \max_{z \in X} \sum_{k \in I} w_k(z) \right]^{-1} \min_{z \in X} \min_{k \in I} w_k(z) \in \left(0, \frac{1}{2} \right].$$

Последовательность множеств (4) в каждой опорной точке $x \in X_t$ ветвится, причем степень

ее ветвления $|\mathbf{M}_t(x)|$ определяет множество не вложенных друг в друга подмножеств $\mathbf{M}_t(x) \subset 2^I$, заданное соотношениями:

$$\mathbf{M}_t(x) = \left\{ \begin{array}{l} \mathbf{N}_t(x), \quad \text{если } \mathbf{N}_t(x) \neq \emptyset; \\ \{\emptyset\}, \quad \text{если } \mathbf{N}_t(x) = \emptyset; \end{array} \right\} \quad (7)$$

$$\mathbf{N}_t(x) = \left\{ \begin{array}{l} J \subset I \mid X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon) \neq \emptyset; \\ X \cap Y_M(x, \varepsilon) \cap Y(x, \varepsilon) \neq \emptyset, \\ M \neq J \subset M \subset I. \end{array} \right\}$$

Наибольшая степень ветвления последовательности (4) совпадает с наибольшим числом не вложенных друг в друга подмножеств множества номеров частных критериев эффективности $I = \{k \mid 1 \leq k \leq m\}$, так что

$$|\mathbf{M}_t(x)| \leq \frac{m!}{\lfloor m/2 \rfloor! (m - \lfloor m/2 \rfloor)!},$$

где $|A|$ — число элементов в конечном множестве A ; $\lfloor z \rfloor$ — целая часть числа z . Соотношения (6) и (7) включают величину $\varepsilon = \varepsilon_t(x) \in (0, 1)$ — параметр возмущения, который в начальной точке x^1 и в любых последующих точках $x^t \in X_t$, $x^{t+1} \in X_{t+1}(x^t)$ последовательности (4) таких, что

$$x^{t+1} = x^t + h(x^t, J_t), \quad J_t \in M_t(x), \quad (8)$$

удовлетворяет условиям:

$$\varepsilon_1(x^1) = \kappa \in (0, 1),$$

$$\varepsilon_{t+1}(x^{t+1}) = \begin{cases} \kappa \varepsilon_t(x^t), & \text{если } Q_t = \emptyset; \\ \varepsilon_t(x^t), & \text{если } Q_t \neq \emptyset, \end{cases}$$

$$Q_t = \bigcap_{q \leq t} \varepsilon_q(x^q) = \varepsilon_t(x^t) J_q, \quad (9)$$

где величина κ определяет степень дробления параметра возмущения. Эта величина может быть выбрана любой на интервале $(0, 1)$, но остается фиксированной на протяжении всей вычислительной процедуры.

Замечание 1. Комментируя процедуру (4)–(9), следует подчеркнуть:

- включение $y \in Y_J(x, \varepsilon)$ доставляет относительное (порядка ε^2) *увеличение* значений критериев из множества $\{w_k\}_{k \in J}$ за счет возможного относительного *уменьшения* (на величину порядка ε) остальных критериев;
- включение $y \in Y(x, \varepsilon)$ препятствует слишком резкому росту значений частных критериев, что на поздних этапах вычислительной процедуры не позволяет переходить от одной эффективной векторной оценки к другой и возвращаться обратно через все множество достижимых векторных оценок;

- пересечение $X \cap Y_J(x, \varepsilon)$ — выпуклый компакт, поскольку X — выпуклый компакт, $Y_J(x, \varepsilon)$ — замкнуто и выпукло в согласии с (6) и следствием 1;
- пересечение множеств $X \cap Y(x, \varepsilon)$ компактно, но не выпукло, поскольку согласно (6) множество $Y(x, \varepsilon)$ выпукло лишь в некоторых частных случаях (например, если псевдодогнуты компоненты вектор-функций $\pm w(x)$).

В согласии с доказательством, предложенным в работе [1], заданная соотношениями (4)–(9) последовательность множеств $\{w(X_t)\}_{t=1}^\infty$ аппроксимирует множество Парето в следующем смысле.

Теорема 1. Пусть в (1) компоненты вектор-функции $w \in \mathbb{R}^m$ положительно определены и псевдодогнуты в открытой окрестности непустого выпуклого компакта X . Тогда отклонение множества Парето $w(X_e)$ от аппроксимирующего множества $w(X_t)$ и отклонение аппроксимирующего множества $w(X_t)$ от множества Слейтера $w(X_0)$ стремятся к нулю с ростом номера аппроксимации t :

$$\lim_{t \rightarrow \infty} D(w(X_e), w(X_t)) = \lim_{t \rightarrow \infty} D(w(X_t), w(X_0)) = 0.$$

4 Численные методы на основе процедуры

Пусть условия теоремы 1 выполняются. Универсальная процедура (4)–(9) не содержит однозначного рецепта построения последовательности $\{w(X_t)\}_{t=1}^\infty$, аппроксимирующей множество Парето, поскольку не дает прямого указания, в какую именно новую опорную точку $y = x + h(x, J) \in X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon)$ следует переходить из исходной опорной точки $x \in X_t$.

Началу работы вычислительной процедуры предшествует процедура выбора начального приближения. В согласии с утверждением теоремы 1, сходимость процедуры (4)–(9) обеспечена при старте из любых точек X . Вместе с тем эффективность процедуры существенным образом зависит от выбора начальной точки (либо состава начального множества $X_1 \subset X$).

Процедура (4)–(9) становится реальной вычислительной процедурой (численным методом), если она дополнена *правилом выбора*:

- (а) начальной точки $\{x^1\} \subset X$ (множества начальных точек $X_1 \subset X$, $|X_1| > 1$, если в работе численного метода предусмотрен мультистарт);
- (б) каждой последующей опорной точки $y \in X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon)$.

5 Выбор начального приближения

Согласно замечанию 1, для формирования начального приближения X_1 можно использовать упрощенный вариант процедуры (4)–(9), где параметр возмущения $\varepsilon \in (0, 1)$ фиксирован, ограничения сверху на приращения частных критериев $\{w_k\}_{k \in I}$ отменяются — условие (6) заменяется требованием

$$\left. \begin{aligned} & y = x + h(x, J) \in X \cap Y_J(x, \varepsilon), \\ & \emptyset \neq J \subset I = \{k | 1 \leq k \leq m\}; \\ & Y_J(x, \varepsilon) = \left\{ y \in \mathbb{R}^s \mid \frac{w_k(y)}{w_k(x)} \geq 1 + \sigma\varepsilon^2, \right. \\ & \left. k \in J, \frac{w_k(y)}{w_k(x)} \geq 1 - \sigma\varepsilon, k \in I \setminus J \right\}, \\ & \sigma = \left[2 \max_{z \in X} \sum_{k \in I} w_k(z) \right]^{-1} \min_{z \in X} \min_{k \in I} w_k(z) \in \left(0, \frac{1}{2} \right], \end{aligned} \right\} (10)$$

тогда как множество подмножеств в (7) приобретает вид:

$$M_t(x) = \{ J \subset I | \emptyset \neq J, X \cap Y_J(x, \varepsilon) \neq \emptyset \} \quad (11)$$

и содержит все такие подмножества $J \subset I = \{k | 1 \leq k \leq m\}$, что каждый из критериев $\{w_k\}_{k \in J}$ допускает относительное (порядка ε^2) увеличение.

Сформулируем на основе упрощенной процедуры (4), (5), (10) и (11) правило выбора начального приближения.

Правило А. В качестве решения $y^J \in X \cap Y_J(x, \varepsilon)$ принимается проекция внешней точки $x \in X$, $x \notin Y_J(x, \varepsilon)$, на выпуклый компакт $X \cap Y_J(x, \varepsilon)$:

$$y^J = \arg \min_{y \in X \cap Y_J(x, \varepsilon)} \|x - y\|, \quad J \in M_t(x), \quad (12)$$

для чего следует решить задачу выпуклого программирования (минимизировать квадратичную функцию $\|x - y\|^2$ на выпуклом компакте).

Если компактное выпуклое допустимое множество X задано дифференцируемыми функциональными ограничениями, приближенное решение задачи (12) может быть получено следующим образом.

Согласно определению 4 и следствию 2, в каждой опорной точке $x \in X$ задана прямоугольная матрица $V(x)$ такая, что замыкание соответствующего конуса внутренних направлений $H(x)$ удовлетворяет соотношениям:

$$\begin{aligned} \overline{H}(x) &= \{h \in \mathbb{R}^s | V(x)h \geq 0\}, \quad X \subset \{x\} + \overline{H}(x), \\ X \setminus \{\{x\} + \overline{H}(x)\} &= \emptyset. \end{aligned} \quad (13)$$

Определим многогранный конус

$$H_J(x) = \{h \in \mathbb{R}^s | W_J(x)h \geq 0\}, \quad (14)$$

где строками матрицы $W_J(x)$ служат векторы $\nabla w_k(x)$, $k \in J$. Согласно (13), (14) можно утверждать, что

$$X \cap Y_J(x, \varepsilon) \subset \{x\} + \overline{H}(x) \cap H_J(x), \quad (15)$$

поскольку для всякого вектора $y = x + h \in X$, удовлетворяющего при некотором номере $k \in J$ неравенству $\langle \nabla w_k(x), h \rangle < 0$, выполняется, согласно определению 3, условие $w_k(x) \geq w_k(y)$, что ввиду (10) влечет $y \notin Y_J(x, \varepsilon)$.

Многогранный конус $\overline{H}(x) \cap H_J(x)$ можно представить в виде конической оболочки некоторого остова $B(x)$:

$$\begin{aligned} \overline{H}(x) \cap H_J(x) &= \{h \in \mathbb{R}^s | V(x)h \geq 0, W_J(x)h \geq 0\} = \\ &= \{h \in \mathbb{R}^s | h = B(x)\lambda, \lambda \geq 0\}, \end{aligned} \quad (16)$$

где для составленной из единичных вектор-столбцов матрицы

$$\begin{aligned} B(x) &= [b^1(x), b^2(x), \dots, b^r(x)], \\ \|b^j(x)\| &= 1, \quad 1 \leq j \leq r, \end{aligned} \quad (17)$$

существуют конструктивные методы построения [3].

Согласно включениям (13), (15), пересечение $\{X - \{x\}\} \cap H_J(x)$ можно аппроксимировать, выбрав из конуса $\overline{H}(x) \cap H_J(x)$ конечный «пучок» лучей.

Лемма 2. Если для малого $\rho > 0$ выполняются соотношения

$$\begin{aligned} \Delta(\Lambda_r, \Lambda) &= D(\Lambda_r, \Lambda) = \sup_{\alpha \in \Lambda_r} \inf_{\beta \in \Lambda} \|\alpha - \beta\| \leq \rho, \\ \Lambda_r &= \left\{ \lambda \in \mathbb{R}^s \mid \sum_{j=1}^r \lambda_j = 1, \lambda \geq 0 \right\}, \\ \Lambda &= \{\lambda^q\}_{q=1}^Q \subset \Lambda_r, \end{aligned} \quad (18)$$

где конечное множество Λ является ρ -аппроксимацией $(r - 1)$ -мерного стандартного симплекса Λ_r , то ввиду (16)–(18) вложенный «пучок» лучей

$$\begin{aligned} L_q(x) &= \left\{ q \in \mathbb{R}^s \mid g = \gamma \sum_{j=1}^r \lambda_j^q b^j(x), \gamma \geq 0 \right\} \subset \\ &\subset \overline{H}(x) \cap H_J(x), \quad 1 \leq q \leq Q, \end{aligned} \quad (19)$$

аппроксимирует выпуклый компакт $\{X - \{x\}\} \cap H_J(x)$ в следующем смысле.

Если можно указать величину $\delta > 0$ такую, что ненулевой вектор $h \in \mathbb{R}^s$ содержится в множестве $\{X - \{x\}\} \cap H_J(x)$ вместе с шаром

$$U_\delta(h) = \{g \in \mathbb{R}^s | \|h - g\| \leq \delta\} \subset \{X - \{x\}\} \cap H_J(x),$$

то в (19) найдется луч, удовлетворяющий соотношению $L_q(x) \cap U_\delta(h) \perp = \emptyset$ при условии $\rho r d_X \leq \delta$, где $d_X = \max_{x^1, x^2 \in X} \|x^1 - x^2\|$ — диаметр компакта X .

Следовательно, если «мишень» $X \cap Y_J(x, \varepsilon)$ конечного размера, то один из заданных соотношением (19) лучей $\{x\} + L_1(x)$ в нее попадет. Строгое утверждение состоит в следующем: при условии $\text{int}\{X \cap Y_J(x, \varepsilon)\} \neq \emptyset$ из включения (15) по лемме 2 следует, что в конечном множестве лучей (19) всегда можно указать луч $L_q(x)$ такой, что

$$\{\{x\} + L_q(x)\} \cap X \cap Y_J(x, \varepsilon) \neq \emptyset, \quad (20)$$

если конечное множество $\{\lambda^q\}_{q=1}^Q$ аппроксимирует стандартный симплекс Λ_r с требуемой точностью. Метод равномерной аппроксимации стандартного симплекса конечным множеством точек предложен в работе [4].

Но тогда хотя бы одна из проекций

$$p^q = \arg \min_{y \in \{\{x\} + L_q(x)\} \cap Y_J(x, \varepsilon)} \|x - y\|, \quad 1 \leq q \leq Q, \quad (21)$$

внешней точки $x \in X \setminus Y_J(x, \varepsilon)$ на пересечение $\{\{x\} + L_q(x)\} \cap Y_J(x, \varepsilon)$ удовлетворяет включению $p^q \in X \cap Y_J(x, \varepsilon)$, поскольку включения $x \in X$ и $p^q \in Y_J(x, \varepsilon)$ выполняются, так что отрицание $p^q \notin X$ противоречит утверждению (20) ввиду выпуклости множества X . Следовательно, по правилу А выбор опорной точки $y^J = p^q \in X \cap Y_J(x, \varepsilon)$ требует решения не более чем Q задач одномерного квадратичного программирования (21).

Если применение упрощенной процедуры на этом заканчивается, множество $\{y^J\}_{J \in M_t(x)}$, составленное из векторов, удовлетворяющих условиям (10), (11), может быть принято в качестве начального множества X_1 . Тем самым согласно правилу А по упрощенной процедуре через несколько шагов (число которых определяется спецификой решаемой задачи) будет построено начальное приближение $X_1 \subset \mathbb{R}^s$, образ которого — множество $w(X_1) \subset \mathbb{R}^m$ — может представлять собой грубую аппроксимацию множества Парето по всему «фронту».

6 Правило выбора шага универсальной процедуры

Если задано начальное множество X_1 , в согласии с результатом теоремы 1 универсальная процедура (4)–(9) обеспечивает построение последовательности множеств $\{w(X_t)\}_{t=1}^\infty$, аппроксимирующих множество Парето. Следует лишь принять правило выбора шага процедуры, для чего в соотношениях (6), (7) требуется установить конкретное

правило перехода от опорного решения $x \in X_t$ к новым опорным решениям

$$y^J \in X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon), \quad J \in M_t(x). \quad (22)$$

Сформулируем конструктивный способ выбора опорного решения (22).

Правило Б. Пусть множество допустимых решений — выпуклый компакт X — задано дифференцируемыми функциональными ограничениями. Согласно определению 4 и следствию 2, в каждой опорной точке $x \in X$ известна прямоугольная матрица $V(x)$ такая, что замыкание соответствующего конуса внутренних направлений $H(x)$ удовлетворяет условиям (13), а его пересечение с многогранным конусом (14) удовлетворяет условиям (15).

Если в правой части включения (22) пересечение множеств имеет внутренность $\text{int}\{X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon)\} \neq \emptyset$, из включения (15) по лемме 2 следует, что в конечном множестве лучей (19) можно указать луч $L_q(x)$ такой, что луч $\{x\} + L_q(x)$ содержит решение (22):

$$Z_J = \{\{x\} + L_q(x)\} \cap X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon) \neq \emptyset. \quad (23)$$

Найти опорное решение (22) на множестве (23) помогает

Лемма 3. Условие (23) в соответствии с леммой 1 влечет утверждение

$$\begin{aligned} \{p^q, g^q\} \cap X \cap Y_J(x, \varepsilon) \cap Y(x, \varepsilon) &\neq \emptyset, \\ p^q &= \arg \min_{y \in \{\{x\} + L_q(x)\} \cap Y_J(x, \varepsilon)} \|x - y\|, \\ g^q &= \arg \max_{y \in \{\{x\} + L_q(x)\} \cap X} \|x - y\|. \end{aligned}$$

Тем самым по правилу Б для определения опорной точки (22) достаточно проверить выполнение включения (22) не более чем для $2Q$ точек, выбранных на множестве, состоящем из Q лучей:

$$\begin{aligned} p^q, g^q &\in \{x\} + L_q(x), \quad 1 \leq q \leq Q, \\ p^q &= \arg \min_{y \in \{\{x\} + L_q(x)\} \cap Y_J(x, \varepsilon)} \|x - y\|, \\ g^q &= \arg \max_{y \in \{\{x\} + L_q(x)\} \cap X} \|x - y\|, \quad (24) \end{aligned}$$

где согласно соотношению (6) точка p^q есть ближайшая к опорной точке x точка отрезка $\{x\} + L_q(x) \cap Y_J(x, \varepsilon)$, тогда как g^q — наиболее удаленная от x точка отрезка $\{x\} + L_q(x) \cap X$.

7 Заключение

Задача синтеза численных методов аппроксимации множества Парето исследовалась на основе предложенной в работе [1] универсальной вычисли-

тельной процедуры при следующих предположениях: множество допустимых решений X — непустой выпуклый компакт (заданный, если потребуется, дифференцируемыми функциональными ограничениями), компонентами вектора частных критериев эффективности служат функции, псевдогогнутые в некоторой открытой окрестности X .

Разработанный численный метод аппроксимации множества Парето определяется: правилом А выбора начального приближения и правилом Б выбора каждого последующего опорного решения.

Построение каждой из точек начального приближения сводится по правилу А к решению одномерной задачи квадратичного программирования на каждом луче из конечного «пучка» лучей, аппроксимирующих многогранный конус, вложенный в конус возможных направлений.

По правилу Б на каждом этапе вычислительной процедуры переход от текущего опорного решения к последующему предполагает аппроксимацию (согласно включению (19)) многогранного конуса конечным множеством лучей $L_q(x)$, $1 \leq q \leq Q$, и выбор опорного решения среди конечного чис-

ла пар решений $\{p^q, g^q\} \subset \{x\} + L_q(x)$ в согласии с соотношениями (24).

Литература

1. Рабинович Я. И. Универсальная процедура построения множества Парето // Ж. вычисл. матем. матем. физ., 2017. Т. 57. № 1. С. 28–47.
2. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы / Пер. с англ. — М.: Мир, 1986. 583 с. (Bazaraa M. S., Shetty C. M. Nonlinear programming: Theory and algorithms. — New York, NY, USA: Wiley, 1979. 872 p.)
3. Линейные неравенства и смежные вопросы / Под ред. Г. У. Куна, А. У. Таккера; пер. с англ. — Annals of mathematics studies ser. — М.: ИЛ, 1959. 470 с. (Linear inequalities and related systems / Eds. H. W. Kuhn, A. W. Tucker. — Annals of mathematics studies ser. — Princeton, NJ, USA: Princeton University Press, 1956. 322 p.)
4. Рабинович Я. И. Численные методы оценивания приближенных решений в задачах многокритериальной оптимизации // Докл. Акад. наук, 2015. Т. 462. № 2. С. 151–153.

Поступила в редакцию 21.10.22

PROCEDURE OF CONSTRUCTING A PARETO SET FOR DIFFERENTIABLE CRITERIA FUNCTIONS

Ya. I. Rabinovich

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

Abstract: A ubiquitous computational procedure for the multicriteria optimization allows one to approximate the Pareto set under different requirements to the vector of particular efficiency criteria and the set of feasible solutions. In the paper, it is assumed that particular efficiency criteria are pseudoconcave in an open neighborhood of a compact convex set of feasible solutions which can be given by differentiable functional constraints. To build specific numerical methods for approximating the Pareto set, a rule for choosing the initial approximation and a rule for moving from the current reference solution to the next one are proposed.

Keywords: multicriteria optimization; Pareto set; numerical methods of approximation; universal procedure

DOI: 10.14357/19922264230403

EDN: NEZRGD

References

1. Rabinovich, Ya. I. 2017. Universal procedure for constructing a Pareto set. *Comp. Math. Math. Phys.* 57(1):45–63. doi: 10.1134/S0965542517010122.
2. Bazaraa, M. S., and C. M. Shetty. 1979. *Nonlinear programming: Theory and algorithms*. New York, NY: Wiley. 872 p.
3. Kuhn, H. W., and A. W. Tucker, eds. 1956. *Linear inequalities and related systems*. Annals of mathematics studies ser. Princeton, NJ: Princeton University Press. 322 p.
4. Rabinovich, Ya. I. 2015. Numerical methods for estimating approximate solutions of multicriteria optimization problems. *Dokl. Math.* 91(3):384–386. doi: 10.1134/S1064562415030114. EDN: UFAQXF.

Received October 21, 2022

Contributor

Rabinovich Yaacov I. (b. 1947) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; jacrabin@rambler.ru

НЕПАРАМЕТРИЧЕСКИЙ АЛГОРИТМ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

В. П. Тубольцев¹, А. В. Лапко², В. А. Лапко³

Аннотация: Предлагается непараметрический алгоритм автоматической классификации статистических данных большого объема. Эта задача возникает при обработке данных дистанционного зондирования природных объектов. Рассматриваемый алгоритм предполагает сжатие исходной информации на основе декомпозиции многомерного пространства признаков. В результате статистическая выборка большого объема преобразуется в массив данных, составленный из центров многомерных интервалов дискретизации и соответствующих им частот принадлежности случайных величин. Полученная информация используется при синтезе регрессионной оценки плотности вероятности. Под классом понимается компактная группа наблюдений случайной величины, соответствующая одномерному фрагменту плотности вероятности. На этой основе разрабатывается непараметрический алгоритм автоматической классификации, который основан на последовательной процедуре проверки близости центров многомерных интервалов дискретизации и соотношений между частотами принадлежности случайных величин из исходной выборки этим интервалам. Для повышения вычислительной эффективности предлагаемого алгоритма автоматической классификации используется многопоточный метод его программной реализации. Практическая значимость разработанного алгоритма автоматической классификации подтверждается результатами его применения при оценивании состояния лесных массивов по данным дистанционного зондирования.

Ключевые слова: автоматическая классификация; выборки большого объема; дискретизация области значений случайных величин; регрессионная оценка плотности вероятности; данные дистанционного зондирования

DOI: 10.14357/19922264230404

EDN: MPEWAW

1 Введение

Обнаружение компактных групп наблюдений — одна из основных задач системного анализа данных дистанционного зондирования объектов различной природы [1, 2].

Обзор методов автоматической классификации представлен в работах [3, 4]. Особое внимание уделяется разработке алгоритмов автоматической классификации для обнаружения компактных групп наблюдений, соответствующих одномерным фрагментам плотности вероятности признаков исследуемых объектов. Подобное определение класса было введено академиком Я. З. Цыпкиным [5] и развито в работах профессора В. И. Васильева [6] с использованием непараметрической оценки плотности вероятности Розенблатта—Парзена [7–10].

В этих условиях обоснована возможность решения задачи автоматической классификации путем итерационной процедуры последовательного не-

параметрического оценивания байесовского уравнения разделяющей поверхности между классами [11].

Предложенный подход развит при решении задачи автоматической классификации в условиях больших объемов статистических данных [12]. Его идея состоит в сжатии исходной информации путем декомпозиции пространства признаков в массив данных, составленный из центров многомерных интервалов дискретизации и соответствующих им частот принадлежности значений случайных величин. Основу этого подхода составляет анализ многомерной гистограммы признаков классифицируемых объектов.

Цель статьи состоит в совершенствовании непараметрического алгоритма автоматической классификации статистической информации большого объема на примере данных дистанционного зондирования путем использования регрессионной оценки плотности вероятности, применение которой

¹Сибирский государственный университет науки и технологий им. академика М. Ф. Решетнёва, vitalya.98@mail.ru

²Сибирский государственный университет науки и технологий им. академика М. Ф. Решетнёва; Институт вычислительного моделирования Сибирского отделения Российской академии наук, lapko@icm.krasn.ru

³Сибирский государственный университет науки и технологий им. академика М. Ф. Решетнёва; Институт вычислительного моделирования Сибирского отделения Российской академии наук, valapko@yandex.ru

позволяет сгладить многомерную гистограмму признаков классифицируемых объектов.

2 Оценивание плотности вероятности в условиях статистических данных большого объема

Имеется выборка

$$V = (x_{vi}, v = \overline{1, k}, i = \overline{1, n}), \quad (1)$$

составленная из статистически независимых наблюдений многомерной случайной величины $x = (x_v, v = \overline{1, k})$ размерности k большого объема n . Подобная информация, например, отражает данные дистанционного зондирования n элементов земной поверхности в пространстве спектральных признаков размерности k . Вычислительная эффективность непараметрических алгоритмов принятия решений ядерного типа во многом зависит от объема n статистических данных. При оценивании плотности вероятности $p(x)$ эта проблема решается путем сжатия исходной информации, заданной выборкой V вида (1), что предполагает выполнение следующих действий [10].

1. Разбить область определения плотности вероятности $p(x)$ случайной величины x по каждой ее компоненте x_v на N непересекающихся интервалов длиной $2\beta_v, v = \overline{1, k}$. Для выбора числа N интервалов дискретизации можно использовать результаты исследований [10, 13, 14].
2. Определить оценки вероятностей попадания элементов выборки V в каждый j -й многомерный интервал

$$\overline{P}_j = \frac{n(j)}{n}, \quad j = \overline{1, N^k},$$

где $n(j)$ — число наблюдений из выборки V , принадлежащих многомерному интервалу $D(j) = [z_v \pm \beta_v, v = \overline{1, k}]$, а $z_j = (z_{vj}, v = \overline{1, k})$ — координаты центра j -го интервала.

3. Положить, что в каждом многомерном интервале $D(j), j = \overline{1, N^k}$, случайная величина $x = (x_v, v = \overline{1, k})$ имеет равномерный закон распределения. С этих позиций вычислить оценки плотности вероятности

$$\overline{p}(x_j) = \overline{p}_j = \frac{\overline{P}_j}{\prod_{v=1}^k (2\beta_v)} \quad \forall x \in D(j), \quad j = \overline{1, N^k}.$$

4. На основе полученной информации сформировать массив данных

$$\overline{V} = (z_j, \overline{P}_j, j = \overline{1, N^k}) \quad (2)$$

и осуществить синтез регрессионной оценки плотности вероятности $p(x)$ [10]:

$$\overline{p}(x) = \frac{1}{\prod_{v=1}^k c_v} \sum_{j=1}^{N^k} \overline{P}_j \prod_{v=1}^k \Phi\left(\frac{x_v - z_{vj}}{c_v}\right). \quad (3)$$

Здесь ядерные функции $\Phi(u_v)$ удовлетворяют условиям [8, 10]:

$$\begin{aligned} \Phi(u_v) &= \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty, \\ \int_{-\infty}^{\infty} \Phi(u_v) du_v &= 1, \quad \int_{-\infty}^{\infty} u_v^2 \Phi(u_v) du_v = 1, \\ \int_{-\infty}^{\infty} u_v^m \Phi(u_v) du_v &< \infty, \quad 0 \leq m < \infty, \quad v = \overline{1, k}. \end{aligned}$$

Коэффициенты размытости c_v ядерных функций $\Phi(u_v)$ в статистике (3) убывают с ростом N [10]. Значения c_v зависят от длины интервала изменения случайных величин $x_v, v = \overline{1, k}$. Поэтому будем полагать, что $c_v = c\overline{\sigma}_v$, где $\overline{\sigma}_v$ — оценка среднеквадратичного отклонения случайной величины $x_v, v = \overline{1, k}$. Значения $\overline{\sigma}_v$ оцениваются по данным массива \overline{V} .

Выбор оптимального значения c параметра c определяется из условия минимума среднеквадратичной ошибки аппроксимации регрессионной оценкой (3) плотности вероятности $p(x)$:

$$W(c) = \frac{1}{N^k} \sum_{i=1}^{N^k} (\overline{p}_i - \overline{p}(z_i))^2.$$

При вычислении $\overline{p}(z_i)$ значение z_i исключается из статистики $\overline{p}(x)$.

Процедура оптимизации регрессионной оценки плотности вероятности случайной величины x повышает вычислительную эффективность процедуры выбора коэффициентов размытости ядерных функций по сравнению с традиционной методикой [10].

Традиционная методика основана на выборе коэффициента размытости ядерной функции из условия минимума оценки среднеквадратического отклонения $\overline{p}(x)$ от $p(x)$ [10]:

$$W_1(c) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \bar{p}^2(x_1, \dots, x_k) dx_1 \dots dx_k - 2 \sum_{i=1}^{N^k} \bar{P}_i p(x_{1i}, \dots, x_{ki}),$$

где

$$\bar{p}(x_{1i}, \dots, x_{ki}) = \frac{1}{\prod_{v=1}^k c_v} \sum_{\substack{j=1 \\ j \neq i}}^{N^k} \bar{P}_j \prod_{v=1}^k \Phi\left(\frac{x_{vi} - z_{vj}}{c_v}\right).$$

Нетрудно заметить, что процедура вычисления критерия $W(c)$ более проста по сравнению с критерием $W_1(c)$.

Из условия минимума среднеквадратичного отклонения $\bar{p}(x)$ от $p(x)$ в работах [10, 13] определено оптимальное число N интервалов дискретизации области значений одномерной случайной величины x , которое соответствует целому числу выражения

$$\left(n \Delta \int_{-\infty}^{\infty} p^2(x) dx \right)^{1/2}. \quad (4)$$

Значение N определяется видом плотности вероятности и не зависит от ее параметров. Для равномерного закона распределения одномерной случайной величины с плотностью вероятности $p(x)$ выражение для N совпадает с формулой дискретизации Хайнхольда–Гаеде [15]. В этих условиях выражение

$$\Delta \int_a^b p^2(x) dx = 1$$

не зависит от конечных пределов интегрирования a и b ($\Delta = b - a$). При нормальном законе распределения

$$\Delta \int_a^b p^2(x) dx = 1,693,$$

где $a = M(x) - 3\sigma$; $b = M(x) + 3\sigma$. Здесь $M(x)$ — математическое ожидание случайной величины a ; σ — среднеквадратичное отклонение x . Вероятность попадания случайной величины x в эти пределы равна 0,997.

В работе [10] обоснована возможность оценивания произведения Δ на интеграл от квадрата плотности вероятности одномерной случайной величины в выражении (4) по значению коэффициента контрэкссесса. На этой основе предложена формула оценивания числа интервалов дискретизации

многомерной случайной величины $x = (x_v, v = \overline{1, k})$.

Регрессионная оценка плотности вероятности (3) лежит в основе синтеза непараметрического алгоритма автоматической классификации в условиях большого объема статистических данных.

3 Алгоритм автоматической классификации

Пусть имеется выборка многомерных статистических данных V (1) большого объема n , распределенная с неизвестной плотностью вероятности $p(x) = p(x_1, \dots, x_k)$. Необходимо выборку V разбить на группы компактных наблюдений V_j , $j = \overline{1, M}$, число M которых неизвестно. Под компактной группой наблюдений (классом) случайной величины x будем подразумевать область ее значений, которая соответствует одномодальному фрагменту многомерной плотности вероятности $p(x)$ [11, 12].

Следуя методике синтеза регрессионной оценки плотности вероятности, определим статистику $\bar{p}(x)$ по формуле (3). Для обнаружения наблюдений первого класса по информации массива данных \bar{V} вида (2) выберем элемент (z_j, \bar{p}_j) , который определяется условием:

$$\bar{p}_j = \max_{i=1, N^k} \bar{p}_i. \quad (5)$$

Этот элемент (z_j, \bar{p}_j) соответствует максимальному значению оценки плотности вероятности $\bar{p}(x)$. Тогда начальный этап формирования элементов массива данных \bar{V}_1 из \bar{V} , принадлежащих первому классу Ω_1 , определяется правилом:

$$D(j) \subset \Omega_1, \text{ если } \prod_{v=1}^k \Phi\left(\frac{z_{vi} - z_{vj}}{c_v}\right) 1(\bar{p}_i, \bar{p}_j) > 0, \\ i = \overline{1, N^k}, i \neq j. \quad (6)$$

Здесь, например, ядерная и индикаторная функции имеют вид:

$$\Phi\left(\frac{z_{vi} - z_{vj}}{c_v}\right) = \begin{cases} 0,5, & \text{если } |z_{vi} - z_{vj}| \leq c_v; \\ 0, & \text{если } |z_{vi} - z_{vj}| > c_v; \end{cases} \\ 1(\bar{p}_i, \bar{p}_j) = \begin{cases} 1, & \text{если } \bar{p}_i \leq \bar{p}_j; \\ 0, & \text{если } \bar{p}_i > \bar{p}_j, \end{cases}$$

где $c_v \geq 2\beta_v$, так как при конкретных значениях c_v ядерные функции определяют элементы массива \bar{V} , смежные с элементом (z_j, \bar{p}_j) . Индикаторная функция $1(\bar{p}_i, \bar{p}_j)$ позволяет определять элементы (z_i, \bar{p}_i)

массива данных \bar{V} , которые характеризуются убывающими значениями плотности вероятности при изменении значений x в окрестности z_j . На этом этапе индикаторная функция принимает значение $1(\bar{p}_i, \bar{p}_j) = 1$, так как выполняется условие (5).

Элементы выборки V , принадлежащие интервалу $D(j)$, включаются в выборку наблюдений V_1 первого класса. Данное правило сохраняется и для последующих этапов классификации.

Обозначим через I_1 множество номеров элементов массива данных \bar{V} , отнесенных на первом этапе автоматической классификации в соответствии с правилом (6) к первому классу Ω_1 . Множество I_1 включает номер j элемента (z_j, \bar{p}_j) .

Второй этап классификации реализует решающее правило

$$D(i) \subset \Omega_1, \\ \text{если } \sum_{j \in I_1} \prod_{v=1}^k \Phi \left(\frac{z_{vi} - z_{vj}}{c_v} \right) 1(\bar{p}_i, \bar{p}_j) > 0, \\ i \in I \setminus I_1, I = \overline{1, N^k}, j \in I_1. \quad (7)$$

По аналогии на t -м этапе автоматической классификации элементы массива данных \bar{V} , принадлежащие первому классу Ω_1 , определяются правилом:

$$D(i) \subset \Omega_1, \\ \text{если } \sum_{j \in I_{t-1}} \prod_{v=1}^k \Phi \left(\frac{z_{vi} - z_{vj}}{c_v} \right) 1(\bar{p}_i, \bar{p}_j) > 0, \\ i \in I \setminus (I_1 \cup I_2 \cup \dots \cup I_{t-1}), j \in I_{t-1}. \quad (8)$$

Последовательная процедура автоматической классификации (6)–(8) продолжается до выполнения условия $I_t = \emptyset$, когда прекратится обнаружение новых элементов массива данных \bar{V} из класса Ω_1 .

Обнаруженный первый класс характеризуется регрессионной оценкой плотности вероятности

$$\bar{p}_1(x) = \frac{1}{\prod_{v=1}^k c_v} \sum_{i \in I_{t-1}} \bar{P}_i \prod_{v=1}^k \Phi \left(\frac{x_v - z_{vi}}{c_v} \right),$$

которая восстанавливается по данным \bar{V}_1 . В этом случае оптимальные коэффициенты размытости c_v соответствуют значениям $\bar{c}_v = \bar{c} \bar{\sigma}_v, v = \overline{1, k}$. Массив данных $\bar{V}_1 \in \bar{V}$ определяется процедурами автоматической классификации, заданными соотношениями (6)–(8).

При обнаружении элементов выборки V_1 исходной статистической информации V , принадлежащих первому классу Ω_1 , используются результаты автоматической классификации массива данных \bar{V} .

Для обнаружения элементов массива $\bar{V}_2 \in \bar{V}$, принадлежащих второму классу Ω_2 , из оставшихся данных $\bar{V} \setminus \bar{V}_1$ выбирается элемент (z_j, \bar{p}_j) с максимальным значением \bar{p}_j и предложенная выше процедура автоматической классификации используется для определения класса Ω_2 . По аналогии формируются оставшиеся классы, число которых априори не определено.

4 Оценивание состояния лесных массивов, поврежденных полиграфом, по данным дистанционного зондирования

Территория исследования определялась юго-западной частью Дзержинского района Красноярского края. На этой территории преобладают пихтовые, кедровые и еловые древостои, встречаются березы и осины. Высота территории варьирует в диапазоне от 370 до 610 м над уровнем моря.

Исходная информация формировалась по данным дистанционного зондирования аппаратом Европейского космического агентства Sentinel-2A за 4 августа 2018 г. и 23 мая 2022 г. Снимки получены с геопортала Earth Explorer, из которых вырезаны тестовые участки в 6000 га (рис. 1). Каждый из них определяется 605 414 пикселями. Каждый пиксель характеризуется шестью спектральными признаками $x = (x_1, \dots, x_6)$, которым соответствуют длины волн (нм): 492,7 (x_1); 559,8 (x_2); 664,6 (x_3); 740,5 (x_4); 832,8 (x_5); 2202,4 (x_6). На рис. 1, а и 1, б представлены RGB-изображения исходных данных за 2018 и 2022 г. соответственно. Каналам R, G и B соответствуют признаки x_6, x_5 и x_3 .

В центральной части исследуемой территории располагались лесные массивы усохших темнохвойных древостоев, поврежденных полиграфом уссурийским. Их отличительная особенность — куртинный характер повреждений. Они отображаются на рисунке ярким фиолетовым цветом и имеют гладкие границы, приближенные к овалу или кругу. Ярким розовым цветом отображаются сплошные рубки лесных насаждений, которые отличаются правильной формой, близкой к прямоугольнику. Части изображения в темных оттенках фиолетового и зеленого соответствуют хвойным лесным насаждениям. Фрагменты рисунка зеленого цвета, от темных к светлым тонам, соответствуют лиственным древостоям и травяно-кустарниковым сообществам. На изображении отчетливо различаются дороги. Они представляют собой линии различной толщины с ярким зеленым или желтым оттенком.

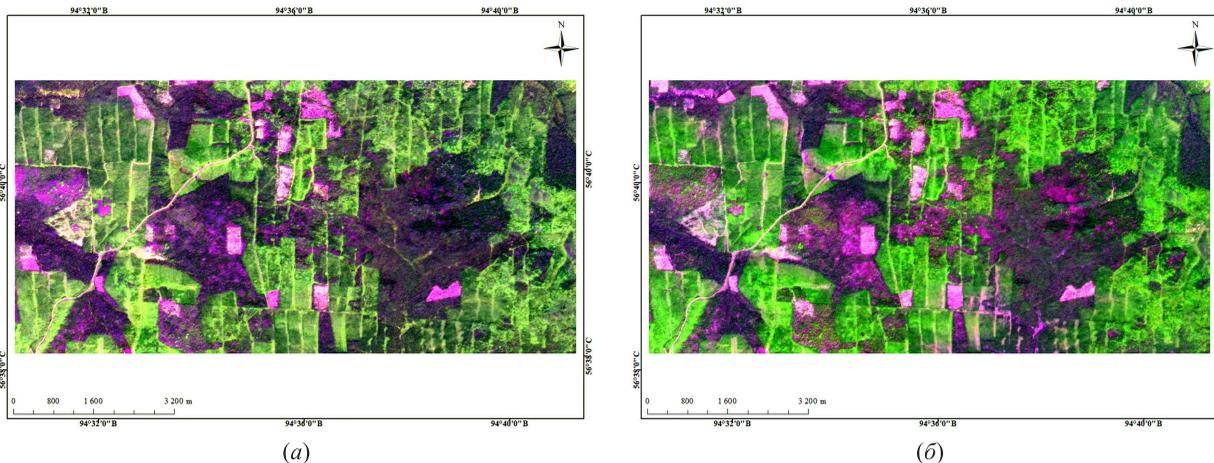


Рис. 1 Отображение исходных данных, составленных из материалов дистанционного зондирования космическим аппаратом Sentinel-2A за 4 августа 2018 г. (а) и 23 мая 2022 г. (б)

Для обнаружения компактных групп наблюдений (классов) в пространстве спектральных признаков $x = (x_1, \dots, x_6)$ использовался метод анализа данных ISODATA (Iterative Self-Organizing Data Analysis) и предлагаемый непараметрический алгоритм автоматической классификации. Его программная реализация NAC v.2.0 позволяет выполнять следующие функции: загружать изображения в формате GeoTIFF; проводить классификацию с возможностью выбора метода расчета числа интервалов дискретизации исходного пространства признаков; получать изображение для дальнейшей геообработки [16]. В этой программе реализованы методы многопоточных вычислений, которые повышают скорость обработки данных большого объема.

Применение программы NAC v.2.0. позволило в исходных изображениях за 2018 и 2022 гг. обнаружить соответственно 84 и 177 компактных групп точек (рис. 2). При выборе метода расчета числа интервалов дискретизации был установлен флажок «агрегация», что позволило избежать в итоговом классифицированном изображении большого числа классов. С увеличением порядкового номера компактной группы наблюдений число пикселей, ей принадлежащих, уменьшается. Это можно объяснить тем, что алгоритм основан на последовательном определении класса как одномодального фрагмента плотности вероятности спектральных признаков исследуемого объекта. Обнаруженные классы соответствуют лесным массивам с различной степенью поражения полиграфом уссурий-

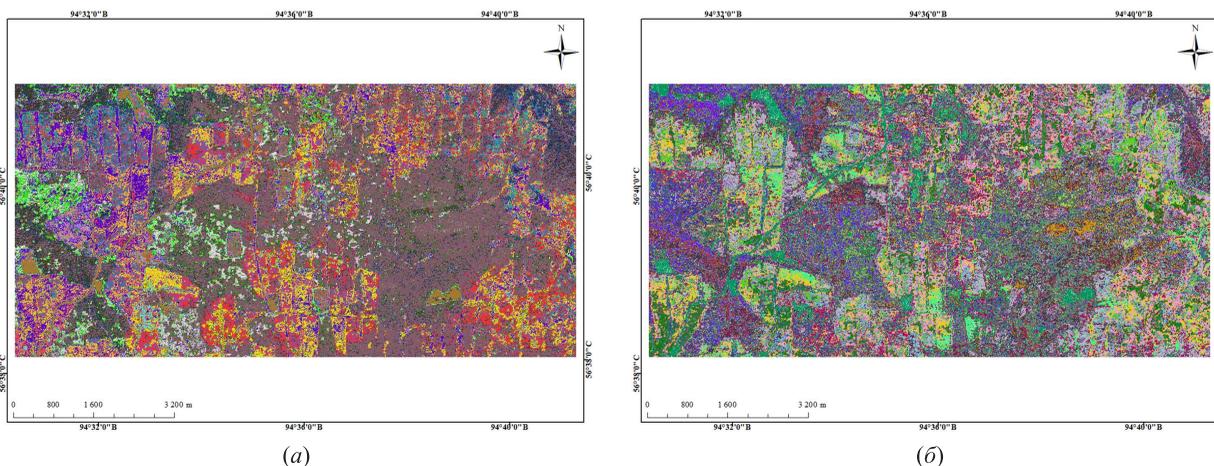


Рис. 2 Отображение результатов классификации непараметрическим алгоритмом автоматической классификации исходных данных за 4 августа 2018 г. (а) и 23 мая 2022 г. (б)

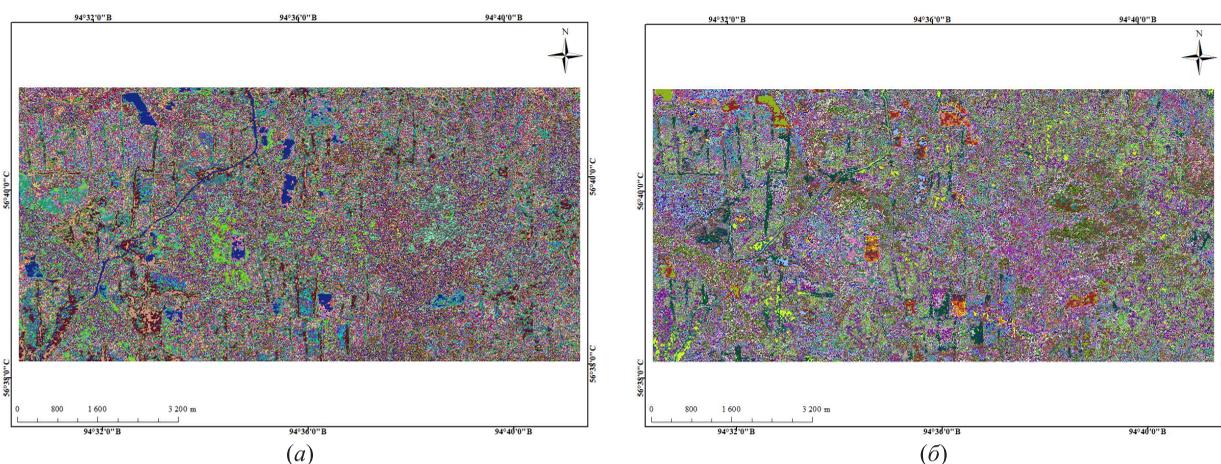


Рис. 3 Отображение результатов автоматической классификации алгоритмом ISODATA спектральных данных, полученных 4 августа 2018 г. (а) и 23 мая 2022 г. (б)

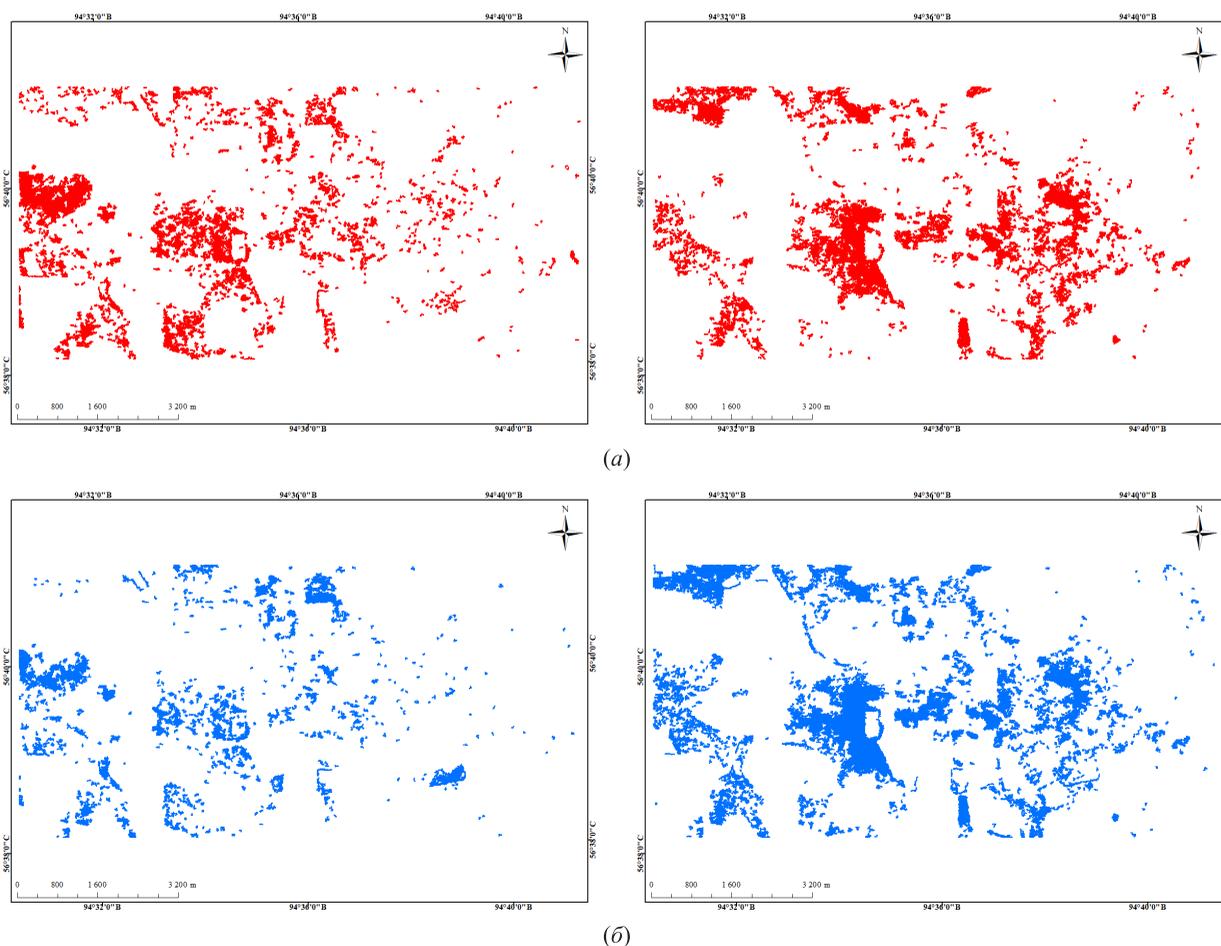


Рис. 4 Отображение классов, наиболее точно характеризующих пиксели, соответствующие лесным насаждениям различной степени повреждения полиграфом уссурийским. Результаты классификации по данным 4 августа 2018 г. (левый столбец) и 23 мая 2022 г. (правый столбец) предлагаемым непараметрическим алгоритмом (а) и алгоритмом ISODATA (б)

ским, усохшим и лиственным древостоям, травяно-кустарниковым сообществам и вырубкам различной давности.

Применение алгоритма ISODATA, реализованного в программном продукте ArcGIS ArcMap, требует указания необходимого числа классов. При классификации исходных изображений за 2018 и 2022 гг. алгоритмом ISODATA число классов принималось равным 84 и 177 соответственно. Это число классов было обнаружено при использовании программы NAC v.2.0. В результате обработки данных 2018 и 2022 гг. алгоритмом ISODATA были выявлены только 73 и 128 классов соответственно (рис. 3). В обоих результатах число пикселей в классах распределено равномерно и не зависит от порядкового номера класса.

По результатам классификации экспертами определены классы (рис. 4), которые характеризуют пиксели, соответствующие лесным насаждениям с разной степенью повреждения полиграфом уссурийским.

Полученные классы позволили рассчитать площадь повреждений полиграфом уссурийским лесных насаждений исследуемой территории. Программа NAC v.2.0 позволила определить площадь повреждений в 2018 и 2022 гг. в размере 503,5 и 635,0 га соответственно. При использовании алгоритма ISODATA поврежденные площади составили 323,6 и 795,4 га. Пространственное распределение обнаруженных классов рассматриваемыми методами автоматической классификации различается на участках изображения, характеризующих вырубки, переход от поврежденных лесных насаждений к здоровым и открытые почвы. Например, на изображении 2018 г. в центральной части алгоритм ISODATA объединил в один класс пиксели, характеризующие все темнохвойные насаждения и поврежденные лесные массивы. В юго-восточной части изображения этот алгоритм отнес вырубку к поврежденным древостоям. При этом пиксели рассматриваемого класса характеризуют поврежденные насаждения полиграфом уссурийским в северной части изображения.

Экспертный анализ показал, что предлагаемый метод автоматической классификации обладает преимуществом по сравнению с алгоритмом ISODATA, что следует из анализа рис. 2–4. Предлагаемый алгоритм автоматической классификации достаточно полно выделяет зоны поврежденных древостоев, разделяет темнохвойные и лиственные породы, определяет участки с открытой почвой, территории вырубок различной давности.

Алгоритм ISODATA в этих условиях показал схожий результат классификации. Отличия наблюдаются на участках вырубок и поврежденных лесных

насаждений, что вызывает разницу в их итоговой расчетной площади.

Сравниваемые алгоритмы определили одинаковые участки изображения, соответствующие поврежденным лесным насаждениям. По результатам автоматической классификации алгоритмом ISODATA и непараметрическим классификатором рассчитаны площади территорий поврежденных древостоев полиграфом уссурийским (см. рис. 4). По данным 2022 г. они составили 795,4 и 635,0 га лесных массивов соответственно, что указывает на их различие в 25%.

5 Заключение

Разработанный непараметрический алгоритм автоматической классификации статистических данных большого объема основан на их сжатии путем декомпозиции многомерного пространства признаков исследуемых объектов и алгоритмизации традиционной процедуры классификации. В предлагаемом алгоритме автоматической классификации использование многомерной гистограммы заменено на анализ регрессионной оценки плотности вероятности случайных величин. Его применение позволяет обнаруживать классы, соответствующие одномерным фрагментам плотности вероятности. Использование многопоточной технологии обработки данных позволяет в 2 раза сократить время автоматической классификации, что подтверждается результатами обработки спектральных данных дистанционного зондирования лесных массивов. Установлены условия преимущества непараметрического алгоритма по сравнению с методом ISODATA.

Литература

1. Abbas A. W., Minallh N., Ahmad N., Abid S. A. R., Khan M. A. A. K-means and ISODATA clustering algorithms for landcover classification using remote sensing // Sindh University Research J. (Science Series), 2016. Vol. 48. No. 2. P. 315–318.
2. Manthena N. R., Kumaran N., Chandra S. V. Remote sensing image classification using CNN–LSTM model // Revue d'Intelligence Artificielle, 2022. Vol. 36. No. 1. P. 147–153. doi: 10.18280/ria.360117.
3. Дорофеев А. А. Алгоритмы автоматической классификации (обзор) // Автоматика и телемеханика, 1971. № 12. С. 78–113.
4. Дорофеев А. А. Методология экспертно-классификационного анализа в задачах управления и обработки сложноорганизованных данных (история и перспективы развития) // Проблемы управления, 2009. № 3.1. С. 19–28. EDN: KJUOIN.

5. Цыпкин Я. З. Основы теории обучающихся систем. — М.: Наука, 1970. 252 с.
6. Васильев В. И., Эш С. Н. Особенности алгоритмов самообучения и кластеризации // Управляющие системы и машины, 2011. № 3. С. 3–9.
7. Parzen E. On estimation of a probability density function and mode // Ann. Math. Stat., 1962. Vol. 33. No. 3. P. 1065–1076. doi: 10.1214/aoms/1177704472.
8. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее применения, 1969. Т. 14. № 1. С. 156–161.
9. Тарарушкин Е. В. Восстановление плотности распределения частиц дисперсных материалов методом окна Парзена–Розенблатта // Вестник МГСУ, 2018. Т. 13. Вып. 7(118). С. 855–862. doi: 10.22227/1997-0935.2018.7.855-862. EDN: UVNCCV.
10. Лапко А. В., Лапко В. А. Ядерные оценки плотности вероятности и их применение. — Красноярск: СибГУ им. М. Ф. Решетнева, 2021. 308 с.
11. Лапко А. В., Лапко В. А. Непараметрический алгоритм автоматической классификации в условиях статистических данных большого объема // Информатика и системы управления, 2018. Т. 57. № 3. С. 59–70. doi: 10.22250/isu.2018.57.59-70.
12. Зеньков И. В., Лапко А. В., Лапко В. А., Им С. Т., Тубольцев В. П., Авдеенок В. Л. Непараметрический алгоритм автоматической классификации многомерных статистических данных большого объема и его применение // Компьютерная оптика, 2021. Т. 45. № 2. С. 253–260. doi: 10.18287/2412-6179-CO-801. EDN: WUOYYA.
13. Scott D. W. Multivariate density estimation: Theory, practice, and visualization. — Hoboken, NJ, USA: John Wiley & Sons, 2015. 384 p.
14. Fushimi T., Saito K., Motoda H. Constructing outlier-free histograms with variable bin-width based on distance minimization // Intell. Data Anal., 2023. Vol. 27. No. 1. P. 5–29.
15. Heinhold I., Gaede K. W. Ingenieur statistic. — München, Wien: Springer-Verlag, 1964. 352 p.
16. Лапко А. В., Лапко В. А., Им С. Т., Тубольцев В. П., Авдеенок В. Л. Программа автоматической классификации данных дистанционного зондирования Земли на основе непараметрических алгоритмов принятия решений (NAC v. 2.0). Свидетельство о государственной регистрации программы для ЭВМ № 2022619023 от 18.05.2022.

Поступила в редакцию 16.01.23

NONPARAMETRIC ALGORITHM FOR AUTOMATIC CLASSIFICATION OF REMOTE SENSING DATA

V. P. Tuboltsev¹, A. V. Lapko^{1,2}, and V. A. Lapko^{1,2}

¹M. F. Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russian Federation

²Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk 660036, Russian Federation

Abstract: A nonparametric algorithm for automatic classification of large-volume statistical data is proposed. The algorithm under consideration assumes compression of initial information based on decomposition of multidimensional feature space. As a result, a large statistical sample is transformed into a data array composed of the centers of multidimensional sampling intervals and their corresponding frequencies of random variables. The information obtained is used in the synthesis of the regression estimate of the probability density. A class is understood as a compact group of observations of a random variable corresponding to a unimodal fragment of the probability density function. On this basis, a nonparametric automatic classification algorithm is developed which is based on the sequential procedure for checking the proximity of the centers of multidimensional sampling intervals and the ratios between the frequencies of belonging of random variables from the original sample to these intervals. To improve the computational efficiency of the proposed automatic classification algorithm, a multithreaded method of its software implementation is used. The practical significance of the developed algorithm for automatic classification is confirmed by the results of its application for assessing the state of the forests areas using remote sensing data.

Keywords: automatic classification; large-volume samples; sampling of the range of values of random variables; regression estimation of probability density; remote sensing data

DOI: 10.14357/19922264230404

EDN: MPEWAW

References

1. Abbas, A. W., N. Minallh, N. Ahmad, S. A. R. Abid, and M. A. A. Khan. 2016. K-means and ISODATA clustering algorithms for landcover classification using remote sensing. *Sindh University Research J. (Science Series)* 48(2):315–318.
2. Manthena, N. R., N. Kumaran, and S. V. Chandra. 2022. Remote sensing image classification using CNN–LSTM

- model. *Revue d'Intelligence Artificielle* 36(1):147–153. doi: 10.18280/ria.360117.
3. Dorofeyuk, A. A. 1971. Algoritmy avtomaticheskoy klassifikatsii (obzor) [Algorithms of automatic classification (review)]. *Automat. Rem. Contr.* 12:78–113.
 4. Dorofeyuk, A. A. 2009. Metodologiya ekspertno-klassifikatsionnogo analiza v zadachakh upravleniya i obrabotki slozhnoorganizovannykh dannykh (istoriya i perspektivy razvitiya) [Expert-rangin analysis methodology in complex organized data processing and control problems (history of development and perspectives)]. *Problemy upravleniya* [Control Sciences] 3S1:19–28. EDN: KJUOIN.
 5. Tsyppin, Ya. Z. 1970. *Osnovy teorii obuchayushchikhsya sistem* [Foundations of the theory of learning systems]. Moscow: Nauka. 252 p.
 6. Vasil'ev, V. I., and S. N. Esh. 2011. Osobennosti algoritmov samoobucheniya i klasterizatsii [Features of self-learning algorithms and clustering]. *Upravlyayushchie sistemy i mashiny* [Control Systems and Computers] 3:3–9.
 7. Parzen, E. 1962. On estimation of a probability density function and mode. *Ann. Math. Stat.* 33(3):1065–1076. doi: 10.1214/aoms/1177704472.
 8. Epanechnikov, V. A. 1969. Non-parametric estimation of a multivariate probability density. *Theor. Probab. Appl.* 14(1):153–158. doi: 10.1137/1114019.
 9. Tararushkin, E. V. 2018. Vosstanovlenie plotnosti raspredeleniya chastits dispersnykh materialov metodom okna Parzena–Rozenblatta [Reconstructing distribution density of particles for disperse materials by the Parzen–Rozenblatt window method]. *Vestnik MGSU* 13(7):855–862. doi: 10.22227/1997-0935.2018.7.855-862. EDN: UVNCCV.
 10. Lapko, A. V., and V. A. Lapko. 2021. *Yadernye otsenki plotnosti veroyatnosti i ikh primenenie* [Kernel probability density estimates and their application]. Krasnoyarsk: Reshetnev University Pubs. 308 p.
 11. Lapko, A. V., and V. A. Lapko. 2018. Neparame-tricheskiy algoritm avtomaticheskoy klassifikatsii v usloviyakh statisticheskikh dannykh bol'shogo ob"ema [Nonparametric algorithm of automatic classification under conditions of large-scale statistical data]. *Informatika i sistemy upravleniya* [Information Science and Control Systems] 57(3):59–70. doi: 10.22250/isu.2018.57.59-70. EDN: YACMRN.
 12. Zenkov, I. V., A. V. Lapko, V. A. Lapko, S. T. Im, V. P. Tuboltsev, and V. L. Avdeenok. 2021. A nonparametric algorithm for automatic classification of large multivariate statistical data sets and its application. *Computer Optics* 45(2):253–260. doi: 10.18287/2412-6179-CO-801. EDN: WUOYYA.
 13. Scott, D. W. 2015. *Multivariate density estimation: Theory, practice, and visualization*. Hoboken, NJ: John Wiley & Sons. 384 p.
 14. Fushimi, T., K. Saito, and H. Motoda. 2023. Constructing outlier-free histograms with variable bin-width based on distance minimization. *Intell. Data Anal.* 27(1):5–29.
 15. Heinhold, I., and K. W. Gaede. 1964. *Ingenieur statistik*. München, Wien: Springer Verlag. 352 p.
 16. Lapko, A. V., V. A. Lapko, S. T. Im, V. P. Tuboltsev, and V. L. Avdeenok. 2022. Programma avtomaticheskoy klassifikatsii dannykh distantsionnogo zondirovaniya Zemli na osnove neparame-tricheskikh algoritmov prinyatiya resheniy (NAC v. 2.0) [The program for automatic classification of Earth remote sensing data based on nonparametric decision-making algorithms (NAC v. 2.0)]. Certificate of State Registration of the Computer Program RF No. 2022619023.

Received January 16, 2023

Contributors

Tuboltsev Vitaly P. (b. 1998) — PhD student, M. F. Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russian Federation; vitalya.98@mail.ru

Lapko Alexander V. (b. 1949) — Doctor of Science in technology, professor, Honored Scientist of the RF, principal scientist, Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk 660036, Russian Federation; professor, Department of Space Facilities and Technologies, M. F. Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russian Federation; lapko@icm.krasn.ru

Lapko Vasilij A. (b. 1974) — Doctor of Science in technology, professor, leading scientist, Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk 660036, Russian Federation; head of department, M. F. Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russian Federation; valapko@yandex.ru

ТЕХНОЛОГИЯ МНОГОФАКТОРНОЙ КЛАССИФИКАЦИИ МАТЕМАТИЧЕСКОГО КОНТЕНТА ЭЛЕКТРОННОЙ СИСТЕМЫ ОБУЧЕНИЯ*

А. В. Босов¹, А. В. Иванов²

Аннотация: Статья продолжает исследование задачи классификации контента электронной обучающей системы (ЭОС). Разработанная ранее технология тематической классификации математического контента, содержащегося в блоках задач и примеров ЭОС, усовершенствована и дополнена новыми функциями. Для этого использованная ранее модель контента с двумя свойствами — традиционной текстовой формулировкой и формульной частью в формате $\text{T}_\text{E}\text{X}$ — дополнена рядом дополнительных числовых атрибутов, таких как наличие трансцендентных и производных функций, число формул в задании. Этот блок атрибутов позволил повысить качество имеющегося тематического классификатора и реализовать два новых, а именно: в дополнение к тематической классификации задач реализована возможность автоматического определения, во-первых, сложности задания и, во-вторых, набора компетенций, формируемых у обучающегося заданием. Такая многофакторная классификация представляется важным этапом перспективного направления развития ЭОС — автоматизированного оценивания качества образовательного контента. Проверка работоспособности предлагаемых алгоритмов, обучение классификаторов и анализ качества классификации выполнены по той же дисциплине теории функций комплексного переменного, но на существенно расширенном материале, включая задания для самостоятельной работы студентов — расчетные и экзаменационные работы.

Ключевые слова: электронная обучающая система; математический контент; машинное обучение; многофакторная классификация; оценка качества контента

DOI: 10.14357/19922264230405

EDN: LISHHZ

1 Введение

Электронные обучающие системы используются повсеместно, достаточно давно и стали привычными, в том числе и в качестве средств дистанционного обучения. Несмотря на их разнообразие, от простого электронного учебника до интегрированной обучающей среды вуза, неотъемлемой и наиболее ответственной частью ЭОС остается учебный контент, т. е. сам материал, предоставляемый системой учащемуся и преподавателю. И актуальным был и остается вопрос качества этого контента. В статье не затрагивается «ручной» экспертный анализ качества учебного контента, в фокусе — вопрос автоматического анализа. Этот вопрос довольно сложный и со временем вполне возможно сформирует самостоятельное направление [1, 2]. На данный момент ясно, что здесь есть богатый источник прикладных задач для методов машинного обучения, что продемонстрировано в предыдущих работах по этой тематике [3, 4]. В этих работах подтверждена принципиальная возможность авто-

матизировать тематическую классификацию математических задач. По комбинации математических формул, размеченных в формате $\text{T}_\text{E}\text{X}$, и традиционных текстовых формулировок заданий оказалось возможным очень эффективно определять тематику задания. С исследовательской позиции полезной оказалась смешанная модель заданий, объединяющая формулы и текст. С прикладной — средство, автоматически определяющее тему (раздел курса) конкретного задания. Это средство позволяет сделать первый шаг в автоматизации анализа качества обучающего контента, позволив автоматизировать определение объемных и сравнительных показателей распределения элементов контента по темам, делать выводы о достаточности примеров и самостоятельных заданий и т. п.

На основе этой технологии в данной статье исследуется возможность анализа других показателей. При классификации по более сложным, чем тематика, показателям обнаружилось недостаточное качество модели. Поэтому она была расшире-

* Исследование выполнено за счет гранта Российского научного фонда № 22-28-00588 (<https://rscf.ru/project/22-28-00588/>). Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика» ФИЦ ИУ РАН, Москва).

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

² Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, aivanov@ipiran.ru

на третьим блоком свойств, уточняющих разметку формульной части. Уточненной модели посвящен следующий раздел статьи. Здесь же обсуждается и главный методический вопрос — определение показателей для классификации. Самый простой — атрибут сложности задания — не вызывает сомнений. Как и в постановке с тематикой, этот атрибут позволит определять понятные объемные и сравнительные показатели распределения элементов контента по уровням сложности. Для содержательного анализа было использовано понятие компетентностного подхода в образовании, очень популярного и активно развиваемого, в том числе и с целями определения содержания образования и оценки образовательных результатов [5, 6]. Применительно к имеющемуся в распоряжении контенту — заданиям по курсу теории функций комплексного переменного — эта идея свелась к разметке заданий перечнем формируемых компетенций. Полный перечень компетенций нетрудно определить в соответствии с федеральными государственными образовательными стандартами высшего образования.

Задача классификации, которую надо решать, — это типичная задача машинного обучения. Принципиальное условие успеха — качество данных, используемых для обучения алгоритмов. Если предыдущая задача тематической классификации успешно решалась на данных, извлеченных по большей части из единственного электронного учебного пособия, то первые же эксперименты с многофакторными моделями, включающими наборы компетенций, показали недостаточность имеющегося набора. Практический вклад в подготовку обучающей выборки сделали авторы пособия [7], предоставив для исследования массив задач, используемых для самостоятельной работы студентов и при проведении экзаменов. Этот расширенный массив был ими же размечен для выбранных показателей сложности и компетенций. Таким образом сформирован объемный, размеченный компетентными экспертами массив классифицированных данных. Соответственно, для рассматриваемой задачи стал доступен практически весь имеющийся методический и алгоритмический аппарат в области классификации [8–11]. К настоящему времени этот богатейший аппарат нашел множество приложений, порой довольно неочевидных, в экономике, производстве, телекоммуникациях [12–17]. Поэтому главные усилия были направлены на совершенствование модели и качества обучающей выборки, проблемы формирования которой решить часто сложнее, чем подобрать хороший алгоритм [18]. Помимо модели и обучающего набора улучшены процедура стратификации обучающей

выборки и метрика обучения, проведены эксперименты с распространенными многоклассовыми классификаторами.

2 Исходные данные и модель задания

Основной учебный материал, использованный в статьях [3, 4], включал 200 заданий из 9 разделов курса «Теория функций комплексного переменного». Модель задания содержала два независимо формируемых вектора признаков: первый формировался на основе текстовой части задачи с помощью типовых процедур обработки текстов на естественном языке, второй — на основе формул в формате TeX с помощью специальных процедур обработки формульных описаний. На рис. 1 эта часть модели отмечена как «модель задания TeX».

Поскольку для классификации применяется парадигма «обучение с учителем», чтобы применить модель для классификации дополнительных категорий, потребовалось выполнить разметку данных дополнительными целевыми атрибутами. Необходимо иметь достаточное число примеров по каждому из новых атрибутов потребовала существенного расширения обучающей выборки. Также была предпринята попытка расширить модель за счет дополнительных признаков, поскольку используемые подходы к формированию векторов признаков сопровождаются потерей части информации о структуре задания. Дополнена модель задания следующими элементами (на рис. 1 отмечены как «модель задания эксперт»):

- (1) тремя дополнительными признаками формульной части задания: числом формул в задании, наличием трансцендентных функций и числом производных;
- (2) атрибутом сложности задания — целым числом из диапазона 1–7;
- (3) 20 атрибутами формируемых компетенций — набором индикаторов, принимающих значение 0 или 1 и соответствующих номеру компетенции, определенному образовательным стандартом для данной дисциплины.

Надо отметить, что информация первого раздела может быть получена и получается автоматически. Собственно, она была доступна в процессе предварительной обработки заданий, описанном в [3], но не использовалась. Разметка заданий по признакам 2 и 3 — это задача, выполняемая исключительно экспертом. В данном исследовании эту разметку выполнили авторы пособия [7],

№9. Восстановить оригинал по изображению $F(p) = \frac{e^{-3p}}{p+2}$.

№5. Вычислить интеграл с помощью вычетов: $\int_{-\infty}^{+\infty} \frac{x^2}{(x^2+1)(x^2+9)} dx$.

№10. Восстановить оригинал по изображению $F(p) = \frac{3p-1}{p^2+4p+6}$.

№6. Найти вычет $\operatorname{res}_0 \left(\frac{1}{z} + \frac{1}{z^2} \right)$.

№11. Восстановить оригинал по изображению $F(p) = \frac{p}{(p^2+1)^2}$.

№8. Определить характер особой точки $z_0 = 0$ для следующих функций: $f(z) = \frac{1}{1-\operatorname{ch} z}$.

№9. Определить характер особой точки $z_0 = 0$ для следующих функций: $g(z) = \frac{z^2}{1-\operatorname{ch} z}$.



Рис. 1 Обновленная модель задания

а главным вызовом представленных далее расчетов было подтверждение возможности автоматизации навыков экспертов по этим признакам (сложность + компетенции) типовыми средствами машинного обучения.

Список компетенций (рис. 2) содержит 20 навыков, которые должны быть сформированы у учащегося в результате прохождения курса. Эксперт без особых трудностей определяет тот перечень компетенций, который отвечает конкретному заданию. Трудность классификации размеченного материала состоит в том, что число компетенций, назначаемых заданиям, переменное: одному заданию могут отвечать от 1 до 4 разных навыков. Соответственно, при таком характере образовательного материала важно не только правильно выбрать и настроить алгоритмы классификации, но и получить сбалансированную обучающую выборку.

Уже при выполнении расчетов в [3, 4] с единственным целевым атрибутом объем в 200 заданий с учетом 9 классов (тематик) не представлялся достаточным. Успех тематической классификации был обеспечен только на смешанной модели, соединяющей две группы разнотипных признаков. В случае 20 целевых атрибутов число классов, определяемых их сочетаниями, выросло более чем до 100, при этом количество примеров в каждом классе не превышало нескольких штук. Поэтому увеличение выборки было критически необходимо. При этом ресурсы существующих пособий уже были практически исчерпаны. Решение позволили получить авторы [7], предоставившие значительный объем заданий, используемых для самостоятельной работы студентов, т. е. для проведения рас-

1. Умение выполнять арифметические действия над комплексными числами в алгебраической форме с использованием основных определений, формул и правил.
2. Умение изображать комплексные числа на плоскости.
3. Умение находить модуль и аргумент комплексного числа, выполнять арифметические действия с комплексными числами в тригонометрической форме.
4. Умение вычислять значения элементарных функций комплексного переменного (ФКП) с использованием свойств ФКП и формул для вычисления.
5. Способность исследовать ФКП на дифференцируемость и аналитичность с использованием определения дифференцируемой, аналитической функции комплексного переменного, условий дифференцируемости ФКП в точке, свойств действительной и мнимой части аналитической функции
6. Умение восстанавливать аналитическую функцию по её действительной или мнимой части.
7. Способность вычислять интеграл ФКП путем сведения к двум криволинейным интегралам второго рода.
8. Способность вычислять интеграл ФКП производя замену переменной при параметризации кривых в комплексной плоскости.
9. Способность вычислять интеграл по формуле Ньютона – Лейбница в случае аналитической подынтегральной функции.
10. Способность вычислять интеграл ФКП с помощью интегральных формул Коши в различных случаях.
11. Способность раскладывать функции аналитические в круге в ряд Тейлора, аналитические в кольце – в ряд Лорана, используя при этом табличные разложения.
12. Умение находить изолированные особые точки (ИОТ), определять их тип, определять порядок нуля функции и, с его помощью, порядок полюса.
13. Умение находить значение вычетов относительно ИОТ.
14. Умение находить значение контурных, несобственных и определенных интегралов с помощью вычетов.
15. Умение различными способами находить изображение при использовании преобразования Лапласа.
16. Умение различными способами восстанавливать оригинал изображения при использовании преобразования Лапласа.
17. Способность находить решение задачи Коши для линейных дифференциальных уравнений операционным методом.
18. Способность находить решение задачи Коши для систем дифференциальных уравнений операционным методом.
19. Решение интегральных уравнений операционным методом.
20. Решение разностных уравнений с помощью дискретного преобразования Лапласа.

Рис. 2 Выписка компетенций из стандарта

четных, контрольных и экзаменационных работ. По понятной причине эти задания не публиковались и не будут публиковаться в открытом доступе. Но благодаря этому материалу общий объем обучающей выборки после исключения дубликатов был доведен до 750 заданий, что позволило получить удовлетворительные результаты классификации.

3 Алгоритмы

За исключением специфической модели, рассматриваемая задача классификации представляется достаточно типичной, поэтому применялись самые распространенные алгоритмы. В данном разделе они указаны для каждого этапа решения задачи: подготовки обучающей выборки; выбора классификаторов; выбора методов многозначной классификации; интерпретации результатов.

Этап *подготовки обучающей выборки* в основном повторяет описанный в [3]. Отличия состоят в добавлении процедуры генерации дополнительных признаков формульной части задания, написанной на языке Python, и замене алгоритма формирования обучающей и тестовой выборок. При неограниченном или хотя бы достаточно большом объеме обучающей выборки вопрос формирования выборки можно решать предельно просто — задать размеры и набирать кандидатов случайным образом, выбирая образцы с равными вероятностями и независимо (случайный выбор). Если при распределении примеров учитывать их число для каждого из имеющихся классов, например делить примеры каждого класса пополам, направляя 50% для обучения и 50% для тестирования и выбирая образцы с равными вероятностями, то стратифицированный характер такого разбиения обеспечен. Такой вариант использовался в [3, 4] и показал свою эффективность. Однако в случае классификации задач по компетенциям стратификация должна выполняться по 20 целевым атрибутам, комбинации значений которых образуют 109 классов. Стандартные алгоритмы используемой библиотеки `scikit-learn` не позволяют в этом случае добиваться сбалансированных разбиений. Нужен более гибкий алгоритм, учитывающий распределение примеров в обучающей выборке. Если это распределение не учитывать, то для каких-то классов можно получить разбиение, либо содержащее избыток тестовых примеров, либо не содержащее их вообще, что исказит оценку качества классификации. Хороший вариант балансировки обучающей выборки предложен в [19]. Этот алгоритм итеративной

стратификации последовательно выбирает признаки и образцы из выборки так, чтобы для каждого сочетания компетенций получалось примерно одинаковое число примеров в обучающей и тестовой группах. Его результативность иллюстрируется на рис. 3. Здесь показаны некоторые наиболее неудачные результаты, возникшие при формировании обучающей и тестовой выборок по признакам компетенций. В двух левых колонках показаны распределения для некоторых наборов компетенций, сформированные типовым алгоритмом случайного выбора, и видно, что число примеров в обучающей выборке оказывается меньше, местами — значительно меньше, чем в тестовой выборке. При этом есть случаи, когда тестовая выборка пустая. Две правых колонки — результаты итеративной стратификации, которая со всеми имеющимися данными справляется успешно. Этот же алгоритм теперь использовался и при формировании выборок для классификации задач по тематике и сложности.

Следующий тип алгоритмов — это *многоклассовые классификаторы*, позволяющие отнести образец к одному из классов по одному признаку. Для атрибутов тематики и сложности алгоритм дает окончательный результат, для атрибута компетенций — промежуточный, используемый на следующем шаге для определения набора компетенций алгоритмом многозначной классификации.

Для классификации были выбраны четыре типовых алгоритма классификации.

1. Логистическая регрессия или модель логит (Logit) [20], т. е. обобщенная линейная регрессионная модель, адаптированная для классификации (детали реализации — https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression). Для тематической классификации и классификации по компетенциям использовался вариант с l_1 -регуляризацией, обратным коэффициентом регуляризации 100 и алгоритмом координатного спуска. Для классификации по сложности использовался вариант с l_2 -регуляризацией, обратным коэффициентом регуляризации 10 и алгоритмом Ньютона–Рафсона, использующим метод сопряженных градиентов.
2. Метод k -ближайших соседей (kNN) [11] (детали реализации — <https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbors-classification>). Для тематической классификации число соседей равно 6, веса распределены равномерно. Для классификации по сложности число соседей равно 10, веса распределены обратно пропорционально расстоянию между образцами. Для классификации по компетен-

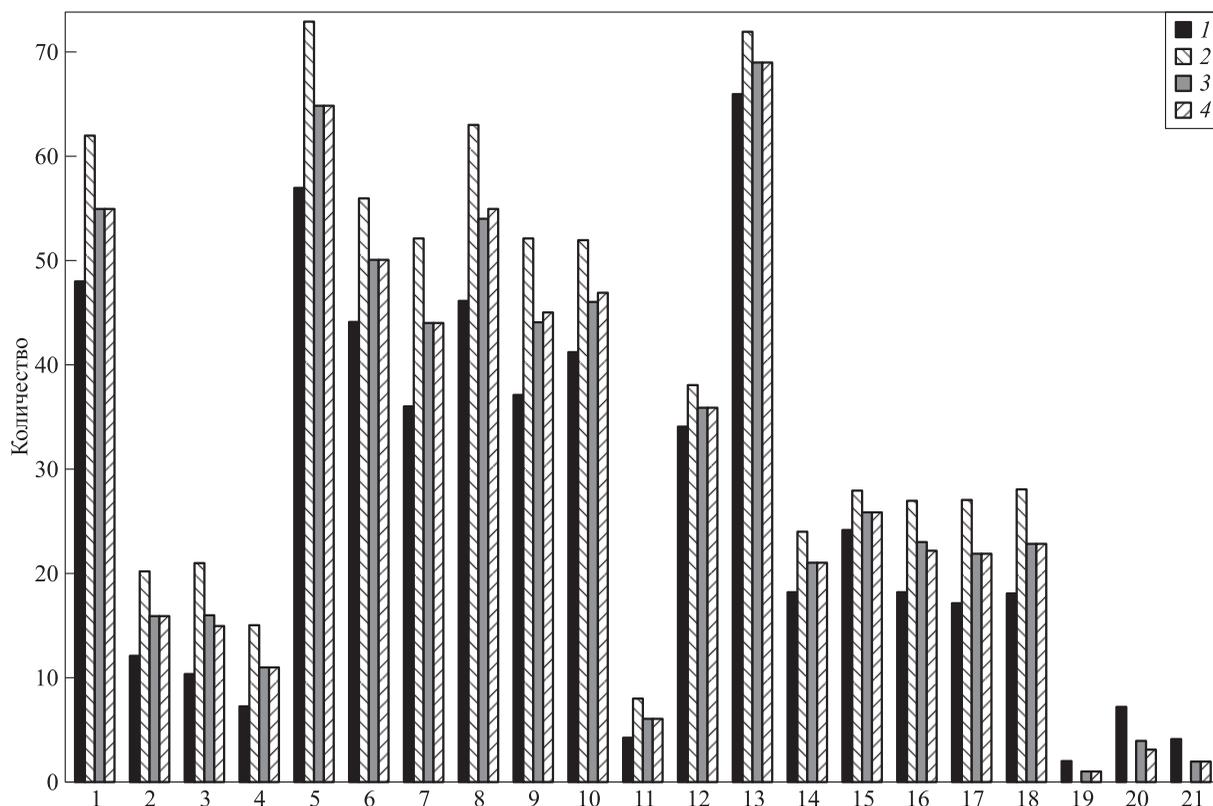


Рис. 3 Результаты распределения задач по группам обучения и тестирования: 1 и 2 — случайный выбор примеров для обучения и тестирования соответственно, 3 и 4 — стратифицированный выбор примеров для обучения и тестирования соответственно

циям число соседей равно 8, веса распределены обратно пропорционально расстоянию между образцами. Все классификаторы для вычисления расстояний между образцами используют евклидову норму.

- Метод чрезвычайно рандомизированных деревьев (ExtraTrees) [21] (детали реализации — <https://scikit-learn.org/stable/modules/ensemble.html#random-forests-and-other-randomized-tree-ensembles>). Для тематической классификации число деревьев в ансамбле равно 200, критерием классификации служит энтропия, при разбиении пространства признаков рассматривается все множество признаков. Для классификации по сложности число деревьев в ансамбле равно 200, критерием классификации служит энтропия, при разбиении пространства признаков рассматривается подмножество с размером, равным логарифму по основанию 2 от числа признаков. Для классификации по компетенциям число деревьев в ансамбле равно 30, критерием классификации служит примесь Джини, при разбиении пространства призна-

ков рассматривается подмножество с размером, равным корню из числа признаков.

- Многослойная нейронная сеть прямого распространения (MLP — multilayer perceptron) [22] с двумя скрытыми слоями и функцией активации $f(x) = \max(0, x)$ (детали реализации — https://scikit-learn.org/stable/modules/neural_networks_supervised.html#neural-networks-supervised). Для тематической классификации число нейронов в первом и втором скрытых слоях равно соответственно 20 и 28. Для классификации по сложности число нейронов в первом и втором скрытых слоях равно соответственно 20 и 24. Для классификации по компетенциям число нейронов в первом и втором скрытых слоях равно соответственно 120 и 8.

Следующий этап касается только классификации задания по компетенциям, где требуемый результат не может быть получен с помощью единичного классификатора. Существует довольно много методов *многозначной классификации* [23], хотя не все из них могут быть использованы в рассматри-

ваемой задаче. Распространенный прием заключается в трансформации задачи, когда для каждого целевого атрибута создается свой классификатор. Это и было сделано для определения набора компетенций. Для целей данной работы выбраны два метода. Первый, наиболее простой и «справедливый», определяется стратегией OneVsRest — «один против остальных» — и состоит в том, что для каждой компетенции обучается и используется свой классификатор, не зависящий от остальных, и набор компетенций получается как комбинация 20 решений по каждому из имеющихся вариантов. «Справедливость» этой стратегии очевидным образом компрометируется неучетом взаимного влияния однофакторных классификаций друг на друга. А влияние это, судя по данным и по содержанию компетенций (см. рис. 2), весьма велико. Достаточно сказать, что имеющиеся размеченные 750 заданий формируют всего 109 разных групп компетенций, что, конечно, существенно меньше возможного числа 2^{20} множества всех подмножеств имеющихся компетенций. Это подтвердили и результаты классификации: хотя существенная доля классификаторов нижнего уровня имела точность 1, общий результат оказался существенно ниже. Для учета этих влияний был опробован метод классификационных цепочек (Classifier Chains), при котором результирующий набор формируется цепочкой решений однофакторных классификаторов, каждое из которых учитывает результат предстоящего в цепочке классификатора. Формировался ансамбль из 15 таких цепочек, порядок классификаторов в каждой цепочке выбирался случайным неповторяющимся образом, все варианты компетенции предполагались равновероятными.

Наконец, последний этап — интерпретация результатов, т.е. оценка качества реализованных алгоритмов, их сравнение. Этот этап требует метрики качества. Первый вариант дает обычная «точность» (accuracy), т.е. доля правильных классификаций. Недостатки этой «прямолинейной» оценки проявляются в случае классификации по компе-

тениям, поскольку правильной классификацией считается только полное совпадение всего списка компетенций. В результате оценка оказывается занижена. Также на эту оценку негативно влияет несбалансированность классов, что ярко проявляется в рассматриваемой задаче (достаточно с этой целью посмотреть на рис. 3). Более точное представление о работе классификатора дает метрика LogLoss, т.е. функция потерь логистической регрессии, основанная на суммировании логарифмов потерь от ошибочной классификации. Этот вариант представляется более объективным, но использовать его без ассигасы, наверное, не вполне правильно из-за отсутствия физической интерпретации абсолютного значения оценки потерь.

4 Экспериментальная классификация

Итак, для проведения расчетов имеются 750 заданий, подготовленных в соответствии с предложенной моделью (рис. 1) и размеченных атрибутами тематики, сложности и набором компетенций. Размеры обучающей и тестовой выборок определены равными, т.е. по 375 заданий. Стратификация по всем группам атрибутов выполнена итеративным алгоритмом. Обучены четыре алгоритма классификации (обозначены Logit, kNN, ExtraTrees и MLP) и два алгоритма многозначной классификации (обозначены OneVsRest и ClassifierChains). Метрики Accuracy и LogLoss вычислены для всех 16 возможных комбинаций алгоритмов классификации, результаты приведены в таблице (полужирным выделены «победители»).

Отметим, что гиперпараметры для всех использованных алгоритмов оптимизированы путем поиска по сетке с перекрестной валидацией по критерию минимизации метрики LogLoss. Несмотря на то что процедура оптимизации требует значительных затрат вычислительных ресурсов, эффективность оптимизации гиперпараметров алгоритмов

Анализ качества классификации

Классификатор	Атрибут			
	Тематика	Сложность	Компетенции + OneVsRest	Компетенции + ClassifierChains
Logit	Accuracy = 0,981 LogLoss = 0,055	Accuracy = 0,869 LogLoss = 0,403	Accuracy = 0,851 LogLoss = 1,343	Accuracy = 0,856 LogLoss = 1,328
kNN	Accuracy = 0,965 LogLoss = 0,099	Accuracy = 0,837 LogLoss = 0,817	Accuracy = 0,795 LogLoss = 2,073	Accuracy = 0,785 LogLoss = 1,872
ExtraTrees	Accuracy = 0,987 LogLoss = 0,090	Accuracy = 0,877 LogLoss = 0,367	Accuracy = 0,827 LogLoss = 1,718	Accuracy = 0,830 LogLoss = 1,537
MLP	Accuracy = 0,981 LogLoss = 0,046	Accuracy = 0,848 LogLoss = 0,914	Accuracy = 0,851 LogLoss = 1,401	Accuracy = 0,856 LogLoss = 1,459

классификации весьма существенна: метрики, вычисленные при использовании оптимизированных гиперпараметров, против значений по умолчанию отличались в 1,5–2 раза. Исключение здесь наблюдалось только для многослойной нейронной сети прямого распространения. Этот метод демонстрирует низкую чувствительность к выбору параметров, что, по-видимому, вполне объяснимо тем, что сам по себе алгоритм содержит большое число самонастраиваемых параметров, которые в любом случае позволяют извлечь из имеющегося контента всю информацию, содержательную с точки зрения классификации. Собственно, такое поведение перцептронов — это именно то, что от них ожидалось. Но с позиции затрат вычислительных ресурсов этот метод безусловно и значительно проигрывает всем остальным.

5 Выводы

Основной полученный результат — это доказательство принципиальной возможности автоматизировать классификацию математических задач не только по признаку тематики [3, 4], но и по другому признаку, причем гораздо более сложному, — многозначному признаку компетенций. Попутно была подтверждена эффективность предложенной модели математического задания (цифрового образовательного контента на языке научной разметки \TeX) для другого содержания. Усилия по созданию модели и подготовке обучающей выборки привели к тому, что все выбранные типовые методы классификации работают вполне эффективно, и нет сомнений, что будут работать и другие известные алгоритмы. По-видимому, это означает, что наиболее ценными в задачах такого типа оказываются модель и обучающий контент, а существующих на данный момент методов и алгоритмов достаточно.

Важно еще обратить внимание на перспективу дальнейшего развития и использования представленных результатов. Описанная выше технология обладает несомненным потенциалом в вопросах автоматизации анализа качества образовательного контента. На первый взгляд, анализ качества, т. е. экспертиза, — вопрос довольно сложный, поскольку в случае с экспертом-человеком возникает проблема объективности, учет которой крайне сложен. Но, с другой стороны, как видим, средства машинного обучения, которые в последние годы продемонстрировали способности решения самых разных задач, прекрасно справляются и с контентом данного типа. При этом объективность таких решений не вызывает сомнений. Более того, им под силу объединить (обобщить, усреднить, при-

вести к общему знаменателю) мнения многих экспертов, что при «ручной» экспертизе сделать практически невозможно. Таким образом, получается, что проблематика автоматизированного анализа качества цифрового образовательного контента является богатым источником прикладных задач для методов машинного обучения, а представленная в статье технология хорошо вписывается в это прикладное направление.

Литература

1. Мартюшова Я. Г. Теоретические основания конструирования электронных учебников для студентов технических университетов // Отечественная и зарубежная педагогика, 2018. Т. 1. № 5. С. 151–165. doi: 10.24411/2224-0772-2018-10010. EDN: YOCAOT.
2. Босов А. В., Мартюшова Я. Г., Наумов А. В. Выбор направлений оценивания качества электронных средств обучения для организации учебного процесса вуза // Сибирский педагогический ж., 2022. № 2. С. 54–63. doi: 10.15293/1813-4718.2202.05. EDN: KNPIQX.
3. Босов А. В., Иванов А. В. Технология классификации типов контента электронного учебника // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 63–72. doi: 10.14357/19922264220410. EDN: YERCNH.
4. Bosov A. V., Ivanov A. V. Two approaches to E-book content classification // Artificial intelligence application in networks and systems / Eds. R. Silhavy, P. Silhavy. — Lecture notes in networks and systems ser. — Cham, Switzerland: Springer, 2023. Vol. 724. P. 77–87. doi: 10.1007/978-3-031-35314-7_6.
5. Хуторской А. В. Дидактическая эвристика: Теория и технология креативного обучения. — М.: МГУ, 2003. 416 с.
6. Модернизация образовательного процесса в начальной, основной и старшей школе: варианты решения / Под ред. А. Г. Каспржака. — М.: Просвещение, 2004. 415 с. EDN: QTGWNX.
7. Битюков Ю. И., Мартюшова Я. Г. Решение задач по теории функций комплексного переменного. — М.: МАИ, 2022. 87 с.
8. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мещалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989. 608 с.
9. Mitchell T. Machine learning. — New York, NY, USA: McGraw-Hill Education, 1997. 432 p.
10. Журавлев Ю. И., Рязанов В. В., Сенько О. В. Распознавание. Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. 159 с.
11. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. — New York, NY, USA: Springer, 2009. 745 p. doi: 10.1007/978-0-387-84858-7.

12. *Goebel P. C., Palik B. J., Kirkman L. K., Drew M. B., West L., Pederson D. C.* Forest ecosystems of a Lower Gulf Coastal Plain landscape: Multifactor classification and analysis // *J. Torrey Bot. Soc.*, 2001. Vol. 128. No. 1. P. 47–75. doi: 10.2307/3088659.
13. *Щербакова С. А.* Международный туризм: экономика и география. — М.: Финансы и статистика, 2007. 144 с.
14. *Кулешков И. В., Колбиков В. С., Басниев К. С.* Классификация нефтегазовых эксплуатационных объектов с оценкой степени сложности выработки запасов углеводородов // *Технологии нефти и газа*, 2010. № 2(67). С. 29–38. EDN: LKYHCP.
15. *Задорожнева Ю. В., Калинина А. Э.* Многофакторная оценка эффективности реализации социально-экономической политики региона // *Современные проблемы науки и образования*, 2012. № 4. С. 192–192. EDN: PBIRUD.
16. *Казарин О. В., Шаряпов Р. А., Яценко В. В.* Многофакторная классификация угроз информационной безопасности киберфизических систем // *Вестник РГГУ. Сер. Информатика. Информационная безопасность*, 2018. № 1. С. 39–55. EDN: FNXPAP.
17. *Xiong Y., Wang L., Wang Q., Liu S., Kou B.* Improved convolutional neural network with feature selection for imbalanced ECG multi-factor classification // *Measurement*, 2022. Vol. 189. Art. 110471. doi: 10.1016/j.measurement.2021.110471.
18. *Кафтанников И. Л., Парасич А. В.* Проблемы формирования обучающей выборки в задачах машинного обучения // *Вестник ЮУрГУ. Сер. Компьютерные технологии, управление, радиоэлектроника*, 2016. Т. 16. № 3. С. 15–24. doi: 10.14529/ct«sr160302. EDN: TZWHTS.
19. *Sechidis K., Tsoumakas G., Vlahavas I.* On the stratification of multi-label data // *Machine learning and knowledge discovery in databases / Eds. D. Gunopulos, T. Hofmann, D. Malerba, M. Vazirgiannis.* — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer, 2011. Vol. 6913. P. 145–158. doi: 10.1007/978-3-642-23808-6_10.
20. *Hosmer D. W., Lemeshow S., Sturdivant R. X.* Applied logistic regression. — 3rd ed. — New York, NY, USA: Wiley, 2013. 528 p. doi: 10.1002/9781118548387.
21. *Breiman L.* Random forests // *Mach. Learn.*, 2001. No. 45. P. 5–32. doi: 10.1023/A:1010933404324.
22. *Haykin S.* Neural networks and learning machines. — 3rd ed. — Cranbury, NJ, USA: Pearson Education, 2009. 906 p.
23. *Bogatinovski J., Todorovski L., Džeroski S., Kocev D.* Comprehensive comparative study of multi-label classification methods // *Expert Syst. Appl.*, 2022. Vol. 203. Art. 117215. doi: 10.1016/j.eswa.2022.117215.

Поступила в редакцию 15.10.23

MULTIFACTOR CLASSIFICATION TECHNOLOGY OF MATHEMATICAL CONTENT OF E-LEARNING SYSTEM

A. V. Bosov and A. V. Ivanov

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

Abstract: The article continues the study of the problem of classifying the content of an e-learning system. The previously developed technology for thematic classification of mathematical content contained in the blocks of tasks and examples of e-learning system has been improved and supplemented with new functions. For this purpose, the previously used content model with two properties — a text description of a task and its formula part in \TeX format — has been supplemented with a number of formal numerical attributes, such as the presence of transcendental and derived functions, and the number of formulas in the task. This block of attributes made it possible to improve the quality of the existing thematic classifier and to implement two new ones. The first classifier determines the level of complexity of the task. The second multilabel classifier determines the set of student competencies that the task should form. Such a multifactorial classification is an important stage in the promising direction of the development of e-learning system — automated assessment of the quality of educational content. Performance testing of the proposed algorithms, training of classifiers, and analysis of classification quality were carried out using the tasks from the same discipline of the theory of functions of a complex variable but on significantly expanded set of data, including tasks for independent work of students — calculation and examination tasks.

Keywords: e-learning system; mathematical content; machine learning; multifactor classification; content quality assessment

DOI: 10.14357/19922264230405

EDN: LISHHZ

Acknowledgments

The research has been supported by the Russian Science Foundation according to the research project No. 22-28-00588 (<https://rscf.ru/project/22-28-00588/>). The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

References

- Martyushova, Ya. G. 2018. Teoreticheskie osnovaniya konstruirovaniya elektronnykh uchebnikov dlya studentov tekhnicheskikh universitetov [Theoretical foundations of designing electronic textbooks for students of technical universities]. *Otechestvennaya i zarubezhnaya pedagogika* [Domestic and Foreign Pedagogy] 1(5):151–165. doi: 10.24411/2224-0772-2018-10010. EDN: YOCAOT.
- Bosov, A. V., Ya. G. Martyushova, and A. V. Naumov. 2022. Vybory napravleniy otsenivaniya kachestva elektronnykh sredstv obucheniya dlya organizatsii uchebnogo protsesssa vuza [Directions selection for assessing the quality of electronic learning tools for the organization of the educational process of the higher educational process]. *Sibirskiy pedagogicheskiy zh.* [Siberian Pedagogical J.] 2:54–63. doi: 10.15293/1813-4718.2202.05. EDN: KNPIQX.
- Bosov, A. V., and A. V. Ivanov. 2022. Tekhnologiya klassifikatsii tipov kontenta elektronnoy uchebnika [Technology for classification of content types of e-textbooks]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):63–72. doi: 10.14357/19922264220410. EDN: YERCNH.
- Bosov, A. V., and A. V. Ivanov. 2023. Two approaches to E-Book content classification. *Artificial intelligence application in networks and systems*. Eds. R. Silhavy and P. Silhavy. Lecture notes in networks and systems ser. Cham, Switzerland: Springer. 724:77–87. doi: 10.1007/978-3-031-35314-7_6.
- Khutorskoy, A. V. 2003. *Didakticheskaya evristika: Teoriya i tekhnologiya kreativnogo obucheniya* [Didactic heuristics: Theory and technology of creative learning]. Moscow: MSU. 416 p.
- Kasprzhak, A. G., ed. 2004. *Modernizatsiya obrazovatel'nogo protsesssa v nachal'noy, osnovnoy i starshey shkole: varianty resheniya* [Modernization of the educational process in primary, secondary and high schools: Solution options]. Moscow: Prosveshchenie. 415 p. EDN: QTGWXN.
- Bitukov, Yu. I., and Ya. G. Martyushova. 2022. *Reshenie zadach po teorii funktsiy kompleksnogo peremennogo* [Solving problems in the theory of functions of a complex variable]. Moscow: MAI. 87 p.
- Ayvazyan, S. A., V. M. Bukhshtaber, I. S. Enyukov, and L. D. Meshalkin. 1989. *Prikladnaya statistika: klassifikatsiya i snizhenie razmernosti* [Applied statistics: Classification and dimensionality reduction]. Moscow: Finansy i statistika. 608 p.
- Mitchell, T. 1997. *Machine learning*. New York, NY: McGraw-Hill Education. 432 p.
- Zhuravlev, Yu. I., V. V. Ryazanov, and O. V. Sen'ko. 2006. *Raspoznavanie. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya* [Recognition. Mathematical methods. Software system. Practical applications]. Moscow: Fazis. 159 p.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning*. 2nd ed. New York, NY: Springer. 745 p. doi: 10.1007/978-0-387-84858-7.
- Goebel, P. C., B. J. Palik, L. K. Kirkman, M. B. Drew, L. West, and D. C. Pederson. 2001. Forest ecosystems of a Lower Gulf Coastal Plain landscape: Multifactor classification and analysis. *J. Torrey Bot. Soc.* 128(1):47–75. doi: 10.2307/3088659.
- Shcherbakova, S. A. 2007. *Mezhdunarodnyy turizm: ekonomika i geografiya* [International tourism: Economics and geography]. Moscow: Finansy i statistika. 144 p.
- Kuleshkov, I. V., V. S. Kolbikov, and K. S. Basniev. 2010. Klassifikatsiya neftegazovykh ekspluatatsionnykh ob'ektov s otsenkoy stepeni slozhnosti vyrabotki zapasov uglevodorodov [Oil and gas production facilities classification with complexity of hydrocarbon reserves production assessment]. *Tekhnologii nefti i gaza* [Oil and Gas Technologies] 2(67):29–38. EDN: LKYHCP.
- Zadorozhneva, Yu. V., and A. E. Kalinina. 2012. Mnogofaktornaya otsenka effektivnosti realizatsii sotsial'no-ekonomicheskoy politiki regiona [Multifactor assessment of efficiency realization of socio-economic policy of the region]. *Sovremennyye problemy nauki i obrazovaniya* [Modern Problems of Science and Education] 4:192–192. EDN: PBIRUD.
- Kazarin, O. V., R. A. Sharyapov, and V. V. Yashchenko. 2018. Mnogofaktornaya klassifikatsiya ugroz informatsionnoy bezopasnosti kiberfizicheskikh sistem [Multifactorial classification of threats to information security of cyber-physical systems]. *Vestnik RGGU. Ser. Informatika. Informatsionnaya bezopasnost'. Matematika* [RGGU Bulletin. Ser. Information science. Information security. Mathematics] 1:39–55. EDN: FNXPAP.
- Xiong, Y., L. Wang, Q. Wang, S. Liu, and B. Kou. 2022. Improved convolutional neural network with feature selection for imbalanced ECG multi-factor classification. *Measurement* 189:110471. doi: 10.1016/j.measurement.2021.110471.
- Kaftannikov, I. L., and A. V. Parasich. 2016. Problemy formirovaniya obuchayushchey vyborki v zadachakh mashinnogo obucheniya [Problems of training set's formation in machine learning tasks]. *Vestnik YuUrGU. Ser. Komp'yuternye tekhnologii, upravlenie, radioelektronika* [Bulletin of the South Ural State University. Ser. Computer technologies, automatic control, radio electronics] 16(3):15–24. doi: 10.14529/ct&sr160302. EDN: TZWHTS.
- Sechidis, K., G. Tsoumakas, and I. Vlahavas. 2011. On the stratification of multi-label data. *Machine learning*

- and knowledge discovery in databases*. Eds. D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis. Lecture notes in computer science ser. Berlin, Heidelberg: Springer. 6913:145–158. doi: 10.1007/978-3-642-23808-6_10.
20. Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied logistic regression*. 3rd ed. New York, NY: Wiley. 528 p. doi: 10.1002/9781118548387.
21. Breiman, L. 2001. Random forests. *Mach. Learn.* 45:5–32. doi: 10.1023/A:1010933404324.
22. Haykin, S. 2009. *Neural networks and learning machines*. 3rd ed. Cranbury, NJ: Pearson Education. 906 p.
23. Bogatinovski, J., L. Todorovski, S. Džeroski, and D. Koccev. 2022. Comprehensive comparative study of multi-label classification methods. *Expert Syst. Appl.* 203:117215. 23 p. doi: 10.1016/j.eswa.2022.117215.

Received October 15, 2023

Contributors

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

Ivanov Alexey V. (b. 1976) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; aivanov@ipiran.ru

РАСШИРЯЕМЫЙ ПОДХОД К СЛИЯНИЮ ДАННЫХ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СРЕДАХ*

В. В. Сазонтьев¹, С. А. Ступников², В. Н. Захаров³

Аннотация: Статья относится к области разработки методов и средств интеграции данных. Один из важнейших этапов интеграции данных — это слияние данных, объединение записей, относящихся к одной и той же сущности реального мира, в одну запись с разрешением конфликтов для каждого из атрибутов. Рассмотрена формальная постановка задачи слияния данных, приведен краткий обзор основных существующих групп методов слияния данных. Предложен подход к реализации этапа слияния данных в расширяемой системе интеграции разнородных источников данных в распределенной вычислительной среде, рассмотрена его программная архитектура и основные идеи реализации.

Ключевые слова: слияние данных; распределенные вычислительные среды

DOI: 10.14357/19922264230406

EDN: LCQAZX

1 Введение

Как только в мире появились наборы данных, которые было необходимо связывать и соединять, возникло понятие интеграции данных. Даже до появления компьютеров перед учеными-статистиками вставали задачи сопоставления и анализа данных переписей населения, собранных в разные периоды времени. Задачи интеграции данных — обеспечения унифицированного доступа к множеству разнородных автономных источников данных — сложны не только ввиду способности человека представлять информацию о сущностях реального мира крайне разнообразным образом, но и ввиду возможности отслеживания практически любых событий и взаимодействий в цифровом мире, а также желания человека проводить анализ данных и применять его результаты для принятия решений [1].

В интеграции данных можно выделить четыре основных этапа: извлечение данных; сопоставление схем; связывание записей; слияние данных [2].

Извлечение данных предполагает преобразование неструктурированных данных (например, текстов) или полуструктурированных данных из веб-источников в структурированный вид [3].

Сопоставление схем предполагает определение некоторой унифицированной схемы данных для предметной области и установление семантических отношений между атрибутами и сущностями схем источников и атрибутами и сущностями унифицированной схемы.

Связывание записей состоит в группировке записей из разных источников, относящихся к одной и той же сущности реального мира.

Наконец, *слияние данных* означает объединение сгруппированных записей в одну с разрешением конфликтов для каждого из атрибутов.

Данная статья посвящена описанию подхода к реализации этапа слияния данных в разрабатываемой системе интеграции разнородных источников данных в распределенной вычислительной среде [4, 5].

2 Методы слияния данных

Даже когда источники данных предоставляют информацию об одном и том же атрибуте одной и той же сущности, значения атрибута в разных источниках могут не совпадать. Эти конфликты возникают из-за неправильного ввода, некорректных вычислений, устаревшей информации, противоречивой интерпретации семантики и т. д. Цель слияния данных — определить, какое значение отражает положение дел в реальном мире.

Формально задачу слияния данных можно определить следующим образом. Пусть S — множество источников данных, D — множество элементов данных. *Элемент данных* представляет отдельный аспект сущности реального мира (например, дату рождения человека), а в реляционной базе данных элемент данных соответствует атрибуту кортежа из некоторого отношения. Для каждого $d \in D$ источ-

* Исследование выполнено за счет гранта Российского научного фонда № 22-21-00692 (<https://rscf.ru/project/22-21-00692/>).

¹ Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; vladimirsvsite@gmail.com

² Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; sstupnikov@ipiran.ru

³ Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; vzakharov@ipiran.ru

ник $s \in S$ может предоставлять значение. Значение может быть атомарным (дата рождения), множеством (телефонные номера), списком (авторы статьи). Среди различных значений элемента данных одно соответствует реальному миру (истинное значение), остальные — нет (ложные значения). Задача слияния данных — определить истинное значение для каждого элемента $d \in D$ [1].

В настоящее время разработано множество методов решения задачи слияния данных. Разрешение конфликтов и удаление ошибочных значений опирается на коллективный «разум» источников данных, определение надежных источников, определение копирования данных между источниками.

При разрешении конфликтов возможно использование различных высокоуровневых стратегий [6]: *игнорирование конфликтов* (сохранение всех вариантов значений или создание всех возможных комбинаций значений, когда разрешение конфликта фактически остается за пользователем); *избегание конфликтов* (когда значения берутся из предпочитаемого источника либо отбираются лишь непротиворечивые данные); *разрешение конфликтов* (выбор наиболее часто встречающегося, случайного, среднего, наиболее актуального значения и т. д.).

Реализация стратегий возможна при помощи различных реляционных операций [6]. Использование различных вариантов *соединения* (join) отношений опирается на комбинацию кортежей из отношений при обращении в истину некоторых предикатов над атрибутами. Варианты операции *объединения* (union) предполагают добавление кортежей из исходных отношений в заранее определенное отношение унифицирующей схемы.

Современным базовым методом определения истинного значения элемента данных обычно считается голосование: значение, представляемое наибольшим числом источников, объявляется истинным. *Надежность источника* определяется степенью корректности предоставляемых им значений. Более надежный источник обладает большим весом при голосовании. При обнаружении *копирования значения* из другого источника вес этого значения при голосовании может быть уменьшен. К подходам, сочетающим вычисление вероятности копирования, вычисление надежности источников и голосование, относится, например, алгоритм *AccuSoru* и его модификации [1].

Другие существующие методы вычисления надежности источников и корректности значений опираются на алгоритм вычисления важности веб-страниц *PageRank*; метрики схожести, используемые в области информационного поиска; вероятностные графические модели; вычисление точности и полноты (precision/recall) [1].

Все большее значение в области слияния данных приобретают методы машинного обучения [2].

Многообещающим выглядит применение дискриминативных вероятностных моделей для слияния данных (в простейшем случае — логистической регрессии) [7]. При этом процесс слияния данных делится на два этапа:

- (1) статистическое обучение для вычисления параметров модели, используемой для оценки надежности источников;
- (2) использование вероятностного вывода для предсказания значений элементов данных.

Для обучения моделей привлекаются признаки, характерные для предметной области: например, для источников научных статей характерны *число цитирований* и *год публикации*, а для веб-источников — статистика трафика (*число посещений*, *число ежедневных просмотров* и т. д.). Оптимизация при обучении моделей основывается на максимизации ожидания (expectation maximization, EM), опирающейся на пересечения между значениями в источниках и среднюю надежность источников, и минимизации эмпирического риска (empirical risk minimization, ERM), опирающейся на истинные значения элементов данных. Для автоматизации выбора между EM и ERM применяется оптимизатор, определяющий, какой из алгоритмов ведет к более точным результатам слияния данных.

3 Подход к реализации этапа слияния данных в системе интеграции данных

3.1 Архитектура и основные положения подхода

Предлагаемый подход предназначен для реализации в рамках расширяемой системы интеграции данных в распределенной вычислительной среде [3–5]. Архитектура подхода (основные компоненты, связь с базой данных системы интеграции) представлена на рис. 1.

В системе интеграции возможно одновременное решение нескольких *задач интеграции* в различных предметных областях. На этапе слияния данных при решении конкретной задачи интеграции предполагается, что этапы извлечения данных и связывания записей уже пройдены и их результаты помещены в *базу данных извлеченных записей* и *базу данных связанных записей* соответственно. Основные положения подхода к реализации слияния данных в системе состоят в следующем:

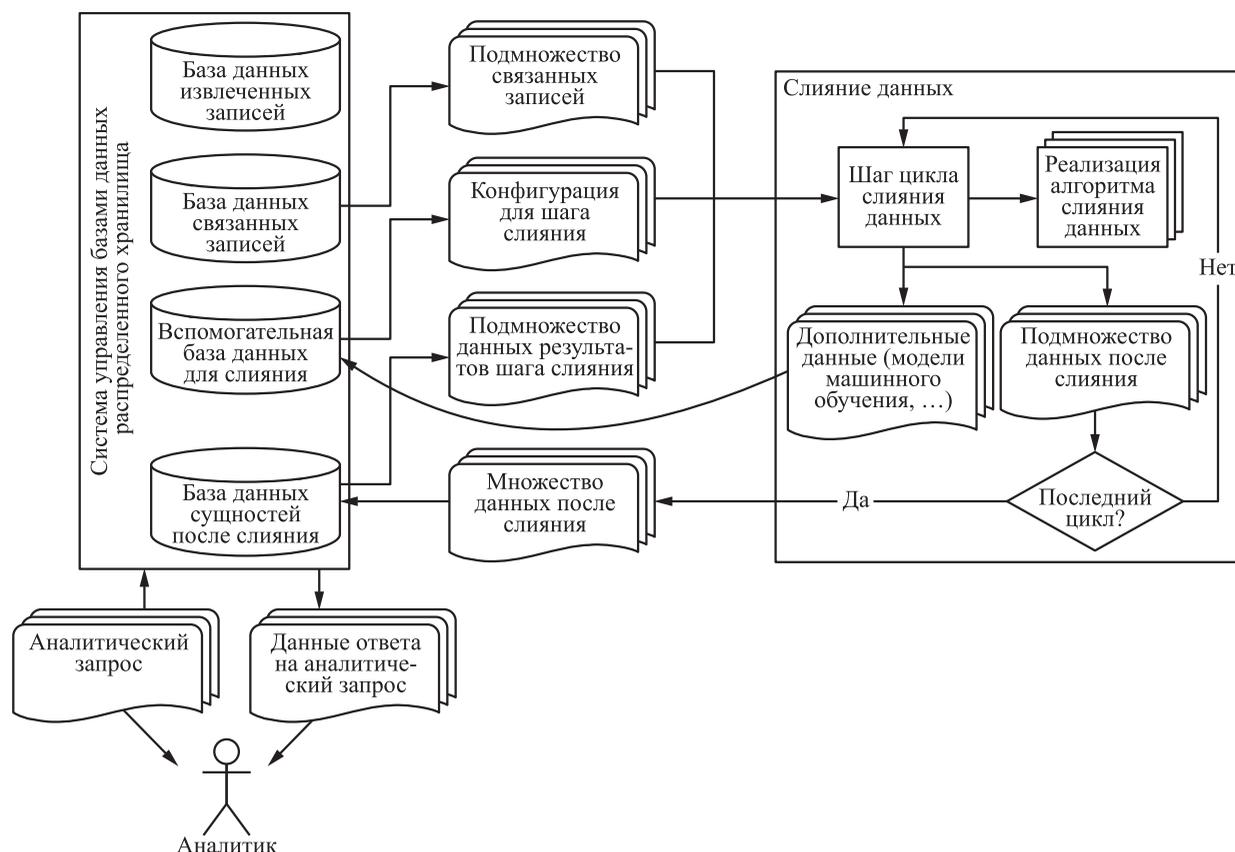


Рис. 1 Программная архитектура компонента слияния данных

- процесс слияния данных состоит из конечного числа шагов, последовательность которых определяется *конфигурацией* процесса слияния и задается экспертом в предопределенном формате;
 - конфигурация шага слияния включает список *входных* атрибутов, список *выходных* атрибутов, имя *ключевого атрибута*, а также название и список значений параметров алгоритма слияния данных, применяемого на данном шаге. Выходные атрибуты могут быть входными для последующих шагов;
 - на каждом шаге слияния происходит применение некоторого алгоритма слияния на тех записях из базы данных связанных записей, в которых есть входные атрибуты. Если среди входных атрибутов присутствуют выходные атрибуты предыдущих шагов, то они соединяются (используется операция реляционной алгебры *соединение* — join) с данными из базы данных связанных записей по ключевому атрибуту. Набор записей с выходными атрибутами, сформированный в результате работы алгоритма, сохраняется в *базе данных соединенных сущностей*;
 - вспомогательные данные для шагов слияния сохраняются во *вспомогательной базе данных для слияния*. К таким данным, например, относятся обученные модели машинного обучения (если шаг предполагает использование соответствующих методов обучения на данных из базы данных связанных записей), которые могут быть повторно использованы в других процессах слияния данных;
 - по завершении всех шагов процесса слияния данных пользователю предоставляется возможность задания аналитических запросов к базе данных соединенных сущностей.
- Подход расширяем в следующем смысле: в систему могут унифицированным образом встраиваться различные новые алгоритмы слияния данных (основанные на реляционных операциях, определении истинных значений элементов данных и надежности источников, применении моделей машинного обучения). При этом не требуется внесение изменений в общую архитектуру систе-

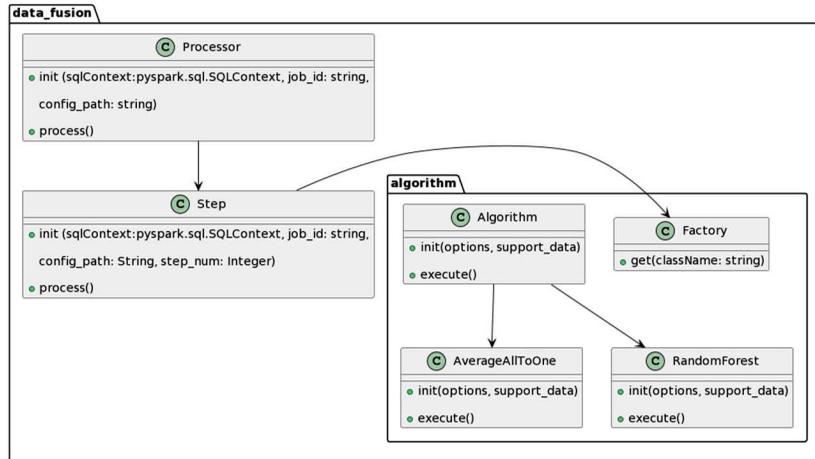


Рис. 2 Основные классы реализации слияния данных

мы. На каждом из этапов интеграции данных, как и на каждом из шагов слияния данных, происходят чтение входных данных и запись выходных данных в соответствующие базы данных без передачи данных напрямую от одного программного компонента к другому. При этом реализации методов, использующих одни и те же входные данные, становятся взаимозаменяемыми в системе.

3.2 Основные идеи программной реализации подхода

Рассматриваемый подход реализуется в составе прототипа системы интеграции данных¹ [3, 5]. Прототип реализован с использованием языков программирования Python и Javascript в распределенной вычислительной среде Hadoop с использованием распределенной файловой системы HDFS и вычислительной модели Spark. Отдельные шаги этапов интеграции данных реализуются в виде различных классов языка Python. При этом для каждого шага определен единый родительский класс, чтобы путем наследования от него можно было определять различные реализации шагов с сохранением методов чтения и записи. Это позволяет предоставлять унифицированный интерфейс для отдельных шагов и обеспечивает взаимозаменяемость классов реализации. Результаты каждого этапа и шага сохраняются в базе данных, развернутой в распределенной файловой системе.

Основные классы, используемые для реализации слияния данных, — это *Processor* и *Step* из пакета *data_fusion*, а также *Factory* и *Algorithm* из пакета *algorithm* (рис. 2). Класс *Processor* отвечает за исполнение цикла слияния данных для задачи системы

интеграции с идентификатором *job_id*. Значение параметра *config_path* — это путь в хранилище к файлу с конфигурацией процесса слияния. Класс *Step* отвечает за исполнение отдельного шага цикла слияния с номером *step_num*, реализует чтение и запись данных в базу данных.

Класс *Factory* отвечает за выбор класса алгоритма слияния данных для шага и передачу настроек в экземпляр класса алгоритма. Метод *get* возвращает экземпляр класса с названием *className*. Класс *Algorithm* — это родительский класс для всех классов, реализующих алгоритмы слияния данных. Параметр *options* класса *Algorithm* — это словарь параметров для конкретного алгоритма (например, для алгоритма *случайного леса* — число деревьев решений, глубина дерева и т. д.); параметр *support_data* — это путь к данным конкретного шага в хранилище (например, временно создаваемые файлы, готовые модели машинного обучения). Примеры классов, реализующих конкретные алгоритмы слияния, — *AverageAllToOne* (усреднение значений атрибутов записей), *RandomForest* (обучение и применение модели случайного леса для определения значения атрибута).

4 Заключение

Слияние данных — объединение записей, относящихся к одной и той же сущности реального мира, в одну запись с разрешением конфликтов для каждого из атрибутов — важный этап интеграции данных, нацеленной на обеспечение унифицированного доступа к множеству разнородных автономных источников данных. В работе рассмотрена формальная постановка задачи слияния данных,

¹https://bitbucket.org/VladimirSaz/integration_system.

приведен краткий обзор основных существующих групп методов слияния данных. Предложен подход к реализации этапа слияния данных в системе интеграции данных, разрабатываемой авторами, рассмотрены его программная архитектура и основные идеи реализации.

Литература

1. Dong X. L., Srivastava D. Big Data integration. — Synthesis lectures on data management ser. — Morgan & Claypool Pubs., 2015. 178 p.
2. Dong X. L., Rekatsinas T. Data integration and machine learning: A natural synergy // Conference (International) on Management of Data. — New York, NY, USA: ACM, 2018. P. 1645–1650. doi: 10.1145/3183713.3197387.
3. Sazontev V. V., Stupnikov S. A. An extensible approach to searching and selecting data sources for materialized Big Data integration in distributed computing environments // Pattern Recognition Image Analysis, 2023. Vol. 33. No. 2. P. 147–156. doi: 10.1134/S1054661823020141. EDN: YXUMDO.
4. Sazontev V. V. Methods for Big Data integration in distributed computation environments // CEUR Workshop Procee. Vol. 2277. P. 238–244.
5. Sazontev V., Stupnikov S. An extensible approach for materialized Big Data integration in distributed computation environments // Ivannikov Memorial Workshop. — IEEE, 2019. P. 33–38. doi: 10.1109/IVMEM.2019.00011. EDN: BWILRL.
6. Bleiholder J., Naumann F. Data fusion // ACM Comput. Surv., 2008. Vol. 41. No. 1. P. 1–41. doi: 10.1145/1456650.1456651.
7. Rekatsinas T., Joglekar M., Garcia-Molina H., Parameswaran A., Re C. Slimfast: Guaranteed results for data fusion and source reliability // Conference (International) on Management of Data Proceedings. — New York, NY, USA: ACM, 2017. P. 1399–1414. doi: 10.1145/3035918.3035951.

Поступила в редакцию 29.09.23

AN EXTENSIBLE APPROACH TO DATA FUSION IN DISTRIBUTED COMPUTING ENVIRONMENTS

V. V. Sazontev, S. A. Stupnikov, and V. N. Zakharov

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

Abstract: The paper belongs to the area of development of methods and tools for data integration. One of the most important stages of data integration is data fusion, i. e., the combination of records relating to the same real-world entity into a single record with conflict resolution for each of the attributes. The paper considers the formal statement of the data fusion problem, provides a brief review of major groups of data fusion methods. An approach for implementation of the data fusion stage within an extensible heterogeneous data integration system in a distributed computing environment is proposed. Software architecture and basic implementation ideas of the approach are considered.

Keywords: data fusion; distributed computing environment

DOI: 10.14357/19922264230406

EDN: LCQAZX

Acknowledgments

The work was supported by the Russian Science Foundation, project No. 22-21-00692 (<https://rscf.ru/project/22-21-00692/>).

References

1. Dong, X. L., and D. Srivastava. 2015. *Big Data integration*. Synthesis lectures on data management ser. Morgan & Claypool Pubs. 178 p.
2. Dong, X. L., and T. Rekatsinas. 2018. Data integration and machine learning: A natural synergy. *Conference (International) on Management of Data*. New York, NY: ACM. 1645–1650. doi: 10.1145/3183713.3197387.
3. Sazontev, V. V., and S. A. Stupnikov. 2023. An extensible approach to searching and selecting data sources for materialized Big Data integration in distributed computing environments. *Pattern Recognition Image Analysis* 33(2):147–156. doi: 10.1134/S1054661823020141. EDN: YXUMDO.
4. Sazontev, V. V. 2018. Methods for Big Data integration in distributed computation environments. *CEUR Workshop Procee*. 2277:238–244.
5. Sazontev, V., and S. Stupnikov. 2019. An extensible approach for materialized Big Data integration in distributed computation environments. *Ivannikov Memorial Workshop*. IEEE. 33–38. doi: 10.1109/IVMEM.2019.00011. EDN: BWILRL.

6. Bleiholder, J., and F. Naumann. 2008. Data fusion. *ACM Comput. Surv.* 41(1):1–41. doi: 10.1145/1456650.1456651.
7. Rekatsinas, T., M. Joglekar, H. Garcia-Molina, A. Parameswaran, and C. Re. 2017. SLiMFast: Guaranteed results for data fusion and source reliability. *Conference (International) on Management of Data Proceedings*. New York, NY: ACM. 1399–1414. doi: 10.1145/3035918.3035951.

Received September 29, 2023

Contributors

Sazontev Vladimir V. (b. 1990) — programmer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vladimirsvsite@gmail.com

Stupnikov Sergey A. (b. 1978) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sstupnikov@ipiran.ru

Zakharov Victor N. (b. 1948) — Doctor of Science in technology, associate professor, scientific secretary, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vzakharov@ipiran.ru

РЕШЕНИЕ ЗАДАЧИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ЗАПАСОМ НЕПРЕРЫВНОГО ПРОДУКТА В СТОХАСТИЧЕСКОЙ МОДЕЛИ РЕГЕНЕРАЦИИ СО СЛУЧАЙНЫМИ СТОИМОСТНЫМИ ХАРАКТЕРИСТИКАМИ

П. В. Шнурков¹

Аннотация: Работа посвящена исследованию проблемы управления запасом определенного непрерывного продукта, эволюция объема которого описывается регенерирующим случайным процессом. Основная особенность рассматриваемой математической модели заключается в том, что стоимостные характеристики, определяющие цену поставки продукта потребителю и затраты, связанные с обеспечением функционирования системы, зависят от случайных внешних факторов. Случайный параметр управления представляет собой время от момента очередного пополнения запаса до момента последующего заказа на пополнение. Доказано, что стационарный стоимостный показатель эффективности управления в рассматриваемой задаче оптимизации по своей аналитической структуре представляет собой дробно-линейный интегральный функционал, зависящий от функции распределения параметра управления. Теоретическое решение задачи оптимизации основано на использовании теоремы об экстремуме дробно-линейных интегральных функционалов.

Ключевые слова: проблема управления запасом непрерывного продукта; случайные стоимостные характеристики системы; управляемые регенерирующие случайные процессы; дробно-линейные интегральные функционалы в задачах стохастического оптимального управления

DOI: 10.14357/19922264230407

EDN: WHZRNY

1 Введение

Применение теории регенерирующих случайных процессов и процессов восстановления для создания и исследования моделей управления запасами известно достаточно давно. Соответствующие динамические стохастические модели в той или иной форме рассматриваются не только в классических изданиях прошлых лет, но и в современных научных изданиях, посвященных теории логистики и управления запасами. Упомянем в этой связи работы [1, 2]. Ряд стохастических моделей управления запасами, рассмотренных в этих изданиях, имеет характерные признаки регенерирующих и полумарковских процессов. К сожалению, даже в упомянутых фундаментальных и весьма содержательных изданиях отсутствуют специальные разделы, посвященные применению регенерирующих и полумарковских случайных процессов в теории управления запасами.

Среди специальных научных исследований, посвященных применению регенерирующих процессов в теории управления запасами, выделим работы П. Айзигера и его соавторов [3, 4]. В частности, в работе [3] построена общая теоретическая конструкция использования регенерирующего процесса для создания моделей управления запасами и определе-

ния их вероятностных характеристик. В работе [4] рассмотрены примеры и численные иллюстрации применения полученных теоретических результатов.

В своих предшествующих исследованиях автор настоящей работы также получил ряд результатов, связанных с применением регенерирующих процессов к задачам управления запасами в различных конкретных стохастических моделях [5–8].

Основная отличительная особенность данного исследования заключается в стохастическом характере стоимостных характеристик модели, определяющих затраты и доходы, возникающие в ходе функционирования системы. По своей аналитической структуре все эти характеристики представляют собой суммы заданных детерминированных функций, которые можно назвать трендами, или направлениями эволюции, и случайных отклонений от трендов, которые в каждый момент времени имеют усеченные нормальные распределения с заданными параметрами. Проводя аналогию с известными физическими явлениями, можно сказать, что в данной модели стоимостные характеристики представляют собой случайные флуктуации, эволюционирующие в заданных границах, которые зависят от времени, прошедшего с момента

¹Национальный исследовательский университет «Высшая школа экономики», pshnurkov@hse.ru

начала очередного периода регенерации и объема запаса в эти моменты времени. В работе получены аналитические представления для классического стационарного стоимостного показателя эффективности, который по своему экономическому содержанию представляет собой среднюю удельную прибыль. Установлено, что этот показатель имеет структуру дробно-линейного интегрального функционала, зависящего от распределения случайного параметра управления. Как и в предшествующих работах [5–8], этот параметр управления совпадает с длительностью периода времени от момента очередного пополнения запаса (момента регенерации основного процесса) до момента следующего заказа на пополнение запаса. Теоретическое решение поставленной задачи оптимизации основывается на утверждении об экстремуме дробно-линейного интегрального функционала.

2 Описание математической модели исследуемой системы

В настоящем исследовании рассматривается стохастическая модель некоторой системы, предназначенной для хранения и поставки потребителю определенного продукта. Объем продукта измеряется вещественными числами и может принимать значения из множества $(-\infty; \tau]$, где τ — максимально допустимый объем продукта, который может находиться в системе. Следуя традиции, сложившейся в математической теории управления запасами, будем считать, что положительные значения объема продукта соответствуют реальному запасу, а отрицательные значения означают наличие неудовлетворенного спроса, или дефицита продукта.

Введем основное вероятностное пространство (Ω, \mathcal{A}, P) , на котором будут определены все рассматриваемые в дальнейшем стохастические объекты. Это пространство будет формально описывать реальный случайный эксперимент, в ходе которого реализуется исследуемая вероятностная модель управления запасом.

Обозначим через $x(t)$ объем запаса продукта в системе в момент времени t . Переменная величина $x(t) = x(\omega, t)$, $\omega \in \Omega$, $t \in [0, \infty)$, представляет собой случайный процесс с непрерывным временем и множеством состояний $(-\infty; \tau]$, свойства которого будут установлены в дальнейшем.

Спрос на продукт детерминирован. Товар реализуется со склада с постоянной скоростью $\alpha > 0$. Через время τ/α после начала потребления, когда уровень запаса продукта на складе достигает нуля, реализация товара продолжается с запасного хранилища с той же постоянной скоростью α .

Система начинает свое функционирование при максимальном объеме продукта: $x(0) = \tau$. В начальный момент работы системы принимается решение, когда провести заказ на следующую поставку продукта. Промежуток времени с момента начала работы до заказа на поставку описывается случайной величиной η с функцией распределения $G(u) = P(\eta < u)$. После проведенного заказа продукт проступает в систему через случайное время δ , имеющее распределение $H(z) = P(\delta < z)$.

Пополнение запаса осуществляется мгновенно. После пополнения запаса до максимального уровня τ эволюция системы происходит по закономерностям, описанным выше, независимо от прошлого.

Рассмотрим случайные моменты времени $t_n = t_n(\omega)$, $\omega \in \Omega$, $n = 0, 1, 2, \dots$, $t_0 = 0$, в которые происходит пополнение запаса до максимального уровня τ . Зафиксируем некоторый момент t_n . Период времени от момента пополнения запаса t_n до следующего заказа на пополнение является случайной величиной η_n с функцией распределения $G(u) = P(\eta_n < u)$.

Рассмотрим далее случайную величину $t'_n = t_n + \eta_n$, которая представляет собой момент заказа на поставку продукции. В момент t'_n начинается так называемый период задержки поставки. Обозначим случайную длительность этого периода через δ_n . Следующее пополнение запаса до максимального уровня τ осуществляется в момент времени $t_{n+1} = t'_n + \delta_n = t_n + \eta_n + \delta_n$. Далее эволюция процесса $x(t)$ происходит по тем же закономерностям, независимо от прошлого. Отсюда следует, что случайный процесс $x(t)$ относится к виду регенерирующих процессов, а моменты t_n , $n = 0, 1, 2, \dots$, представляют собой моменты его регенерации. Интервал времени $[t_n; t_{n+1})$ называется периодом регенерации данного процесса.

Траектория процесса $x(t)$ на интервале времени $[t_n; t_{n+1})$ определяется соотношением:

$$x(t) = x(t_n) - \alpha(t - t_n), \quad t_n \leq t < t_{n+1}.$$

Управление процессом (решение) принимается в начальный момент каждого периода регенерации $[t_n; t_{n+1})$, т.е. в момент времени t_n . Управление является случайной величиной $u_n = u_n(\omega)$, $\omega \in \Omega$, принимающей значения из множества допустимых управлений $U = [0, \infty)$, и совпадает со случайной величиной η_n , т.е. выполняется соотношение $P(\omega \in \Omega : u_n = \eta_n) = 1$. Функция распределения $G(u) = P(\eta_n < u)$ называется управляющим распределением.

Задача оптимального управления регенерирующим процессом $x(t)$ формально заключается в нахождении вероятностного распределения $G^*(u)$,

которое бы доставляло глобальный экстремум стоимостному показателю средней удельной прибыли вида $I(G)$. Аналитическое представление для него будет получено в дальнейшем.

3 Стоимостные характеристики стохастической модели

Предположим, что с каждым периодом регенерации $[t_n, t_{n+1})$ связаны семейства случайных величин $s_1^{(n)}(t), s_2^{(n)}(t), r_1^{(n)}(t), r_2^{(n)}(t), n = 0, 1, 2, \dots$, заданных на исходном вероятностном пространстве (Ω, \mathcal{A}, P) , зависящих от параметра времени t , которые описывают внешние факторы, влияющие на стоимостные характеристики модели. Предположим также, что указанные факторы удовлетворяют следующим условиям.

1. Время эволюции каждого фактора с номером n связано с соответствующим периодом регенерации $[t_n, t_{n+1})$ и изменяется от 0 до случайной длительности этого периода $\Delta t_n = t_{n+1} - t_n$.
2. Все рассматриваемые стохастические семейства $\{s_1^{(n)}(t), n = 0, 1, 2, \dots\}, \{s_2^{(n)}(t), n = 0, 1, 2, \dots\}; \{r_1^{(n)}(t), n = 0, 1, 2, \dots\}$ и $\{r_2^{(n)}(t), n = 0, 1, 2, \dots\}$ состоят из случайных процессов, которые независимы при различных значениях n и имеют одинаковые конечномерные распределения. Таким образом, указанные случайные факторы действуют независимо на различных периодах регенерации.

Заметим, что все введенные факторы могут совпадать. Тогда стоимостные характеристики модели будут зависеть от одного случайного фактора, описывающего влияние внешней среды.

Будем предполагать, что стоимостные характеристики модели зависят от указанных внешних факторов. Введем следующие обозначения для стоимостных характеристик:

$c_1(x, s_1)$ — затраты в единицу времени, связанные с хранением запаса объема $x \geq 0$ при условии, что внешний фактор $s_1(t)$ принимает значение s_1 ;

$c_2(x, s_2)$ — затраты (штрафы) в единицу времени, связанные с реализацией продукции объема $|x|$ с запасного склада, $x < 0$, при условии, что внешний фактор $s_2(t)$ принимает значение s_2 ;

$d_1(x, r_1)$ — цена, которую платит потребитель в единицу времени за единицу продукта при условии, что объем запаса принимает значение $x > 0$, а внешний фактор $r_1(t)$ принял значение r_1 ;

$d_2(x, r_2)$ — цена, которую платит потребитель в единицу времени за единицу продукта в период дефицита, т. е. при условиях, что объем запаса принимает отрицательное значение $x < 0$, а внешний фактор $r_2(t)$ принял значение r_2 .

Теперь перейдем к описанию конкретного влияния случайных факторов на стоимостные характеристики. Будем предполагать, что в рассматриваемой стохастической модели все стоимостные характеристики представляются в виде суммы детерминированной составляющей, или тренда, и случайной составляющей, или флуктуации, описывающей отклонения от тренда.

Пусть заданы следующие детерминированные функции: $c_1^{\max}(x(t)), c_1^{\min}(x(t)), c_1^{\text{trend}}(x(t)), c_2^{\max}(x(t)), c_2^{\min}(x(t)), c_2^{\text{trend}}(x(t)), d_1^{\max}(x(t)), d_1^{\min}(x(t)), d_1^{\text{trend}}(x(t)), d_2^{\max}(x(t)), d_2^{\min}(x(t)), d_2^{\text{trend}}(x(t))$, которые определены при всех значениях аргумента $x(t) \in (-\infty; \tau]$, принимают только неотрицательные значения и для которых справедливы следующие соотношения:

$$\begin{aligned} c_1^{\min}(x(t)) &< c_1^{\text{trend}}(x(t)) < c_1^{\max}(x(t)); \\ c_2^{\min}(x(t)) &< c_2^{\text{trend}}(x(t)) < c_2^{\max}(x(t)); \\ d_1^{\min}(x(t)) &< d_1^{\text{trend}}(x(t)) < d_1^{\max}(x(t)); \\ d_2^{\min}(x(t)) &< d_2^{\text{trend}}(x(t)) < d_2^{\max}(x(t)), \end{aligned}$$

где $x(t)$ — объем запаса в системе через время t после начала очередного периода регенерации:

$$x(t) = \tau - \alpha t, \quad t \geq 0.$$

Указанные функции определяют переменные границы областей, в которых могут принимать значения стоимостные характеристики модели, а также линии трендов этих характеристик.

Обозначим:

$$\begin{aligned} \Delta c_1^{(+)}(x(t)) &= c_1^{\max}(x(t)) - c_1^{\text{trend}}(x(t)); \\ \Delta c_1^{(-)}(x(t)) &= c_1^{\text{trend}}(x(t)) - c_1^{\min}(x(t)). \end{aligned}$$

Соответствующие функции $\Delta c_2^{(+)}(x(t)), \Delta c_2^{(-)}(x(t)), \Delta d_1^{(+)}(x(t)), \Delta d_1^{(-)}(x(t)), \Delta d_2^{(+)}(x(t))$ и $\Delta d_2^{(-)}(x(t))$ определяются аналогично.

Будем предполагать, что в любой момент времени, отстоящий на величину $t \geq 0$ от начала очередного периода регенерации, случайные стоимостные характеристики определяются следующими соотношениями:

$$c_1(x(t), s_1^{(n)}(t)) = c_1^{\text{trend}}(x(t)) + \Delta \beta_1(x(t)); \quad (1)$$

$$c_2(x(t), s_2^{(n)}(t)) = c_2^{\text{trend}}(x(t)) + \Delta \beta_2(x(t)); \quad (2)$$

$$d_1(x(t), r_1^{(n)}(t)) = d_1^{\text{trend}}(x(t)) + \Delta \gamma_1(x(t)); \quad (3)$$

$$d_2(x(t), r_2^{(n)}(t)) = d_2^{\text{trend}}(x(t)) + \Delta \gamma_2(x(t)). \quad (4)$$

Случайные функции $\Delta\beta_1(x(t))$, $\Delta\beta_2(x(t))$, $\Delta\gamma_1(x(t))$ и $\Delta\gamma_2(x(t))$, фигурирующие в правых частях соотношений (1)–(4), представляют собой отклонения значений стоимостных характеристик от линий трендов. Предполагается, что при каждом значении t эти величины имеют усеченные нормальные распределения на отрезках $[-\Delta c_k^{(-)}(x(t)); \Delta c_k^{(+)}(x(t))]$, $k = 1, 2$, и $[-\Delta d_l^{(-)}(x(t)); \Delta d_l^{(+)}(x(t))]$, $l = 1, 2$, соответственно. Параметры этих распределений считаются заданными, причем

$$\begin{aligned} E \Delta\beta_1(x(t)) &= a_1^{(1)}(t); & E \Delta\beta_2(x(t)) &= a_2^{(1)}(t); \\ E \Delta\gamma_1(x(t)) &= a_1^{(2)}(t); & E \Delta\gamma_2(x(t)) &= a_2^{(2)}(t). \end{aligned}$$

Замечание 1. Случайные отклонения от линий трендов могут иметь произвольные усеченные распределения, сосредоточенные на конечных интервалах с переменными границами, для которых известны их вероятностные характеристики.

При сделанных предположениях математические ожидания стоимостных характеристик задаются равенствами:

$$\begin{aligned} \bar{c}_1(x(t)) &= E c_1(x(t), s_1^{(n)}(t)) = \\ &= c_1^{\text{trend}}(x(t)) + a_1^{(1)}(t), \quad x(t) \geq 0; \end{aligned} \quad (5)$$

$$\begin{aligned} \bar{c}_2(x(t)) &= E c_2(x(t), s_2^{(n)}(t)) = \\ &= c_2^{\text{trend}}(x(t)) + a_2^{(1)}(t), \quad x(t) < 0; \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{d}_1(x(t)) &= E d_1(x(t), r_1^{(n)}(t)) = \\ &= d_1^{\text{trend}}(x(t)) + a_1^{(2)}(t), \quad x(t) \geq 0; \end{aligned} \quad (7)$$

$$\begin{aligned} \bar{d}_2(x(t)) &= E d_2(x(t), r_2^{(n)}(t)) = \\ &= d_2^{\text{trend}}(x(t)) + a_2^{(2)}(t), \quad x(t) < 0. \end{aligned} \quad (8)$$

4 Численное моделирование стоимостных характеристик

В данном разделе рассмотрены два примера наборов стоимостных характеристик построенной модели. В первом примере математические ожидания случайных флуктуаций равны нулю и математические ожидания стоимостных характеристик совпадают с линиями трендов. Во втором примере параметры случайных флуктуаций сконструированы таким образом, что математические ожидания стоимостных характеристик смещены относительно линий трендов.

Пример 1. Зададим функции, определяющие цену продукта:

$$d_1^{\text{max}}(x) = \begin{cases} 100 - \frac{x}{100}, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$d_1^{\text{min}}(x) = \begin{cases} 90 - \frac{x}{100}, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$d_1^{\text{trend}}(x) = \begin{cases} 95 - \frac{x}{100}, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$d_2^{\text{max}}(x) = \begin{cases} 0, & x \geq 0; \\ 100e^{x/100}, & x < 0; \end{cases}$$

$$d_2^{\text{min}}(x) = \begin{cases} 0, & x \geq 0; \\ 90e^{x/100}, & x < 0; \end{cases}$$

$$d_2^{\text{trend}}(x) = \begin{cases} 0, & x \geq 0; \\ 95e^{x/100}, & x < 0. \end{cases}$$

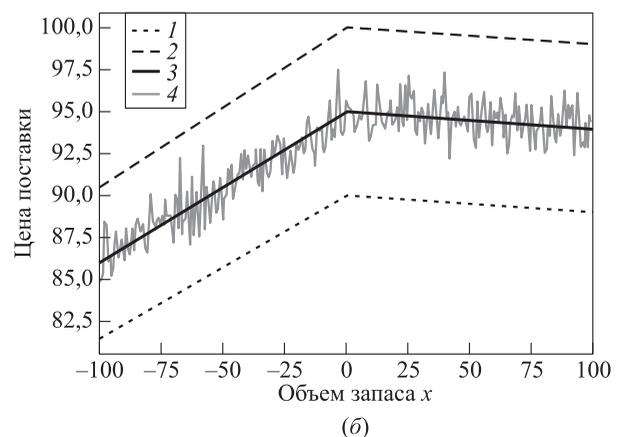
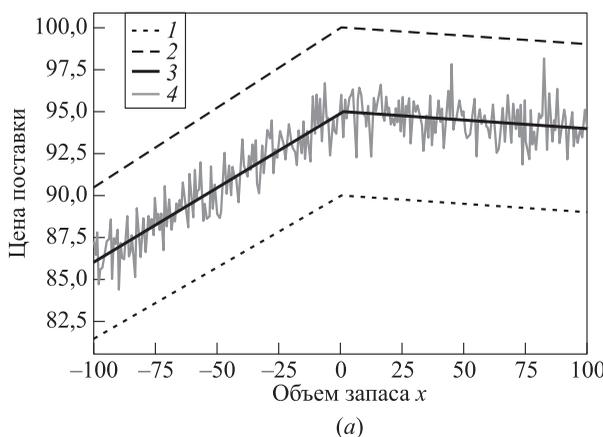


Рис. 1 Пример 1. Возможные траектории изменения цены единицы продукта на периоде регенерации для реализаций 1 (а) и 2 (б): 1 — минимальные значения; 2 — максимальные значения; 3 — тренд; 4 — истинные значения

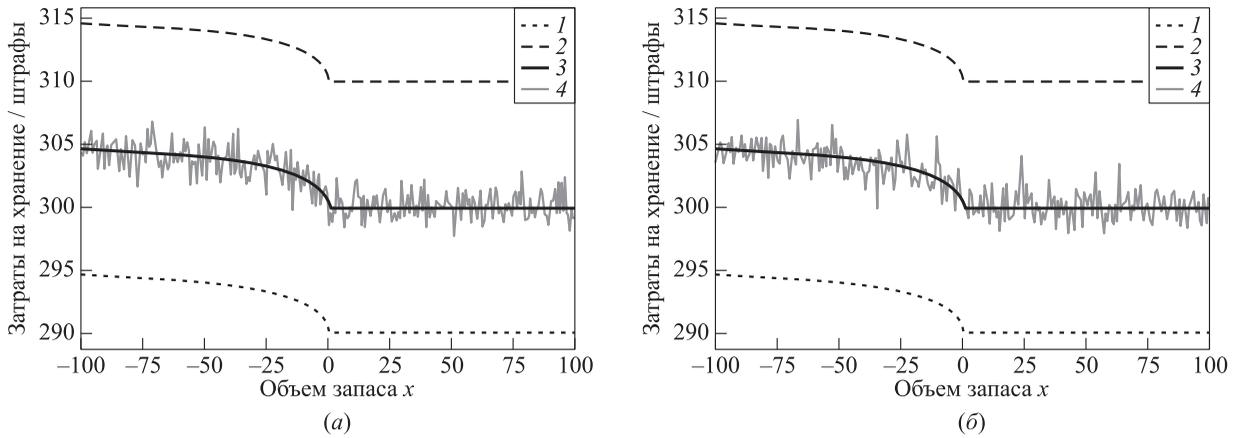


Рис. 2 Пример 1. Возможные траектории изменения затрат на хранение и штрафов, связанных с дефицитом единицы продукта для реализаций 1 (а) и 2 (б): 1 — минимальные значения; 2 — максимальные значения; 3 — тренд; 4 — истинные значения

Зададим далее функции, определяющие величины затрат, связанных с функционированием системы, при помощи следующих соотношений:

$$c_1^{\max}(x) = \begin{cases} 310, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$c_1^{\min}(x) = \begin{cases} 290, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$c_1^{\text{trend}}(x) = \begin{cases} 300, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

$$c_2^{\max}(x) = \begin{cases} 0, & x \geq 0; \\ 310 + \ln(1 - x), & x < 0; \end{cases}$$

$$c_2^{\min}(x) = \begin{cases} 0, & x \geq 0; \\ 290 + \ln(1 - x), & x < 0; \end{cases}$$

$$c_2^{\text{trend}}(x) = \begin{cases} 0, & x \geq 0; \\ 300 + \ln(1 - x), & x < 0. \end{cases}$$

Предположим также, что в данном примере математические ожидания уклонений от линии трендов равны 0:

$$a_1^{(1)}(t) = a_2^{(1)}(t) = a_1^{(2)}(t) = a_2^{(2)}(t) = 0,$$

а соответствующие дисперсии равны 1.

Траектории (реализации) случайной цены представлены на рис. 1, траектории случайных затрат на рис. 2. Заметим, что для каждой из указанных случайных функций приведены по две возможные реализации.

Пример 2. Предположим, что функции, определяющие границы областей изменения и тренды для

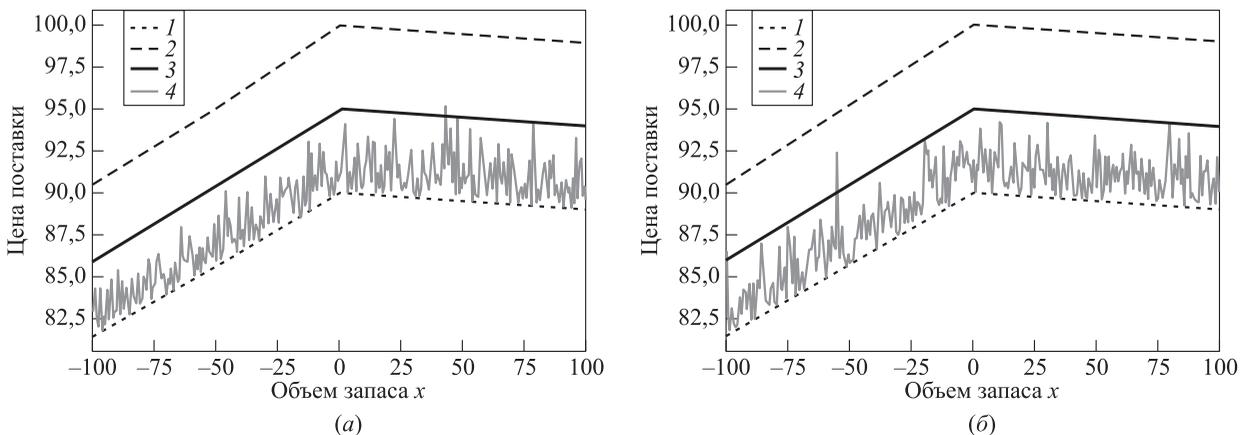


Рис. 3 Пример 2. Возможные траектории цены единицы товара на периоде регенерации для реализаций 1 (а) и 2 (б): 1 — минимальные значения; 2 — максимальные значения; 3 — тренд; 4 — истинные значения

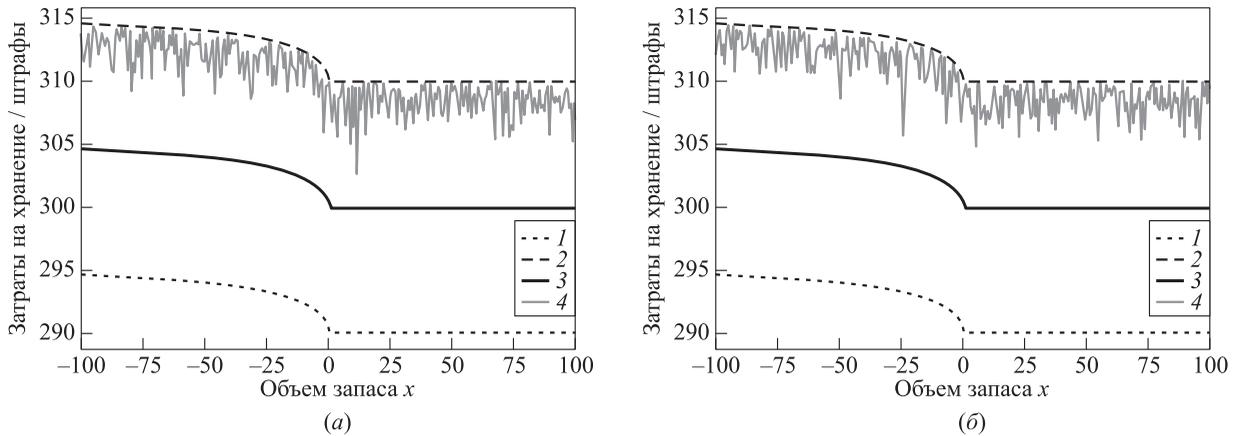


Рис. 4 Пример 2. Возможные траектории изменения затрат на периоде регенерации для реализаций 1 (а) и 2 (б): 1 — минимальные значения; 2 — максимальные значения; 3 — тренд; 4 — истинные значения

стоимостных характеристик, совпадают с теми, которые были заданы в примере 1. В то же время математические ожидания случайных отклонений от трендов (флуктуаций) задаются следующим образом:

$$\begin{aligned}
 a_1^{(1)}(t) &= \Delta c_1^{(+)} - \frac{\Delta c_1^{(+)} + \Delta c_1^{(-)}}{3}; \\
 a_2^{(1)}(t) &= \Delta c_2^{(+)} - \frac{\Delta c_2^{(+)} + \Delta c_2^{(-)}}{3}; \\
 a_1^{(2)}(t) &= \Delta d_1^{(-)} + \frac{\Delta d_1^{(+)} + \Delta d_1^{(-)}}{4}; \\
 a_2^{(2)}(t) &= \Delta d_2^{(-)} + \frac{\Delta d_2^{(+)} + \Delta d_2^{(-)}}{4}.
 \end{aligned}$$

Дисперсии соответствующих отклонений предполагаются равными единице.

В данном варианте траектории случайной цены, образующиеся при заданных характеристиках модели, представлены на рис. 3, а траектории случайных затрат — на рис. 4. Как и в примере 1, для каждой из указанных случайных функций приведены по две возможные реализации.

5 Аналитическое представление стационарного стоимостного показателя эффективности

В данном разделе будет получено явное аналитическое представление для стационарного стоимостного показателя эффективности управления — средней удельной прибыли на периоде регенерации. Данный показатель представим в виде дробно-линейного интегрального функционала от вероятностного распределения $G(\cdot)$, характеризующего случайный параметр управления.

Приведем сначала вспомогательный результат, связанный с представлением математического ожидания случайной прибыли, образующейся на произвольном фиксированном конечном интервале времени, вложенном в некоторый период регенерации.

Как и ранее, обозначим через $t_n, n = 0, 1, 2, \dots, t_0 = 0$, — случайные моменты пополнения запаса (моменты регенерации), $\Delta t_n = t_{n+1} - t_n, n = 0, 1, 2, \dots$, — случайные длительности периодов регенерации. Зафиксируем произвольные числа $t^{(0)}$ и $\Delta^{(0)}$, для которых $0 \leq t^{(0)} < t^{(0)} + \Delta^{(0)}$. Предположим, что для некоторого периода регенерации $[t_n, t_{n+1}]$ выполняется условие

$$t^{(0)} + \Delta^{(0)} < \Delta t_n.$$

Обозначим через $\tilde{S}(t^{(0)}, t^{(0)} + \Delta^{(0)})$ случайную прибыль, связанную с рассматриваемой моделью и образующуюся на интервале времени $[t_n + t^{(0)}, t_n + t^{(0)} + \Delta^{(0)}]$.

Обозначим также

$$S(t^{(0)}, t^{(0)} + \Delta^{(0)}) = \mathbb{E} \tilde{S}(t^{(0)}, t^{(0)} + \Delta^{(0)}).$$

Лемма 1. Математическое ожидание случайной прибыли, образующейся на заданном конечном интервале времени, вложенном в некоторый период регенерации, может быть аналитически выражено следующим образом:

$$\begin{aligned}
 S(t^{(0)}, t^{(0)} + \Delta^{(0)}) &= \\
 &= \int_{t^{(0)}}^{t^{(0)} + \Delta^{(0)}} [\alpha \bar{d}_1(\tau - \alpha t) - \bar{c}_1(\tau - \alpha t)] dt, \\
 &0 \leq t^{(0)} < t^{(0)} + \Delta^{(0)} \leq \frac{\tau}{\alpha};
 \end{aligned}$$

$$S(t^{(0)}, t^{(0)} + \Delta^{(0)}) = \int_{t^{(0)}}^{\tau/\alpha} [\alpha \bar{d}_1(\tau - \alpha t) - \bar{c}_1(\tau - \alpha t)] dt + \int_{\tau/\alpha}^{t^{(0)} + \Delta^{(0)}} [\alpha \bar{d}_2(\tau - \alpha t) - \bar{c}_2(\tau - \alpha t)] dt, \\ 0 \leq t^{(0)} < \frac{\tau}{\alpha} < t^{(0)} + \Delta^{(0)}.$$

Доказательство леммы 1 основано на использовании утверждения о математическом ожидании интеграла от случайной функции, приведенного в монографии Липцера и Ширяева [9, гл. 1, § 1], и интегральном представлении для величины прибыли, образующейся на заданном конечном интервале времени.

Теперь возникает возможность найти явное представление для стационарного показателя эффективности управления — средней удельной прибыли. Соответствующий результат можно сформулировать в виде теоремы.

Теорема 1. *Предположим, что для любого управляющего вероятностного распределения $G(\cdot)$ математическое ожидание длительности периода регенерации строго положительно. Тогда стационарный показатель средней удельной прибыли в рассматриваемой стохастической модели регенерации представляется в виде дробно-линейного интегрального функционала от распределения $G(\cdot)$, который имеет вид:*

$$I(G) = \frac{\int_0^\infty A(u) dG(u)}{\int_0^\infty B(u) dG(u)}. \quad (9)$$

Здесь подынтегральные функции числителя и знаменателя задаются следующими формулами:

$$A(u) = \int_0^{\tau/\alpha - u} \left[\int_0^{u+z} [\alpha \bar{d}_1(\tau - \alpha t) - \bar{c}_1(\tau - \alpha t)] dt \right] dH(z) + \int_{\tau/\alpha - u}^\infty \left[\int_0^{\tau/\alpha} [\alpha \bar{d}_1(\tau - \alpha t) - \bar{c}_1(\tau - \alpha t)] dt + \int_{\tau/\alpha}^{u+z} [\alpha \bar{d}_2(\tau - \alpha t) - \bar{c}_2(\tau - \alpha t)] dt \right] dH(z), \\ 0 \leq u \leq \frac{\tau}{\alpha}; \quad (10)$$

$$A(u) = \int_0^\infty \left[\int_0^{\tau/\alpha} [\alpha \bar{d}_1(\tau - \alpha t) - \bar{c}_1(\tau - \alpha t)] dt + \int_{\tau/\alpha}^{u+z} [\alpha \bar{d}_2(\tau - \alpha t) - \bar{c}_2(\tau - \alpha t)] dt \right] dH(z), \\ u > \frac{\tau}{\alpha}; \quad (11)$$

$$B(u) = u + \int_0^\infty z dH(z) = u + T_0, \quad u \geq 0, \quad (12)$$

где $T_0 = E \delta$ — математическое ожидание случайной длительности задержки поставки (заданная величина).

Доказательство теоремы 1 основывается на использовании известного теоретического утверждения для регенерирующих процессов, а именно: эргодической теоремы для стоимостных аддитивных функционалов, связанных с этими процессами. Кроме того, при доказательстве используется утверждение леммы 1 и интегральные представления для величины прибыли при различных соотношениях между параметрами.

Замечание 2. Стоимостные характеристики модели $\bar{c}_1(x)$, $\bar{d}_1(x)$, $x \geq 0$, и $\bar{c}_2(x)$, $\bar{d}_2(x)$, $x < 0$, фигурирующие в правых частях соотношений (10) и (11), определяются равенствами (5)–(8). При заданных вероятностных характеристиках случайных цен и затрат эти функции становятся известными.

6 Общее решение задачи оптимального управления запасом

Проблема оптимального управления запасом в рассматриваемой стохастической модели регенерации может быть сформулирована как задача безусловного экстремума для дробно-линейного интегрального функционала (9):

$$I(G) \xrightarrow{G \in \Gamma} \max, \quad (13)$$

где Γ — множество функций распределения неотрицательных случайных величин, принимающих значения из множества допустимых управлений $U = [0, \infty)$.

Для решения экстремальной задачи (13) применим теорему о безусловном экстремуме дробно-линейного интегрального функционала, сформулированную и доказанную в работах автора настоящего исследования [10, 11]. Заметим, что

условия упомянутой теоремы выполнены. Действительно, если $T_0 > 0$, то имеют место соотношения:

$$B(u) = u + T_0 > 0, \quad u \in U;$$

$$\int_U B(u) dG(u) = \int_U [u + T_0] dG(u) > 0, \quad G \in \Gamma.$$

Согласно основному утверждению указанной теоремы, если основная функция данного дробно-линейного интегрального функционала

$$C(u) = \frac{A(u)}{B(u)}, \quad u \in U, \quad (14)$$

достигает своего глобального максимума на множестве $U = [0, \infty)$ в некоторой фиксированной точке $u_* \in U$, то решение экстремальной задачи (13) существует и достигается на вырожденном вероятностном распределении $G_{u_*}^*(u)$, сосредоточенном в точке u_* :

$$G_{u_*}^*(u) = \begin{cases} 0, & u \leq u_*; \\ 1, & u > u_*. \end{cases}$$

Таким образом, проблема оптимального управления в данной стохастической модели регенерации сводится к исследованию на глобальный максимум функции $C(u)$ вида (14), где функции $A(u)$ и $B(u)$ заданы явными аналитическими формулами (10)–(12).

7 Заключение

Подводя итог проведенного исследования проблемы оптимального управления запасом непрерывного продукта в стохастической модели регенерации со случайными стоимостными характеристиками, можно утверждать, что, как и в предшествующих исследованиях проблем управления запасом в различных регенерационных моделях [5–8], решение задачи оптимального управления сводится к исследованию на глобальный экстремум так называемой основной функции дробно-линейного интегрального функционала, выражающего стационарный стоимостный показатель эффективности управления. При этом указанная основная функция допускает явное аналитическое представление через математические ожидания исходных стоимостных характеристик модели.

Литература

1. *Porteus E. L.* Foundations of stochastic inventory theory. — Stanford, CA, USA: Stanford Business Books, 2002. 320 p.
2. *Axsäter S.* Inventory control. — International ser. in operations research & management science. — Springer, 2015. Vol. 225. 281 p. doi: 10.1007/978-3-319-15729-0.
3. *Iseger P., Oldenkamp E., Frenk J. B. G.* Inventory control and regenerative processes: Theory // SSRN Electronic J., 1999. 23 p. doi: 10.2139/ssrn.1014782.
4. *Iseger P., Oldenkamp E., Frenk J. B. G.* Inventory control and regenerative processes: Computations // SSRN Electronic J., 1999. 25 p. doi: 10.2139/ssrn.1014783.
5. *Шнурков П. В., Мельников П. В.* Исследование проблемы управления запасом непрерывного продукта при детерминированной задержке поставки // Автоматика и телемеханика, 2008. № 10. С. 93–113.
6. *Шнурков П. В., Пименова Е. Ю.* Оптимальное управление запасом непрерывного продукта в схеме регенерации с детерминированной задержкой поставки и периодом реального пополнения // Системы и средства информатики, 2017. Т. 27. № 4. С. 80–94. doi: 10.14357/08696527170406. EDN: ZSUSNP.
7. *Шнурков П. В., Вахтанов Н. А.* Исследование проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением и случайной задержкой поставки // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 54–61. doi: 10.14357/19922264190208. EDN: ZTINLF.
8. *Шнурков П. В., Вахтанов Н. А.* О решении проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 50–57. doi: 10.14357/19922264190308. EDN: BPKRGC.
9. *Липцер Р. Ш., Ширяев А. Н.* Статистика случайных процессов (нелинейная фильтрация и смежные вопросы). — М.: Наука, 1974. 696 с.
10. *Шнурков П. В.* О решении задачи безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. Акад. наук. Сер.: Математика, 2016. Т. 470. № 4. С. 387–392. doi: 10.7868/S0869565216280045. EDN: WLNMTN.
11. *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. doi: 10.14357/19922264160408. EDN: XGSITZ.

Поступила в редакцию 24.09.23

SOLUTION OF THE PROBLEM OF OPTIMAL CONTROL OF THE STOCK OF A CONTINUOUS PRODUCT IN A STOCHASTIC MODEL OF REGENERATION WITH RANDOM COST CHARACTERISTICS

P. V. Shnurkov

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

Abstract: The work is devoted to the study of the problem of managing the stock of a certain continuous product, the evolution of the volume of which is described by a regenerating stochastic process. The main feature of the considered mathematical model is that the cost characteristics that determine the price of supplying the product to the consumer and the costs associated with ensuring the functioning of the system depend on random external factors. The random control parameter is the time from the moment of the next replenishment of the stock to the moment of the next order for replenishment. It is proved that the stationary cost indicator of control efficiency in the optimization problem under consideration in its analytical structure is a fractional-linear integral functional depending on the distribution function of the control parameter. The theoretical solution of the optimization problem is based on the use of the extremum theorem for linear-fractional integral functionals.

Keywords: continuous product inventory control problem; random cost characteristics of the system; controlled regenerative stochastic processes; linear-fractional integral functionals in problems of stochastic optimal control

DOI: 10.14357/19922264230407

EDN: WHZRN Y

References

1. Porteus, E. L. 2002. *Foundations of stochastic inventory theory*. Stanford, CA: Stanford Business Book. 320 p.
2. Axsäter, S. 2015. *Inventory control*. International ser. in operations research & management science. Springer. Vol. 225. 281 p. doi: 10.1007/978-3-319-15729-0.
3. Iseger, P., E. Oldenkamp, and J. B. G. Frenk. 1999. Inventory control and regenerative processes: Theory. *SSRN Electronic J.* 23 p. doi: 10.2139/ssrn.1014782.
4. Iseger, P., E. Oldenkamp, and J. B. G. Frenk. 1999. Inventory control and regenerative processes: Computations. *SSRN Electronic J.* 25 p. doi: 10.2139/ssrn.1014783.
5. Shnurkov, P. V., and R. V. Mel'nikov. 2008. Analysis of the problem of continuous-product inventory control under deterministic lead time. *Automat. Rem. Contr.* 69(10):1734–1751. doi: 10.1134/S0005117908100081.
6. Shnurkov, P. V., and E. Yu. Pimenova. 2017. Optimal'noe upravleniye zapasom nepreryvnogo produkta v skheme regeneratsii s determinirovannoy zaderzhkoy postavki i periodom real'nogo popolneniya [Optimal inventory control of continuous product in regeneration theory with determinate delay of the delivery and the period of real replenishment]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):80–94. doi: 10.14357/08696527170406. EDN: ZSUSNP.
7. Shnurkov, P. V., and N. A. Vakhtanov. 2019. Issledovanie problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoy modeli regeneratsii s nepreryvno proiskhodyashchim potrebleniem i sluchaynoy zaderzhkoy postavki [Research of the optimal control problem of inventory of a discrete product in the stochastic regeneration model with continuously occurring consumption and random delivery delay]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):54–61. doi: 10.14357/19922264190208. EDN: ZTINLF.
8. Shnurkov, P. V., and N. A. Vakhtanov. 2019. O reshenii problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoy modeli regeneratsii s nepreryvno proiskhodyashchim potrebleniem [On the solution of the optimal control problem of inventory of a discrete product in the stochastic model of regeneration with continuously occurring consumption]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):50–57. doi: 10.14357/19922264190308. EDN: BPKRGC.
9. Liptser, R. S., and A. N. Shiryaev. 1974. *Statistika sluchaynykh protsessov (nelineynaya fil'tratsiya i smezhnyye voprosy)* [Statistics of random processes (nonlinear filtering and related problems)]. Moscow: Nauka. 696 p.
10. 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. doi: 10.1134/S1064562416050161. EDN: XFUKJH.
11. Shnurkov, P. V., A. K. Gorshenin, and V. V. Belousov. 2016. Analiticheskoe reshenie zadachi optimal'nogo upravleniya polumarkovskim protsessom s konechnym mnozhestvom sostoyaniy [Analytical solution of the optimal control task of a semi-Markov process with finite set of states]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):72–88. doi: 10.14357/19922264160408. EDN: XGSITZ.

Received September 24, 2023

Contributor

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

ГРАНИЦЫ НЕЗАВЕРШЕННОЙ РАБОТЫ В СИСТЕМЕ С ПОВТОРНЫМИ ВЫЗОВАМИ РАЗНЫХ КЛАССОВ И ПОКАЗАТЕЛЬНЫМ ВРЕМЕНЕМ ОБСЛУЖИВАНИЯ*

И. В. Пешкова¹

Аннотация: Исследуется односерверная система с повторными вызовами и пуассоновским входным потоком, в которую поступает M классов заявок. Для системы с временами обслуживания, имеющими показательное распределение, получены верхняя и нижняя границы для стационарной незавершенной работы. В качестве нижней границы выступает стационарная незавершенная работа в классической системе $M/H_M/1$ с гиперэкспоненциальным временем обслуживания. Верхней границей служит незавершенная работа в системе, в которой к времени обслуживания добавляется дополнительное время, равное интервалу между попытками попасть на сервер с самой «медленной орбиты». Полученные результаты численного моделирования подтверждают теоретические выводы.

Ключевые слова: система с повторными вызовами; незавершенная работа; стохастическая упорядоченность

DOI: 10.14357/19922264230408

EDN: UOKQRD

1 Введение

Модели систем с повторными вызовами широко используются для моделирования телефонных станций, кол-центров, систем связи, телекоммуникационных сетей. В работах [1, 2] изложены приложения и математические методы анализа систем с повторными вызовами, включая системы с постоянной интенсивностью повторов. В работе [3] телефонная станция была смоделирована с помощью такой системы. Аналогичная модель используется для описания широкого класса протоколов множественного доступа [4, 5]. В частности, в работе [6] показано, что постоянная интенсивность повторных вызовов снижает интенсивность попыток (при незапланированном множественном доступе) обратно пропорционально числу задержанных пакетов. В результате общая скорость повторной обработки становится нечувствительной к виртуальному «размеру орбиты» (числу отложенных пакетов). Более того, системы с повторными вызовами с постоянной интенсивностью вызовов использовались для описания ТСП-трафика с короткими НТТР-соединениями [7, 8] и оптико-электрической гибридной схемой разрешения конфликтов [9, 10].

Большинство же современных моделей повторных вызовов имеют сложную природу или конфигурацию, и поэтому для их исследования при-

меняются численные методы или имитационное моделирование.

Ранее в работе [11] была доказана теорема о верхней и нижней границах стационарной незавершенной работы для исходной системы с повторными вызовами с постоянной интенсивностью вызовов (см. теорему 1 ниже). Эта теорема стала основой анализа, развитого в данной статье.

В данном исследовании рассматривается частный случай односерверной системы с повторными вызовами с пуассоновским входным потоком и показательным распределением времени обслуживания, при этом время обслуживания и время нахождения на орбите зависят от класса заявки k . Для такой системы предлагается строить две классические системы с неограниченной очередью (с ожиданием) типа $M/G/1$: в первой (минорантной) системе время обслуживания представляет собой конечную смесь времен обслуживания заявок всех классов (т. е. имеет гиперэкспоненциальное распределение), во второй (мажорантной) системе ко времени обслуживания первой системы добавляется дополнительное время, равное интервалу между вызовами с самой «медленной орбиты». Более того, для минорантной системы получено распределение стационарного времени ожидания в явном виде для трех классов ($M = 3$). Системы, в которых время обслуживания задано в виде конечной смеси распределений, обсуждались ранее в работах [12, 13].

* Работа выполнена при финансовой поддержке РФФИ (проект 21-71-10135).

¹Петрозаводский государственный университет; Институт прикладных математических исследований Карельского научного центра Российской академии наук, iaminova@petsu.ru

Структура статьи следующая. В разд. 2 приведено описание модели системы с повторными вызовами, минорантной и мажорантной систем, а также основная теорема, полученная авторами в работе [11]. В разд. 3 получены коэффициенты загрузки, математические ожидания незавершенной нагрузки, а также преобразования Лапласа—Стилтьеса для незавершенной нагрузки в минорантной и мажорантной системах с показательным распределением времени обслуживания. В разд. 4 приведены результаты численного эксперимента для случая трех классов. При этом параметры для минорантной системы были использованы такие же, как в работе [14], в которой получена функция распределения стационарного времени пребывания. Отметим, что в работе [14] неверно указано, что полученное распределение — это распределение стационарного времени ожидания. В статье получена функция распределения стационарного времени ожидания для минорантной системы в явном виде. Для исходной системы с повторными вызовами и мажорантной системы проведены численные эксперименты и построены эмпирические функции распределения. Полученные результаты численного эксперимента иллюстрируют доказанную стохастическую упорядоченность стационарной незавершенной работы в рассмотренных системах.

2 Описание модели

Рассмотрим односерверную систему с повторными вызовами Σ , в которой обслуживаются M классов заявок. Заявки k -го класса поступают в систему в соответствии с пуассоновским потоком с параметром λ_k , $k = 1, \dots, M$. Если заявка застает сервер пустым, то она немедленно обслуживается, в противном случае, если сервер занят, заявка уходит на k -ю орбиту бесконечного объема, образуя очередь в порядке поступления на орбиту, $k = 1, \dots, M$. Первая в очереди на k -й орбите заявка производит независимые попытки присоединиться к обслуживанию на сервере через экспоненциальное время η_k . Интенсивность вызовов с орбиты не зависит от размера орбиты (т. е. от числа заявок на орбите), но может зависеть от класса орбиты k . Такие системы называют системами с постоянной интенсивностью вызовов.

Обозначим через t_n момент прихода n -й заявки в общий пуассоновский входной поток (образованный суперпозицией M входных потоков, $t_1 = 0$), $S_n(k)$ — время обслуживания n -й заявки k -го класса, $k = 1, \dots, M$, $n \geq 1$. Пусть последовательность независимых одинаково распределенных (н. о. р.) интервалов между приходами заявок $\{T_n := t_{n+1} -$

$- t_n, n \geq 1\}$ и последовательность времен обслуживания $\{S_n(k), n \geq 1\}$ независимы для каждого k -го класса. Предположим, что интервалы между приходами заявок с (непустой) k -й орбиты η_k распределены показательным и не зависят от размера орбиты (числа заявок на k -й орбите). Время обслуживания заявок k -го класса $S(k)$ имеет произвольное распределение с функцией распределения (ф. р.) $F_{S(k)}$, $k = 1, \dots, M$. (Далее в обозначениях опускаем индекс номера заявки.) Обозначим

$$\lambda = \sum_{k=1}^M \lambda_k; \quad \rho_k = \lambda_k \mathbb{E}S(k); \quad \rho = \sum_{k=1}^M \rho_k.$$

Пусть $W(t)$ есть незавершенная работа в момент времени t^- , и предположим, что система в начальный момент времени пуста: $W(0) = 0$. Обозначим $W_n = W(t_n)$, $n \geq 1$. Известно [15], что неравенство

$$\rho + \max_{k=1, \dots, M} \frac{\lambda}{\lambda + \eta_k} < 1 \quad (1)$$

служит достаточным условием стационарности системы. При таком условии существует предел

$$W_n \Rightarrow W, \quad n \rightarrow \infty$$

(где обозначим \Rightarrow сходимость по распределению). Функция распределения F_W стационарной незавершенной работы W для исходной системы Σ неизвестна. С другой стороны, W служит важной характеристикой качества обслуживания системы. Ниже мы построим верхнюю и нижнюю границы F_W , используя классические $M/G/1$ системы с неограниченной очередью, в которых время обслуживания представляется конечной смесью распределений.

Рассмотрим две новые системы: *минорантную систему* $\Sigma^{(1)}$ и *мажорантную систему* $\Sigma^{(2)}$. (Далее индекс (i) означает номер системы.) Входной поток во все три системы — пуассоновский с параметром λ (это суперпозиция входных потоков, образованных заявками разных классов).

Пусть в минорантной системе $\Sigma^{(1)}$ время обслуживания $S^{(1)} = S$ задано конечной смесью распределений вида

$$S = \sum_{k=1}^M I(k)S(k), \quad n \geq 1, \quad (2)$$

где $I(k)$ — индикатор, такой что $\mathbb{E}I(k) = p_k = \lambda_k/\lambda$; $S(k)$ — время обслуживания заявки k -го класса. Заметим, что система $\Sigma^{(1)}$ представляет собой классическую (с дисциплиной обслуживания первым пришел — первым обслужен) односерверную систему типа $M/G/1$ с неограниченной очередью, в которой время обслуживания (2) является конечной

смесью времен обслуживания заявок всех классов исходной системы.

В мажорантной системе $\Sigma^{(2)}$ — классической односерверной системе типа $M/G/1$ с неограниченной очередью — каждая заявка обслуживается на сервере в течение времени S , заданного соотношением (2), плюс время ξ_0 , имеющее показательное распределение с параметром

$$\mu_0 = \min_{1 \leq k \leq M} (\lambda + \eta_k),$$

т.е. $S^{(2)} = S + \xi_0$. Таким образом, случайная величина (с.в.) ξ_0 соответствует самой «медленной» орбите (с наибольшими интервалами между попытками). Заметим, что мажорантная система $\Sigma^{(2)}$ имеет другой коэффициент загрузки,

$$\rho^{(2)} = \lambda \mathbb{E}S + \frac{\lambda}{\mu_0} = \rho + \frac{\lambda}{\mu_0},$$

и условие стационарности (1) для нее принимает вид:

$$\rho^{(2)} < 1.$$

В работе [11] доказана следующая теорема, в которой даны верхняя и нижняя границы незавершенной работы W в исходной системе с повторными вызовами Σ .

Теорема 1. Пусть системы $\Sigma^{(1)}$, $\Sigma^{(2)}$ и Σ в начальный момент времени пусты, т.е.

$$W_1^{(1)} = W_1 = W_1^{(2)} = 0.$$

Тогда при выполнении условия (1) стационарные времена незавершенной работы стохастически упорядочены:

$$W^{(1)} \underset{\text{st}}{\leq} W \underset{\text{st}}{\leq} W^{(2)}, \quad (3)$$

где $W^{(1)} \underset{\text{st}}{\leq} W$ означает $\overline{F}_{W^{(1)}}(x) \leq \overline{F}_W(x)$ для любого $x \geq 0$, $\overline{F}_{W^{(1)}}(x) = 1 - F_{W^{(1)}}(x)$.

В следующем разделе применим данный результат для системы, в которой M классов заявок, имеющих показательное распределение времени обслуживания.

3 Границы незавершенной работы W в системе с показательным обслуживанием разных классов

Пусть в исходной системе с повторными вызовами Σ времена обслуживания k -го класса $S(k)$ имеют показательное распределение с ф.р.

$$F_{S(k)}(x) = 1 - e^{-\mu_k x}, \quad x \geq 0, \quad \mu_k > 0. \quad (4)$$

В качестве минорантной системы $\Sigma^{(1)}$ рассмотрим систему с неограниченной очередью $M/H_M/1$, в которой времена обслуживания $S^{(1)} = S$ имеют гиперэкспоненциальное распределение (представляются M -компонентной смесью показательных распределенных с.в. $S(k)$) с ф.р.

$$F_{S^{(1)}}(x) = 1 - \sum_{k=1}^M p_k e^{-\mu_k x}, \quad \mu_k > 0,$$

$$p_k \geq 0, \quad k = 1, \dots, M, \quad \sum_{i=k}^M p_k = 1.$$

Обозначим коэффициент загрузки $\rho^{(1)} = \sum_{k=1}^M \lambda p_k / \mu_k$ в системе $\Sigma^{(1)}$. Поскольку

$$\rho^{(1)} \leq \rho + \frac{\lambda}{\mu_0},$$

то, если условие стационарности (1) выполнено, система $\Sigma^{(1)}$ также стационарна.

Рассмотрим преобразование Лапласа–Стилтьеса:

$$\psi_{S_e}(z) = \int_0^{\infty} e^{-zt} dF_{S_e}(t),$$

где F_{S_e} — так называемый интегрированный хвост распределения с плотностью

$$f_{S_e}(x) = \frac{1}{\mathbb{E}S} \overline{F}_S(x), \quad x \geq 0.$$

Распределение F_{S_e} соответствует распределению стационарного перескока процесса восстановления, формируемого последовательностью н.о.р. времен обслуживания $\{S_n\}$ [16].

В работе [11] доказано, что преобразование Лапласа–Стилтьеса стационарной незавершенной работы $W^{(1)}$ выражается через преобразования Лапласа–Стилтьеса компонент смеси времен обслуживания в следующем виде:

$$\begin{aligned} \psi_{W^{(1)}}(z) &= \frac{1 - \rho}{z \left(1 - \rho \sum_{k=1}^M (\rho_k / \rho) \psi_{S_e(k)}(z) \right)} = \\ &= \frac{1 - \rho}{z \left(1 - \sum_{k=1}^M \rho_k \psi_{S_e(k)}(z) \right)}. \end{aligned} \quad (5)$$

Преобразование Лапласа–Стилтьеса для показательного распределения хорошо известно:

$$\psi_{S_e(k)}(z) = \frac{\mu_k}{\mu_k + z}.$$

Подставляя его в соотношение (5), получим

$$\psi_{W^{(1)}}(z) = \frac{1 - \sum_{k=1}^M \lambda_k / \mu_k}{z \left(1 - \sum_{k=1}^M \lambda_k / (\mu_k + z) \right)}.$$

Применяя формулу Полачика–Хинчина, получим среднюю стационарную незавершенную работу в системе $\Sigma^{(1)}$ в виде:

$$\mathbb{E}W^{(1)} = \frac{\lambda \mathbb{E}(S^{(1)})^2}{2(1 - \rho^{(1)})} = \frac{\sum_{k=1}^M \rho_k^2 + \rho^2}{2\lambda(1 - \rho)}. \quad (6)$$

Рассмотрим теперь мажорантную систему $\Sigma^{(2)}$, время обслуживания в которой равно сумме с. в. S с гиперэкспоненциальным распределением (4) и с. в. ξ_0 , т. е. $S^{(2)} = S + \xi_0$. Обозначим для простоты $\mu_0 = \min_{1 \leq k \leq M} (\lambda + \eta_k)$, и пусть с. в. ξ_0 имеет показательное распределение с параметром μ_0 . Условие стационарности в такой системе совпадает с (1).

Известно [11], что в системе $\Sigma^{(2)}$ преобразование Лапласа–Стилтьеса для стационарной незавершенной работы $W^{(2)}$ имеет следующий вид:

$$\psi_{W^{(2)}}(z) = \left(1 - \rho - \frac{\lambda}{\mu_0} \right) / \left(z \left(1 - \frac{\mu_0}{\mu_0 + z} \left(\sum_{k=1}^M \rho_k \psi_{S_e(k)}(z) + \frac{\lambda}{\mu_0} \right) \right) \right). \quad (7)$$

Подставив $\psi_{S_e(k)}(z) = \mu_k / (\mu_k + z)$ в (7), получим

$$\psi_{W^{(2)}}(z) = \left(1 - \sum_{k=1}^M \frac{\lambda_k}{\mu_k} - \frac{\lambda}{\mu_0} \right) / \left(z \left(1 - \frac{\mu_0}{\mu_0 + z} \left(\sum_{k=1}^M \frac{\lambda_k}{\mu_k + z} + \frac{\lambda}{\mu_0} \right) \right) \right).$$

Аналогично формуле (6) можно получить среднюю стационарную незавершенную работу

$$\begin{aligned} \mathbb{E}W^{(2)} &= \frac{\lambda \mathbb{E}(S^{(2)})^2}{2(1 - \rho^{(2)})} = \\ &= \frac{\mu_0^2 \left(\rho^2 + \sum_{k=1}^M \rho_k^2 \right) + 2\lambda(\lambda + \rho\mu_0)}{2\lambda\mu_0(\mu_0 - \rho\mu_0 - \lambda)}. \quad (8) \end{aligned}$$

Из теоремы 1 следует, что стационарные времена незавершенной работы стохастически упорядочены:

$$W^{(1)} \underset{st}{\leq} W \underset{st}{\leq} W^{(2)},$$

а следовательно, их математические ожидания также упорядочены [17]:

$$\mathbb{E}W^{(1)} \leq \mathbb{E}W \leq \mathbb{E}W^{(2)}.$$

Действительно, легко проверить, что

$$\begin{aligned} \mathbb{E}W^{(2)} &= \frac{\lambda \mathbb{E}(S^{(2)})^2}{2(1 - \rho^{(2)})} = \\ &= \frac{\mu_0^2 \lambda \mathbb{E}(S^{(1)})^2 + 2\lambda(\lambda + \rho\mu_0)}{\mu_0^2 2(1 - \rho^{(1)}) - 2\lambda^2 \mu_0} \geq \mathbb{E}W^{(1)}. \end{aligned}$$

4 Численный эксперимент

В качестве примера рассмотрим систему с повторными вызовами Σ с тремя классами заявок ($M = 3$), в которую поступает пуассоновский поток с интенсивностью $\lambda = 10$. Пусть $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$, $\mu_1 = 10$, $\mu_2 = 30$ и $\mu_3 = 60$. Будем полагать, что η_k принимают значения $\eta_1 = 50$, $\eta_2 = 100$ и $\eta_3 = 150$. В этом случае $\mu_0 = \lambda + \eta_1 = 60$, коэффициенты загрузки равны

$$\rho^{(1)} = \rho = \sum_{k=1}^3 \frac{\lambda p_k}{\mu_k} = \frac{23}{36}; \quad \rho^{(2)} = \rho + \frac{\lambda}{\mu_0} = \frac{29}{36} < 1$$

и условие стационарности (1) выполнено. По формулам (6)–(8) находим математические ожидания $\mathbb{E}W^{(1)} \approx 0,093$ и $\mathbb{E}W^{(2)} \approx 0,106$.

Рассмотрим минорантную систему $\Sigma^{(1)}$, в которой время обслуживания имеет гиперэкспоненциальное распределение

$$\bar{F}_S(x) = \sum_{k=1}^3 p_k e^{-\mu_k x}.$$

Для такой классической системы $M/G/1$ в работе [14] получено распределение числа клиентов в системе в стационарном режиме $N^{(1)}$ в следующем виде:

$$\pi_n^* = \mathbb{P} \{ N^{(1)} = n \} = \sum_{k=1}^3 \beta_k (r_k)^n,$$

где параметры β_k и r_k вычислены и равны

$$\begin{aligned} \beta_1 &= 0,040; \quad \beta_2 = 0,075; \quad \beta_3 = 0,245; \\ r_1 &= 0,146; \quad r_2 = 0,268; \quad r_3 = 0,711. \end{aligned}$$

В работе [14] с помощью результата [18] получено стационарное распределение времени пребывания клиента в системе $V^{(1)}$ с хвостом ф. р. вида

$$\bar{F}_{V^{(1)}}(x) = \sum_{k=1}^3 \gamma_k e^{-\theta_k x},$$

где коэффициенты γ_k и параметры θ_k для исходных параметров равны

$$\begin{aligned} \gamma_1 &= 0,047; & \gamma_2 &= 0,103; & \gamma_3 &= 0,850; \\ \theta_1 &= 58,633; & \theta_2 &= 27,307; & \theta_3 &= 4,059. \end{aligned}$$

При этом в утверждении теоремы 3 из работы [14] было ошибочно указано, что $\sum_{k=1}^3 \gamma_k = 1 - \rho^{(1)}$ (что было бы верно, если бы распределение $F_V^{(1)}$ соответствовало времени ожидания). На самом деле легко проверить, что $\sum_{k=1}^3 \gamma_k = 1$. Для исправления этой неточности повторим вывод, получив выражение для стационарного времени ожидания в системе (что соответствует незавершенной работе в системе в момент прихода клиента). Заметим, что π_{n+1}^* есть стационарная вероятность наблюдать n клиентов в очереди, т. е.

$$\mathbb{P}\{Q^{(1)} = n\} = \pi_{n+1}^*,$$

где $Q^{(1)}$ есть число клиентов в очереди в стационарном режиме.

Вычислив производящую функцию вероятностей $\pi(z)$ для $Q^{(1)}$, получим

$$\begin{aligned} \pi(z) &= \sum_{n=0}^{\infty} z^n \pi_{n+1}^* = \sum_{k=1}^3 \frac{\beta_k}{z} \sum_{n=1}^{\infty} (r_k z)^n = \\ &= \sum_{k=1}^3 \frac{\beta_k}{z} \left(\frac{1}{1 - r_k z} - 1 \right) = \sum_{k=1}^3 \frac{\beta_k r_k}{1 - r_k z}. \end{aligned}$$

С другой стороны, производящая функция стационарной очереди $\pi(z)$ и преобразование Лапласа–Стилтьеса для стационарного времени ожидания $\psi_{W^{(1)}}(z)$ связаны формулой:

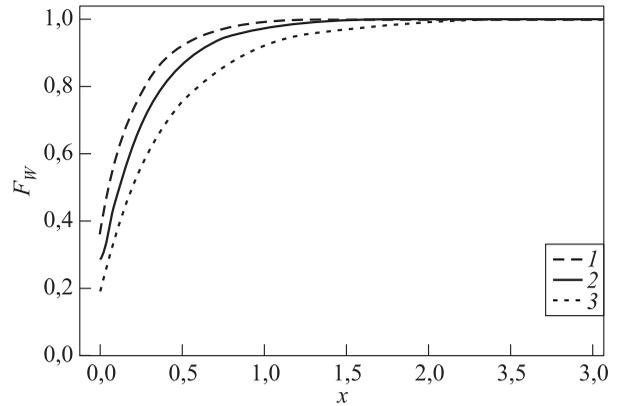
$$\begin{aligned} \pi(z) &= \sum_{n=0}^{\infty} \int_{0^-}^{\infty} z^n e^{-\lambda x} \frac{(\lambda x)^n}{n!} dF_{W^{(1)}}(x) = \\ &= \int_{0^-}^{\infty} e^{-(\lambda - \lambda z)x} dF_{W^{(1)}}(x) = \\ &= \psi_{W^{(1)}}(\lambda - \lambda z) + (1 - \rho^{(1)}), \end{aligned}$$

где $F_{W^{(1)}}(0) = (1 - \rho^{(1)})$ — скачок ф. р. в нуле. Сделав замену переменной $s = \lambda - \lambda z$, получим

$$\psi_{W^{(1)}}(s) = \sum_{k=1}^3 \frac{\beta_k r_k}{1 - r_k(1 - s/\lambda)} = \sum_{k=1}^3 \frac{\beta_k r_k}{1 - r_k} \frac{\theta_k}{\theta_k + s},$$

где, как и в работе [14],

$$\theta_k = \frac{\lambda(1 - r_k)}{r_k}.$$



Функции распределения в нижней $\Sigma^{(1)}$ (1), исходной Σ (2) и верхней $\Sigma^{(2)}$ (3) системах при $\lambda = 10$, $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$, $\mu_1 = 10$, $\mu_2 = 30$, $\mu_3 = 60$, $\eta_1 = 50$, $\eta_2 = 100$ и $\eta_3 = 150$

Таким образом, $\psi_{W^{(1)}}$ соответствует взвешенной сумме показательных распределений. Отметим при этом, что, в отличие от [14], коэффициенты смеси имеют вид:

$$\hat{\gamma}_k = \frac{\beta_k r_k}{1 - r_k}.$$

Таким образом,

$$\bar{F}_{W^{(1)}}(x) = \sum_{k=1}^3 \hat{\gamma}_k e^{-\theta_k x} + (1 - \rho^{(1)}), \quad (9)$$

где $\hat{\gamma}_1 = 0,007$, $\hat{\gamma}_2 = 0,027$ и $\hat{\gamma}_3 = 0,604$. Заметим, что $\sum_{k=1}^3 \hat{\gamma}_k = \rho^{(1)} \approx 0,638 < 1$.

Воспользуемся выражением (9) для построения ф. р. в системе $\Sigma^{(1)}$, а для построения оценок в исходной Σ и верхней $\Sigma^{(2)}$ системах воспользуемся имитационным моделированием. Построим графики (эмпирических) ф. р. для незавершенной работы в трех системах. Как видно на рисунке, стохастический порядок (3) для стационарных времен ожидания выполнен, что и следовало ожидать.

5 Заключение

В работе показано, что для исходной системы с повторными вызовами можно построить минорантную и мажорантную системы так, что стационарная незавершенная нагрузка во всех трех системах будет стохастически упорядочена. Численный эксперимент для системы с показательными временами обслуживания подтверждает теоретические выводы. При этом в качестве примера рассмотрены такие параметры (как в работе [14]), для которых получена ф. р. стационарного времени ожидания в явном виде в минорантной системе.

Литература

1. *Artalejo J. R.* Accessible bibliography on retrial queues // *Math. Comput. Model.*, 1999. Vol. 30. Iss. 3-4. P. 1–6. doi: 10.1016/S0895-7177(99)00128-4.
2. *Artalejo J., Gomez-Corral A.* Retrial queueing systems: A computational approach. — Springer, 2008. 318 p. doi: 10.1007/978-3-540-78725-9.
3. *Fayolle G.* A simple telephone exchange with delayed feedbacks // *Seminar (International) on Teletraffic Analysis and Computer Performance Evaluation Proceedings.* — Elsevier Science, 1986. P. 245–253.
4. *Choi B. D., Shin Y. W., Ahn W. C.* Retrial queues with collision arising from unslotted CSMA/CD protocol // *Queueing Syst.*, 1992. Vol. 11. P. 335–356. doi: 10.1007/BF01163860.
5. *Choi B. D., Rhee K. H., Park K. K.* The $M/G/1$ retrial queue with retrial rate control policy // *Probab. Eng. Inform. Sc.*, 1993. Vol. 7. P. 29–46. doi: 10.1017/S0269964800002771.
6. *Bertsekas D., Gallager R.* Data networks. — Athena Scientific, 2021. 570 p.
7. *Avrachenkov K., Yechiali U.* Retrial networks with finite buffers and their application to Internet data traffic // *Probab. Eng. Inform. Sc.*, 2008. Vol. 22. P. 519–536. doi: 10.1017/S0269964808000314.
8. *Avrachenkov K., Yechiali U.* On tandem blocking queues with a common retrial queue // *Comput. Oper. Res.*, 2010. Vol. 37. Iss. 7. P. 1174–1180. doi: 10.1016/j.cor.2009.10.004.
9. *Yao S., Xue F., Mukherjee B., Yoo S. J. B., Dixit S.* Electrical ingress buffering and traffic aggregation for optical packet switching and their effect on TCP-level performance in optical mesh networks // *IEEE Commun. Mag.*, 2002. Vol. 40. Iss. 9. P. 66–72. doi: 10.1109/MCOM.2002.1031831.
10. *Wong E. W. M., Andrew L. L. H., Cui T., Moran B., Zalesky A., Tucker R. S., Zukerman M.* Towards a bufferless optical internet // *J. Lightwave Technol.*, 2009. Vol. 27. P. 2817–2833. doi: 10.1109/JLT.2009.2017211.
11. *Morozov E. V., Peshkova I. V., Rummyantsev A. S.* Bounds and maxima for the workload in a multiclass orbit queue // *Mathematics*, 2023. Vol. 11. Iss. 3. Art. 564. doi: 10.3390/math11030564.
12. *Peshkova I., Morozov E.* On comparison of multiserver systems with multicomponent mixture distributions // *J. Math. Sci.*, 2022. Vol. 267. No. 2. P. 260–272. doi: 10.1007/s10958-022-06132-z.
13. *Пешкова И. В.* Границы экстремального индекса времени ожидания в системе $M/G/1$ с распределением времени обслуживания в виде конечной смеси // *Информатика и её применения*, 2022. Т. 16. Вып. 2. С. 26–33. doi: 10.14357/19922264220405. EDN: VFKRKT.
14. *Rego V.* Some explicit formulas for mixed exponential service systems // *Computers Operations Research*, 1988. Vol. 15. Iss. 6. P. 509–520. doi: 10.1016/0305-0548(88)90047-0.
15. *Morozov E. V., Rummyantsev A. S., Dey S., Deepak T. G.* Performance analysis and stability of multiclass orbit queue with constant retrial rates and balking // *Perform. Evaluation*, 2019. Vol. 134. Art. 102005. doi: 10.1016/J.PEVA.2019.102005.
16. *Asmussen S.* Applied probability and queues. Stochastic modelling and applied probability. — New York, NY, USA: Springer-Verlag, 2003. 438 p.
17. *Ross S., Shanthikumar J., Zhu Z.* On increasing-failure-rate random variables // *J. Appl. Probab.*, 2005. Vol. 42. P. 797–809. doi: 10.1239/jap/1127322028.
18. *Haji R., Newell G. F.* A relation between stationary queue and waiting time distributions // *J. Appl. Probab.*, 1971. Vol. 8. P. 617–620. doi: 10.2307/3212186.

Поступила в редакцию 26.08.23

BOUNDS OF THE WORKLOAD IN A MULTICLASS RETRIAL QUEUE WITH EXPONENTIAL SERVICES

I. V. Peshkova^{1,2}

¹Petrozavodsk State University, 33 Lenina Pr., Petrozavodsk 185910, Russian Federation

²Karelian Research Center of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation

Abstract: A multiclass retrial queue with Poisson input and M classes of customers is investigated. For the given retrial system with exponential service times, the lower and upper bounds of the workload are derived. It is shown that the workload in the classical system $M/H_M/1$ with hyperexponential service times is the lower bound for the workload of the given retrial system. The upper bound is the workload in the classical $M/G/1$ system where each customer occupies the server for the given service time and additional time corresponding to the inter-retrial time from the “slowest” orbit. The presented simulation results confirm the theoretical conclusions.

Keywords: retrial queue; workload; stochastic ordering

DOI: 10.14357/19922264230408

EDN: UOKQRD

Acknowledgments

The research has been prepared with the support of the Russian Science Foundation according to the research project No. 21-71-10135.

References

1. Artalejo, J.R. 1999. Accessible bibliography on retrial queues. *Math. Comput. Model.* 30(3-4):1–6. doi: 10.1016/S0895-7177(99)00128-4.
2. Artalejo, J., and A. Gomez-Corral. 2008. *Retrial queueing systems: A computational approach*. Springer. 318 p. doi: 10.1007/978-3-540-78725-9.
3. Fayolle, G. 1986. A simple telephone exchange with delayed feedbacks. *Seminar (International) on Teletraffic Analysis and Computer Performance Evaluation Proceedings*. Elsevier Science. 245–253.
4. Choi, B.D., Y.W. Shin, and W.C. Ahn. 1992. Retrial queues with collision arising from unslotted CSMA/CD protocol. *Queueing Syst.* 11:335–356. doi: 10.1007/BF01163860.
5. Choi, B.D., K.H. Rhee, and K.K. Park. 1993. The $M/G/1$ retrial queue with retrial rate control policy. *Probab. Eng. Inform. Sc.* 7(1):29–46. doi: 10.1017/S0269964800002771.
6. Bertsekas, D., and R. Gallager. 2021. *Data networks*. Athena Scientific. 570 p.
7. Avrachenkov, K., and U. Yechiali. 2008. Retrial networks with finite buffers and their application to Internet data traffic. *Probab. Eng. Inform. Sc.* 22(4):519–536. doi: 10.1017/S0269964808000314.
8. Avrachenkov, K., and U. Yechiali. 2010. On tandem blocking queues with a common retrial queue. *Comput. Oper. Res.* 37(7):1174–1180. doi: 10.1016/j.cor.2009.10.004.
9. Yao, S., F. Xue, B. Mukherjee, S. J. B. Yoo, and S. Dixit. 2002. Electrical ingress buffering and traffic aggregation for optical packet switching and their effect on TCP-level performance in optical mesh networks. *IEEE Commun. Mag.* 40(9):66–72. doi: 10.1109/MCOM.2002.1031831.
10. Wong, E.W.M., L.L.H. Andrew, T. Cui, B. Moran, A. Zalesky, R.S. Tucker, and M. Zukerman. 2009. Towards a bufferless optical internet. *J. Lightwave Technol.* 27(14):2817–2833. doi: 10.1109/JLT.2009.2017211.
11. Morozov, E. V., I. V. Peshkova, and A. S. Rumyantsev. 2023. Bounds and maxima for the workload in a multiclass orbit queue. *Mathematics* 11(3):564. doi: 10.3390/math11030564.
12. Peshkova, I., and E. Morozov. 2022. On comparison of multiserver systems with multicomponent mixture distributions. *J. Math. Sci.* 267(2):260–272. doi: 10.1007/s10958-022-06132-z.
13. Peshkova, I. V. 2022. Granitsy ekstremal'nogo indeksa vremeni ozhidaniya v sisteme $M/G/1$ s raspredeleniem vremeni obsluzhivaniya v vide konechnoy smesi [On bounds of the stationary waiting time extremal index in $M/G/1$ system with mixture service times]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):26–33. doi: 10.14357/19922264220405. EDN: VFKRKT.
14. Rego, V. 1988. Some explicit formulas for mixed exponential service systems. *Comput. Oper. Res.* 15(6):509–520. doi: 10.1016/0305-0548(88)90047-0.
15. Morozov, E. V., A. S. Rumyantsev, S. Dey, and T. G. Deepak. 2019. Performance analysis and stability of multiclass orbit queue with constant retrial rates and balking. *Perform. Evaluation* 134:102005. doi: 10.1016/J.PEVA.2019.102005.
16. Asmussen, S. 2003. *Applied probability and queues. Stochastic modelling and applied probability*. New York, NY: Springer. 438 p.
17. Ross, S., J. Shanthikumar, and Z. Zhu. 2005. On increasing-failure-rate random variables. *J. Appl. Probab.* 42(3):797–809. doi: 10.1239/jap/1127322028.
18. Haji, R., and G.F. Newell. 1971. A relation between stationary queue and waiting time distributions. *J. Appl. Probab.* 8(3):617–620. doi: 10.2307/3212186.

Received August 26, 2023

Contributor

Peshkova Irina V. (b. 1975) — Candidate of Science (PhD) in physics and mathematics, associate professor, Petrozavodsk State University, 33 Lenina Pr., Petrozavodsk 185910, Russian Federation; senior scientist, Karelian Research Center of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation; iaminova@petsu.ru

МОДЕЛИ СОВМЕСТНОГО ОБСЛУЖИВАНИЯ ТРАФИКА eMBB И URLLC НА ОСНОВЕ ПРИОРИТЕТОВ В ПРОМЫШЛЕННЫХ РАЗВЕРТЫВАНИЯХ 5G NR*

Д. В. Иванова¹, Е. В. Маркова², С. Я. Шоргин³, Ю. В. Гайдамака⁴

Аннотация: Технология радиодоступа 5G NR способна осуществлять одновременную поддержку как сверхнадежной доставки с малой задержкой (Ultra-Reliable Low Latency Communication, URLLC), так и расширенного мобильного широкополосного доступа (enhanced Mobile Broadband, eMBB). Из-за критических требований к задержке и надежности, предъявляемых при предоставлении услуг двух классов обслуживания, возникает необходимость введения приоритетов. В статье рассмотрена промышленная среда, в которой производственное оборудование при управлении движением и синхронизации работы генерирует URLLC-трафик, а при удаленном мониторинге — eMBB-трафик. Предложена модель с приоритетным обслуживанием на базовой станции (БС) с прямой связью между устройствами (device-to-device, D2D) и без нее. Полученные численные результаты показывают, что введение приоритетов позволяет эффективно изолировать трафик URLLC и eMBB. При этом стратегия с поддержкой D2D, в которой БС явно резервирует ресурсы для прямой связи, значительно превосходит по показателям качества обслуживания стратегии, в которых явное резервирование не используется, а также стратегию, в которой весь трафик проходит через БС.

Ключевые слова: 5G; NR (New Radio); D2D; URLLC; eMBB; управление ресурсами; приоритетное обслуживание

DOI: 10.14357/19922264230409

EDN: JXCGXQ

1 Введение

Мобильные системы пятого поколения (5G), характеризующиеся высокими требованиями к качеству обслуживания, разработаны с учетом стремительного роста числа новых приложений. Помимо классических сценариев использования технологии 5G NR (New Radio) обещает поддержку в сфере промышленной автоматизации таких приложений, как совместное управление мобильными роботами, синхронизация, позиционирование, сервисы дополненной реальности, а также обслуживание на основе технологий телеприсутствия [1].

Системы, управляющие движущимися элементами производственного оборудования, требуют сверхнадежной доставки с низкими задержками (URLLC). В то же время для видеонаблюдения требуется расширенный мобильный широкополосный доступ (eMBB). Таким образом, БС NR должны поддерживать предоставление услуг двух классов обслуживания eMBB и URLLC одновременно. Механизмы их отдельной поддержки на БС NR в мил-

лиметровом диапазоне (mmWave) широко исследованы (см., например, [2–4] для eMBB и [5, 6] для URLLC). Однако в области совместной поддержки исследования практически не проводились.

В статье исследованы три схемы (стратегии) одновременного предоставления услуг eMBB и URLLC с реализацией явного приоритета [7]: (1) базовая стратегия — передача трафика через БС NR; (2) стратегия D2D-aware — использование D2D для передачи с полной координацией через БС (дополнительная интерференция не создается); (3) стратегия D2D-unaware — использование D2D без координации через БС (уменьшает задержку, но создает дополнительную интерференцию). Для построения и анализа модели использованы методы стохастической геометрии и теории массового обслуживания [8].

2 Системная модель

Рассмотрим модель развертывания 5G NR в промышленной среде, например на автомати-

* Исследование выполнено за счет гранта Российского научного фонда № 22-79-10053.

¹ Российский университет дружбы народов имени Патриса Лумумбы, ivanova-dv@rudn.ru

² Российский университет дружбы народов имени Патриса Лумумбы, markova-ev@rudn.ru

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

⁴ Российский университет дружбы народов имени Патриса Лумумбы; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaydamaka-yuv@rudn.ru

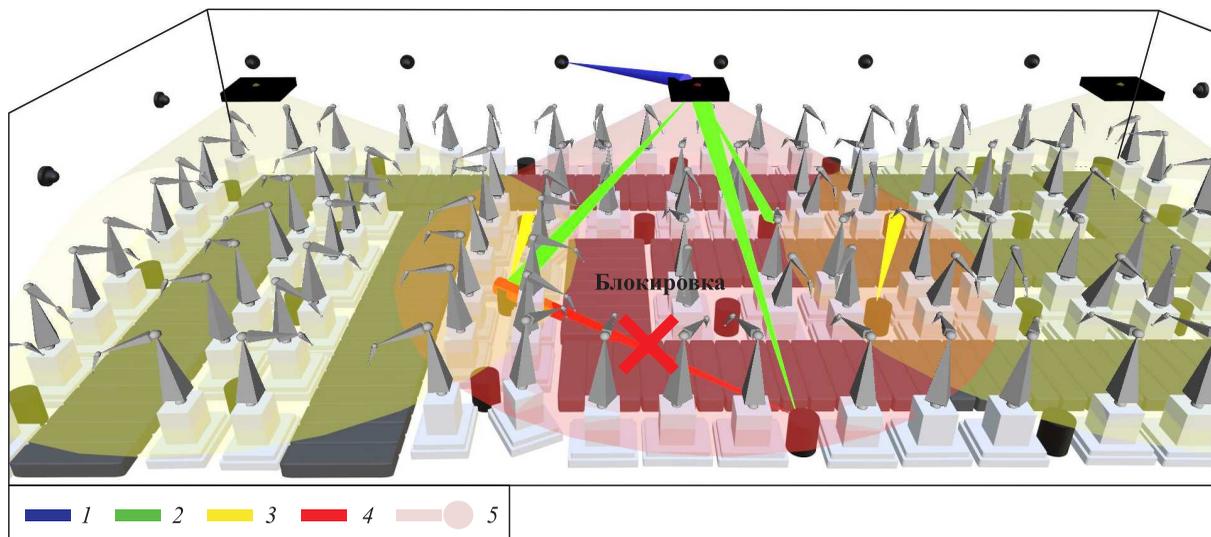


Рис. 1 Рассматриваемый сценарий развертывания: 1 — передача от камеры до БС; 2 — D2D-передача с координацией через БС; 3 — прямая D2D-передача; 4 — заблокированная D2D-передача; 5 — зона покрытия рассматриваемой БС

зированном заводе с несколькими производственными линиями (рис. 1). Автоматизация производства предполагает регулярный мониторинг процесса с помощью расположенных на станках датчиков с контролем посредством камер видеонаблюдения. Предположим, что БС NR монтируются на потолке на высоте h_A , образуя пуассоновский точечный процесс с плотностью χ БС/м². Пользовательское оборудование, состоящее из датчиков и камер, расположено на высоте h_U в узлах регулярной сетки с шагом l м. Ширина полосы пропускания каждой БС составляет W Гц, что соответствует емкости соты сети связи C условных единиц ресурса.

Датчики, связанные с автоматизированным оборудованием, генерируют потоковый трафик, соответствующий URLLC-услуге и характеризующийся гарантированной скоростью передачи данных $c_1 \geq 1$. Эластичный трафик, соответствующий eMBB, характеризуется минимальной скоростью $c_2^{\min} \geq 1$ и генерируется видеокерами, использующимися для удаленного мониторинга. Известна средняя интенсивность генерации запросов на передачу данных от датчиков и камер λ_i , а также средняя длительность сессий по передаче данных μ_i^{-1} , $i = 1, 2$. Объем ресурса для обслуживания запроса URLLC через БС $b_{1,B}$ и в режиме D2D $b_{1,D}$, а также объем ресурса для обслуживания запроса eMBB b_2^{\min} зависит от плотности БС и рассчитывается по формулам:

$$b_2^{\min} = \frac{c_2^{\min}}{E[S_{e,B}]}; \quad b_{1,B} = \frac{c_1}{E[S_{e,B}]}; \quad b_{1,D} = \frac{c_1}{E[S_{e,D}]},$$

где $E[S_{e,B}]$ — средняя спектральная эффективность при передаче трафика через БС, а $E[S_{e,D}]$ — при D2D-передаче.

Рассчитаем среднюю спектральную эффективность для разных схем обслуживания. Плотности вероятности расстояний D между двумя случайно выбранными пользовательскими устройствами и B от случайно выбранного устройства до БС определяются как [10]

$$f_B(x) = \frac{2x}{r_N};$$

$$f_D(x) = \frac{2x}{r_N^2} \left[\frac{2}{\pi} \cos^{-1} \left(\frac{x}{2r_N} \right) - \frac{x}{r_N \pi} \sqrt{1 - \frac{x^2}{4r_N^2}} \right].$$

Далее воспользуемся моделью распространения и блокировки сигнала, рассмотренной в [9]. Эффективный радиус покрытия r_N определяется как $\min(r_{N,S}, r_{N,V})$, где $r_{N,S}$ — максимально возможное расстояние между пользовательским оборудованием и БС NR; $r_{N,V}$ — половина расстояния между БС NR.

Отношение сигнала к шуму (SNR) на приемнике, расположенном на расстоянии x от БС

$$S(x) = \frac{P_U G_A G_U}{N_0 W + M_I} x^{-\zeta},$$

где P_U — излучаемая мощность; G_A и G_U — коэффициент усиления антенны на базовой станции и на пользовательском оборудовании соответственно; N_0 — спектральная плотность мощности шума;

M_I — средняя мощность помех; ζ — коэффициент распространения. Тогда

$$E[S_{e,B}] = \int_0^{r_N} f_B(x) \log_2[1 + S(x)] dx;$$

$$E[S_{e,D}] = \int_0^{2r_N} f_D(x) \log_2[1 + S(x)] dx.$$

Для расчета вероятности блокировки D2D-передачи сначала найдем $p_{B,1}(x)$ — вероятность того, что путь прямой видимости длиной x между двумя устройствами перекрывается одним станком. Воспользуемся методами интегральной геометрии и получим вероятность перекрытия [11]:

$$p_{B,1}(x) = \frac{2w(\pi w + 4x)(1 - \kappa)}{\pi(2\pi r_N^2 - 4r_N^2 \sin^{-1}(x/(2r_N)) - x\sqrt{4r_N^2 - x^2})},$$

где κ — прозрачность пользовательского оборудования; w — ширина пользовательского оборудования. Согласно рассматриваемой модели, максимальное число станков N_R , находящихся в зоне покрытия БС NR, можно найти с помощью аппроксимации задачи о круге Гаусса.

Общее число станков имеет биномиальное распределение с параметрами N_R и ν (вероятность нахождения станка в точке сетки). Таким образом, вероятность того, что путь прямой видимости заблокирован, имеет вид:

$$p_B(x) = \sum_{j=1}^{N_R} \binom{N_R}{j} \nu^j (1 - \nu)^{N_R - j} [1 - (1 - p_{B,1}(x))]^j.$$

Тогда искомая вероятность блокировки будет рассчитываться как

$$p_B = \int_0^{2r_N} f_D(x) p_B(x) dx.$$

3 Математическая модель

Функционирование рассматриваемой системы описывает двумерный марковский случайный процесс $(N_1(t), N_2(t))$, $t \geq 0$, где $N_i(t)$, $i = 1, 2$, — случайное число обслуживаемых системой запросов типа i в момент t . Обозначим максимальное число запросов на предоставление услуг URLLC (запросы 1-го типа) и eMBB (запросы 2-го типа), которое

может находиться в системе, $N_1 = \lfloor C/b_1 \rfloor$ и $N_2 = \lfloor C/b_2^{\min} \rfloor$ соответственно, тогда $n_i = 0, \dots, N_i$ — число обслуживаемых системой запросов типа i , $i = 1, 2$. Состояние системы описывает двумерный вектор $\mathbf{n} = (n_1, n_2)$ над пространством состояний

$$\mathbf{X} = \{(n_1, n_2) : n_1 \geq 0, n_2 \geq 0, n_1 b_1 + n_2 b_2^{\min} \leq C\}.$$

Обозначим через

$$k(n_1) = \left\lfloor \frac{C - n_1 b_1}{b_2^{\min}} \right\rfloor$$

максимальное число запросов на предоставление услуги eMBB, которое может быть принято в системе, когда в ней уже обслуживаются n_1 запросов на предоставление услуги URLLC. При этом число единиц ресурса, выделяемое для обслуживания каждого запроса на предоставление услуги eMBB, может меняться в зависимости от состояния системы:

$$b_2(n_1, n_2) = \left\lfloor \frac{C - n_1 b_1}{n_2} \right\rfloor \geq b_2^{\min}.$$

Сформулируем правила приема и обслуживания запросов:

- если число обслуживаемых системой запросов типа i меньше максимально возможного числа запросов данного типа N_i и число свободных единиц ресурса, доступных для данных запросов, не меньше b_1 и b_2^{\min} для 1-го и 2-го типа соответственно, то поступающий в систему запрос будет принят на обслуживание;
- если число обслуживаемых системой запросов на предоставление услуги URLLC меньше максимально возможного числа запросов данного типа N_1 , число свободных единиц ресурса, доступных для запросов данного типа, меньше b_1 , а число обслуживаемых запросов на предоставление услуги eMBB не меньше 1, то поступающий запрос на предоставление услуги URLLC будет принят на обслуживание за счет прерывания обслуживания $q(n_1, n_2) = \lceil (b_1 - C + (n_1 b_1 + n_2 b_2^{\min})) / b_2^{\min} \rceil$ запросов eMBB;
- в противном случае поступающие в систему запросы будут заблокированы.

На основе сформулированных правил составим диаграмму интенсивностей переходов (рис. 2).

Стационарное распределение вероятностей состояний системы $p(\mathbf{n})$, $\mathbf{n} \in \mathbf{X}$, может быть получено путем численного решения системы уравнений равновесия

$$\mathbf{p}^T \mathbf{A} = \mathbf{0}^T; \quad \mathbf{p}^T \mathbf{1} = 1,$$

где \mathbf{A} — инфинитезимальная матрица, элементы которой определяются следующим образом:

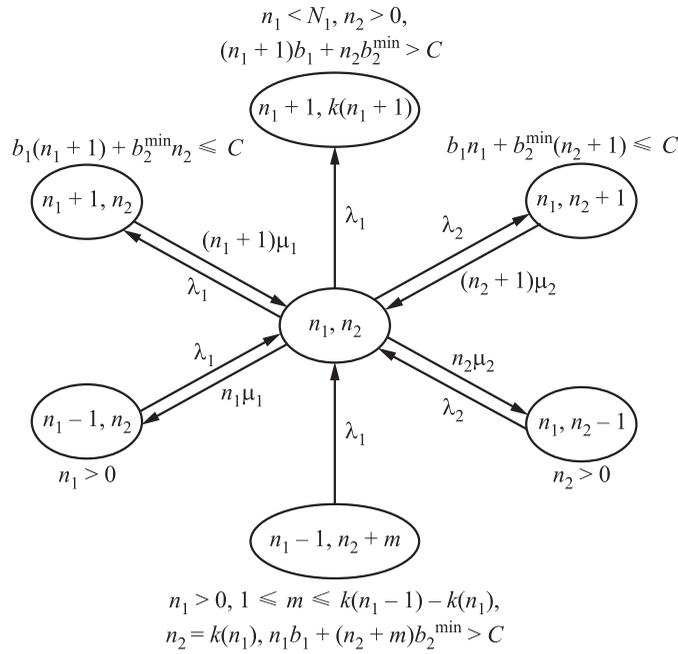


Рис. 2 Диаграмма интенсивностей переходов для центрального состояния

$$a(\mathbf{n}, \mathbf{n}') =$$

$$= \begin{cases} \lambda_1, & \mathbf{n}' = \mathbf{n} + \mathbf{e}_1, \\ & n_1 < N_1, b_1(n_1 + 1) + b_2^{\min} n_2 \leq C, \\ & \text{или } n'_1 = n_1 + 1, n'_2 = n_2 - q(n_1, n_2), \\ & n_1 < N_1, n_2 > 0, \\ & b_1(n_1 + 1) + b_2^{\min} n_2 > C; \\ \lambda_2, & \mathbf{n}' = \mathbf{n} + \mathbf{e}_2, \\ & n_2 < N_2, b_1 n_1 + b_2^{\min}(n_2 + 1) \leq C; \\ n_1 \mu_1, & \mathbf{n}' = \mathbf{n} - \mathbf{e}_1, n_1 > 0; \\ n_2 \mu_2, & \mathbf{n}' = \mathbf{n} - \mathbf{e}_2, n_2 > 0; \\ \phi, & \mathbf{n}' = \mathbf{n}; \\ 0 & \text{в ином случае;} \end{cases}$$

$$\phi = -[\lambda_1 I\{n_1 < N_1, b_1(n_1 + 1) + b_2^{\min} n_2 \leq C\} + \lambda_1 I\{n_1 < N_1, n_2 > 0, b_1(n_1 + 1) + b_2^{\min} n_2 > C\} + \lambda_2 I\{n_2 < N_2, b_1 n_1 + b_2^{\min}(n_2 + 1) \leq C\} + n_1 \mu_1 + n_2 \mu_2],$$

где $I\{x\}$ — индикаторная функция.

Рассчитав распределение вероятностей $p(\mathbf{n})$, можно вычислить основные показатели эффективности модели: вероятность потери URLLC-запросов

$$B_1 = \sum_{n_1=0}^{k(N_1)} p(N_1, n_1);$$

вероятность потери eMBB-запросов

$$B_2 = \sum_{n_1=0}^{N_1} p(n_1, k(n_1));$$

вероятность прерывания обслуживания eMBB-запроса

$$\Pi = \sum_{n_1=0}^{N_1-1} \sum_{\substack{n_2=k(n_1+1)+1 \\ k(n_1) \neq k(n_1+1)}}^{k(n_1)} \lambda_1 p(n_1, n_2) / (\lambda_1 + \lambda_2 I\{n_2 < k(n_1)\} + n_1 \mu_1 + n_2 \mu_2).$$

4 Численные результаты

Для проведения численного эксперимента исследуем вероятность потери запроса на передачу данных, представленную как функцию от плотности пользовательского оборудования (вероятности наличия станка в точке сетки) ν и от минимальной скорости передачи данных c_2^{\min} на рис. 3.

На рис. 3, а, где $c_1 = 2$ Мбит/с, $c_2^{\min} = 1$ Мбит/с, $G_A = 16 \times 4$, $G_U = 4 \times 4$, $\xi = 5 \cdot 10^{-4}$, $\mu_1 = 10^3$ и $\mu_2 = 1/120$, можно увидеть, что увеличение ν приводит к росту вероятности потери. Очевидно, что базовая стратегия, в которой все запросы проходят через БС, характеризуется постоянной вероятностью потери запросов на предоставление услуг eMBB и URLLC. В свою очередь, стратегия D2D-aware характеризуется меньшей вероятностью потери запросов на передачу трафика URLLC, которая не превышает 10^{-5} . Рассматривая стратегию D2D-unaware, отметим, что преимущества от использования прямой передачи данных незначи-

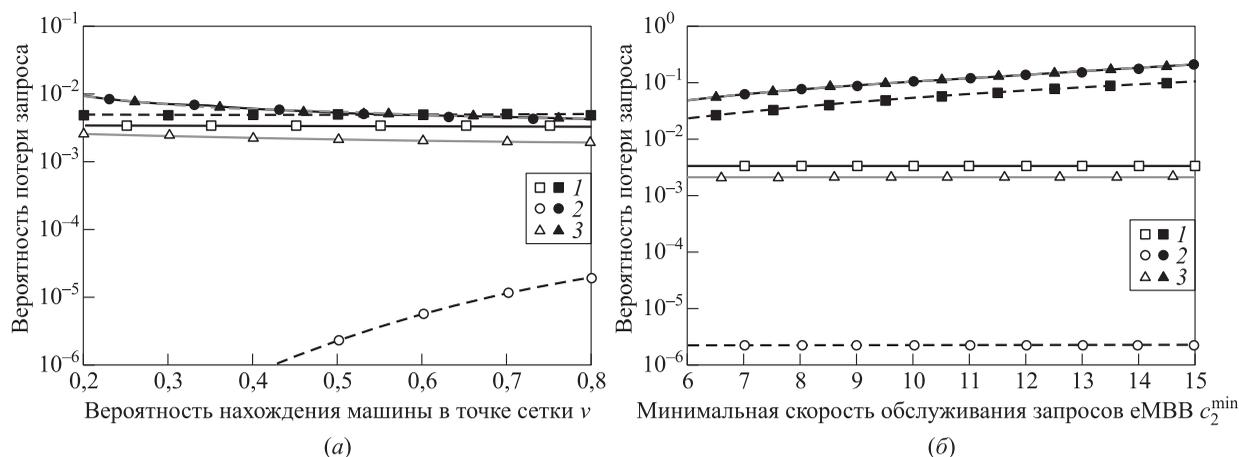


Рис. 3 Вероятность потери запроса в зависимости от плотности ν (а) и скорости c_2^{\min} (б): 1 — базовая стратегия; 2 — стратегия D2D-aware; 3 — стратегия D2D-unaware; пустые значки — URLLC; залитые значки — eMBB

тельны, поскольку неконтролируемая интерференция отрицательно влияет на передачу в дополнение к вероятности потери запросов из-за нехватки ресурсов, что в условиях высокой нагрузки может оказаться критичным.

Результаты показывают, что передача eMBB-трафика оказывает влияние на обслуживание URLLC-трафика. На рис. 3, б, где $\nu = 0,5$, проиллюстрирована зависимость вероятности потери запросов на передачу данных от минимальной требуемой скорости передачи eMBB-трафика. Таким образом, обслуживание на основе приоритетов эффективно в том случае, когда требования к запросам на предоставление услуги eMBB растут, так как вероятность потери запроса на предоставление услуги URLLC при этом сохраняется практически неизменной. Предлагаемый метод на основе приоритетов гарантирует, что обслуживание запросов на предоставление услуги URLLC будет защищено от потенциально изменяющейся нагрузки, создаваемой запросами на предоставление услуги eMBB. В данном случае также наблюдается превосходство стратегии D2D-aware над другими стратегиями для любых рассматриваемых значений c_2^{\min} .

5 Заключение

В статье предложена модель одновременной поддержки услуг двух классов обслуживания — eMBB и URLLC, основанная на реализации явного приоритета. Для трех рассмотренных схем передачи трафика на примере развертывания системы на автоматизированном предприятии проведена оценка ключевых вероятностных характеристик, в частности вероятности потери запросов на передачу данных, на основе которой сделан вывод

о наиболее эффективной стратегии совместного обслуживания.

Литература

1. Ghosh A., Ratasuk R., Rao A. M. Industrial IoT networks powered by 5G New Radio // *Microwave J.*, 2019. Vol. 62. No. 12. P. 24–40.
2. Moltchanov D., Samuylov A., Petrov V., Gapeyenko M., Himayat N., Andreev S., Koucheryavy Y. Improving session continuity with bandwidth reservation in mmWave communications // *IEEE Wirel. Commun. Lett.*, 2018. Vol. 8. No. 1. P. 105–108. doi: 10.1109/LWC.2018.2859988.
3. Begishev V., Moltchanov D., Sopin E., Samuylov A., Andreev S., Koucheryavy Y., Samouylov K. Quantifying the impact of guard capacity on session continuity in 3GPP New Radio systems // *IEEE T. Veh. Technol.*, 2019. Vol. 68. No. 12. P. 12345–12359. doi: 10.1109/TVT.2019.2948702.
4. Samuylov A., Moltchanov D., Kovalchukov R., Pirmagomedov R., Gaidamaka Y., Andreev S., Koucheryavy Y., Samouylov K. Characterizing resource allocation trade-offs in 5G NR serving multicast and unicast traffic // *IEEE T. Wirel. Commun.*, 2020. Vol. 19. No. 5. P. 3421–3434. doi: 10.1109/TWC.2020.2973375.
5. Rao J., Vrzic S. Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis // *IEEE Network*, 2018. Vol. 32. No. 2. P. 32–40. doi: 10.1109/MNET.2018.1700227.
6. Mahmood N. H., Karimi A., Berardinelli G., Pedersen K. I., Laselva D. On the resource utilization of multi-connectivity transmission for URLLC services in 5G New Radio // *IEEE Wireless Communications and Networking Conference Workshop*. — Piscataway, NJ, USA: IEEE, 2019. Art. 8902865. 6 p. doi: 10.1109/WCNCW.2019.8902865.
7. Кочеткова И. А., Влашкина А. С., Ву Н. Н., Шоргин В. С. Система массового обслуживания с управляемым по сигналам перераспределением приборов для анализа

- нарезки ресурсов сети 5G // Информатика и её применения, 2021. Т. 15. Вып. 3. С. 91–97. doi: 10.14357/19922264210312. EDN: JJENVV.
8. Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 48–55. doi: 10.14357/19922264180307. EDN: YAMDIL.
 9. Ivanova D., Markova E., Moltchanov D., Pirmagomedov R., Koucheryavy Y., Samouylov K. Performance of priority-based traffic coexistence strategies in 5G mmWave industrial deployments // IEEE Access, 2022. Vol. 10. P. 9241–9256. doi: 10.1109/ACCESS.2022.3143583.
 10. Moltchanov D. Distance distributions in random networks // Ad Hoc Netw., 2012. Vol. 10. No. 6. P. 1146–1166. doi: 10.1016/j.adhoc.2012.02.005.
 11. Santalo L. A. Integral geometry and geometric probability. — Cambridge: Cambridge University Press, 2004. 428 p. doi: 10.1017/CBO9780511617331.

Поступила в редакцию 25.09.23

PRIORITY-BASED eMBB AND URLLC TRAFFIC COEXISTENCE MODELS IN 5G NR INDUSTRIAL DEPLOYMENTS

D. V. Ivanova¹, E. V. Markova¹, S. Ya. Shorgin², and Yu. V. Gaidamaka^{1,2}

¹RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

Abstract: The technology 5G New Radio simultaneously supports both Ultra-Reliable Low-Latency Service (URLLC) and enhanced Mobile Broadband Service (eMBB). Owing to extreme latency and reliability requirements of both types of services, a prioritization needs to be provided. The present authors consider an industrial environment where production equipment utilizes URLLC service for controlling motion and synchronous operation while eMBB service is used for remote monitoring. The authors proposed the model with priority service at base station (BS) with and without direct device-to-device (D2D) communications. The obtained numerical results indicate that priorities allow one to isolate URLLC and eMBB traffic efficiently. The D2D-aware strategy where the BS explicitly reserves resources for direct communications significantly outperforms strategies where explicit reservation is not utilized as well as the strategy where all the traffic goes through the BS.

Keywords: 5G; NR (New Radio); D2D; URLLC; eMBB; resource allocation; priority service

DOI: 10.14357/19922264230409

EDN: JXCGXQ

Acknowledgments

The reported study was funded by the Russian Science Foundation, project No. 22-79-10053.

References

1. Ghosh, A., R. Ratasuk, and A. M. Rao. 2019. Industrial IoT networks powered by 5G New Radio. *Microwave J.* 62(12):24–40.
2. Moltchanov, D., A. Samuylov, V. Petrov, M. Gapeyenko, N. Himayat, S. Andreev, and Ye. Koucheryavy. 2019. Improving session continuity with bandwidth reservation in mmWave communications. *IEEE Wirel. Commun. Lett.* 8(1):105–108. doi: 10.1109/LWC.2018.2859988.
3. Begishev, V., D. Moltchanov, E. Sopin, A. Samuylov, S. Andreev, Y. Koucheryavy, and K. Samouylov. 2019. Quantifying the impact of guard capacity on session continuity in 3GPP New Radio systems. *IEEE T. Veh. Technol.* 68(12):12345–12359. doi: 10.1109/TVT.2019.2948702.
4. Samuylov, A., D. Moltchanov, R. Kovalchukov, R. Pirmagomedov, Y. Gaidamaka, S. Andreev, Y. Koucheryavy, and K. Samouylov. 2020. Characterizing resource allocation trade-offs in 5G NR serving multicast and unicast traffic. *IEEE T. Wirel. Commun.* 19(5):3421–3434. doi: 10.1109/TWC.2020.2973375.
5. Rao, J., and S. Vrzic. 2018. Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis. *IEEE Network* 32(2):32–40. doi: 10.1109/MNET.2018.1700227.
6. Mahmood, N. H., A. Karimi, G. Berardinelli, K. I. Pedersen, and D. Laselva. 2019. On the resource utilization of multi-connectivity transmission for URLLC services in 5G New Radio. *IEEE Wireless Communications and Networking Conference Workshop*. Piscataway, NJ: IEEE. 6 p. doi: 10.1109/WCNCW.2019.8902865.
7. Kochetkova, I. A., A. S. Vlaskina, N. N. Vu, and V. S. Shorgin. 2021. Sistema massovogo obsluzhivaniya s upravlyаемым po signalam pereraspredeleniem priborov.

- rov dlya analiza narezki resursov seti 5G [Queuing system with signals for dynamic resource allocation for analyzing network slicing in 5G networks]. *Informatika i ee Primeneniya — Inform. Appl.* 15(3):91–97. doi: 10.14357/19922264210312. EDN: JJENVV.
8. Gorbunova, A. V., V. A. Naumov, Y. V. Gaidamaka, and K. E. Samouylov. 2018. Resursnye sistemy massovogo obsluzhivaniya kak modeli besprovodnykh sistem svyazi [Resource queuing systems as models of wireless communication systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):48–55. doi: 10.14357/19922264180307. EDN: YAMDIL.
 9. Ivanova, D., E. Markova, D. Moltchanov, R. Pirmagomedov, Y. Koucheryavy, and K. Samouylov. 2022. Performance of priority-based traffic coexistence strategies in 5G mmWave industrial deployments. *IEEE Access* 10:9241–9256. doi: 10.1109/ACCESS.2022.3143583.
 10. Moltchanov, D. 2012. Distance distributions in random networks. *AD Hoc Netw.* 10(6):1146–1166. doi: 10.1016/j.adhoc.2012.02.005.
 11. Santalo, L. A. 2004. *Integral geometry and geometric probability*. Cambridge: Cambridge University Press. 428 p. doi: 10.1017/CBO9780511617331.

Received September 25, 2023

Contributors

Ivanova Daria V. (b. 1996) — PhD student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ivanova-dv@rudn.ru

Markova Ekaterina V. (b. 1987) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; markova-ev@rudn.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

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

МОДЕЛИ УЧЕТА ВЛИЯНИЯ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ТРАФИКА ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ НА ЭФФЕКТИВНОСТЬ ПРОГНОЗИРОВАНИЯ СРЕДСТВАМИ МАШИННОГО ОБУЧЕНИЯ

С. Л. Френкель¹, В. Н. Захаров²

Аннотация: Статья представляет собой попытку некоторого упорядочения и категоризации огромного потока публикаций по современным методам, техникам и моделям прогнозов данных различной природы с точки зрения их применимости для прогнозирования трафика в вычислительных сетях (ВС). Указанное упорядочение выполняется в рамках предложенной концептуальной модели (КМ) алгоритмов прогнозирования. В рамках этой КМ выделены характеристики как моделей трафика ВС, так и методов управления трафиком, которые явно или неявно могут быть используемы в современных программных инструментах предсказания. Показано, что анализ таких вероятностных аспектов описания данных, как наличие существенной нестационарности, некоторых нелинейных эффектов в моделях данных, а также специфики законов распределения данных, позволяет воздействовать на эффективность обучения предикторов.

Ключевые слова: сетевой трафик; вероятностные модели прогнозирования

DOI: 10.14357/19922264230410

EDN: CQYNFJ

1 Введение

Значительные успехи в развитии систем прогнозирования, достигнутые в рамках работ по машинному обучению (МО), прежде всего на базе нейронных сетей, не устраняют тем не менее трудности их использования. Среди многих причин этого можно указать также на то, что современные методы МО не учитывают некоторые математические свойства предсказываемых случайных данных. Это происходит вследствие того, что в системах МО формирование прогноза значений по обучающим выборкам данных часто выполняется как поиск внутренних закономерностей и связей без использования каких-либо математических вероятностных моделей данных. Это относится и к методам, использующим явные математические модели связи прогнозируемого значения в будущем, такие как методы линейной, полиномиальной или экспоненциальной регрессии [1], но не учитывающим математические свойства входных данных, например автокорреляционные. Эти обстоятельства часто дополняются непрозрачностью преобразования входных данных в пространство признаков (features), например в алгоритмах глубокого обучения (deep learning) [1], что затрудняет априорную оценку влияния на результат тех или иных свойств входных данных.

Статья представляет собой анализ публикаций по современным методам прогнозирования поведения вычислительных сетей, основанных на использовании программных инструментов прогнозирования (ИП) МО, в котором основное внимание уделяется способам учета влияния характеристик прогнозируемых данных на эффективность прогноза трафика с учетом математических свойств моделей предсказания.

Для структуризации огромного массива публикаций по современным методам, техникам и моделям прогнозов, на котором выполнен данный анализ, предлагается классификация моделей предсказания, используемых в современных программных МО, с использованием предложенного в [2] обобщенного представления *концептуальной модели* задачи предсказания, широко используемого в практике больших данных и программной инженерии. *Концептуальная модель* — это набор «точек зрения» об элементах и целях моделируемой системы, а также отношений между ними, выраженных в терминах той или иной теории [3].

В статье дается классификация подходов к задаче предсказания [2], выделяются основные ее элементы, дается общая характеристика задачи предсказания поведения ВС на основе инструментов МО, отмечается связь задачи предсказания будущих

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, fsergei51@gmail.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vzhakharov@ipiran.ru

значений временного ряда с общей задачей МО [4] (разд. 2). Формулируются требования к **математическим моделям предсказания** (разд. 3). Анализируется влияние вероятностных свойств случайных временных рядов на эффективность предсказания, в частности возможность предсказания без точного знания вероятностных распределений предсказываемых данных (разд. 4). Показано, что ряд характеристик отражает нелинейную связь прошлых и будущих значений временных рядов, моделирующих трафик, поэтому рассматриваем нелинейные алгоритмы предикторов (разд. 5).

2 Постановка задачи предсказания

Хотя вероятностные модели предсказания случайных последовательностей, рядов и процессов имеют более чем восьмидесятилетнюю историю (Колмогоров, Винер [5]), интенсивные исследования в этой области продолжают до сих пор. Это связано с огромным разнообразием задач, с которыми сталкиваются при обработке больших массивов данных. Для этих задач могут требоваться различные критерии точности предсказания, различные модели данных, различные требования к времени (оперативности) получения прогноза. Например, для многих задач предсказания значений финансовых индексов используется критерий минимума среднеквадратичной ошибки (MSE, mean squared error) предсказанных данных относительно истинных значений (на которой основаны критерий Колмогорова–Винера и его многочисленные линейные модификации). Но этот критерий заведомо не подходит, когда требуется предсказать знак приращения. В одних задачах вполне разумно использовать в качестве критерия эффективности предсказания некоторые усредненные по большим прошлым выборкам значения, в других нужны локальные оценки. Есть алгоритмы предсказания, которые используют числовые (измеряемые) данные без каких-либо предположений об их статистических и вероятностных свойствах (например, нейронные сети [1]), а есть неявно рассматривающие данные как выборку из нормального стационарного процесса [1] или марковской последовательности [6]. В [2] была обоснована классификация подходов к задаче предсказания, которой будем пользоваться в дальнейшем:

- математическая задача оценки неизвестных условных распределений вероятности (теоретико-вероятностный);

- математическая задача выбора оптимального решения (DM, decision making);
- алгоритмическая задача управления.

В *теоретико-вероятностном* подходе [7] задача предсказания состоит в следующем: пусть в момент времени t выполнена, по реализации процесса x_1, x_2, \dots, x_t , рассматриваемого (здесь и в дальнейшем) как *временной ряд* [8], оценка условных вероятностей $\gamma(x_{t+1}|x_1, x_2, \dots, x_t)$ того, что в момент $t + 1$ увидим значение, которое предсказываем. Чем точнее (в принятой метрике) эта оценка, тем лучше прогноз. Математическая модель оценки $\gamma(x_{t+1}|x_1, x_2, \dots, x_t)$, как и сама условная вероятность $\gamma(\cdot)$, называется в литературе предиктором [7].

Предикторами наряду с $\gamma(\cdot)$ также называют программные продукты (например, некоторые модули облачных MS AWS Azure ML, Google Cloud ML и т. д.), которые используются для предсказания.

В *DM-постановке* формулируется следующая задача предсказания с учетом требования эффективности (точности) прогноза: для заданного пространства стратегий B возможных предсказываемых значений X и функции потерь $l(b, x)$, $b \in B$, $x \in X$, выбрать как предсказываемое следующее значение \hat{b}_t при знании прошлых состояний x_1, \dots, x_{t-1} так, чтобы общие потери от ошибки предсказания («проигрыш») были бы асимптотически близки к проигрышу, полученному наилучшей фиксированной стратегией, известной апостериорно после просмотра всей последовательности $x = x_1, \dots, x_{t-1}$, т. е. предсказанное \hat{b}_t должно минимизировать возможные апостериорные потери. *Стратегия* представляет собой функцию, возвращающую выходной вектор y для каждой последовательности x в соответствии с функцией условного распределения $P(y|x)$ (или $\gamma(\cdot)$) [7, 9].

Третья постановка задачи прогнозирования, которую назовем *алгоритмической моделью предсказания*, состоит в использовании различных концептуальных моделей получения прогноза, не содержащих явно каких-либо основных элементов математических моделей указанных выше двух математических формулировок проблемы предсказания. Содержательно указанные концептуальные модели лежат в основе входного языка описания задачи прогнозирования, в частности параметров вызываемых функций из тех или иных библиотек машинного обучения [1, 10, 11]. Например, это может быть дерево поиска в известном предикторе eXtreme Gradient Boosting Regressor (XGBRegressor) [11], если пользовательская концептуальная модель рассматривает процесс построения прогноза некоторого процесса как поиск приемлемого (по выбранному критерию эффектив-

ности) предсказания среди набора возможных его значений в фиксированной области данных. XGB применяет к обучающему набору различные способы предсказания будущих значений временных рядов. Это можно рассматривать как последовательный поиск «моделей предсказания», улучшающих прошлые результаты. Эти модели могут быть разработаны на основе любых из указанных выше классов математических моделей, но во входном языке алгоритмической модели те или иные параметры и переменные этих моделей никак не присутствуют.

Поскольку предсказание выполняется по конечной выборке, естественной оценкой качества для всех трех типов моделей служит некоторая статистическая оценка, которая зависит от статистических характеристик данных, что должно быть отражено в моделях предсказания.

3 Концептуальная модель задачи предсказания

В [2] предложена следующая обобщенная форма представления концептуальной модели проблемы предсказания:

$$FM = \{M_D^\theta(S), M_L, F_{LM}\}. \quad (1)$$

Здесь $M_D^\theta(S)$ — модель предсказания, определяемая типом данных D (двоичные, действительные), с вероятностной мерой θ на D и структурой S на D , под которой понимается способ структурирования данных типа D как входных переменных предиктора (например, разделение данных на обучающую и тестовую выборку, скалярные либо векторные данные и т.д.); M_L — модель потерь («штрафов») от принятого предсказания с функцией потерь $L(X_S, Y_S)$, где $X_S, Y_S \in D$ — соответственно наблюдаемые и предсказанные данные со структурой S ; $F_{LM} = \{f_1, \dots, f_m\}$ — множество предикторов, основанных на моделях $M_D^\theta(S)$ и M_L , т.е. способов вычисления оценки будущего значения с распределением условной вероятности предсказанного значения $\gamma(\cdot)$.

Предполагается, что предиктор $f_p, p = 1, \dots, m$, всегда соответствует *допустимой* модели предсказания, т.е. все его входные данные и параметры могут быть однозначно определены в рамках рассматриваемой в данный момент модели.

FM позволяет представлять классы моделей согласно типам предсказываемых данных, используемым вероятностным мерам, функциям потерь при принятии решений о приемлемости полученного

прогноза. При этом необходимость учета вероятностных характеристик данных следует из сделанного выше замечания о статистическом характере оценки качества предсказания.

Пример 1.

Пусть $\theta = \text{Prob}(x_i = 1)$ — вероятностная мера Бернулли на $D = \{0, 1\}^n$ — множестве последовательностей длиной n . С точки зрения (1) n — это параметр структуры S для последовательности $X^t = \{x_1, x_2, \dots, x_i, \dots, x_{t-1}, x_t\} \in D$.

Рассмотрим последовательность нулей и единиц как последовательность независимых испытаний Бернулли с функцией потерь «0/1loss» — меры потерь пользователя предиктора от ошибок предсказания, равной доле неправильных предсказаний на последовательности длины n [6], т.е.

$$e_\Theta = E_\Theta \left(\sum_{i=1}^n \frac{I(b_t \neq x_t)}{n} \right), \quad (2)$$

где $\Theta = \text{Prob}(x_t = 1)$ — параметр распределения Бернулли; $I(\cdot)$ — индикаторная функция; b_t и x_t — предсказанное и истинное значение соответственно; E_Θ означает математическое ожидание по распределению Θ .

Заметим, что формула (2) универсальна в том смысле, что она не зависит от распределения на D . В литературе критерии, подобные (2), называют критериями *минимизации эмпирического риска* (ERM, empirical risk minimization) [9].

Можно рассмотреть следующий предиктор $b_t = f(x_{t-1})$, где $f \in F_{LM}$ и функция потерь определяется как «оптимальное правило предсказания по критерию минимума потерь по мере e_n » [6]:

$$b_t = \begin{cases} 1, & \text{если } \text{Prob}(1) > \frac{1}{2}; \\ 0, & \text{если } \text{Prob}(1) < \frac{1}{2}, \end{cases} \quad (3)$$

а в случае $\text{Prob}(0) = \text{Prob}(1) = 1/2$ прогноз не выполняется.

В этом случае, как легко видеть, средние потери равны вероятности ошибки $1 - \Theta$ при $\Theta > 1/2$, и Θ , если $\Theta < 1/2$, что можно выразить как

$$L_p = \min(\Theta, 1 - \Theta).$$

Поскольку речь идет об использовании МО для предсказания, рассмотрим, существует ли необходимость трансформации сформулированной концептуальной модели предсказания для учета специфики МО.

Концептуальная модель машинного обучения. Одна из хорошо известных и используемых формализаций МО — концепция PAC, Probably Approximate-

ly Correct, называемая так потому, что она оценивает *вероятность* достижения заданной *точности* классификации после обучения [4].

В РАС предсказание b_t рассматривается как принятие гипотезы $h \in H$, H — множество гипотез о предсказываемом значении случайной на множестве данных X величины с неизвестным распределением, для которой целевая функция L вычисляет характеристики качества предсказанных значений Y (*меток* [4]), по которым должен быть выбран предиктор $h_S : X \rightarrow Y$, где S — конечная выборка из X , соответствующая структуре S (которая может, например, определять, какие элементы из X следует включать в выборку — см. пример ниже). В предыдущих обозначениях принятие гипотезы h означает выбор предиктора f , реализующего предсказание некоторого значения из X . Например, в таком инструменте МО, как линейная регрессия, задача обучения может определяться на подмножестве действительных чисел R для некоторого числа переменных (коэффициентов регрессии) d при вычисленном предиктором $f(\cdot)$ наборе известных наблюдаемых значений Y («меток» в терминологии РАС) в обучающей выборке. Гипотеза предсказания соответствует в этом случае поиску линейной функции h , которая лучше всего аппроксимирует отношения между переменными X и Y (например, объем трафика в зависимости от времени и интенсивности атак).

Итак, оценка качества модели обучения требует использования тех же математических инструментов (вероятностных мер, функций потерь) на множествах данных тех же типов, что и задача предсказания. При этом структуризация S может разделять множество наблюдаемых данных на подмножества обучения предиктора и его оценки (тестирования). Поэтому концептуальная модель (1) покрывает также аспект обучения.

Ниже приведен более сложный пример задания и анализа влияния структуры S .

Пример 2.

Пусть в двоичной последовательности Бернулли предсказание по правилу (3) делается с учетом того, что k предшествующих предсказываемому моменту t значений единичные. Если параметр Θ (т. е. вероятность $\text{Prob}(1)$) неизвестен, то его оценивают как долю успехов («единиц») Θ_n на некоторой конечной подпоследовательности длиной n и Θ представляет собой математическое ожидание этой оценки (на множестве таких n -подпоследовательностей). В [12] показано, что условное математическое ожидание оценки Θ_n при условии, что оценка проводится по подпоследовательности, следующей за указанными k единицами, меньше истинного

значения Θ . Это значит, что в рамках правила (3) априорная оценка ошибки предсказания в момент, следующий за указанными k успехами, при использовании Θ_n в качестве $\text{Prob}(1)$ будет выше, чем в отсутствие фиксации указанных k событий.

Таким образом, свойства структуры данных S в модели (1) (в данном примере — наличие k выделенных наблюдений) влияют на эффективность прогноза. При этом незнание распределения данных D в (1) означает невозможность априорной оценки точности предсказания при данном алгоритме предсказания. Поэтому необходима теоретическая модель, позволяющая опираться не на конкретные знания распределений, а на знания о тех или иных *свойствах* распределений.

4 Влияние вероятностных свойств случайных временных рядов на эффективность предсказания

Условное распределение предсказываемых данных $\gamma(\cdot)$ (предиктора) определяется статистической зависимостью данных в прошлом и будущем (линейные или нелинейные модели, корреляционные структуры) и стационарностью данных в том смысле, что некоторые вероятностные законы, управляющие изменением данных во время наблюдения, остаются неизменными, по крайней мере на обучающей выборке, по которой должен быть сделан прогноз. Например, для сильно коррелированных временных рядов линейные регрессионные модели (например, ARIMA — autoregressive integrated moving average [1]) дают лучший прогноз, чем в случае их слабой коррелированности. Иными словами, если при рассмотрении модельных свойств временного ряда ограничиваться только корреляционными свойствами последовательности и рассмотренная выше условная вероятность прогноза $\gamma(x_{t+1}|x_1, x_2, \dots, x_t)$ явно связана с коррелированностью (как это имеет место, например, для гауссовского распределения, когда коррелированность эквивалентна независимости), то по степени коррелированности можно ожидать то или иное качество предсказания линейным регрессионным предиктором.

Рассмотрим, какие из измеряемых характеристик процессов можно использовать для оценки предсказуемости данных тем или иным предиктором. Согласно предложенной концептуальной модели задачи предсказания (1), в которой интегрируются вероятностные модели как данных, так

и принятия решения (через функцию потерь), рассмотрим следующие классы измеряемых характеристик выборок случайных временных рядов, связанных с вероятностными законами поведения:

- свойства автокорреляционных функций удаленных по времени событий и определяемые ими структурные свойства траекторий соответствующих процессов,
- нелинейность связей прошлых и будущих значений временных рядов,
- возможное влияние свойств тяжелых хвостов распределений,
- оценки степени стационарности рассматриваемых временных рядов.

Автокорреляционные свойства данных и их влияние на предсказуемость. Возможность предсказания по имеющимся данным, очевидно, зависит от степени корреляционной связи между их значениями, соответствующими удаленным событиям. Известной моделью такой зависимости, причем, как показали многолетние исследования, наблюдаемой в трафиках для многих реальных сетей [13, 14], является так называемое LRD-свойство (Long-Range Dependence) случайного процесса. Временной ряд X_t , $t \in Z$, называется *дальнозависимым* (LRD), если его ковариационная функция [14–16]

$$r(t) = E((X_0 - EX_0)(X_t - EX_t)) \sim c_\gamma |t|^{2-2H}$$

при $t \rightarrow \infty$.

Здесь $c_\gamma > 0$ — некоторая константа; H — показатель Херста, который вычисляется как математическое ожидание:

$$H = \lim_{n \rightarrow \infty} E \left[\frac{R(n)}{S(n)} \right],$$

где $R(n)$ — диапазон (размах) накопленных отклонений первых значений от среднего значения ряда, $S(n)$ — стандартное отклонение, n — величина промежутка времени (число точек на отрезке временного ряда).

Показатель Херста H характеризует время, отделяющее друг от друга коррелированные (например, сильнее, чем некоторое пороговое значение) отсчеты трафика.

Наиболее известная модель процесса с LRD — модель дробного броуновского движения fBM (fractal Brownian Movement) [16], которая обладает указанной корреляционной функцией.

Естественно предположить, что прогнозирование трафика со свойствами LRD может быть довольно эффективным (точным) с точки зрения используемой функции потерь (см. разд. 3), поскольку

информация о прошлом явно влияет на предсказываемое будущее. Существенно, что траектории случайных процессов с такими свойствами обладают «самоподобием», т. е. повторяемостью паттернов в разных временных масштабах [10, 17]. Известно, что это свойство наиболее сильно проявляется при $H \in (1/2, 1)$ [16, 18].

При этом важно, что при нестационарности исходного дробного броуновского процесса (моделирующего данный трафик с LRD-свойствами) его приращения положительно коррелированы, если $H \in (1/2, 1)$, некоррелированы, если $H = 1/2$, и отрицательно коррелированы, если $H \in (0, 1/2)$. Это свойство оказывается важным для предсказания знака приращений.

Если $H \notin (1/2, 1)$, то fBM утрачивает свойства самоподобия и оно в этом случае не может использоваться для предсказания.

Примером метода, зависящего от знака автокорреляции значений временного ряда служит метод [19]. В [19, 20] показана эффективность предсказания знака разности $y_{i+1} - y_i$:

$$\text{sign}(y_{i+1} - y_i) = -\text{sign}(y_i), \quad (4)$$

где $y_i = x_i - (Ex_i)$, $i = 1, \dots, t$, — последовательность центрированных случайных величин с нулевым средним.

Признаком возможности использования (4) как предиктора служит некоррелированность (или крайне слабая корреляция) последовательных разностей значений рассматриваемого временного ряда (процесса) с тенденцией к отрицательной корреляции значений $y_{i+1} - y_i$ и y_i , так как «антикорреляция» двух случайных величин означает тенденцию к изменению в противоположную сторону [16].

Напомним, что при $H < 1/2$ приращения имеют отрицательную корреляцию и демонстрируют короткодействующую зависимость, для $H > 1/2$ автокорреляция приращений положительна. Таким образом, два последовательных приращения, как правило, имеют одно и то же направление.

Примером полезности прогнозирования знака приращений трафика служит, например, ситуация, когда атаки на сеть осуществляются способом, который не может быть обнаружен антивирусным программным обеспечением, и аналитикам приходится полагаться на анализ изменений объема трафика или направления изменений в интенсивности записи в файле журнала [14].

Стационарность. Большинство рассматриваемых реальных процессов в ВС формально не относятся к стационарным, и о стационарности можно говорить уверенно лишь до известной степени. Для оценки этой степени хорошо известны ста-

статистические тесты на стационарность [17], однако включить их результаты в модель предсказания затруднительно, поскольку кроме интуитивных качественных рассуждений о том, что стационарные процессы предсказывать лучше, чем нестационарные, ничего использовать в модели нельзя. Однако для процессов с LRD можно получить более обоснованные эвристические правила связи стационарности и предсказуемости. В [15] показано, что сравнительно короткие наблюдения можно рассматривать как стационарные, по крайней мере по дисперсии, поскольку при самоподобии дисперсия выборочного среднего уменьшается медленнее, чем величина, обратная размеру выборки, т. е.

$$\text{var} \left(X^{(m)} \right) \sim a_2 m^{-\beta}, \quad m \rightarrow \infty, \quad 0 < \beta < 1,$$

где a_2 — положительная константа.

Существенно, что автокорреляции затухают как гиперболические функции, а не экспоненциально. Соответственно, с известной долей условности можно говорить и о стационарности по автокорреляционным функциям.

Показатель Херста H как показатель стационарности. В [18] показано, что в качестве критерия стационарности для самоподобного процесса может выступать Hurst exponent H . Значения $H > 1$ указывают (с учетом статистической трудности его оценки по конечной выборке в окрестности $H = 1$) на невозможность рассматривать предсказываемый процесс как стационарный. Для стационарного самоподобного процесса $H \in (0,5, 1)$ чем ближе значение параметра Херста к 1, тем медленнее затухает дисперсия по мере увеличения временного масштаба, и трафик становится более пульсирующим и, соответственно, нестационарным.

Влияние свойств LRD и стационарности на технику решения задач предсказания. Если используется среднеквадратичный критерий точности прогноза и ряд можно считать стационарным, то квадрат ошибки линейного предсказания для стационарного временного ряда пропорционален $1 - \rho^2$ [8], где $\rho = \text{ACF}(1)$, $\text{ACF}(1)$ означает корреляционную функцию в лаге 1.

С учетом того, что для LRD-процесса автокорреляционная функция экспоненциально зависит от H [15, 16], очевидна монотонная связь между среднеквадратичной ошибкой линейного предсказания и экспонентой Херста в интервале $(1/2, 1)$.

Этот анализ может быть полезным, например при использовании таких программных инструментов предсказания, как авторегрессионная модель ARMA (autoregressive moving average) со скользящим средним и авторегрессионное интегрированное скользящее среднее ARIMA.

Для предсказания чистого случайного блуждания может быть полезным предсказание изменения знака (4), как говорилось выше.

Теория линейного предсказания с критерием MSE, базирующаяся на модели Колмогорова–Винера, исходит из того, что временной ряд имеет конечное математическое ожидание и дисперсию. Однако для временных рядов со свойствами LRD это не всегда может быть выполнено, в частности если потоки данных имеют распределение Парето [21].

Влияние параметров вероятностных распределений процессов с LRD на предсказание. Многочисленные исследования LRD-рядов показали, что их распределения как распределения с тяжелыми хвостами близки к распределению Парето:

$$\text{Pareto}(X) = \frac{ab^a}{x^{a+1}},$$

где $x \geq a$. При этом среднее и дисперсия:

$$\begin{aligned} \mu_{\text{Pareto}}(x) &= \frac{ab}{a-1}; \\ \text{Var}_{\text{Pareto}} &= \frac{ab^2}{(a-1)^2(a-2)}, \end{aligned}$$

где $x \geq a > 0, b > 0$.

Очевидно, что μ_{Pareto} и $\text{Var}_{\text{Pareto}}$ не существуют при $a = 1$ и 2 и использование предикторов Винера–Колмогорова для вывода о предсказуемости ряда LRD относительно MSE неприемлемо.

Учет нелинейных свойств траекторий трафика. Рассмотренное выше самоподобие траектории трафика отражает нелинейную связь прошлого и будущего, где степень нелинейности зависит от H [2, 14, 16].

Как отмечалось, при $H = 1/2$ LRD-свойство утрачивается [20]. Значит, показатель H может указать, обладает ли прогнозируемый временной ряд свойством чистого случайного блуждания, для которого, очевидно, затруднительно эффективное использование МО, или же имеет некоторую корреляционную структуру, что означало бы потенциальную возможность эффективного прогноза. Например, в [22] для интернета вещей (IoT, Internet of Things) параметр H для сетевого трафика используется как «априорное знание» при обучении. Поскольку свойство самоподобия хорошо интерпретируемо в терминах описания трафика IT-специалистами, его использование повышает интерпретируемость модели, а именно: возможность понять, как влияют на качество прогноза свойства данных, в частности самоподобие.

В ряде инструментов предсказания прогностическим признаком служит разрушение самоподобия — методы, основанные на том наблюдении, что наличие DDoS-атаки (distributed denial of service) снижает степень самоподобия нормального трафика, поскольку инструменты DDoS не генерируют самоподобный трафик [14].

5 Нелинейные модели и нейронные алгоритмы

Прежде всего заметим, что очевидным примером применения нелинейного алгоритма служит рассмотренный предиктор (4). Однако нелинейная модель прогнозирования в современных ИП реализуется, как правило, либо в моделях нелинейной регрессии, либо в искусственных нейронных сетях [1, 23, 24], например в моделях глубокого обучения, управляемых данными, в дополнение к алгоритмам машинного обучения, таким как регрессия опорных векторов (SVR, Support Vector Regression) [10, 24].

Нейронная сеть может иметь проблемы из-за *переобучения*, или *оверфиттинга* (overfitting) [24], когда модель хорошо работает **только** с данными из обучающей выборки, адаптируясь к обучающим примерам, вместо того чтобы учиться классифицировать данные, не участвовавшие в обучении. Нетрудно видеть, что этому может способствовать такое нелинейное явление, как самоподобие. При этом такой распространенный способ снижения оверфиттинга, как dropout — выключение некоторых нейронов с некоторой вероятностью на каком-то интервале данных из процесса обучения, может не работать из-за того, что обучающая выборка будет подобна предыдущей. Внезапные изменения [25] в трафике облачных сред нейронная сеть может легко спутать с аномалиями трафика, что также ведет к неэффективности обучения. Выраженность такой нестационарности можно оценить по значениям H .

В качестве нелинейного подхода к прогнозированию трафика также применяется регрессия опорных векторов (SVR) [23], однако выбор подходящих функций ядра и оптимальных параметров очень сложен. Примеры такого выбора кратко рассматриваются в [26] (интересные результаты изучения влияния моделей данных на использование свойства самоподобия приведены в [27]).

Из приведенного анализа зависимости эффективности использования моделей предсказания от конкретных свойств временных рядов следует, что естественным путем учета зависимости свойств предикторов от поведения данных (стационар-

ность, нелинейность зависимости) должен быть их онлайн-выбор в процессе решения задачи управления сетью. Например, в [27] для прогнозирования времени отклика и пропускной способности облачных сервисов на разных этапах использования ресурсов рассматривались алгоритмы искусственных нейронных сетей (ANN, artificial neural networks) и линейной регрессии.

6 Заключение

Как показывает анализ литературы, многие современные инструменты предсказания, основанные на принципах МО, работают недостаточно эффективно ввиду выраженной нелинейности изменения трафика, его нестационарности, необходимости большого объема предыдущих наблюдений. Данная статья представляет собой попытку некоторого упорядочения и категоризации огромного потока публикаций по современным методам, техникам и моделям прогнозов данных различной природы с точки зрения возможности априорной оценки эффективности предсказания значений данных соответствующими ИП. Для этого выделены характеристики вероятностных моделей, явно или неявно используемых в современных программных инструментах предсказания, позволяющие оценивать наличие существенной нестационарности в потоках данных, некоторых нелинейных эффектов, специфики законов распределения, влияющих на эффективность обучения предикторов.

Литература

1. *Chen A., Law J., Aibin M.* A survey on traffic prediction techniques using artificial Intelligence for communication networks // *Telecom*, 2021. Vol. 2. Iss. 4. P. 518–535. doi: 10.3390/telecom2040029.
2. *Frenkel S.* On impact of data models on predictability assessment of time series // *Data analysis and optimization / Eds. B. Goldengorin, S. Kuznetsov.* — Optimization and its applications ser. — Springer, 2023. Vol. 202. P. 103–134. doi: 10.1007/978-3-031-31654-8_7.
3. *Fettke P.* Conceptual modelling and artificial intelligence: Overview and research challenges from the perspective of predictive business process management // *CEUR Workshop Procee.*, 2020. Vol. 2542. P. 157–164.
4. *Shalev-Shwartz S.* Understanding machine learning: From theory to algorithms. — Cambridge University Press, 2014. 449 p.
5. *Brovelli M., Sanso F., Venuti G.* A discussion on the Wiener–Kolmogorov prediction principle with easy-to compute and robust variants // *J. Geodesy*, 2003. Vol. 76. P. 673–683. doi: 10.1007/s00190-002-0292-3.

6. *Merhav N., Feder M.* Universal prediction // IEEE T. Inform. Theory, 1998. Vol. 44. Iss. 6. P. 2124–2147. doi: 10.1109/18.720534.
7. *Ryabko B.* Compression-based methods for nonparametric prediction and estimation of some characteristics of time series // IEEE T. Inform. Theory, 2009. Vol. 55. Iss. 9. P. 4309–4315. doi: 10.1109/TIT.2009.2025546.
8. *Box G. E. P., Jenkins G. M., Reinsel G.* Time series analysis: Forecasting and control. — New York, NY, USA: John Wiley & Sons, Inc, 2008. 709 p.
9. *Aryan M.* Efficient methods for large-scale empirical risk minimization: D.Sc. Thesis. — Philadelphia, PA, USA: University of Pennsylvania, 2017. 303 p.
10. *Sharma S.* Activation functions in neural networks, 2017. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
11. *Weiwei J.* Cellular traffic prediction with machine learning: A survey // Expert Syst. Appl., 2022. Vol. 201. P. 117–163. doi: 10.1016/j.eswa.2022.117163.
12. *Miller J., Sanjurjo A.* Surprised by the hot hand fallacy? A truth in the law of small numbers. — Cornell University, 2019. arXiv:1902.01265v1. 49 p.
13. *Jiang M., Nikolic M., Hardy S., Trajkovic L.* Impact of self-similarity on wireless data Network performance // Conference (International) on Communications Proceedings. — Piscataway, NJ, USA: IEEE, 2001. Vol. 2. P. 477–481. doi: 10.1109/ICC.2001.936986.
14. *Brignoli D.* DDOS detection based on traffic self-similarity: Master Thesis. — Christchurch, New Zealand: University Canterbury, 2008. 100 p.
15. *Lleland W. E., Taqqu M. S., Willinger W., Wilson D. V.* On the self-similar nature of Ethernet traffic (extended version) // IEEE ACM T. Network., 1994. Vol. 2. Iss. 1. P. 1–15. doi: 10.1109/90.282603.
16. *Coskun B., Vardar-Acar B., Demirtas C.* A generalized correlated random walk, converging to fractional Brownian motion. — Cornell University, 2019. arXiv:1903.05424v3. 21 p.
17. *Riya J., Chetty P.* What is a stationarity test and how to do it? // PG, 2020. <https://www.projectguru.in/what-is-a-stationarity-test-how-to-do-it>.
18. *Park C., Hernandez F., Long L., Marron J.* Long-range dependence analysis of internet traffic // J. Appl. Stat., 2011. Vol. 38. Iss. 7. P. 1407–1433. doi: 10.1080/02664763.2010.505949.
19. *Sornette D., Andersen J.* Increments of uncorrelated time series can be predicted with a universal 75% probability of success // Int. J. Mod. Phys. C, 2000. Vol. 11. Iss. 4. P. 713–720. doi: 10.1142/S0129183100000626.
20. *Frenkel S.* Predicting the direction of changes in the values of time series for relatively small training samples // Cyber security, cryptography, and machine learning / Eds. S. Dolev, J. Katz, A. Meisels. — Lecture notes in computer science ser. — Cham, Switzerland: Springer, 2022. Vol. 13301. P. 118–134. doi: 10.1007/978-3-031-07689-3_9.
21. *Song W., Duan S., Chen D., Zio E., Yan W., Cai F.* Finite iterative forecasting model based on fractional generalized Pareto motion // Fractal, 2022. Vol. 6. Iss. 9. P. 471–490. doi: 10.3390/fractalfract6090471.
22. *Pan C., Wang Y., Shi H., Shi J., Cai R.* Network traffic prediction incorporating prior knowledge for an intelligent network // Sensors, 2022. Vol. 22. Iss. 7. P. 2674–2690. doi: 10.3390/s22072674.
23. *Tong H., Li C., He J.* A boosting-based framework for self-similar and non-linear internet traffic prediction // Advances in neural networks / Eds. F. Yin, J. Wang, C. Guo. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer, 2004. Vol. 3174. P. 931–936. doi: 10.1007/978-3-540-28648-6_148.
24. *Zhang J., Tan D., Zhu H.* From machine learning to deep learning: Progress in machine intelligence for rational drug discovery // Drug Discov. Today, 2017. Vol. 22. Iss. 11. P. 1680–1685. doi: 10.1016/j.drudis.2017.08.010.
25. *Nikravesh A. Y., Ajila S. A., Lung C.-H.* An autonomic prediction suite for cloud resource provisioning // J. Cloud Computing, 2017. Vol. 6. Iss. 3. doi: 10.1186/s13677-017-0073-4.
26. *Dolev S., Frenkel S., Zakharov V.* Forecasting tools in practical applications: Selection and evaluation methodology // Conference (International) on Engineering Technologies and Computer Science Proceedings. — Piscataway, NJ, USA: IEEE, 2021. P. 37–44. doi: 10.1109/EnT52731.2021.00013.
27. *Lukashenko O., Morozov E., Pagano M.* A Gaussian approximation of the distributed computing process // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 109–116. doi: 10.14357/19922264190215. EDN: PKNUSZ.

Поступила в редакцию 22.08.23

MODELS FOR STUDY OF THE INFLUENCE OF STATISTICAL CHARACTERISTICS OF COMPUTER NETWORKS TRAFFIC ON THE EFFICIENCY OF PREDICTION BY MACHINE LEARNING TOOLS

S. L. Frenkel and V. N. Zakharov

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

Abstract: The article is an attempt to streamline and categorize a huge stream of publications on modern methods, techniques, and models of data forecasting of various nature in terms of their applicability for traffic forecasting in computer networks. The specified ordering is performed within the framework of the proposed conceptual model of forecasting algorithms. Within the framework of this conceptual model, the characteristics of both computer network traffic models and traffic control methods that can be explicitly or implicitly used in modern prediction software tools are highlighted. It is shown that the analysis of such probabilistic aspects of data description as the presence of significant nonstationarity, some nonlinear effects in data models, as well as the specifics of data distribution laws can influence the efficiency of learning predictors.

Keywords: network traffic prediction; probabilistic models

DOI: 10.14357/19922264230410

EDN: CQYNFJ

References

- Chen, A., J. Law, and M. Aibin. 2021. A survey on traffic prediction techniques using artificial intelligence for communication networks. *Telecom* 2(4):518–535. doi: 10.3390/telecom2040029.
- Frenkel, S. 2023. On impact of data models on predictability assessment of time series. *Data analysis and optimization*. Eds. B. Goldengorin and S. Kuznetsov. Optimization and its applications ser. Springer. 202:103–134. doi: 10.1007/978-3-031-31654-8.7.
- Fettke, P. 2020. Conceptual modelling and artificial intelligence: Overview and research challenges from the perspective of predictive business process management. *CEUR Workshop Procee.* 2542:157–164.
- Shalev-Shwartz, S. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge University Press. 449 p.
- Brovelli, M., F. Sanso, and G. Venuti. 2003. A discussion on the Wiener–Kolmogorov prediction principle with easy-to compute and robust variants. *J. Geodesy* 76:673–683. doi: 10.1007/s00190-002-0292-3.
- Merhav, N., and M. Feder. 1998. Universal prediction. *IEEE T. Inform. Theory* 44(6):2124–2147. doi: 10.1109/18.720534.
- Ryabko, B. 2009. Compression-based methods for non-parametric prediction and estimation of some characteristics of time series. *IEEE T. Inform. Theory* 55(9):4309–4315. doi: 10.1109/TIT.2009.2025546.
- Box, G. E. P., G. M. Jenkins, and G. Reinsel. 2008. *Time series analysis: Forecasting and control*. New York, NY: John Wiley & Sons, Inc. 709 p.
- Aryan, M. 2017. Efficient methods for large-scale empirical risk minimization. Philadelphia, PA: University of Pennsylvania. PhD Thesis. 303 p.
- Sharma, S. 2017. Activation functions in neural networks. Available at: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6> (accessed December 4, 2023).
- Weiwei, J. 2022. Cellular traffic prediction with machine learning: A survey. *Expert Syst. Appl.* 201:117–163. doi: 10.1016/j.eswa.2022.117163.
- Miller, J., and A. Sanjurjo. 2019. Surprised by the hot hand fallacy? A truth in the law of small numbers. *arXiv.org*. 49 p. Available at: <https://arxiv.org/abs/1902.01265v1> (accessed December 4, 2023).
- Jiang, M., M. Nikolic, S. Hardy, and L. Trajkovic. 2001. Impact of self-similarity on wireless data Network performance. *Conference (International) on Communications Proceedings*. Piscataway, NJ: IEEE. 2:477–481. doi: 10.1109/ICC.2001.936986.
- Brignoli, D. 2008. DDOS detection based on traffic self-similarity. Christchurch, New Zealand: University Canterbury. Master Thesis. 100 p.
- Lleland, W. E., M. S. Taqqu, W. Willinger, and D. V. Willson. 1994. On the self-similar nature of Ethernet traffic (extended version). *IEEE ACM T. Network.* 2(1):1–15. doi: 10.1109/90.282603.
- Coskun, B., B. Vardar-Acar, and C. Demirtas. 2019. A generalized correlated random walk, converging to fractional Brownian motion. *arXiv.org*. 21 p. Available at: <https://arxiv.org/abs/1903.05424v3> (accessed December 4, 2023).
- Riya, J., and P. Chetty. 2020. What is a stationarity test and how to do it? *PG*. Available at: <https://www>.

- projectguru.in/what-is-a-stationarity-test-how-to-do-it (accessed December 4, 2023).
18. Park, C., F. Hernandez, L. Long, and J. Maron. 2011. Long-range dependence analysis of internet traffic. *J. Appl. Stat.* 38(7):1407–1433. doi: 10.1080/02664763.2010.505949.
 19. Sornette, D., and J. Andersen. 2000. Increments of uncorrelated time series can be predicted with a universal 75% probability of success. *Int. J. Mod. Phys. C* 11(4):713–720. doi: 10.1142/S0129183100000626.
 20. Frenkel, S. 2022. Predicting the direction of changes in the values of time series for relatively small training samples. *Cyber security, cryptography, and machine learning*. Eds. S. Dolev, J. Katz, and A. Meisels. Lecture notes in computer science ser. Cham, Switzerland: Springer. 13301:118–134. doi: 10.1007/978-3-031-07689-3_9.
 21. Song, W., S. Duan, D. Chen, E. Zio, W. Yan, and F. Cai. 2022. Finite iterative forecasting model based on fractional generalized Pareto motion. *Fractal* 6(9):471–490. doi: 10.3390/fractalfract6090471.
 22. Pan, C., Y. Wang, H. Shi, J. Shi, and R. Cai. 2022. Network traffic prediction incorporating prior knowledge for an intelligent network. *Sensors* 22(7):2674–2690. doi: 10.3390/s22072674.
 23. Tong, H., C. Li, and J. He. 2004. A boosting-based framework for self-similar and non-linear internet traffic prediction. *Advances in neural networks*. Eds. F. Yin, J. Wang, and C. Guo. Lecture notes in computer science ser. Berlin, Heidelberg: Springer. 3174:931–936. doi: 10.1007/978-3-540-28648-6_148.
 24. Zhang, J., D. Tan, and H. Zhu. 2017. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov. Today* 22(11):1680–1685. doi: 10.1016/j.drudis.2017.08.010.
 25. Nikraves, A. Y., S. A. Ajila, and C.-H. Lung. 2017. An autonomous prediction suite for cloud resource provisioning. *J. Cloud Computing* 6(3). 20 p. doi: 10.1186/s13677-017-0073-4.
 26. Dolev, S., S. Frenkel, and V. Zakharov. 2021. Forecasting tools in practical applications: Selection and evaluation methodology. *Conference (International) on Engineering Technologies and Computer Science Proceedings*. Piscataway, NJ: IEEE. 37–44. doi: 10.1109/EnT52731.2021.00013.
 27. Lukashenko, O., E. Morozov, and M. Pagano. 2019. A Gaussian approximation of the distributed computing process. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):109–116. doi: 10.14357/19922264190215. EDN: PKNUSZ.

Received August 22, 2023

Contributors

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; fsergei51@gmail.com

Zakharov Victor N. (b. 1948) — Doctor of Science in technology, associate professor; Scientific Secretary, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vzakharov@ipiran.ru

АННОТИРОВАНИЕ ПАРАЛЛЕЛЬНЫХ КОРПУСОВ: ПОДХОДЫ И НАПРАВЛЕНИЯ РАЗВИТИЯ*

А. А. Гончаров¹

Аннотация: Представлены возможные направления развития инструментов для аннотирования параллельных корпусов с учетом актуального положения дел в этой сфере. Рассмотрены основные подходы к проведению исследований на корпусном материале — (1) корпусно-ориентированный, (2) корпусно-управляемый и (3) использующий корпус как источник иллюстративного материала — и кратко описаны различия между ними. Показано, что, несмотря на обилие инструментов для аннотирования корпусов, подавляющее большинство из них предназначено для работы с монопольными корпусами и/или поддерживает очень узкий спектр функций по аннотированию текстовых данных. Наибольшее число функций предоставляют надкорпусные базы данных (НБД) и веб-приложения для доступа к ним, разрабатываемые в ФИЦ ИУ РАН: (1) формирование блоков текста оригинала и перевода, необходимых и достаточных для анализа вхождения изучаемой языковой единицы и варианта ее перевода; (2) выявление вхождения изучаемой языковой единицы и варианта ее перевода; (3) выбор признаков, характеризующих употребление изучаемой языковой единицы и варианта ее перевода; (4) выбор признаков, характеризующих переводное соответствие. Такой спектр функций позволяет решать значительную часть исследовательских задач, однако он может быть расширен. Предлагаются три направления развития имеющегося функционала, способные обеспечить более детализированное описание языкового материала.

Ключевые слова: параллельный корпус; корпусная лингвистика; аннотирование корпуса; лингвистическое аннотирование

DOI: 10.14357/19922264230411

EDN: GDKDOZ

1 Введение

На сегодняшний день обращение к корпусам в лингвистических исследованиях становится скорее стандартом, чем исключением: по частоте использования и значимости эти информационные ресурсы встают в один ряд со словарями. Еще в 2005 г. В. А. Плунгян отмечал: «Вполне возможно, что в недалеком будущем без корпуса изучаемого языка лингвисту будет так же невозможно обходиться, как, например, без словаря этого языка. Более того, корпус, словарь и грамматика, скорее всего, соединятся в один электронный ресурс — или базу данных, на основании которой и можно будет изучать язык» [1, с. 14]. Действительно, во многом развитие идет именно в этом направлении, а данные корпусов вновь и вновь заставляют ставить под сомнение или корректировать утверждения лингвистов, сделанные без использования корпусного материала (см. примеры в [2–4]).

Ввиду столь широкого распространения корпусных исследований перед специалистами по информатике и компьютерной лингвистике встает вопрос о том, каким условиям должны удовлетворять инструменты, используемые при проведении этих

исследований. Цель статьи состоит в том, чтобы (с учетом текущего положения вещей) обозначить некоторые возможные направления развития таких инструментов на примере работы с параллельным корпусом.

2 Подходы к проведению корпусных исследований

Прежде чем перейти к рассмотрению собственно инструментов, стоит отметить, что исследования, выполняемые в русле корпусной лингвистики, различаются в зависимости от подхода. В подавляющем большинстве работ таких подходов выделяется два: (1) корпусно-ориентированный и (2) корпусно-управляемый [5–8], к которым иногда добавляется (3) подход, использующий корпус как источник иллюстративного материала [9]. Для русскоязычных обозначений этих подходов характерна вариативность. Так, в работе [9] подход 1 назван как «основанный на корпусе», а подход 2 — «направляемый корпусом»; в настоящей статье для их именовании используются термины из [8, с. 14]. Что касается подхода 3, то он в работе [9] назван

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

¹ Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

как «использующий корпус». Это представляется не вполне удачным, так как любой из подходов к работе с корпусом по определению подразумевает его использование; в настоящей статье за основу взято русскоязычное обозначение этого подхода из работы [10, с. 399].

Границы перечисленных подходов размыты, а некоторые исследователи и вовсе ставят под сомнение необходимость подобного разделения. Например, в работе [11] с говорящим названием «Corpus-based and corpus-driven approaches to linguistic analysis: One and the same?» ставится под сомнение необходимость противопоставления подходов к проведению корпусных исследований в пользу общего обозначения *corpus approach*. В работе [12] также утверждается, что четкое разграничение между подходами преувеличено. Автор работы [13], напротив, считает, что это разграничение преуменьшено, так как исследование, выполненное в строгом соответствии с корпусно-управляемым подходом, по его мнению, не что иное, как миф: «[the distinction between corpus-based and corpus-driven approaches] is understated given that *truly* corpus-driven work seems a myth at best».

Кратко укажем, в чем состоят основные характеристики каждого из подходов.

1. Корпусно-ориентированный подход (*corpus-based*). Корпусные данные анализируются качественно и количественно, а исходной точкой исследования оказывается сформулированная заранее гипотеза и/или допущение (т. е. движение осуществляется от теории к данным).

2. Корпусно-управляемый подход (*corpus-driven*). Корпусные данные анализируются качественно и количественно, причем они сами же служат исходной точкой исследования. Теоретические положения заранее не формулируются или это делается в минимальной степени (т. е. движение осуществляется от данных к теории). Говоря об этом подходе, Дж. Синклер заявляет: «В корпусно-управляемом лингвистическом исследовании вы не используете заранее размеченный текст, а работаете непосредственно с текстом как он есть, и только тогда вы можете увидеть, как устроен этот текст, не содержащий посторонних примесей» [14, с. 191]¹.

3. Подход, использующий корпус как источник иллюстративного материала (*corpus-informed/corpus-illustrated*). Корпус служит лишь для поиска примеров на естественном языке, количественный анализ данных не проводится.

Значительная часть публикуемых исследований, содержащих отсылки к корпусу, относится имен-

но к третьему подходу. Безусловно, эти исследования гораздо менее субъективны по сравнению с теми, где языковые примеры конструируются лингвистом. Однако — и в пользу этого свидетельствует в том числе то, что третий подход даже не упоминается в абсолютном большинстве работ по корпусной лингвистике, — исследования такого типа, вероятно, не могут считаться по-настоящему корпусными, так как корпус в них выступает лишь в качестве удобного инструмента для быстрого поиска иллюстративного материала. В рамках настоящей статьи такие исследования не рассматриваются, поскольку они не подразумевают какого-либо аннотирования, а задействуют лишь поисковый функционал корпусного менеджера.

3 Инструменты для аннотирования параллельных корпусов

Если для проведения исследований, где корпус используется как источник иллюстративного материала, необходим лишь инструмент для поиска релевантных примеров в текстах корпуса, то такие исследования, которые выполняются в рамках корпусно-ориентированного или корпусно-управляемого подхода, требуют широкого применения информационных технологий. Так, чтобы коллектив исследователей (особенно если он распределенный) мог работать над изучением какой-либо категории языковых единиц на корпусном материале в рамках продолжительного научного проекта, необходимо обеспечить поддержку многопользовательского режима работы с данными. Этот режим работы может обеспечиваться при помощи веб-интерфейса.

Однако, если работа с текстами проводится только в режиме чтения, это не позволяет последовательно анализировать примеры употребления исследуемых языковых единиц и сохранять результаты анализа в форме аннотированных переводных соответствий (в случае моноязычного корпуса — аннотированных блоков текста). Традиционно лингвисты составляли картотеки примеров, на смену которым сегодня пришли электронные документы, создаваемые при помощи текстовых или табличных процессоров, входящих в состав офисных пакетов приложений. Однако подобное использование электронных документов, как правило, снова переводит работу с данными в однопользовательский режим, а также затрудняет как сам процесс

¹ Перевод мой — А. Г. Оригинальная цитата: «In corpus-driven linguistics you do not use pre-tagged text, but you process the raw text directly and then the patterns of this uncontaminated text are able to be observed».

исследования, так и получение непротиворечивых (*consistent*) и легко проверяемых результатов.

Ввиду этого на протяжении последних десятилетий непрерывно создаются новые программные продукты для создания корпусов и последующей работы с ними, в том числе аннотирования. Так, на постоянно обновляемом интернет-ресурсе Tools for Corpus Linguistics [15] по состоянию на 10.10.2023 содержится информация о 277 инструментах для работы с корпусами. Каждый инструмент описан при помощи набора меток, которые можно использовать для поиска инструментов, удовлетворяющих заданным критериям. Шестьдесят пять из них доступны через интернет (метка «Web»), но если добавить критерий поддержки параллельных корпусов (метка «multilingual»), то в поисковой выдаче останутся лишь две записи: ACTRES Corpus Manager [16] и корпусный менеджер Sketch Engine [17]. Впрочем, даже это не говорит о том, что упомянутые инструменты предоставляют возможность аннотирования найденных в них примеров в соответствии с целями исследования: автору настоящей статьи не удалось обнаружить подобного функционала (доступны лишь функции формирования параллельных корпусов и последующего поиска по ним).

В этом отношении, по всей видимости, уникальным инструментом оказываются НБД и веб-приложения для работы с ними, разрабатываемые в ФИЦ ИУ РАН с 2013 г. [18, 19]. Помимо разнообразных функций поиска по параллельным текстам (см., например, [20, 21]) они дают возможность многоаспектного аннотирования переводных соответствий, содержащих вхождение исследуемой языковой единицы. Именно это сочетание характеристик — поддержка аннотирования примеров из параллельных текстов с использованием веб-приложения — отличает НБД как от параллельных корпусов (в них отсутствует функционал для аннотирования), так и от других известных автору инструментов для аннотирования (они не позволяют работать с параллельными текстами).

4 Аннотирование параллельных корпусов с использованием надкорпусных баз данных

Функционал для аннотирования примеров употребления изучаемой языковой единицы с исполь-

зованием НБД предоставляет следующие возможности¹.

1. Формирование блоков текста оригинала и перевода, необходимых и достаточных для анализа вхождения изучаемой языковой единицы и варианта ее перевода. Одной пары фрагментов параллельного текста далеко не всегда достаточно для анализа вхождения изучаемой языковой единицы. Так, в примере 1 содержимое пары не позволяет определить семантику изучаемой языковой единицы — формы *soll* (букв. ‘должен’), а также объяснить причины, которые заставили переводчика использовать в качестве варианта ее перевода форму *хочешь*.

Пример 1. Soll ich? [Н. Böll. Ansichten eines Clowns (1963)] Хочешь? [Глазами клоуна (пер. Л. Черная, 1964)]

Поэтому при аннотировании блок текста должен быть расширен за счет содержимого предыдущей пары фрагментов (пример 2).

Пример 2. Ich könnte jetzt vor deinen Augen von hier bis zur Tür humpeln, daß du vor Schmerz und Mitleid aufschreien und sofort einen Arzt anrufen würdest, den besten Chirurgen der Welt, Fretzer. Soll ich? [Н. Böll. Ansichten eines Clowns (1963)] Хочешь, я проковыляю сейчас до двери так, что ты закричишь от боли и жалости и кинешься звонить врачу, самому лучшему в мире хирургу, Фретцеру? Хочешь? [Глазами клоуна (пер. Л. Черная, 1964)]

В случае если пара, содержащая вхождение изучаемой языковой единицы, напротив, включает слишком длинные предложения, то при формировании блоков текста оригинала и перевода они могут быть сокращены. Так, в примере 2 нет необходимости включать в состав блоков текста всю предыдущую пару, поскольку в таком случае объем блоков текста будет хотя и достаточным для анализа вхождения языковой единицы и варианта ее перевода, но не необходимым. Поэтому часть первого предложения в оригинале и в переводе может быть опущена (пример 3), что обозначено при помощи «[...]».

Пример 3. Ich könnte jetzt vor deinen Augen von hier bis zur Tür humpeln [...]. Soll ich? [Н. Böll. Ansichten eines Clowns (1963)] Хочешь, я проковыляю сейчас до двери [...]? Хочешь? [Глазами клоуна (пер. Л. Черная, 1964)]

¹Для упрощения изложения здесь и далее исходим из того, что объект изучения — языковая единица, содержащаяся в текстах оригинала. По этой причине при перечислении возможностей аннотирования говорится об изучаемой языковой единице и варианте ее перевода. Объектом изучения может быть и языковая единица, содержащаяся в текстах перевода. В таком случае та языковая единица, которая соответствует ей в оригинале, может быть названа стимулом перевода.

2. Выявление вхождения изучаемой языковой единицы и варианта ее перевода. Не всегда объектом изучения оказываются лексические единицы: это могут быть, например, глагольные формы. И тогда особенно актуальным становится выявление тех слов, которые представляют собой вхождение изучаемой формы, их графическое выделение и выбор подходящей метки. Так, в примере 4 в русском языке исследуемая языковая единица — *видал* (глагол несовершенного вида в форме прошедшего времени изъявительного наклонения действительного залога), которой во французском переводе соответствует неоднословная единица *ai vu* (глагол в форме прошедшего сложного времени изъявительного наклонения — *passé composé* — действительного залога). При аннотировании эти словоформы графически выделяются (в интерфейсе НБД для этого используется полужирный шрифт), после чего добавляются метки, указывающие на то, к какой языковой единице они относятся: «Past-IPF (ind, act)» для русского и «PasséComp (ind, act)» для французского.

Пример 4. — Какое письмо? Я никакого письма не **видал**, — сказал Захар [И. Гончаров. Обломов (1848–1859)]. — Quelle lettre, je n'**ai pas vu** de lettre, dit Zakhar [Obloinov (пер. L. Jurgenson, 1988)].

3. Выбор признаков, характеризующих употребление изучаемой языковой единицы и варианта ее перевода. В зависимости от класса изучаемых языковых единиц при аннотировании примеров из корпуса необходимо добавлять сведения о тех признаках употребления этих единиц, которые значимы для их анализа.

Так, при анализе вхождений глагольных форм необходимо фиксировать, в частности, то, не сопровождается ли анализируемая форма отрицанием и не употреблена ли она в составе прямой речи. В примере 4 это именно так (и в оригинале, и в переводе), поэтому аннотатор должен добавить метки «Neg» для отрицания и «DialRepl» для прямой речи.

4. Выбор признаков, характеризующих переводное соответствие. В то время как признаки, рассматриваемые в п. 3, характеризуют либо оригинал, либо перевод, иногда возникает необходимость добавления меток, описывающих переводное соответствие в целом. Так, в примере 5 переводчик переформулировал исходный текст (не искажая его смысла): в оригинале буквально сказано 'Я хотел бы поговорить с госпожой Шнир'. Поскольку подобные изменения влияют на выбор варианта перевода, переводное соответствие следует снабдить меткой «Paragraph» для переформулирования.

Пример 5. "Ich möchte Frau Schnier sprechen", sagte ich. [H. Böll. Ansichten eines Clowns (1963)] — Позовите, пожалуйста, госпожу Шнир, — сказал я. [Глазами клоуна (пер. Л. Черная, 1964)]

Таким образом, НБД позволяет работать с четырьмя основными видами разметки примеров, найденных в параллельном корпусе.

5 Направления развития функционала для аннотирования параллельных корпусов

Несмотря на возможности аннотирования, поддержку которых обеспечивают НБД и веб-приложение для работы с ней, в ходе выполнения научных проектов, предполагающих аннотирование параллельных текстов, стало ясно, что функционал существующих инструментов аннотирования необходимо расширить.

1. Должна быть предусмотрена возможность не только обрабатывать найденные в корпусе релевантные примеры (фиксируя результаты обработки в форме аннотированных переводных соответствий), но и **снабжать специальными метками те примеры, которые нерелевантны поисковому запросу, но попали в выдачу** (так называемый «шум» в данных). Как правило, такие примеры могут попадать в поисковую выдачу либо из-за неснятой омонимии, либо при выполнении запросов на двух языках, когда искомые единицы оригинала и перевода содержатся в одной и той же паре, но не соответствуют друг другу.

Так, пример 6 иллюстрирует ситуацию, когда при осуществлении поиска глагола *faire* 'делать' во всех формах во французско-русском корпусе в поисковую выдачу попала пара, содержащая в оригинале словоформу *faits* 'факты', омонимичную словоформе *faits* 'сделавшие'.

Пример 6. "Les faits sont avoués?" dit le juge [J. Verne. Le tour du monde en quatre-vingts jours (1872)]. — Признаете ли вы факт преступления? — спросил судья [Вокруг света за восемьдесят дней (пер. Н. Габинский, 1939)].

Пример 7, найденный в англо-русском корпусе по запросу «*but* в оригинале, *однако* в переводе», иллюстрирует ситуацию, когда при поиске на двух языках в выдачу попадает пара, которая, хотя и содержит искомые единицы, не иллюстрирует нужное переводное соответствие: в этом примере союз *but*

переведен при помощи *но*, а слову *однако* в тексте оригинала ничего не соответствует.

Пример 7. “I’m getting nervous,” said Kemp. But it was five minutes before he went to the window again. “It must have been a sparrow,” he said [H. G. Wells. *The Invisible Man* (1897)]. — Однако нервы у меня расхотелись, — проговорил он про себя, но добрых пять минут не решался подойти к окну. — Воробей, должно быть, — сказал он. [Человек-невидимка (пер. Д. Вейс, 1935)].

Подобные примеры необходимо снабжать соответствующими метками, для того чтобы иметь возможность получить (1) общее число пар, найденных по некоторому запросу в корпусе, (2) число пар, содержащих релевантные соответствия, (3) число пар, содержащих нерелевантные соответствия. Все это — неотъемлемая часть корпусно-ориентированного исследования.

2. Необходимо предоставить более широкие возможности разметки блоков текста, содержащих вхождение изучаемой языковой единицы. Например, при аннотировании союза *потому что* в примере 8 может возникнуть потребность в выделении границ фрагментов, связанных при помощи этого союза (выделены одинарным и двойным подчеркиванием).

Пример 8. Обломов всегда ходил дома без галстука и без жилета, **потому что** любил простор и приволье [И. Гончаров. *Обломов* (1848–1859)]. — Chez lui, Oblomov ne portait jamais ni cravate ni gilet, **car** il aimait la liberté et l’espace [Oblomov (пер. А. Adamov, 1959)].

Еще более ценной такая разметка станет в том случае, если в переводе союз никак не переведен, но смысл предложения не искажен. Так, в примере 9 в оригинале имеется союз *car*, выражающий логико-семантическое отношение причины, а в переводе это отношение оказывается имплицитным.

Пример 9. J’étais très soucieux **car** ma panne commençait de m’apparaître comme très grave, et l’eau à boire qui s’épuisait me faisait craindre le pire [A. de Saint-Exupéry. *Le petit prince* (1942)]. — Мне было не по себе, **положение становилось серьезным, воды почти не осталось, и я начал бояться, что моя вынужденная посадка плохо кончится** [Маленький принц (пер. Н. Галь, 1959)].

Разметка фрагментов текста может быть актуальна и при изучении языковых единиц других

классов (например, для демонстрации сферы действия отрицания, модальной лексики и проч.).

3. Необходимо дать возможность указывать, по какой причине был выбран тот или иной признак употребления изучаемой языковой единицы или варианта ее перевода.

В разд. 4 при описании примера 4 было отмечено, что этот пример следует снабдить меткой, указывающей на наличие отрицания. Однако информация о том, по какой причине аннотатор решил добавить эту метку, никак не зафиксирована, хотя он, вероятно, принял такое решение из-за наличия каких-либо языковых единиц. Для большей полноты описания, а также для его проверяемости следовало бы фиксировать при аннотировании и эту информацию, в явном виде указывая, что добавленные метки «Neg» в оригинале связано с наличием *ne*, а в переводе — с наличием *n'* (усеченная форма слова *ne*) и *pas*.

6 Заключение

В статье показано, что, несмотря на обилие инструментов для работы с корпусами текстов, функционал подавляющего большинства этих инструментов весьма ограничен. С этой точки зрения разработанные в ФИЦ ИУ РАН надкорпусные базы данных и веб-приложения для работы с ними, вероятно, не имеют отечественных и зарубежных аналогов. По этой причине особенно важно анализировать опыт их использования в сфере научных исследований и образования как с целью их дальнейшего совершенствования, так и для разработки новых инструментов аннотирования параллельных корпусов.

Литература

1. Плузган В. А. Зачем нужен Национальный корпус русского языка? Неформальное введение // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. — М.: Индрик, 2005. С. 6–20. EDN: PXFYCP.
2. Перцов Н. В. О роли корпусов в лингвистических исследованиях // Корпусная лингвистика: Труды Междунар. конф. — СПб: Изд-во С.-Петерб. ун-та, 2006. С. 318–331. EDN: RGQPTB.
3. Перцов Н. В. К суждениям о фактах русского языка в свете корпусных данных // Русский язык в научном освещении, 2006. № 1(11). С. 227–245. EDN: PVNQUT.
4. Плузган В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении, 2008. № 2(16). С. 7–20. EDN: MТBALV.

5. *Tognini-Bonelli E.* Corpus linguistics at work. — Amsterdam/Philadelphia: John Benjamins Publishing Co., 2001. 235 p.
6. *Baker P., Hardie A., McEnery T.* A glossary of corpus linguistics. — Edinburgh: Edinburgh University Press, 2006. 187 p.
7. *McEnery T., Hardie A.* Corpus linguistics: Method, theory and practice. — Cambridge: Cambridge University Press, 2012. 310 p.
8. *Захаров В. П., Богданова С. Ю.* Корпусная лингвистика. — 3-е изд. — СПб: Изд-во С.-Петербур. ун-та, 2020. 234 с.
9. *Копотев М. В.* Введение в корпусную лингвистику. — Прага: Animedia Co., 2014.
10. *Добровольский Д. О.* Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов // Вестник Санкт-Петербургского университета. Язык и литература, 2020. Т. 17. № 3. С. 398–411. doi: 10.21638/spbu09.2020.303. EDN: QZIAAB.
11. *Meyer Ch. F.* Corpus-based and corpus-driven approaches to linguistic analysis: One and the same? // Developments in English. Expanding electronic evidence. — Cambridge: Cambridge University Press, 2015. P. 14–28. doi: 10.1017/CBO9781139833882.004.
12. *Xiao R.* Theory-driven corpus research: Using corpora to inform aspect theory // Corpus linguistics: An international handbook / Eds. A. Lüdeling, M. Kytö. — Berlin/New York: Walter de Gruyter, 2009. Vol. 2. P. 987–1008. doi: 10.1515/9783110213881.2.987.
13. *Gries St. Th.* Corpus linguistics and theoretical linguistics. A love–hate relationship? Not necessarily. . . // Int. J. Corpus Linguis., 2010. Vol. 15. Iss. 3. P. 327–343. doi: 10.1075/IJCL.15.3.02GRI.
14. *Sinclair J.* Trust the text: Language, corpus and discourse. — London/New York: Routledge, 2004. 224 p.
15. Tools for corpus linguistics. <https://corpus-analysis.com>.
16. ACTRES corpus manager. <https://actres.unileon.es/ACM2.0/home>.
17. Sketch engine. <https://www.sketchengine.eu>.
18. *Зацман И., Кружков М., Лоцилова Е.* Методы и средства информатики для описания структуры неоднословных коннекторов / Под ред. О. Ю. Иньковой // Структура коннекторов и методы ее описания. — М.: ТОРУС ПРЕСС, 2019. С. 205–230. doi: 10.30826/SEMANTICS19-06. EDN: YVAJWN.
19. *Кружков М. Г.* Концепция построения надкорпусных баз данных // Системы и средства информатики, 2021. Т. 31. № 3. С. 101–112. doi: 10.14357/08696527210309. EDN: UMWNIU.
20. *Гончаров А. А., Инькова О. Ю., Кружков М. Г.* Методология аннотирования в надкорпусных базах данных // Системы и средства информатики, 2019. Т. 29. № 2. С. 148–160. doi: 10.14357/08696527190213. EDN: GNDCJE.
21. *Гончаров А. А.* Поиск с исключением в параллельных текстах // Системы и средства информатики, 2023. Т. 33. № 4. С. 102–114. doi: 10.14357/08696527230410. EDN: CVPFDV.

Поступила в редакцию 15.10.23

PARALLEL CORPUS ANNOTATION: APPROACHES AND DIRECTIONS FOR DEVELOPMENT

A. A. Goncharov

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

Abstract: Possible directions for the development of parallel corpus annotation tools are presented considering the actual situation in this area. The main approaches to conducting research on corpus material — (i) corpus-based; (ii) corpus-driven; and (iii) corpus-illustrated — are considered and the differences between them are briefly described. It is demonstrated that despite the abundance of corpus annotation tools, the vast majority of them are designed to deal with monolingual corpora and/or support a very limited functionality for annotating textual data. The largest number of functions are provided by supracorpora databases and web applications to access them which are being developed at FRC CSC RAS: (i) forming of original and translated text blocks necessary and sufficient for analyzing the occurrence of the studied language unit and its translation variant; (ii) identification of the occurrence of the studied language unit and its translation variant; (iii) selection of features characterizing the use of the studied language unit and its translation variant; and (iv) selection of features characterizing the translation correspondence. This set of functions provides solutions to a significant part of research problems but it can be extended. Three directions for the development of the existing functionality are suggested which can provide a more detailed description of linguistic material.

Keywords: parallel corpus; corpus linguistics; corpus annotation; linguistic annotation

DOI: 10.14357/19922264230411

EDN: GDKDOZ

Acknowledgments

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

References

1. Plungyan, V.A. 2005. Zachem nuzhen Natsional'nyy korpus russkogo yazyka? Neformal'noe vvedenie [What the Russian National Corpus is for? Informal introduction]. *Natsional'nyy korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* [Russian National Corpus: 2003–2005. Results and prospects]. Moscow: Indrik. 6–20. EDN: PXFYCP.
2. Pertsov, N.V. 2006. O roli korpusov v lingvisticheskikh issledovaniyakh [On the role of corpora in linguistic research]. *Scientific Conference (International) “Corpus Linguistics” Proceedings*. St. Petersburg: St. Petersburg University Press. 318–331. EDN: RGQPTB.
3. Pertsov, N.V. 2006. K suzhdeniyam o faktakh russkogo yazyka v svete korpusnykh dannykh [Toward judgments about Russian language facts in the light of corpus data]. *Russkiy yazyk v nauchnom osveshchenii* [Russian Language and Linguistic Theory] 1(11):227–245. EDN: PVNQU.
4. Plungyan, V.A. 2008. Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoy korpusnoy lingvistiki [The corpus as tool and as ideology: On some lessons from modern corpus linguistics]. *Russkiy yazyk v nauchnom osveshchenii* [Russian Language and Linguistic Theory] 2(16):7–20. EDN: MTBALV.
5. Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam/Philadelphia: John Benjamins Publishing Co. 235 p.
6. Baker, P., A. Hardie, and T. McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press. 187 p.
7. McEnery, T., and A. Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press. 310 p.
8. Zakharov, V.P., and S.Yu. Bogdanova. 2020. *Korpusnaya lingvistika* [Corpus linguistics]. 3rd ed. St. Petersburg: St. Petersburg University Press. 234 p.
9. Kopotev, M.V. 2014. *Vvedenie v korpusnuyu lingvistiku* [Introduction to corpus linguistics]. Prague: Animedia Co.
10. Dobrovol'skiy, D.O. 2020. Korpusnyy podkhod k issledovaniyu frazeologii: novye rezul'taty po dannym parallel'nykh korpusov [Corpus-based approach to phraseology research: New evidence from parallel corpora]. *Vestnik Sankt-Petersburgskogo universiteta. Yazyk i literatura* [Vestnik of Saint Petersburg University. Language and Literature] 17(3):398–411. doi: 10.21638/spbu09.2020.303. EDN: QZIAAB.
11. Meyer, Ch.F. 2015. Corpus-based and corpus-driven approaches to linguistic analysis: One and the same? *Developments in English. Expanding electronic evidence*. Cambridge: Cambridge University Press. 14–28. doi: 10.1017/CBO9781139833882.004.
12. Xiao, R. 2009. Theory-driven corpus research: Using corpora to inform aspect theory. *Corpus linguistics: An international handbook*. Eds. A. Lüdeling and M. Kytö. Berlin/New York: Walter de Gruyter. 2:987–1008. doi: 10.1515/9783110213881.2.987.
13. Gries, St.Th. 2010. Corpus linguistics and theoretical linguistics. A love–hate relationship? Not necessarily. . . *Int. J. Corpus Linguist.* 15(3):327–343. doi: 10.1075/IJCL.15.3.02GRI.
14. Sinclair, J. 2004. *Trust the text: Language, corpus and discourse*. London/New York: Routledge. 224 p.
15. Tools for corpus linguistics. Available at: <https://corpus-analysis.com> (accessed November 27, 2023).
16. ACTRES corpus manager. Available at: <https://actres.unileon.es/ACM2.0/home> (accessed November 27, 2023).
17. Sketch engine. Available at: <https://www.sketchengine.eu> (accessed November 27, 2023).
18. Zatsman, I., M. Kruzhkov, and E. Loshchilova. 2019. Metody i sredstva informatiki dlya opisaniya struktury neodnoslovnnykh konnektorov [Methods and means of informatics for multiword connectives structure description]. *Struktura konnektorov i metody ee opisaniya* [Connectives structure and methods of its description]. Ed. O. Yu. Inkova. Moscow: TORUS PRESS. 205–230. doi: 10.30826/SEMANTICS19-06. EDN: YVAJWN.
19. Kruzhkov, M. 2021. Kontseptsiya postroeniya nadkorpusnykh baz dannykh [Conceptual framework for supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(3):101–112. doi: 10.14357/08696527210309. EDN: UMWNIU.
20. Goncharov, A.A., O.Yu. Inkova, and M. Kruzhkov. 2019. Metodologiya annotirovaniya v nadkorpusnykh bazakh dannykh [Annotation methodology of supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):148–160. doi: 10.14357/08696527190213. EDN: GNDCJE.
21. Goncharov, A.A. 2023. Poisk s iskluyucheniem v parallel'nykh tekstakh [Search with exclusion in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):102–114. doi: 10.14357/08696527230410. EDN: CVPFDV.

Received October 15, 2023

Contributor

Goncharov Alexander A. (b. 1994) — scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; a.gonch48@gmail.com

СТЕПЕНЬ СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ДИСКУРСИВНЫХ ОТНОШЕНИЙ: МЕТОДЫ И ИНСТРУМЕНТЫ РАСЧЕТА*

О. Ю. Инькова¹, М. Г. Кружков²

Аннотация: Рассматриваются методы оценки семантической близости дискурсивных отношений. Авторы предлагают несколько подходов к решению этой проблемы с применением двух информационных ресурсов: коллекции сформированных авторами структурированных определений логико-семантических отношений (ЛСО) и Надкорпусной базы данных коннекторов (НБДК), включающей в себя аннотации переводных соответствий текстовых фрагментов с маркерами ЛСО на русском, французском и итальянском языках. Показано, что при оценке семантической близости ЛСО высокий приоритет будут иметь такие факторы, как принадлежность различительных признаков ЛСО к одному семейству в структурированных определениях отношений, соответствия между показателями различных ЛСО в оригинальных и переводных текстах, а также случаи, когда различные ЛСО выражаются одинаковыми показателями в разных контекстах. Менее значим фактор сочетаемости различных ЛСО в рамках одного и того же контекста. Предполагается, что на основе сформулированных методов станет возможным более точно определить, какие различительные признаки ЛСО имеют наибольший вес при определении их семантической близости.

Ключевые слова: надкорпусная база данных; логико-семантические отношения; коннекторы; аннотирование; фасетная классификация

DOI: 10.14357/19922264230412

EDN: FXTSPZ

1 Степень семантической близости дискурсивных отношений

Проблемы классификации дискурсивных отношений, обеспечивающих связность текста, занимают лингвистов и специалистов по автоматической обработке текста не один десяток лет: первые исследования начались в 1970-х гг. [1, 2]. Были предложены их многочисленные классификации (ср. наиболее известные [3–7]), однако никто, насколько известно авторам, не пытался определить степень семантической близости (ССБ) дискурсивных отношений. Это связано прежде всего с тем, что классификации имеют, за редким исключением [7–9], форму списка, и этот вопрос просто не ставился. Однако его решение полезно не только для анализа текста, в том числе автоматического, но и для когнитивных наук и переводоведения, поскольку позволяет выявить общие закономерности человеческого мышления.

Кроме того, сами дискурсивные отношения определены во многом неточно или тавтологично³, схожие или идентичные отношения носят даже в англоязычных классификациях разные названия, а одинаковые названия описывают разную языковую реальность. Например, в теории сегментированного представления дискурса (Segmented Discourse Representation Theory, SDRT [10]) отношение contrast включает как отношения «вопреки ожидаемому», так и уступительные отношения. В классификации Пенсильванского аннотированного корпуса им соответствуют два отношения (opposition и contra-expectation) [7], а в теории риторической структуры — contrast и concession [11] (подробнее см. [9, с. 37]).

В этой связи были сделаны попытки сравнить существующие классификации, чтобы понять, насколько соотносимы выделяемые в них дискурсивные отношения [12–14]. В [14] для этого применяется набор различительных признаков. Этих признаков, однако, недостаточно, чтобы сформулировать уникальное определение отношения,

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

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Женевский университет, olyainkova@yandex.ru

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук, magnit75@yandex.ru

³ См., например, определение отношения альтернативы (disjunction) в теории риторической структуры: (а) элемент представляет собой (не обязательно исключаящую) альтернативу другому; (б) слушающий/читатель распознает, что связанные элементы альтернативны (см. <http://www.sfu.ca/rst>).

и некоторые из них имеют одинаковый набор признаков. Это касается, например, четырех отношений (narration, precondition, background и parallel) в SDRT [14, с. 38].

В работе [15] были заложены основы для разработки структурированных определений дискурсивных, или в терминологии автора логико-семантических, отношений на основе применяемой в НБДК классификации. Каждое ЛСО может быть описано набором различительных признаков (см. примеры в [16] и [17]). Некоторые признаки оказываются общими для нескольких ЛСО, другие — индивидуальны, т. е. свойственны только данному ЛСО. На момент написания статьи в НБДК были описаны 26 ЛСО с использованием 52 различительных признаков. Это позволяет дать каждому ЛСО уникальное определение (см. примеры в разд. 2), а также определить ССБ ЛСО.

2 Критерии, лежащие в основе определения степени семантической близости логико-семантических отношений

В предыдущей работе авторов [17] показано, что не все различительные признаки имеют одинаковый вес при определении семантической близости ЛСО и что, предположительно, наибольшее значение имеет принадлежность общих признаков к одному семейству.

В основе уступительных ЛСО и ЛСО «вопреки ожидаемому» лежат разные базовые операции: импликация — для первого и сравнение, устанавливающее несходство p и q , — для второго (табл. 1).

Это значит, что эти два ЛСО находятся в разных семантических группах. Оба ЛСО при этом установлены на пропозициональном уровне, т. е. непосредственно между положениями дел p и q , которые они связывают, и оба используют отрицательный коррелят одного из положений вещей. Иначе говоря, признаки «как правило, если имеет место q , то не имеет места p » и «положение вещей q служит аргументом в пользу ожидаемого вывода не- r » принадлежат к одному семейству. В примере (1) с ЛСО «вопреки ожидаемому»: *Ему [...] очень неприятно было сталкиваться с народом, но он шел именно туда, где виднелось больше народу.* [Ф. М. Достоевский. Преступление и наказание], положение вещей p = «ему очень неприятно было сталкиваться с народом» ориентирует в пользу вывода r = «он не должен был бы идти к народу». Этот вывод опровергается непосредственно в q (= не- r) = «он шел именно туда, где виднелось больше народу». Семантический механизм, лежащий в основе уступительных отношений (их прототипическим показателем может считаться союз *хотя*), совпадает с этим семантическим механизмом, но в зеркальном отражении:

$$p \text{ хотя } q (q \rightarrow \text{не-}p)$$

$$p \rightarrow r \text{ но } q (q = \text{не-}r), \text{ т. е. } p \rightarrow \text{не-}q \text{ но } q.$$

Отсюда необходимость при замене *хотя* на *но* и наоборот изменить порядок следования фрагментов текста: *Ему неприятно было сталкиваться с народом, но он шел туда, где виднелось больше народу* (ЛСО «вопреки ожидаемому»); *Он шел туда, где виднелось больше народу, хотя ему неприятно было сталкиваться с народом* (ЛСО уступки) [18]. Это позволяет говорить о семантической близости двух ЛСО и, например, в классификации [7] они объединены в одну группу concession.

Таблица 1 Структурированные определения уступительных ЛСО и ЛСО «вопреки ожидаемому»

ЛСО	Базовая семантическая операция	Уровень	Дополнительные характеристики
Уступительные	— операция импликации	— пропозициональный уровень	— p и q — положения вещей; — как правило, если имеет место q , то не имеет места p
«Вопреки ожидаемому»	— операция сравнения, устанавливающая несходство p и q	— пропозициональный уровень	— q имеет большую аргументативную силу, чем p ; — положение вещей p служит аргументом в пользу ожидаемого вывода r ; — положение вещей q служит аргументом в пользу ожидаемого вывода не- r

Кроме того, сформулирована гипотеза, согласно которой при определении ССБ ЛСО могут учитываться также другие факторы:

- (1) соответствия ЛСО в оригинальных и переводных текстах;
- (2) случаи, когда разные ЛСО выражаются одним и тем же показателем;
- (3) сочетаемость показателей ЛСО в одном фрагменте текста.

В НБДК для ЛСО, имеющих структурированные определения, были получены количественные данные по этим трем критериям.

2.1 Соответствие логико-семантических отношений в оригинальных и переводных текстах

Соответствие ЛСО в оригинальных и переводных текстах означает, что некоторому ЛСО в тексте оригинала, точнее, его показателю, соответствует показатель иного ЛСО в тексте перевода. Так, если для перевода на французский язык коннектора *no* в примере (1) был выбран коннектор *mais*, также выражающий ЛСО «вопреки ожидаемому»: (2) *Il lui était désagréable, très désagréable, de rencontrer du monde mais il allait justement là où l'on en voyait le plus* [перевод Élisabeth Guertik], то в примере (3)

тот же коннектор переведен *bien que* — показателем уступительных ЛСО: *C такой поправкой смысл телеграммы становился ясен, но, конечно, трагичен. — Ainsi corrigé, le télégramme prenait un sens parfaitement clair, bien que tragique, naturellement.* [М. Булгаков. Мастер и Маргарита, перевод Claude Ligny].

Количественные данные по ЛСО, соответствующим ЛСО «вопреки ожидаемому» в оригинальных и переводных текстах на русском, французском и итальянском языках, приведены в табл. 2.

Для ЛСО «вопреки ожидаемому» в НБДК сформирована 2141 двуязычная аннотация. В 237 случаях ему соответствует уступительное ЛСО. Это подтверждает важность критерия принадлежности различительных признаков к одному семейству.

Схожую картину можно наблюдать для других отношений (табл. 3): для сопоставительных и соединительных ЛСО (основаны на общей базовой операции и имеют общий различительный признак «сходство p и q относительно некоторого “общего знаменателя”»); для ЛСО оговорки и пропозициональной альтернативы (они имеют общий различительный признак — « p и q — положения вещей, имеющие статус гипотезы»); для ЛСО одновременности и сопоставления (их различительные признаки « Tp включает в себя Tq » и « p и q актуальны для говорящего в момент речи Td » принадлежат к семейству признаков «Единство временного интервала»); для ЛСО одновременности и про-

Таблица 2 Логико-семантические отношения, соответствующие ЛСО «вопреки ожидаемому» в оригинальных и переводных текстах

ЛСО1	ЛСО2	1 + 2	1	2	1 → 2	2 → 1	Сумма
«вопреки ожидаемому»	уступительные	237	2140	853	11,07%	27,78%	38,86%
«вопреки ожидаемому»	одновременность	139	2140	1268	6,50%	10,96%	17,46%
«вопреки ожидаемому»	соединительные	149	2140	2088	6,96%	7,14%	14,10%
«вопреки ожидаемому»	сопоставительные	78	2140	807	3,64%	9,67%	13,31%
«вопреки ожидаемому»	пропозициональное сопутствование	39	2140	378	1,82%	10,32%	12,14%
«вопреки ожидаемому»	исключение из рассмотрения	8	2140	90	0,37%	8,89%	9,26%
«вопреки ожидаемому»	иллокутивное сопутствование	17	2140	471	0,79%	3,61%	4,40%
«вопреки ожидаемому»	интенциональная генерализация	8	2140	248	0,37%	3,23%	3,60%
«вопреки ожидаемому»	замещение	7	2140	294	0,33%	2,38%	2,71%
«вопреки ожидаемому»	пропозициональная коррекция	4	2140	165	0,19%	2,42%	2,61%
«вопреки ожидаемому»	условные	12	2140	1075	0,56%	1,12%	1,68%
«вопреки ожидаемому»	спецификация	11	2140	1608	0,51%	0,68%	1,20%
«вопреки ожидаемому»	исключение	5	2140	615	0,23%	0,81%	1,05%
«вопреки ожидаемому»	отрицательная альтернатива	2	2140	271	0,09%	0,74%	0,83%
«вопреки ожидаемому»	оговорка	1	2140	150	0,05%	0,67%	0,71%
«вопреки ожидаемому»	экстенциональная генерализация	2	2140	588	0,09%	0,34%	0,43%
«вопреки ожидаемому»	переформулирование	2	2140	1183	0,09%	0,17%	0,26%
«вопреки ожидаемому»	пропозициональная альтернатива	1	2140	1238	0,05%	0,08%	0,13%

Расшифровка названий столбцов: 1 + 2 — число переводных аннотаций, в которых ЛСО1 в тексте на одном языке соответствует ЛСО2 в тексте на другом языке; 1 — число аннотаций, в которых в любом из текстов представлено ЛСО1; 2 — число аннотаций, в которых в любом из текстов представлено ЛСО2; 1 → 2 — процент соответствия для ЛСО1 с ЛСО2; 2 → 1 — процент соответствия для ЛСО2 с ЛСО1; сумма — сумма двух предыдущих показателей.

Таблица 3 Соответствия других ЛСО

ЛСО1	ЛСО2	1 + 2	1	2	1 → 2	2 → 1	Сумма
соединительные	сопоставительные	272	2088	807	13,03%	33,71%	46,73%
оговорка	пропозициональная альтернатива	40	150	1238	26,67%	3,23%	29,90%
одновременность	сопоставление	180	1268	807	14,20%	22,30%	36,50%
одновременность	пропозициональное сопутствование	43	1268	378	3,39%	11,38%	14,77%

позиционального сопутствования (общий признак «Тр включает в себя Тq»).

Напротив, ЛСО, соответствующие ЛСО «вопреки ожидаемому» и представленные менее чем в 1% аннотаций (см. табл. 2), не имеют различительных признаков, принадлежащих к одному семейству, и выбор их показателей для перевода показателя ЛСО «вопреки ожидаемому» может быть квалифицирован как авторский и контекстуальный.

2.2 Разные логико-семантические отношения выражаются одним и тем же показателем

Известно, что коннекторы в значительной своей части относятся к многозначным языковым единицам, т.е. могут служить показателями более чем одного ЛСО. Так, для русского союза *и* принято

выделять пять значений: сочинительное, временного следования, добавления, результативно-следственное и несоответствия; для союза *когда* — два: одновременности и условия; у союза *но* выделяются собственно противительное и противительно-уступительное значения, а у *хотя* — уступительное и уступительно-противительное и т.д. [19–21]. Это отражают и данные НБДК, причем с указанием на частотность того или иного значения коннектора в сформированных аннотациях.

В табл. 4 приведены выборочно данные для многозначных коннекторов русского и французского языков.

Приведенные данные подтверждают прежде всего положения теории грамматикализации, согласно которым семантическая эволюция языковых единиц имеет определенные закономерности. Так, было показано, что на основе значения одновременности может развиваться семантика со-

Таблица 4 Количественные данные по ЛСО, выражаемым одним показателем

Язык	Коннектор	ЛСО	Количество аннотаций
RU	а то	отрицательная альтернатива	125
		пропозициональная альтернатива	12
		исключение из рассмотрения	6
RU	если то	условные	183
		сопоставительные	13
RU	когда	одновременность	13
		условные	1
RU	когда то	одновременность	38
		условные	6
RU	между тем	одновременность	126
		«вопреки ожидаемому»	53
		сопоставительные	11
RU	между тем как	сопоставительные	29
		одновременность	6
RU	разве	оговорка	20
		исключение	5
		исключение из рассмотрения	4
		условные	2
FR	cependant	«вопреки ожидаемому»	100
		одновременность	27
FR	en même temps	одновременность	29
		сопоставительные	1
FR	quand	одновременность	197
		условные	10

поставления и противопоставления, а также импликация [22]. Это хорошо видно на примере коннекторов *когда*, *между тем*, а также французских *sependant* 'в то же время, однако', *en même temps* 'в то же время' и *quand* 'когда' (см. табл. 4). С другой стороны, эти данные подтверждают гипотезу авторов о том, что набор ЛСО, которые может маркировать один показатель, не случаен, а включает семантически близкие ЛСО. Так, коннектор *разве* зафиксирован в НБДК как показатель ЛСО оговорок, исключения, исключения из рассмотрения и условия. Эти ЛСО имеют общие различительные признаки. Логико-семантические отношения оговорок и условия — два признака: базовая операция импликации и признаки из семейства гипотетичности; ЛСО условия и исключения устанавливаются на пропозициональном уровне, а ЛСО оговорок и исключения из рассмотрения — на уровне высказывания; ЛСО оговорок, исключения и исключения из рассмотрения обладают общими признаками на уровне семейства признаков (семантика исключения), а ЛСО исключения и исключения из рассмотрения основаны на общей базовой операции (соотнесение элемента и множества).

Таким образом, данный критерий может быть полезен при определении ССБ ЛСО и иметь достаточно высокий приоритет.

2.3 Сочетаемость логико-семантических отношений в рамках одного фрагмента текста

Третий критерий, который можно учитывать при определении ССБ ЛСО, — сочетаемость ЛСО, точнее их показателей. Здесь, однако, возникает ряд сложностей, связанных с тем, что возможность сочетаемости показателей зависит в первую очередь от морфологической природы показателя ЛСО. Как известно, коннекторы относятся к разнообразным морфологическим классам: сочинительные союзы (*и*, *а*, *но*); подчинительные союзы (*хотя*, *потому что*, *как*), так называемые «конкретизаторы союзов», перешедшие в класс коннекторов, как правило, из наречных выражений (*в то же время*, *однако*, *впрочем*); предлоги (*кроме*, *после*). Союзы, например, как сочинительные, так и подчинительные, не могут сочетаться между собой в рамках единого фрагмента текста, и, наоборот, наибольшей легкостью в сочетании именно с союзами обладают «конкретизаторы» (*но* *однако*, *но* *впрочем*, *а* *между тем*, *или* *например*, *и* *в частности*). Если для показателей некоторых ЛСО можно выявить закономерности, то другие менее избирательны в своих сочетаниях. Так, показатель ЛСО спецификации

например сочетается со всеми сочинительными союзами, а показатель ЛСО «вопреки ожидаемому» *впрочем* только с союзами *а* и *но*, т. е. показателями близких ему (*а*) или тех же (*но*) ЛСО. Можно также учитывать двухместные реализации коннекторов, т. е. такие, где компоненты коннектора находятся в каждом из соединяемых фрагментов текста, например *хотя... но*: *Хотя он меня очень уговаривал, но я не согласился*. Но такие сочетания возможны не для всех ЛСО и сужают круг возможностей для получения адекватных количественных данных.

В связи с вышесказанным при подсчете ССБ ЛСО этот критерий может использоваться лишь как дополнительный.

3 Заключение

Из четырех рассмотренных критериев определения ССБ ЛСО: (1) принадлежности различительных признаков ЛСО к одному семейству, (2) соответствия ЛСО в оригинальных и переводных текстах, (3) возможности одного показателя выражать разные ЛСО и (4) сочетаемости показателей ЛСО в одном фрагменте текста — первые три могут иметь достаточно высокий приоритет. Четвертый признак обладает, напротив, наименьшим весом при определении ССБ ЛСО.

Степень детальности разметки, а следовательно, и определений ЛСО не позволяет пока объяснить некоторые явления. Например, семантическую близость ЛСО условия и одновременности, который подтверждается как их соответствиями в оригинальных и переводных текстах, так и возможностью выражаться одним показателем (*когда*). Их общий признак «*Tr* включает в себя *Tq*» не входит в определение условных ЛСО, так как соотношение временных планов положений вещей *p* и *q* может быть самым различным в условном периоде. С другой стороны, при ЛСО одновременности различным может быть их семантическое соотношение (семантическая независимость, противопоставленность, причина, следствие и т. д.). Перевод показателя ЛСО одновременности показателем условных ЛСО наблюдается только при одновременной реализации положений вещей *p* и *q* и при возможности установить между ними отношение импликации. Семантическая близость данных двух ЛСО может быть, следовательно, установлена на более низком иерархическом уровне, а именно: при определении частных случаев его реализации. В НБДК такая возможность предусмотрена, что позволит в дальнейшем более детально описывать каждое ЛСО и его виды, а значит, более точно определить ССБ ЛСО.

Литература

1. *Hobbs J. R.* A computational approach to discourse analysis. — New York, NY, USA: Department of Computer Science, City College, City University of New York, 1976. Research Report 76-2. P. 28–38.
2. *Hobbs J. R.* Why is discourse coherent? — Menlo Park, CA, USA: SRI International, 1978. SRI Technical Note 176. 44 p.
3. *Halliday M. A. K., Hasan R.* Cohesion in English. — London: Longman, 1976. 374 p.
4. *Mann W. C., Thompson S. A.* Rhetorical structure theory: Towards a functional theory of text organization // *Text*, 1988. Vol. 8. No. 3. P. 243–281. doi: 10.1515/text.1.1988.8.3.243.
5. *Asher N.* Reference to abstract objects in discourse. — Dordrecht: Kluwer, 1993. 455 p.
6. *Halliday M. A. K.* An introduction to functional grammar. — 2nd ed. — London: Edward Arnold, 1994. 434 p.
7. PDTB Research Group. The Penn Discourse Treebank 2.0 annotation manual. — Philadelphia, PA, USA: Institute for Research in Cognitive Science, University of Pennsylvania, 2007. Technical Report IRCS-08-01. 104 p. <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf>.
8. *Breindl E., Volodina A., Waßner U. H.* Handbuch der deutschen Konnektoren 2: Semantik der deutschen Satzverknüpfers. — Berlin: Walter de Gruyter, 2014. 1327 p.
9. *Инькова О. Ю.* Логико-семантические отношения: проблемы классификации // *Связность текста: ме-реологические логико-семантические отношения*. — М.: ЯСК, 2019. С. 11–98.
10. *Asher N., Lascarides A.* Logics of conversation. — Cambridge: Cambridge University Press, 2003. 526 p.
11. *Carlson L., Marcu D.* Discourse tagging reference manual. — Marina del Rey, CA, USA: Information Sciences Institute, University of Southern California, 2001. Technical Report ISI-TR-545. 87 p.
12. *Chiaros Ch.* Towards interoperable discourse annotation: Discourse features in the Ontologies of Linguistic Annotation // 9th Conference (International) on Language Resources and Evaluation Proceedings / Eds. N. Calzolari, K. Choukri, T. Declerck, et al. — Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. P. 4569–4577.
13. *Benamara F., Taboada M.* Mapping different rhetorical relation annotations: A proposal // 4th Joint Conference on Lexical and Computational Semantics Proceedings / Eds. M. Palmer, G. Boleda, P. Rosso. — Denver, CO, USA: Association for Computational Linguistics, 2015. P. 147–152. doi: 10.18653/v1/S15-1016.
14. *Sanders T., Demberg V., Hoek J., Scholman M., Asr F. T., Zufferey S., Evers-Vermeul J.* Unifying dimensions in coherence relations: How various annotation frameworks are related // *Corpus Linguist. Ling.*, 2018. Vol. 17. No. 1. P. 1–71. doi: 10.1515/cllt-2016-0078.
15. *Инькова О. Ю.* Определения дискурсивных отношений: опыт Надкорпусной базы данных коннекторов // *Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог»*. — М.: РГГУ, 2021. Вып. 20(27). С. 328–338.
16. *Инькова О. Ю., Кружков М. Г.* Структурированные определения дискурсивных отношений в Надкорпусной базе данных коннекторов // *Информатика и её применения*, 2021. Т. 15. Вып. 4. С. 27–32. doi: 10.14357/19922264210404. EDN: EZJXVI.
17. *Инькова О. Ю., Кружков М. Г.* Критерии определения семантической близости дискурсивных отношений // *Информатика и её применения*, 2023. Т. 17. Вып. 3. С. 100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.
18. *Инькова О. Ю., Нуриев В. А.* Насколько лингвостецифичен союз *хотя*? // *Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог»*. — М.: РГГУ, 2018. Вып. 17(24). С. 254–266.
19. *Словарь современного русского литературного языка: в 17 т. / Под ред. В. И. Чернышева*. — М., Л.: Изд-во Академии наук СССР / Наука, 1950–1965.
20. *Русская грамматика / Под ред. Н. Ю. Шведовой*. — М.: Наука, 1980. Т. 2. 714 с.
21. *Словарь русского языка: в 4 т. / Под ред. А. П. Евгеньевой*. — М.: Русский язык, 1981–1984.
22. *Heine B., Kuteva T.* World lexicon of grammaticalization. — Cambridge: Cambridge University Press, 2002. 387 p.

Поступила в редакцию 15.10.23

EVALUATING THE DEGREE OF DISCOURSE RELATIONS SEMANTIC AFFINITY: METHODS AND INSTRUMENTS

O. Yu. Inkova^{1,2} and M. G. Kruzhkov¹

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

²University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland

Abstract: The methods for evaluating semantic affinity of discourse relations are examined. The authors propose several approaches to this problem using two information resources: a collection of structured definitions of logical-semantic relations (LSRs) formed by the authors and the Supracorpora Database of Connectives incorporating

corpus-based annotations of translation correspondences that include text fragments with LSR markers in Russian, French, and Italian. It is demonstrated that when it comes to assessing the semantic affinity of LSRs, the following factors will be of a higher priority: affiliation of distinctive features of LSRs with the same family in the structured definitions of relations; correspondences between markers of different LSRs in the source and target texts; and cases when different LSRs are regularly expressed by the same markers in different contexts. Of a lesser importance is the factor of compatibility of different LSRs within the same context. It is assumed that based on the proposed methods, it will become possible to specify more precisely which distinguishing features of LSRs have the greatest impact on their potential semantic affinity.

Keywords: supracorpora database; logical-semantic relations; connectives; annotation; faceted classification

DOI: 10.14357/19922264230412

EDN: FXTSPZ

Acknowledgments

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

References

- Hobbs, J. R. 1976. A computational approach to discourse analyses. New York, NY: Department of Computer Science, City College, City University of New York. Research Report 76-2. 28–38.
- Hobbs, J. R. 1978. Why is discourse coherent? Menlo Park, CA: SRI International. SRI Technical Note 176. 44 p.
- Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longman. 374 p.
- Mann, W. C., and S. A. Thompson. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *Text* 8(3):243–281. doi: 10.1515/text.1.1988.8.3.243.
- Asher, N. 1993. *Reference to abstract objects in discourse*. Dordrecht: Kluwer. 455 p.
- Halliday, M. A. K. 1994. *An introduction to functional grammar*. 2nd ed. London: Edward Arnold. 434 p.
- PDTB Research Group. 2007. The Penn Discourse Treebank 2.0 annotation manual. Philadelphia, PA: Institute for Research in Cognitive Science, University of Pennsylvania. Technical Report IRCS-08-01. 104 p. Available at: <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf> (accessed November 28, 2023).
- Breindl, E., A. Volodina, and U. H. Waßner. 2014. *Handbuch der deutschen Konnektoren 2: Semantik der deutschen Satzverknüpfers*. Berlin: Walter de Gruyter. 1327 p.
- Inkova, O. Yu. 2019. Logiko-semanticheskie otnosheniya: problemy klassifikatsii [Logical-semantic relations: Classification problems]. *Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya* [Text coherence: Mereological logical semantic relations]. Moscow: LRC Publishing House. 11–98.
- Asher, N., and A. Lascarides. 2003. *Logics of conversation*. Cambridge: Cambridge University Press. 526 p.
- Carlson, L., and D. Marcu. 2001. Discourse tagging reference manual. Marina del Rey, CA: Information Sciences Institute, University of Southern California. Technical Report ISI-TR-545. 87 p. Available at: <https://www.isi.edu/marcu/discourse/tagging-ref-manual.pdf> (accessed November 28, 2023).
- Chiarcos, Ch. 2014. Towards interoperable discourse annotation: Discourse features in the Ontologies of Linguistic Annotation. *9th Conference (International) on Language Resources and Evaluation Proceedings*. Eds. N. Calzolari, K. Choukri, T. Declerck, et al. Reykjavik, Iceland: European Language Resources Association. 4569–4577.
- Benamara, F., and M. Taboada. 2015. Mapping different rhetorical relation annotations: A proposal. *4th Joint Conference on Lexical and Computational Semantics*. Eds. M. Palmer, G. Boleda, and P. Rosso. Denver, CO, USA: Association for Computational Linguistics. 147–152. doi: 10.18653/v1/S15-1016.
- Sanders, T., V. Demberg, J. Hoek, M. Scholman, F. T. Asr, S. Zufferey, and J. Evers-Vermeul. 2018. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguist. Ling.* 17(1):1–71. doi: 10.1515/cllt-2016-0078.
- Inkova, O. Yu. 2021. Opredeleniya diskursivnykh otnosheniy: opyt Nadkorpustnoy bazy dannykh konnektorov [Definition of discursive relations: The experience of the supracorpora database of connectors]. *Komp'yuternaya lingvistika i intellektual'nye Tekhnologii: Po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog"* [Computational Linguistics and Intellectual Technologies: Papers from the Annual Conference (International) “Dialogue”]. Moscow: RGGU. 20(27):328–338.
- Inkova, O. Yu., and M. G. Kruzhkov. 2021. Strukturirovannye opredeleniya diskursivnykh otnosheniy v Nadkorpustnoy baze dannykh konnektorov [Structured definitions of discourse relations in the Supracorpora Database of Connectives]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):27–32. doi: 10.14357/19922264210404. EDN: EZJXVI.
- Inkova, O. Yu., and M. G. Kruzhkov. 2023. Kriterii opredeleniya semanticheskoy blizosti diskursivnykh otnosheniy [Evaluation criteria for discourse relations semantic affinity]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.

18. Inkova, O. Yu., and V. A. Nuriev. 2018. Naskol'ko lingvospetsifichen soyuz *khotya*? [To what extent is the conjunction *khotya* language-specific?]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog"* [Computational Linguistics and Intellectual Technologies: Papers from the Annual Conference (International) "Dialogue"]. Moscow: RGGU. 17(24):254–266.
19. Chernyshev, V. I., ed. 1950–1965. *Slovar' sovremennogo russkogo literaturnogo yazyka* [Dictionary of modern Russian literary language]. In 17 vols. Moscow, Leningrad: USSR Academy of Sciences Publishing House/Nauka.
20. Shvedova, N. Yu., ed. 1980. *Russkaya grammatika* [Russian grammar]. Moscow: Nauka. Vol. 2. 714 p.
21. Evgen'eva, A. P., ed. 1981–1984. *Slovar' russkogo yazyka* [Dictionary of the Russian language]. Moscow: Russkiy yazyk. 4 vols.
22. Heine, B., and T. Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. 387 p.

Received October 5, 2023

Contributors

Inkova Olga Yu. (b. 1965) — Doctor of Science in philology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; faculty member, University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland; olyainkova@yandex.ru

Kruzhkov Mikhail G. (b. 1975) — senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

НАУЧНАЯ ПАРАДИГМА ИНФОРМАТИКИ: КЛАССИФИКАЦИЯ ОБЪЕКТОВ ПРЕДМЕТНОЙ ОБЛАСТИ*

И. М. Зацман¹

Аннотация: Дано описание трех верхних уровней классификации объектов предметной области информатики, которая позиционируется как составная часть системы научного знания, охватывающая широкий спектр информационных и компьютерных наук. Границы предметной области при таком позиционировании значительно расширяются и во многом соответствуют концепции полиадического компьютеринга Пола Розенблума. В предлагаемой научной парадигме информатики все сущности ее предметной области разделены на два глобальных класса: объекты и их трансформации. Для каждого такого класса в процессе создания парадигмы конструируется своя классификация. С формирования этих классификаций и началось построение парадигмы. В данной статье рассматриваются три верхних уровня классификации объектов предметной области информатики. Основанием для построения первого уровня (самого верхнего) служит предложенное ранее деление предметной области информатики на среды: ментальную, сенсорно воспринимаемую, цифровую и ряд других сред. Основанием для построения второго уровня служит разделение сенсорно воспринимаемых объектов информатики на данные и знаковую информацию как результат преобразования в знаковую форму когнитивных структур человека. Основанием для построения третьего уровня служит типология знаковых систем А. Соломоника. Цель статьи состоит в описании подхода к построению трех верхних уровней классификации объектов информатики и его сопоставлении с ранее использованными подходами к описанию основных сущностей ее предметной области. Также на основе предлагаемого подхода отчасти сформулированы ответы на те вопросы Томаса Куна об основных сущностях предметной области, которые должна содержать парадигма любой науки, а не только информатики.

Ключевые слова: научная парадигма; классификация объектов предметной области информатики; основания классификации; среды предметной области

DOI: 10.14357/19922264230413

EDN: FIUQAT

1 Введение

Вопросы позиционирования информатики в системе научного знания дебатировались на протяжении нескольких десятилетий. Они обсуждаются, например, в работах К. Ньюгора, Д. Харела, К. Колина, Р. Гиляревского, П. Деннинга, П. Розенблума, М. Касперсена и его соавторов [1–8]. Эти работы были опубликованы в период с 1986 по 2019 г. Сегодня эти вопросы по-прежнему сохраняют свою актуальность и активно обсуждаются, в том числе с позиции поиска необходимой совокупности теоретических оснований для разработки стратегии компьютерного образования. В [9] М. Тедре и Д. Паюнен не только рассмотрели камни преткновения и проблемные места в поиске теоретических оснований и в определении предметной области информатики в интересах компьютерного образования. Они также констатировали, что наибольшим препятствием выступает не что иное, как *отсутствие конвенциональной научной парадигмы информатики в трактовке Т. Куна* [10].

Необходимость создания именно куновской парадигмы в интересах компьютерного образования отмечалась и раньше [11]. Неослабевающий интерес именно к парадигмальности теоретических оснований обусловлен тем, что с точки зрения Т. Куна научная парадигма любой науки, чтобы обеспечить эффективные научные исследования, должна содержать или позволять получать ответы на следующие вопросы [10, 11]:

- Из каких основных сущностей состоит ее предметная область?
- Какие трансформации описывают взаимодействие сущностей друг с другом и какие сущности взаимодействуют с органами чувств?
- Какие вопросы допустимо задать об основных ее сущностях и какие методы используются для поиска решений ее задач?

В настоящее время отсутствует конвенциональная парадигма информатики, которая отвечала бы на эти вопросы. Структура предлагаемого варианта научной парадигмы, первоначальная версия

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

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук; izatsman@yandex.ru

которого отчасти позволяет ответить на них, содержит восемь компонентов, перечисленных и специфицированных в работах А. Соломоника [12, 13]. Согласно Соломонику, научная парадигма «зрелой» науки состоит из следующих компонентов, которые могут разрабатываться отдельно, но объединяются в единую и цельную конструкцию: (1) философские основы; (2) предмет изучения; (3) методы изучения; (4) аксиоматика; (5) классификации; (6) система терминов; (7) языки [знаковые системы] науки; (8) методы верификации результатов. Сам термин «научная парадигма» трактуется им в соответствии с теорией Т. Куна, которая описывает процесс смены научных парадигм [10]. При этом А. Соломоник отмечает тот факт, что в книге Куна нет ответа на вопрос: «Из чего должна состоять парадигма любой “зрелой” науки?» [14, с. 23]. Отметим, что в данной статье рассматривается пятый компонент парадигмы (классификации сущностей предметной области информатики) и основное внимание уделяется только классификации объектов ее предметной области.

Цель статьи состоит в описании подхода к построению трех верхних уровней классификации объектов и его сопоставлении с ранее использованными подходами к описанию основных сущностей предметной области информатики (объектный, трансформационный и синтетический). Предлагаемый подход и нижеизложенные результаты первой стадии его реализации уже позволяют отчасти ответить на первые два из трех вопросов, сформулированных Куном. Отметим, что построению верхних уровней классификации *трансформаций объектов предметной области информатики* посвящена работа [15].

2 Три подхода к описанию предметной области информатики

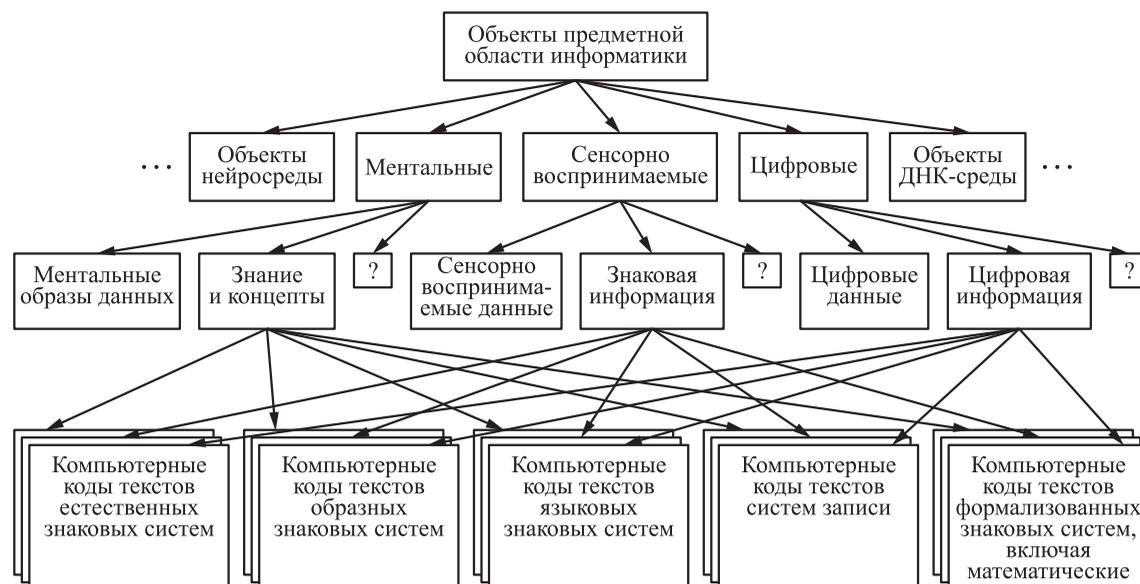
Все сущности информатики в предлагаемой научной парадигме информатики разделены на два глобальных класса: объекты и их трансформации. Для каждого такого класса в процессе создания парадигмы конструируется своя классификация. Рассмотрим три основных подхода, которые ранее использовались в информатике для описания сущностей ее предметной области: объектный, трансформационный и синтетический. Примерами первого подхода, в рамках которого основное внимание уделяется объектам предметной области информатики (например, данные, информация и/или знание), могут служить работы [16–18]; примерами второго подхода, в рамках которого основное

внимание уделяется трансформациям и в меньшей степени трансформируемым объектам — работы [6, 19]; примерами третьего, синтетического, подхода, в котором уделяется внимание и объектам предметной области информатики, и их трансформациям, могут служить работы [20–24].

Для построения двух классификаций (объектов и их трансформаций) наибольший интерес представляют трансформационный и синтетический подходы. В 2009 г. П. Деннинг и П. Розенблум сформулировали суть информатики как компьютеринга следующим образом: «... информатика — это не просто алгоритмы и структуры данных; это трансформации представлений» [6]. Чуть позже, в контексте краткого описания парадигмы информатики как компьютеринга, П. Деннинг и П. Фриман изменили эту формулировку на такую: «*Центральный объект* внимания в информатике можно определить как информационные процессы — естественные или искусственные процессы, *преобразующие информацию* (курсив мой — И. З.)» [19].

На синтетическом подходе остановимся чуть подробнее, так как именно его обобщение использовалось в процессе построения обеих классификаций в процессе создания научной парадигмы информатики. Согласно Р. Гиляревскому, «если *данные воспринимаются и интерпретируются* человеком, то они становятся для него *информацией*» (курсив мой — И. З.) [24, с. 10]. В работах [20–23] внимание также уделяется не только таким объектам предметной области, как информация и знание, но и их трансформациям. Согласно Ю.А. Шрейдеру, «информационно-знаниевые процессы включают преобразование человеческих (в значительной мере личностных) знаний, существующих “здесь” и “теперь”, в социальную информацию, доступную “везде” и “всегда” и *гарантирующую возможность извлечения из нее знаний* (курсив мой — И. З.)» [21].

Главный вывод П. Ингверсена состоит в том, что понятие информации должно удовлетворять двум требованиям. С одной стороны, информация является результатом трансформации в знаковую форму когнитивных структур человека. С другой стороны, информация — это то, что трансформирует (может трансформировать) знания человека, который ее получает [22]. В описании предметной области информатики как информационной науки Дж. Фаррадейн сопоставляет знание и информацию, которые он позиционирует как центральные понятия ее предметной области, а также рассматривает трансформацию информации в знания [20]. По мнению В. Хьорланда, в определении информационной науки в явном виде должна быть выражена соотнесенность информации со знанием [23]. Перечисленные положения из работ [20–24] и стали



Три верхних уровня классификации объектов предметной области информатики

предпосылками формирования двух классификаций сущностей предметной области информатики: ее объектов (см. рисунок) и их трансформаций (см. рисунок на с. 130 в [15]).

В использовании этих положений можно увидеть авторскую позицию при построении обеих классификаций: по максимуму интегрировать уже известные в информатике положения об объектах ее предметной области и их трансформациях, которые не противоречат друг другу.

3 Среды и объекты предметной области информатики

Как отмечено выше, для каждого из двух глобальных классов сущностей предметной области информатики (объекты и их трансформации) в процессе создания ее парадигмы конструируется своя классификация. С описания оснований построения этих классификаций и началось построение парадигмы [25, 26]. В этих работах было дано описание основания для построения верхнего уровня классификации объектов предметной области, основанной на ее *средовом делении*. Были даны определения следующих пяти сред предметной области, каждая из которых включает объекты одной и той же природы [25, 26]:

- (1) *ментальная среда* — это совокупность когнитивных феноменов и их структур, используемых и/или формируемых в процессе по-

знания, происходящего в сознании людей (феномены, формируемые с использованием знаковых систем в процессах познания как смысловые элементы знания, будем называть *концептами*);

- (2) *сенсорно воспринимаемая среда*, которую для краткости иногда будем называть *информационной* — это совокупность сенсорно воспринимаемых объектов, находящихся вне сознания, но взаимодействующих с когнитивными феноменами и их структурами;
- (3) *цифровая среда* — это совокупность компьютерных кодов;
- (4) *нейросреда* — это электрические потенциалы и магнитные поля, генерируемые мозгом, которые используются, например, в информационных технологиях (ИТ) управления роботизированной рукой [27] и в других ИТ, применяющих интерфейсы «мозг–компьютер»;
- (5) *ДНК-среда* — это совокупность цепочек РНК и ДНК¹.

Еще в 1986 г. Кристен Нюгор имплицитно включил в предметную область информатики объекты ментальной среды [1], что фактически можно рассматривать как ее добавление к сенсорно воспринимаемой и цифровой средам. В научной парадигме информатики число сред предлагается увеличить. С одной стороны, добавляются объекты еще двух сред (нейро- и ДНК-среды), с другой стороны, учитывается и возможность включения в будущем

¹ Например, модели трансляции естественных ДНК, созданные микробиологами, используются в информатике при разработке методов записи данных большого объема с применением синтезированных цепочек ДНК.

объектов тех сред, которые не используются сейчас в ИТ, системах и средствах информатики. Поэтому в предлагаемой научной парадигме информатики такая возможность увеличения числа сред устанавливается с самого начала, что в будущем при построении четвертого компонента парадигмы (аксиоматика) планируется сформулировать как одну из аксиом информатики.

4 Верхние уровни классификации объектов

В соответствии с перечисленными пятью средами верхний уровень классификации будет включать как минимум пять классов объектов. Каждый из этих классов содержит объекты только одной среды предметной области: ментальной, информационной, цифровой, нейро- или ДНК-среды. При этом с ростом разнообразия природы объектов верхний уровень классификации может пополняться новыми классами, включающими объекты, природа которых отличается от природы сред, ранее включенных в верхний уровень [26]. Это может произойти, например, в том случае, когда при проектировании будущих ИТ встретятся объекты, которые по своей природе не относятся ни к одной из сред текущей версии классификации [28]. Таким образом, в предлагаемом варианте парадигмы информатики классификацию объектов ее предметной области предлагается сделать открытой, что обусловлено возможным включением в будущем в нее объектов ранее не рассматривавшейся природы.

Перечень классов объектов верхнего уровня (см. рисунок) сформирован согласно числу сред разной природы (объекты нейросреды, ментальной, сенсорно воспринимаемой, цифровой, ДНК-среды и будущих сред, обозначенных на верхнем уровне многоточием) [25]. Далее ограничимся детализацией объектов только трех сред: ментальной, сенсорно воспринимаемой и цифровой.

В каждом из трех классов объектов (ментальных, сенсорно воспринимаемых и цифровых) выделим два подкласса, разместив их на следующем, втором, уровне классификации. Перечислим шесть выделенных подклассов:

- (1) ментальные образы данных;
- (2) знание и составляющие его концепты;
- (3) сенсорно воспринимаемые данные;
- (4) знаковая информация;
- (5) цифровые данные;
- (6) цифровая информация.

Оставшиеся три части от выделения этих подклассов, обозначенные на рисунке знаком «?», в статье не рассматриваются. Предлагаемое выделение шести подклассов соответствует ранее предложенному определению сред предметной области информатики [25]. Необходимость выделения как минимум двух подклассов в каждой из трех сред при решении прикладных задач проиллюстрируем примером кардиограммы, сформированной на основе электрических полей, образующихся при регистрации работы сердца, и сохраненной в компьютере в цифровой форме.

Кардиограмма на экране компьютера до начала ее интерпретации кардиологом — это *сенсорно воспринимаемые данные*. Информация заключения кардиолога, созданная им как результат содержательного анализа данных, является *знаковой* и по определению [25] принадлежит к сенсорно воспринимаемой среде. При компьютерном кодировании (на границе между сенсорно воспринимаемой и цифровой средами) знаковой информации и сенсорно воспринимаемых данных получаем, соответственно, цифровую информацию и цифровые данные. Это два принципиально разных подкласса объектов цифровой среды, так как цифровая информация всегда есть результат кодирования знаковой информации, а для цифровых данных это не так: они могут быть и результатом кодирования сенсорно воспринимаемых данных, и результатом их непосредственной генерации в цифровой среде (например, файл цифровых данных, полученный с помощью магнитно-резонансной томографии).

Процесс содержательного анализа и интерпретации данных кардиограммы, а также подготовки заключения состоит из нескольких этапов трансформаций объектов трех сред: ментальной, сенсорно воспринимаемой и цифровой. Сначала кардиолог сенсорно воспринимает данные кардиограммы, предварительно сохраненной в компьютере (цифровая среда), а затем отображенной на экране (сенсорно воспринимаемая среда). На следующем этапе появляются *ментальные образы данных* в сознании кардиолога как результат их сенсорного восприятия и затем генерируется *знание* (ментальная среда) кардиолога о наличии или отсутствии нарушений в работе сердца пациента как результат креативного процесса, включающего понимание кардиологом ментальных образов данных.

Далее следует этап подготовки заключения на некотором естественном языке, включающий генерацию *концептов* (ментальная среда) как результат деления полученного знания и выражение сгенерированных концептов словами этого языка в сенсорно воспринимаемой среде. Если готовится двуязычное заключение, например на русском языке

и в переводе на английский язык, то деление знания на концепты выполняется по-разному в системах двух языков. Причина разного деления одного и того же знания на концепты состоит в том, что объемы значений (концептов) слов в переводе, как правило, отличаются от объемов значений слов в оригинале. Таким образом, деление одного и того же знания кардиолога (о наличии или отсутствии нарушений в работе сердца пациента) на концепты зависит от используемого естественного языка, т. е. *вербальной знаковой системы*, используемой на границе между ментальной и сенсорно воспринимаемой средами.

В результате второй верхний уровень первоначальной версии классификации включает следующие объекты ментальной, сенсорно воспринимаемой и цифровой сред:

- знание, концепты и ментальные образы сенсорно воспринимаемых данных [29]¹;
- знаковую информацию и сенсорно воспринимаемые данные;
- цифровую информацию и цифровые данные.

При построении третьего уровня классификации объектов в ее первоначальной версии ограничимся только текстовой знаковой информацией, ее компьютерными кодами (цифровой информацией) и тем знанием ментальной среды, которое может быть представлено в текстовой знаковой форме. Основанием для построения этого уровня классификации служит следующая типология знаковых систем А. Соломоника [31, с. 131]: (1) естественные знаковые системы (например, классификатор следов зверей); (2) образные (система дорожных знаков); (3) естественно-языковые; (4) вербально-несловесные системы записи²; (5) формализованные знаковые системы³, включая математические.

Используя эту типологию, введем понятие обобщенного текста — это текст, который может быть создан с использованием любой из пяти перечисленных знаковых систем. Тогда обобщенные тексты по определению могут быть естественными, образными, естественно-языковыми, вербально-несловесными и формализованными. Таким образом, третий уровень классификации в этой статье охватывает не все виды объектов предметной области информатики, а только 15 их видов (по пять видов для каждой из трех сред).

На рисунке приведены надписи только для пяти видов объектов цифровой среды, полученных при

детализации цифровой информации. Под ними следующие пять видов объектов сенсорно воспринимаемой среды (без надписей на рисунке), полученные при детализации знаковой информации — это тексты в каждой из пяти видов знаковых систем. И еще ниже пять видов объектов ментальной среды без надписей — это пять сфер знаний, для представления которых в сенсорно воспринимаемой среде применяются соответственно естественные, образные, естественно-языковые, вербально-несловесные или формализованные знаковые системы. Отметим, что эти сферы могут пересекаться, так как существует такое знание, которое допускает его представление в сенсорно воспринимаемой среде с помощью разных видов знаковых систем.

5 Заключение

Описание трех верхних уровней классификации объектов, относящейся к пятому компоненту (классификации) предлагаемой научной парадигмы, и их визуализация на рисунке представляют собой только отдельные фрагменты этого компонента. Наполнение каждого из уровней классификации, естественно, не претендует на полноту. Но и такое первоначальное описание классификации объектов уже позволяет ответить на первые два из трех вопросов Т. Куна.

На первый вопрос (Из каких основных сущностей состоит предметная область информатики?) предлагается следующий ответ:

«Все основные сущности сначала делятся на два глобальных класса: объекты и их трансформации. Для каждого класса формируется своя классификация, и они в совокупности специфицируют основные сущности предметной области информатики».

Ответ на второй вопрос (Какие трансформации описывают взаимодействие сущностей друг с другом и какие сущности взаимодействуют с органами чувств?) состоит из следующих двух положений:

- (1) с органами чувств человека взаимодействуют объекты сенсорно воспринимаемой среды предметной области информатики;
- (2) виды трансформаций объектов представлены в классификации, первоначальное описание которой приведено в работе [15].

В настоящее время создание парадигмы информатики находится на первоначальной стадии,

¹ Деление на личностное, коллективное, организационное и конвенциональное знание по Ниссену [30] (конвенциональное знание по Ниссену названо социальным у Ю. Шрейдера [21]) в первоначальной версии этой классификации не рассматривается.

² Под системой записи понимается знаковая система, сочетающая вербальные знаки с несловесными (языки нотной записи, карт, таблиц и др.).

³ В двух последующих монографиях А. Соломоник делит формализованные системы на знаковые формализованные системы первого и второго порядка [12, с. 76; 13, с. 64].

на которой формируется только пятый из восьми ее компонентов — классификации информатики. После завершения создания первой версии компонента планируется сформировать второй и третий компоненты парадигмы (предмет и те методы изучения, которые свойственны именно информатике, включая метод информационно-математических трансформаций [32]), описав второй, третий и пятый компоненты в *единой системе терминов* как шестого компонента парадигмы.

В заключение отметим, что сопоставление предлагаемого подхода к построению трех верхних уровней классификации объектов предметной области с ранее использованными подходами позволяет выдвинуть следующую гипотезу: для создания научной парадигмы информатики, охватывающей широкий спектр информационных и компьютерных наук, необходимо выделять среды в ее предметной области, учитывать увеличение числа сред в будущем, определять и объекты каждой среды, и виды их трансформаций.

Литература

1. *Nygaard K.* Program development as a social activity // Information processing / Ed. H.-J. Kugler. — North Holland: Elsevier Science Publs. B. V., IFIP, 1986. P. 189–198.
2. *Harel D.* Algorithmics: The spirit of computing. — Reading, MA, USA: Addison-Wesley, 1987. 514 p.
3. *Коллин К. К.* Становление информатики как фундаментальной науки и комплексной научной проблемы // Системы и средства информатики, 2006. Т. 16. № 3. С. 7–58.
4. *Гиляревский Р. С.* Информатика как наука об информации // Системы и средства информатики, 2006. Т. 16. № 3. С. 59–87.
5. *Denning P.* Computing is a natural science // Commun. ACM, 2007. Vol. 50. Iss. 7. P. 13–18. doi: 10.1145/1272516.1272529.
6. *Denning P., Rosenbloom P.* Computing: The fourth great domain of science // Commun. ACM, 2009. Vol. 52. Iss. 9. P. 27–29. doi: 10.1145/1562164.1562176.
7. *Rosenbloom P. S.* On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 307 p.
8. *Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E.* Informatics as a fundamental discipline for the 21st century // Commun. ACM, 2019. Vol. 62. Iss. 4. P. 58–63. doi: 10.1145/3310330.
9. *Tedre M., Pajunen J.* Grand theories or design guidelines? Perspectives on the role of theory in computing education research // ACM T. Comput. Educ., 2022. Vol. 23. Iss. 1. P. 1–20. doi: 10.1145/3487049.
10. *Кун Т.* Структура научных революций / Пер. с англ. — М.: Прогресс, 1977. 302 с. (*Kuhn T.* The structure of scientific revolutions. — Chicago, IL, USA: University of Chicago Press, 1962. 264 p.).
11. *Fincher S., Tenenberg J.* Using theory to inform capacity-building: Bootstrapping communities of practice in computer science education research // J. Eng. Educ., 2006. Vol. 95. Iss. 4. P. 265–277. doi: 10.1002/j.2168-9830.2006.tb00902.x.
12. *Соломоник А. Б.* Опыт современной теории познания. — СПб: Алетейя, 2019. 232 с.
13. *Solomonick A. B.* The modern theory of cognition. — Newcastle, U.K.: Cambridge Scholars Publishing, 2021. 242 p.
14. *Соломоник А. Б.* Парадигма семиотики. — Минск: МЕТ, 2006. 335 с.
15. *Зацман И. М.* Научная парадигма информатики: классификация трансформаций объектов предметной области // Системы и средства информатики, 2023. Т. 33. № 4. С. 126–138. doi: 10.14357/08696527230412. EDN: ZIKUWO.
16. *Ackoff R.* From data to wisdom // J. Applied Systems Analysis, 1989. Vol. 16. P. 3–9.
17. *Rowley J.* The wisdom hierarchy: Representations of the DIKW hierarchy // J. Inf. Sci., 2007. Vol. 33. Iss. 2. P. 163–180. doi: 10.1177/0165551506070706.
18. *Frické M. H.* Data–Information–Knowledge–Wisdom (DIKW) pyramid, framework, continuum // Encyclopedia of Big Data / Eds. L. Schintler, C. McNeely. — Cham: Springer, 2018. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
19. *Denning P., Freeman P.* Computing’s paradigm // Commun. ACM, 2009. Vol. 52. Iss. 12. P. 28–30. doi: 10.1145/1610252.1610265.
20. *Farradane J.* Knowledge, information, and information science // J. Inf. Sci., 1980. Vol. 2. Iss. 2. P. 75–80. doi: 10.1177/01655515800020020.
21. *Шпрейдер Ю. А.* Информация и знание // Системная концепция информационных процессов. — М.: ВНИИСИ, 1988. С. 47–52.
22. *Ingwersen P.* Information and information science // Encyclopedia of library and information science / Eds. J. D. McDonald, M. Levine-Clark. — New York, NY, USA: Marcel Dekker Inc., 1995. Vol. 56. Sup. 19. P. 137–174.
23. *Hjorland B.* Library and information science: Practice, theory, and philosophical basis // Inform. Process. Manag., 2000. Vol. 36. Iss. 3. P. 501–531. doi: 10.1016/S0306-4573(99)00038-2.
24. Информатика как наука об информации: информационный, документальный, технологический, экономический, социальный и организационный аспекты / Под ред. Р. С. Гиляревского. — М.: Фаир-Пресс, 2006. 592 с.
25. *Зацман И. М.* Теоретические основания компьютерного образования: среды предметной области информатики как основание классификации ее объектов // Системы и средства информатики, 2022. Т. 32. № 4. С. 77–89. doi: 10.14357/08696527220408. EDN: SAUWDF.
26. *Зацман И. М.* О научной парадигме информатики: верхний уровень классификации объектов ее предметной области // Информатика и её применение

- ния, 2022. Т. 16. Вып. 4. С. 73–79. doi: 10.14357/19922264220411. EDN: XZNKVI.
27. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89. doi: 10.14357/19922264190312. EDN: EHRQLF.
 28. Зацман И. М. Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2014. № 11. С. 1–15. EDN: WEPZSZ.
 29. Zatsman I. Digital spiral model of knowledge creation and encoding its dynamics // 18th Forum (International on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2023. P. 581–596.
 30. Nissen M. E. Harnessing knowledge dynamics: Principled organizational knowing & learning. — London: IRM Press, 2006. 278 p.
 31. Соломоник А. Б. Философия знаковых систем и язык. — М.: ЛКИ, 2011. 408 с.
 32. Вакуленко В. В., Зацман И. М. Формализованное описание статистической обработки информации в базах данных // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 93–99. doi: 10.14357/19922264230313. EDN: TYCAEX.

Поступила в редакцию 09.10.23

SCIENTIFIC PARADIGM OF INFORMATICS: CLASSIFICATION OF DOMAIN OBJECTS

I. M. Zatsman

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

Abstract: A description is given of the three top levels of classification of domain’s objects of informatics which is positioned as an integral part of the system of scientific knowledge that covers a wide range of information and computer sciences. With such positioning, the boundaries of the domain expand significantly and largely correspond to the concept of polyadic computing by Paul Rosenbloom. All entities of informatics in the proposed scientific paradigm are divided into two global classes: objects and their transformations. For each class, in the process of creating the paradigm, its own classification is constructed. The paradigm’s creation began with the formation of these classifications. The paper discusses the three top levels of classification of domain’s objects of informatics. The basis for constructing the first (the highest) level is the division of the domain of informatics into the media: mental, sensory, digital, and a number of other media. The basis for constructing the second level of objects’ classification is the division of sensory perceived objects of informatics into data and sign information which is the outcome of transformation of human cognitive structures into a sign form. The basis for constructing the third level of classification of objects is the typology of sign systems by A. Solomonick. The aim of the paper is to describe the approach to constructing the three top levels of classification of domain’s objects of informatics and to compare it with the previously used approaches to describing its subject domain. Also, based on the proposed approach, the answers to those questions of Thomas Kuhn about the basic entities of the subject domain which should contain the paradigm of any science, not just informatics, are partly formulated.

Keywords: scientific paradigm; classification of domain’s objects of informatics; basis of classification; subject domain media

DOI: 10.14357/19922264230413

EDN: FIUQAT

Acknowledgments

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

References

1. Nygaard, K. 1986. Program development as a social activity. *Information processing*. Ed. H.-J. Kugler. North Holland: Elsevier Science Publs. B. V., IFIP. 189–198.
2. Harel, D. 1987. *Algorithmics: The spirit of computing*. Reading, MA: Addison-Wesley. 514 p.
3. Kolin, K. K. 2006. Stanovleniye informatiki kak fundamental’noy nauki i kompleksnoy nauchnoy problemy [Being of informatics as fundamental science and the complex scientific problem]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 3:7–58.
4. Gilyarevskiy, R. S. 2006. Informatika kak nauka ob informatsii [Informatics as science of information]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 3:59–87.
5. Denning, P. 2007. Computing is a natural science. *Commun. ACM* 50(7):13–18. doi: 10.1145/1272516.1272529.

6. Denning, P., and P. Rosenbloom. 2009. Computing: The fourth great domain of science. *Commun. ACM* 52(9):27–29. doi: 10.1145/1562164.1562176.
7. Rosenbloom, P.S. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 307 p.
8. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2019. Informatics as a fundamental discipline for the 21st century. *Commun. ACM* 62(4):58–63. doi: 10.1145/3310330.
9. Tedre, M., and J. Pajunen. 2022. Grand theories or design guidelines? Perspectives on the role of theory in computing education research. *ACM T. Comput. Educ.* 23(1):1–20. doi: 10.1145/3487049.
10. Kuhn, T. 1962. *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press. 264 p.
11. Fincher, S., and J. Tenenbergh. 2006. Using theory to inform capacity-building: Bootstrapping communities of practice in computer science education research. *J. Eng. Educ.* 95(4):265–277. doi: 10.1002/j.2168-9830.2006.tb00902.x.
12. Solomonick, A. B. 2019. *Opyt sovremennoy filosofii poznaniya* [Experience of modern philosophy of cognition]. Saint Petersburg: Aletheia. 232 p.
13. Solomonick, A. B. 2021. *The modern theory of cognition*. Newcastle, U.K.: Cambridge Scholars Publishing. 242 p.
14. Solomonick, A. 2006. *Paradigma semiotiki* [The paradigm of semiotics]. Minsk: MET Publ. 335 p.
15. Zatsman, I. M. 2023. Nauchnaya paradigma informatiki: klassifikatsiya transformatsiy ob"ektov predmetnoy oblasti [Scientific paradigm of informatics: Transformations' classification of domain objects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):126–138. doi: 10.14357/08696527230412. EDN: ZIKUWO.
16. Ackoff, R. 1989. From data to wisdom. *J. Applied Systems Analysis* 16(1):3–9.
17. Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* 33(2):163–180. doi: 10.1177/0165551506070706.
18. Frické, M. H. 2018. Data–Information–Knowledge–Wisdom (DIKW) pyramid, framework, continuum. *Encyclopedia of Big Data*. Eds. L. Schintler and C. McNeely. Cham: Springer. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
19. Denning, P., and P. Freeman. 2009. Computing's paradigm. *Commun. ACM* 52(12):28–30. doi: 10.1145/1610252.1610265.
20. Farradane, J. 1980. Knowledge, information, and information science. *J. Inf. Sci.* 2(2):75–80. doi: 10.1177/01655515800020020.
21. Shreyder, Yu. A. 1988. Informatsiya i znanie [Information and knowledge]. *Sistemnaya kontseptsiya informatsionnykh protsessov* [System concept of information processes]. Moscow: VNIISI. 47–52.
22. Ingwersen, P. 1995. Information and information science. *Encyclopedia of library and information science*. Eds. J. D. McDonald and M. Levine-Clark. New York, NY: Marcel Dekker Inc. 56(19):137–174.
23. Hjørland, B. 2000. Library and information science: Practice, theory, and philosophical basis. *Inform. Process. Manag.* 36(3):501–531. doi: 10.1016/S0306-4573(99)00038-2.
24. Gilyarevskiy, R. S., ed. 2006. *Informatika kak nauka ob informatsii: informatsionnyy, dokumental'nyy, tekhnologicheskyy, ekonomicheskyy, sotsial'nyy i organizatsionnyy aspekty* [Informatics as information science: Informational, documentary, technological, economic, social, and organizational dimensions]. Moscow: FAIR-PRESS. 592 p.
25. Zatsman, I. M. 2022. Teoreticheskie osnovaniya komp'yuternogo obrazovaniya: sredy predmetnoy oblasti informatiki kak osnovanie klassifikatsii ee ob"ektov [Theoretical foundations of digital education: Subject domain media of informatics as the base of its objects' classification]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(4):77–89. doi: 10.14357/08696527220408. EDN: SAUWDF.
26. Zatsman, I. M. 2022. O nauchnoy paradigme informatiki: verkhniy uroven' klassifikatsii ob"ektov ee predmetnoy oblasti [On the scientific paradigm of informatics: The classification high level of its objects]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):73–79. doi: 10.14357/19922264220411. EDN: XZNKVI.
27. Zatsman, I. M. 2019. Interfeysy tret'ego poryadka v informatike [Third-order interfaces in informatics]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):82–89. doi: 10.14357/19922264190312. EDN: EHRQLF.
28. Zatsman, I. 2014. Table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):233–246. EDN: WEPZSZ.
29. Zatsman, I. 2023. Digital spiral model of knowledge creation and encoding its dynamics. *18th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 581–596.
30. Nissen, M. E. 2006. *Harnessing knowledge dynamics: Principled organizational knowing & learning*. London: IRM Press. 278 p.
31. Solomonick, A. B. 2011. *Filosofiya znakovykh sistem i yazyk* [Philosophy of sign systems and language]. Moscow: LKI. 408 p.
32. Vakulenko, V. V., and I. M. Zatsman. 2023. Formalizovannoe opisanie statisticheskoy obrabotki informatsii v bazakh dannykh [Formalized description of statistical information processing in databases]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):93–99. doi: 10.14357/19922264230313. EDN: TYCAEX.

Received October 9, 2023

Contributor

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

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов имени Патриса Лумумбы; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Александр Анатольевич (р. 1994) — научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, ученый секретарь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Федерального исследовательского центра «Информатика и управление» Российской академии наук

Иванов Алексей Владимирович (р. 1976) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Иванова Дарья Вадимовна (р. 1996) — аспирант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов имени Патриса Лумумбы

Инькова Ольга Юрьевна (р. 1965) — доктор филологических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; сотрудник Университета Женевы

Кружков Михаил Григорьевич (р. 1975) — старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Лапко Александр Васильевич (р. 1949) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук; профессор кафедры космических средств и технологий Сибирского государственного университета науки и технологий имени академика М. Ф. Решетнёва

Лапко Василий Александрович (р. 1974) — доктор технических наук, профессор, ведущий научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук; заведующий кафедрой космических средств и технологий Сибирского государственного университета науки и технологий имени академика М. Ф. Решетнёва

Маркова Екатерина Викторовна (р. 1987) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов имени Патриса Лумумбы

Пешкова Ирина Валерьевна (р. 1975) — кандидат физико-математических наук, доцент Петрозаводского государственного университета; старший научный сотрудник Федерального исследовательского центра «Карельский научный центр Российской академии наук»

Рабинович Яков Исаакович (р. 1947) — кандидат физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сазонтьев Владимир Владимирович (р. 1990) — программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Степанов Евгений Павлович (р. 1994) — кандидат физико-математических наук, младший научный сотрудник кафедры автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Ступников Сергей Александрович (р. 1978) — кандидат технических наук, руководитель отдела, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тубольцев Виталий Павлович (р. 1998) — аспирант Сибирского государственного университета науки и технологий имени академика М. Ф. Решетнёва

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, доцент, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник Московского центра фундаментальной и прикладной математики

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 Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2023 г.

	Вып.	Стр.
Агаларов Я. М. Об оптимизации работы резервного прибора в многолинейной системе массового обслуживания	1	89–95
Агаларов Я. М. Оптимизация схемы распределения буферной памяти узла пакетной коммутации	3	39–48
Агасандян Г. А. Многомерные баттерфляи в задачах оптимизации по CC-VaR	1	107–115
Аду К. И. Б., Маркова Е. В., Гайдамака Ю. В., Шоргин С. Я. Анализ схемы доступа с прерыванием при нарезке радиоресурсов сети пятого поколения	1	96–106
Архипов П. О., Филиппских С. Л., Цуканов М. В. Разработка новой модели ступенчатой сверточной нейронной сети для классификации аномалий на панорамах	1	50–56
Бегишев В. О. см. Сопин Э. С.		
Берговин А. К., Ушаков В. Г. Исследование систем обслуживания со смешанными приоритетами	2	57–61
Борисов А. В. Рынок с марковской скачкообразной волатильностью I: мониторинг цены риска как задача оптимальной фильтрации	2	27–33
Борисов А. В. Рынок с марковской скачкообразной волатильностью II: алгоритм вычисления справедливой цены деривативов	3	18–24
Борисов А. В. Рынок с марковской скачкообразной волатильностью III: алгоритм мониторинга цены риска по дискретным наблюдениям цен активов	4	9–16
Босов А. В. Исследование робастности численных аппроксимаций фильтра Вонэма	2	41–49
Босов А. В. Оптимальная фильтрация состояния нелинейной динамической системы по наблюдениям со случайными запаздываниями	3	8–17
Босов А. В., Иванов А. В. Технология многофакторной классификации математического контента электронной системы обучения	4	32–41
Босов А. В., Игнатов А. Н. О задаче оценки и анализа риска транспортных происшествий на рельсовом транспорте	1	73–82
Вакуленко В. В., Зацман И. М. Формализованное описание статистической обработки информации в базах данных	3	93–99
Васильев Н. С. Композициональное представление структуры игры многих лиц в моноидальной категории бинарных отношений	2	18–26
Волканов Д. Ю. см. Горшенин А. К.		
Воронцов М. О., Шестаков О. В. Среднеквадратичный риск FDR-процедуры в условиях слабой зависимости	2	34–40
Гайдамака Ю. В. см. Аду К. И. Б.		
Гайдамака Ю. В. см. Иванова Д. В.		
Гайдамака Ю. В. см. Самуйлов А. К.		
Гаримелла Р. М. см. Разумчик Р. В.		
Гончаров А. А. Аннотирование параллельных корпусов: подходы и направления развития	4	81–87
Горбунов С. А. см. Горшенин А. К.		
Горшенин А. К., Горбунов С. А., Волканов Д. Ю. О кластеризации объектов сетевой вычислительной инфраструктуры на основе анализа статистических аномалий в трафике	3	76–87
Грушо А. А., Грушо Н. А., Забейло М. И., Кульченков В. В., Тимонина Е. Е., Шоргин С. Я. Причинно-следственные связи в задачах классификации	1	43–49
Грушо А. А., Грушо Н. А., Забейло М. И., Смирнов Д. В., Тимонина Е. Е. Классификация с помощью причинно-следственных связей	3	71–75
Грушо А. А., Грушо Н. А., Забейло М. И., Тимонина Е. Е., Шоргин С. Я. Сложные причинно-следственные связи	2	84–89

Грушо Н. А. см. Грушо А. А.		
Грушо Н. А. см. Грушо А. А.		
Грушо Н. А. см. Грушо А. А.		
Дулин С. К. см. Розенберг И. Н.		
Дулина Н. Г. см. Розенберг И. Н.		
Дюкова А. П. см. Дюкова Е. В.		
Дюкова Е. В., Масляков Г. О., Дюкова А. П. Логические методы корректной классификации данных	3	64–70
Забейжайло М. И. см. Грушо А. А.		
Забейжайло М. И. см. Грушо А. А.		
Забейжайло М. И. см. Грушо А. А.		
Захаров В. Н. см. Сазонтьев В. В.		
Захаров В. Н. см. Френкель С. Л.		
Зацман И. М. Данные, информация и знание в научной парадигме информатики	1	116–125
Зацман И. М. Научная парадигма информатики: классификация объектов предметной области	4	96–103
Зацман И. М. Трансформация иерархии Акоффа в научной парадигме информатики	3	107–113
Зацман И. М. см. Вакуленко В. В.		
Зейфман А. И. см. Усов И. А.		
Иванов А. В. см. Босов А. В.		
Иванова Д. В., Маркова Е. В., Шоргин С. Я., Гайдамака Ю. В. Модели совместного обслуживания трафика eMBB и URLLC на основе приоритетов в промышленных развертываниях 5G NR	4	64–70
Игнатов А. Н. см. Босов А. В.		
Инькова О. Ю., Кружков М. Г. Критерии определения семантической близости дискурсивных отношений	3	100–106
Инькова О. Ю., Кружков М. Г. Степень семантической близости дискурсивных отношений: методы и инструменты расчета	4	88–95
Кабанов Ю. М., Сидоренко А. П. Аксиоматический взгляд на модели системного риска Роджерса–Вераарт и Судзуки–Эльсингера	1	11–17
Карпов В. И. см. Нуриев В. А.		
Кириков И. А. см. Листопад С. В.		
Ковалёв С. П. Монада диаграмм как математическая метамодель системной инженерии	2	11–17
Королев Д. О., Малеев О. Г. Исследование эффективности применения бинарных нейронных сетей при детектировании объекта на изображении	3	88–92
Кривенко М. П. Критерии выбора размерности модели факторизации	2	50–56
Кружков М. Г. см. Инькова О. Ю.		
Кружков М. Г. см. Инькова О. Ю.		
Кудрявцев А. А., Шестаков О. В. Метод оценивания параметров гамма-экспоненциального распределения по выборке со слабо зависимыми компонентами	3	58–62
Кульченков В. В. см. Грушо А. А.		
Лапко А. В. см. Тубольцев В. П.		
Лапко В. А. см. Тубольцев В. П.		
Лери М. М. Среднее расстояние в конфигурационных графах со степенным распределением	1	28–34
Листопад С. В., Кириков И. А. Метод на основе нечетких правил для управления конфликтами агентов в гибридных интеллектуальных многоагентных системах	1	66–72
Малашенко Ю. Е., Назарова И. А. Анализ загрузки многопользовательской сети при расщеплении потоков по кратчайшим маршрутам	3	33–38
Малашенко Ю. Е., Назарова И. А. Оценки распределения ресурсов в многопользовательской сети при равных межузловых нагрузках	1	83–88
Малеев О. Г. см. Королев Д. О.		
Маркова Е. В. см. Аду К. И. Б.		
Маркова Е. В. см. Иванова Д. В.		

	Вып.	Стр.
Маслов А. Р. см. Сопин Э. С		
Масляков Г. О. см. Дюкова Е. В.		
Мелехин В. Б., Хачумов В. М., Хачумов М. В. Самообучение автономных интеллектуальных роботов в процессе поисково-исследовательской деятельности	2	78–83
Назарова И. А. см. Малашенко Ю. Е.		
Назарова И. А. см. Малашенко Ю. Е.		
Нейчев Р. Г., Шибяев И. А., Стрижов В. В. Восстановление матрицы суперпозиции в задаче символьной регрессии	1	35–42
Нуриев В. А., Карпов В. И. Методология корпусно-ориентированного исследования в области контрастивной пунктуации	2	90–95
Пешкова И. В. Границы незавершенной работы в системе с повторными вызовами разных классов и показательным временем обслуживания	4	57–63
Платонова А. А. см. Самуйлов А. К.		
Рабинович Я. И. Процедура построения множества Парето для дифференцируемых критериальных функций	4	17–22
Разумчик Р. В., Румянцев А. С., Гаримелла Р. М. Вероятностная модель для оценки основных характеристик производительности марковской модели суперкомпьютера	2	62–70
Розенберг И. Н., Дулин С. К., Дулина Н. Г. Моделирование структуры интероперабельности средствами структурной согласованности	1	57–65
Румовская С. Б. Подходы к подбору специалистов при организации коллективного решения проблем	2	96–103
Румянцев А. С. см. Разумчик Р. В.		
Сазонтьев В. В., Ступников С. А., Захаров В. Н. Расширяемый подход к слиянию данных в распределенных вычислительных средах	4	42–47
Самуйлов А. К., Платонова А. А., Шоргин В. С., Гайдамака Ю. В. К моделированию эффектов обслуживания многоадресного трафика в сетях 5G NR	2	71–77
Сатин Я. А. см. Усов И. А.		
Сидоренко А. П. см. Кабанов Ю. М.		
Синицын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах, не разрешенных относительно производных	1	2–10
Смирнов Д. В. см. Грушо А. А.		
Сопин Э. С., Маслов А. Р., Шоргин В. С., Бегишев В. О. Моделирование настойчивого поведения пользователей в сетях 5G NR с адаптацией скорости и блокировками	3	25–32
Степанов Е. П. см. Шестаков О. В.		
Стрижов В. В. см. Нейчев Р. Г.		
Ступников С. А. см. Сазонтьев В. В.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Торшин И. Ю. О задачах оптимизации, возникающих при применении топологического анализа данных к поиску алгоритмов прогнозирования с фиксированными корректорами	2	2–10
Торшин И. Ю. О формировании множеств прецедентов на основе таблиц разнородных признаковых описаний методами топологической теории анализа данных	3	2–7
Тубольцев В. П., Лапко А. В., Лапко В. А. Непараметрический алгоритм автоматической классификации данных дистанционного зондирования	4	23–31
Усов И. А., Сатин Я. А., Зейфман А. И. О скорости сходимости и предельных характеристиках для одного обобщенного процесса рождения и гибели	3	49–57
Ушаков В. Г., Ушаков Н. Г. Критерии нормальности вероятностного распределения при округленных данных	1	18–27
Ушаков В. Г. см. Берговин А. К.		
Ушаков Н. Г. см. Ушаков В. Г.		
Филиппских С. Л. см. Архипов П. О.		

	Вып.	Стр.
Френкель С. Л., Захаров В. Н. Модели учета влияния статистических характеристик трафика вычислительных сетей на эффективность прогнозирования средствами машинного обучения	4	71–80
Хачумов В. М. см. Мелехин В. Б.		
Хачумов М. В. см. Мелехин В. Б.		
Цуканов М. В. см. Архипов П. О.		
Шестаков О. В., Степанов Е. П. Нелинейная регуляризация обращения линейных однородных операторов с помощью метода блочной пороговой обработки	4	2–8
Шестаков О. В. см. Воронцов М. О.		
Шестаков О. В. см. Кудрявцев А. А.		
Шибает И. А. см. Нейчев Р. Г.		
Шнурков П. В. Решение задачи оптимального управления запасом непрерывного продукта в стохастической модели регенерации со случайными стоимостными характеристиками	4	48–56
Шоргин В. С. см. Самуйлов А. К.		
Шоргин В. С. см. Сопин Э. С.		
Шоргин С. Я. см. Аду К. И. Б.		
Шоргин С. Я. см. Грушо А. А.		
Шоргин С. Я. см. Грушо А. А.		
Шоргин С. Я. см. Иванова Д. В.		

2023 AUTHOR INDEX

	Issue	Page
Adou K. Y. B., Markova E. V., Gaidamaka Yu. V., and Shorgin S. Ya. Preemption-based prioritization scheme for network resources slicing in 5G systems	1	96–106
Agalarov Ya. M. Optimization of a queue-length dependent additional server in the multiserver queue	1	89–95
Agalarov Ya. M. Optimization of the buffer memory allocation scheme of the packet switching node	3	39–48
Agasandyan G. A. Multidimensional butterflies in problems of optimization on CC-VaR	1	107–115
Arkhipov P. O., Philippskih S. L., and Tsukanov M. V. Development of a new model of a step convolutional neural network for classification of anomalies on panoramas	1	50–56
Begishev V. O. see Sopin E. S.		
Bergovin A. K. and Ushakov V. G. Analysis of the queueing systems with mixed priorities	2	57–61
Borisov A. V. Market with Markov jump volatility I: Price of risk monitoring as an optimal filtering problem	2	27–33
Borisov A. V. Market with Markov jump volatility II: Algorithm of derivative fair price calculation	3	18–24
Borisov A. V. Market with Markov jump volatility III: Price of risk monitoring algorithm given discrete-time observations of asset prices	4	9–16
Bosov A. V. Nonlinear dynamic system state optimal filtering by observations with random delays	3	8–17
Bosov A. V. Robustness investigation of the numerical approximation of the Wonham filter	2	41–49
Bosov A. V. and Ignatov A. N. On the problem of assessing and analyzing traffic accidents risk on the rail transport	1	73–82
Bosov A. V. and Ivanov A. V. Multifactor classification technology of mathematical content of e-learning system	4	32–41
Djukova A. P. see Djukova E. V.		
Djukova E. V., Masliakov G. O., and Djukova A. P. Logical methods of correct data classification	3	64–70
Dulin S. K. see Rozenberg I. N.		
Dulina N. G. see Rozenberg I. N.		
Frenkel S. L. and Zakharov V. N. Models for study of the influence of statistical characteristics of computer networks traffic on the efficiency of prediction by machine learning tools	4	71–80
Gaidamaka Yu. V. see Adou K. Y. B.		
Gaidamaka Yu. V. see Ivanova D. V.		
Gaidamaka Yu. V. see Samouylov A. K.		
Garimella R. M. see Razumchik R. V.		
Goncharov A. A. Parallel corpus annotation: Approaches and directions for development	4	81–87
Gorbunov S. A. see Gorshenin A. K.		
Gorshenin A. K., Gorbunov S. A., and Volkanov D. Yu. Toward clustering of network computing infrastructure objects based on analysis of statistical anomalies in network traffic	3	76–87
Grusho A. A., Grusho N. A., Zabezhailo M. I., Kulchenkov V. V., Timonina E. E., and Shorgin S. Ya. Causal relationships in classification problems	1	43–49
Grusho A. A., Grusho N. A., Zabezhailo M. I., Smirnov D. V., and Timonina E. E. Classification by cause-and-effect relationships	3	71–75
Grusho A. A., Grusho N. A., Zabezhailo M. I., Timonina E. E., and Shorgin S. Ya. Complex cause-and-effect relationships	2	84–89
Grusho N. A. see Grusho A. A.		
Grusho N. A. see Grusho A. A.		
Grusho N. A. see Grusho A. A.		

	Issue	Page
Ignatov A. N. see Bosov A. V.		
Inkova O. Yu. and Kruzhkov M. G. Evaluating the degree of discourse relations semantic affinity: Methods and instruments	4	88–95
Inkova O. Yu. and Kruzhkov M. G. Evaluation criteria for discourse relations semantic affinity	3	100–106
Kruzhkov M. G. see Inkova O. Yu.		
Ivanov A. V. see Bosov A. V.		
Ivanova D. V., Markova E. V., Shorgin S. Ya., and Gaidamaka Yu. V. Priority-based eMBB and URLLC traffic coexistence models in 5G NR industrial deployments	4	64–70
Kabanov Yu. M. and Sidorenko A. P. An axiomatic viewpoint on the Rogers–Veraart and Suzuki–Elsinger models of systemic risk	1	11–17
Karpov V. I. see Nuriev V. A.		
Khachumov M. V. see Melekhin V. B.		
Khachumov V. M. see Melekhin V. B.		
Kirikov I. A. see Listopad S. V.		
Korolev D. O. and Maleev O. G. Efficiency of binary neural networks for object detection on an image	3	88–92
Kovalyov S. P. The monad of diagrams as a mathematical metamodel of systems engineering	2	11–17
Krivenko M. P. Criteria for choosing the factorization model dimensionality	2	50–56
Kruzhkov M. G. see Inkova O. Yu.		
Kudryavtsev A. A. and Shestakov O. V. A method for estimating parameters of the gamma-exponential distribution from a sample with weakly dependent components	3	58–63
Kulchenkov V. V. see Grusho A. A.		
Lapko A. V. see Tuboltsev V. P.		
Lapko V. A. see Tuboltsev V. P.		
Leri M. M. An average distance in the power-law configuration graphs	1	28–34
Listopad S. V. and Kirikov I. A. Fuzzy rules based method for agent conflict management in hybrid intelligent multiagent systems	1	66–72
Malashenko Yu. E. and Nazarova I. A. Estimates of the resource distribution in the multiuser network with equal internodal loads	1	83–88
Malashenko Yu. E. and Nazarova I. A. Multiuser network load analysis by splitting flows along the shortest routes	3	33–38
Maleev O. G. see Korolev D. O.		
Markova E. V. see Adou K. Y. B.		
Markova E. V. see Ivanova D. V.		
Masliakov G. O. see Djukova E. V.		
Maslov A. R. see Sopin E. S.		
Melekhin V. B., Khachumov V. M., and Khachumov M. V. Self-learning of autonomous intelligent robots in the process of search and explore activities	2	78–83
Nazarova I. A. see Malashenko Yu. E.		
Nazarova I. A. see Malashenko Yu. E.		
Neychev R. G., Shibaev I. A., and Strijov V. V. Optimal spanning tree reconstruction in symbolic regression	1	35–42
Nuriev V. A. and Karpov V. I. Methodology of the corpus-based studies in the field of contrastive punctuation	2	90–95
Peshkova I. V. Bounds of the workload in a multiclass retrieval queue with exponential services	4	57–63
Philippskih S. L. see Arkhipov P. O.		
Platonova A. A. see Samouylov A. K.		
Rabinovich Ya. I. Procedure of constructing a Pareto set for differentiable criteria functions	4	17–22
Razumchik R. V., Rummyantsev A. S., and Garimella R. M. A queueing system for performance evaluation of a Markovian supercomputer model	2	62–70
Rozenberg I. N., Dulin S. K., and Dulina N. G. Modeling the structure of interoperability by means of structural consistency	1	57–65
Rumovskaya S. B. Selection of specialists in the organization of collective solving problems	2	96–103

	Issue	Page
Rumyantsev A. S. see Razumchik R. V.		
Samouylov A. K., Platonova A. A., Shorgin V. S., and Gaidamaka Yu. V. On modeling the effects of multicast traffic servicing in 5G NR networks	2	71–77
Satin Y. A. see Usov I. A.		
Sazontev V. V., Stupnikov S. A., and Zakharov V. N. An extensible approach to data fusion in distributed computing environments	4	42–47
Shestakov O. V. and Stepanov E. P. Nonlinear regularization of the inversion of linear homogeneous operators using the block thresholding method	4	2–8
Shestakov O. V. see Kudryavtsev A. A.		
Shestakov O. V. see Vorontsov M. O.		
Shibaev I. A. see Neychev R. G.		
Shnurkov P. V. Solution of the problem of optimal control of the stock of a continuous product in a stochastic model of regeneration with random cost characteristics	4	48–56
Shorgin S. Ya. see Adou K. Y. B.		
Shorgin S. Ya. see Grusho A. A.		
Shorgin S. Ya. see Grusho A. A.		
Shorgin S. Ya. see Ivanova D. V.		
Shorgin V. S. see Samouylov A. K.		
Shorgin V. S. see Sopin E. S.		
Sidorenko A. P. see Kabanov Yu. M.		
Sinitsyn I. N. Analytical modeling of distributions with invariant measure in stochastic systems with unsolved derivatives	1	2–10
Smirnov D. V. see Grusho A. A.		
Sopin E. S., Maslov A. R., Shorgin V. S., and Begishev V. O. Modeling insistent user behavior in 5G New Radio networks with rate adaptation and blockage	3	25–32
Stepanov E. P. see Shestakov O. V.		
Strijov V. V. see Neychev R. G.		
Stupnikov S. A. see Sazontev V. V.		
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		
Torshin I. Yu. On optimization problems arising from the application of topological data analysis to the search for forecasting algorithms with fixed correctors	2	2–10
Torshin I. Yu. On the formation of sets of precedents based on tables of heterogeneous feature descriptions by methods of topological theory of data analysis	3	2–7
Tsukanov M. V. see Arkhipov P. O.		
Tuboltsev V. P., Lapko A. V., and Lapko V. A. Nonparametric algorithm for automatic classification of remote sensing data	4	23–31
Ushakov N. G. see Ushakov V. G.		
Ushakov V. G. and Ushakov N. G. Tests for normality of the probabilistic distribution when data are rounded	1	18–27
Ushakov V. G. see Bergovin A. K.		
Usov I. A., Satin Y. A., and Zeifman A. I. On the rate of convergence and limiting characteristics for one quasi-birth–death process	3	49–57
Vakulenko V. V. and Zatsman I. M. Formalized description of statistical information processing in databases	3	93–99
Vasilyev N. S. Multiplayers’ games compositional structure in the monoidal category of binary relations	2	18–26
Volkanov D. Yu. see Gorshenin A. K.		
Vorontsov M. O. and Shestakov O. V. Mean-square risk of the FDR procedure under weak dependence	2	34–40
Zabezhailo M. I. see Grusho A. A.		
Zabezhailo M. I. see Grusho A. A.		

	Issue	Page
Zabzhailo M. I. see Grusho A. A.		
Zakharov V. N. see Frenkel S. L.		
Zakharov V. N. see Sazontev V. V.		
Zatsman I. M. On the scientific paradigm of informatics: Data, information, and knowledge	1	116–125
Zatsman I. M. Scientific paradigm of informatics: Classification of domain objects	4	96–103
Zatsman I. M. Transformation of the Ackoff’s hierarchy in the scientific paradigm of informatics	3	107–113
Zatsman I. M. see Vakulenko V. V.		
Zeifman A. I. see Usov I. A.		

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах: http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp и http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний. Если источник имеет DOI и/или EDN, то их необходимо указывать.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe khozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

Описание неопубликованного документа:

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: ieep@frccsc.ru (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^AT_EX (.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 and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazooitdachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.
- State standards and patents:**
GOST 8.586.5–2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
 - an article and additional materials have to be attached using the “attach” function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: iiep@frccsc.ru (to Svetlana Strigina)
<http://www.ipiran.ru/english/journal.asp>