

Информатика и её применения

Том 14 Выпуск 2 Год 2020

СОДЕРЖАНИЕ

О подходах к построению предельных режимов для некоторых моделей массового обслуживания	
Я. А. Сатин, А. И. Зейфман, Г. Н. Шилова	3
Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям III: случай мультипликативных шумов	
А. В. Борисов	10
Управление выходом стохастической дифференциальной системы по квадратичному критерию. V. Случай неполной информации о состоянии	
А. В. Босов	19
Асимптотика оценки среднеквадратичного риска в задаче обращения преобразования Радона по проекциям, регистрируемыми на случайной сетке	
О. В. Шестаков	26
Методы нахождения причин сбоев в информационных технологиях с помощью метаданных	
Н. А. Грушо, А. А. Грушо, М. И. Забежайло, Е. Е. Тимонина	33
Совместная оценка предсказуемости данных и качества предикторов	
С. Л. Френкель, В. Н. Захаров	40
Аппроксимация распределений размеров частиц лунного реголита на основе метода статистической симуляции выборок	
А. К. Горшенин, В. Ю. Королев	50
Введение отношения порядка на множестве параметров аппроксимирующих моделей	
А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов	58
Стационарные характеристики системы $M/G/2/\infty$ с одним частным случаем дисциплины инверсионного порядка обслуживания с обобщенным вероятностным приоритетом	
Л. А. Мейханаджян, Р. В. Разумчик	66
Оптимизация емкости основного накопителя в системе массового обслуживания типа $G/M/1/K$ с дополнительным накопителем	
Я. М. Агаларов	72
Статистические свойства двоичных неавтономных регистров сдвига с внутренним суммированием	
С. Ю. Мельников, К. Е. Самуйлов	80
Последовательный анализ серий данных на основе многомерных референсных регионов	
М. П. Кривенко	86

Информатика и её применения

Том 14 Выпуск 2 Год 2020

СОДЕРЖАНИЕ

Многофакторные модели полностью связанной линейной регрессии без ограничений на соотношения дисперсий ошибок переменных М. П. Базилевский	92
Решение задачи безусловного экстремума для дробно-линейного интегрального функционала, зависящего от параметра П. В. Шнурков, К. А. Адамова	98
Интеграционная платформа для многомасштабного моделирования нейроморфных систем К. К. Абгарян, Е. С. Гаврилов	104
Выбор моделей оптимальной сложности методами Монте-Карло (на примере моделей производственных функций регионов Российской Федерации) И. Л. Кириллюк, О. В. Сенько	111
Редуцирование спектра моделей перевода в надкорпусных базах данных В. А. Нуриев, И. М. Зацман	119
Об авторах	127
Правила подготовки рукописей	130
Requirements for manuscripts	133

О ПОДХОДАХ К ПОСТРОЕНИЮ ПРЕДЕЛЬНЫХ РЕЖИМОВ ДЛЯ НЕКОТОРЫХ МОДЕЛЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ*

Я. А. Сатин¹, А. И. Зейфман², Г. Н. Шилова³

Аннотация: Рассматриваются нестационарные модели массового обслуживания, число требований в которых описывается конечными марковскими цепями с периодическими интенсивностями. Для многих классов таких моделей в предыдущих статьях разработаны методы получения верхних оценок скорости сходимости к предельному режиму, используя которые можно находить сами предельные характеристики системы, исследовать их устойчивость по отношению к малым возмущениям интенсивностей поступления и обслуживания требований, а также получать информацию о том, насколько в каждый момент времени текущие характеристики системы отличаются от предельных. В настоящей работе изучается другая ситуация, а именно: предполагается, что явные оценки скорости сходимости к предельному режиму получить не удастся. Рассмотрены способы построения предельных режимов таких систем и методики получения информации о скорости сходимости к ним. В качестве примера рассмотрена простая модель нестационарной системы с достаточно медленной скоростью сходимости к предельному режиму.

Ключевые слова: система массового обслуживания; конечные марковские цепи; периодические интенсивности; предельные характеристики; скорость сходимости

DOI: 10.14357/19922264200201

1 Введение

Рассматривается класс нестационарных моделей массового обслуживания, описываемых конечными неоднородными марковскими цепями с 1-периодическими интенсивностями. Одной из важнейших задач при этом становится построение основных предельных характеристик системы. Проще всего проводить такое построение, если удастся получить оценки скорости сходимости к предельному режиму. Общие подходы для получения таких оценок, классы соответствующих моделей, а также возможные приложения для изучения устойчивости и построения усечений таких моделей рассмотрены в работах [1–7]. Однако иногда такие оценки получить в явном виде не удается (см., например, [8]). В этой ситуации обычно или строится решение при конкретных значениях t с заданными начальными условиями (см., например, [9, 10]) — но в таком случае не будет информации о предельном режиме системы; или, наоборот, изучаются предельные характеристики без обсуждения скорости сходимости к ним (см., например, [11]) — в этом случае неизвестно, насколько допустима замена текущих характеристик системы предельными. В на-

стоящей статье обсуждаются возможные подходы к нахождению предельных характеристик системы в такой ситуации.

Число требований $X(t)$ в таких системах описывается конечной марковской цепью с пространством состояний $\{0, 1, \dots, N\}$ и непрерывным временем $t \geq 0$.

Обозначим через

$$p_{ij}(s, t) = \Pr \{X(t) = j | X(s) = i\}, \\ i, j \geq 0, \quad 0 \leq s \leq t,$$

переходные вероятности процесса $X = X(t)$, а через $p_i(t) = \Pr \{X(t) = i\}$ — его вероятности состояний.

Будем обозначать через $\|\bullet\|$ l_1 -норму вектора и матрицы, т. е. $\|\mathbf{x}\| = \sum |x_i|$, $\|B\| = \max_j \sum_i |b_{ij}|$ при $B = (b_{ij})_{i,j=0}^N$, а через Ω — множество всех стохастических векторов, т. е. множество векторов с неотрицательными координатами и единичной l_1 -нормой.

Через $E(t, k) = E \{X(t) | X(0) = k\}$ будем далее обозначать математическое ожидание процесса (среднее число требований) в момент t при условии, что в нулевой момент времени он находится в состоянии k .

* Исследования в разд. 3 и 4 выполнены Я. Сатиным и А. Зейфманом при поддержке Российского научного фонда (проект 19-11-00020.)

¹ Вологодский государственный университет, yasovi@mail.ru

² Вологодский государственный университет; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Вологодский научный центр Российской академии наук, a_zeifman@mail.ru

³ Вологодский государственный университет, shgn@mail.ru

Напомним, что марковская цепь $X(t)$ называется слабо эргодичной, если $\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \rightarrow 0$ при $t \rightarrow \infty$ для любых начальных условий $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$ и любом $s \geq 0$. Марковская цепь $X(t)$ имеет предельное среднее $\phi(t)$, если $E(t, k) - \phi(t) \rightarrow 0$ при $t \rightarrow \infty$ и любом k .

2 Слабая эргодичность конечномерных процессов с периодической матрицей интенсивностей

Определение 1. Состояние i для однородной марковской цепи $X(t)$ называется достижимым, если для любого j существует t , при котором выполняется: $p_{ji}(t) > 0$.

Пусть $A(t) = (a_{ij}(t))_{i,j=1}^N$ — транспонированная матрица интенсивностей рассматриваемой цепи, далее будем предполагать, что все $a_{ij}(t)$ 1-периодичны и интегрируемы на $[0, 1]$. Введем «усредненную» матрицу

$$A_0 = \int_0^1 A(t) dt.$$

Теорема 1. Пусть однородная цепь с матрицей интенсивностей A_0 имеет достижимое состояние i_* . Тогда марковская цепь $X(t)$ с 1-периодической матрицей интенсивностей $A(t)$ слабо эргодична, имеет 1-периодические предельный режим $\pi(t)$ и предельное среднее $\phi(t)$, а также двойное среднее E , причем найдется $\alpha > 0$ такое, что при всех натуральных n справедливы следующие неравенства:

(1) при всех $t \geq n$

$$\|\mathbf{p}(t) - \mathbf{p}_{i_*}(t)\| \leq (1 - \alpha)^n \|\mathbf{p}(0) - \mathbf{p}_{i_*}(0)\| \quad (1)$$

для любого $\mathbf{p}(0)$;

(2) при всех $t \geq n$

$$|E\{X(t) | X(0) = k\} - \phi(t)| \leq 2N(1 - \alpha)^n \quad (2)$$

для любого k .

Доказательство этого утверждения проводится с помощью интегрирования дифференциальных неравенств, как это показано в [12] (см. рассуждения леммы 1, теоремы 19 для стационарного случая и теорем 35 и 36 в случае периодических интенсивностей).

Замечание 1. Условие существования достижимого состояния у однородной цепи с матрицей интенсивностей A_0 является и необходимым для слабой эргодичности $X(t)$.

Замечание 2. В случае непрерывности интенсивностей и наличия достижимого при всех t состояния у рассматриваемой неоднородной цепи можно гарантировать справедливость оценок теоремы уже при $n = 1$ (см. разд. 4).

Следует отметить, что оценки (1) и (2) обычно очень грубые, как показано далее в примере. Они только гарантируют существование предельных характеристик, но не дают реальной возможности их построения, поэтому и приходится использовать другие методы.

3 Построение периодического решения

Опишем способы получения предельного периодического решения произвольной марковской цепи $X(t)$, у которой есть хотя бы одно достижимое состояние и интенсивности 1-периодические.

Итак, рассмотрим прямую систему Колмогорова, описывающую вектор вероятностей состояний процесса $X(t)$:

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}(t). \quad (3)$$

Пусть для определенности достижимым для усредненной цепи является нулевое состояние, т. е. $i^* = 0$. Тогда при некотором n по теореме 1 найдется число $\omega \in (0, 1)$ такое, что верно неравенство $\|\mathbf{p}(n) - \mathbf{p}_{i^*}(n)\| \leq \omega$. Для удобства и простоты вычислений будем считать, что $n = 1$, это условие на практике не будет обременительным.

После подстановки

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

переходим к системе:

$$\frac{d\mathbf{z}}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t). \quad (4)$$

Данная система имеет решение:

$$\mathbf{z}(t) = U(t, 0)\mathbf{z}(0) + \int_0^t U(t, s)\mathbf{f}(s) ds, \quad (5)$$

где $U(t, s)$ — матрица Коши для соответствующей однородной системы

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

Обозначим через $\mathbf{p}^i(t) = (p_{i0}, p_{i1}, p_{i2}, \dots, p_{in})^T$ и $\mathbf{z}^i(t) = (z_{i1}, z_{i2}, \dots, z_{in})^T = (p_{i1}, p_{i2}, \dots, p_{in})^T$ решения систем (3) и (4) соответственно, с началь-

ным условием $\mathbf{p}(0) = e_i$ (все координаты нулевые, кроме i -й, которая равна единице).

Возьмем начальное условие e_0 . Для него $\mathbf{p}(0) = (1, 0, 0, \dots, 0)^T$ и $\mathbf{z}(0) = (0, 0, \dots, 0)^T$. После подстановки в (5) получаем:

$$\mathbf{z}^0(t) = \int_0^t U(t, s)\mathbf{f}(s) ds.$$

Для начального условия e_1 имеем $\mathbf{p}(0) = (0, 1, 0, \dots, 0)^T$ и $\mathbf{z}^1(0) = (1, 0, \dots, 0)^T$, получаем:

$$\begin{aligned} \mathbf{z}^1(t) &= U(t, 0)\mathbf{z}^1(0) + \int_0^t U(t, s)\mathbf{f}(s) ds = \\ &= U(t, 0)\mathbf{z}^1(0) + \mathbf{z}^0(t). \end{aligned}$$

Аналогично для всех $i \leq N$ имеем

$$U(t, 0)\mathbf{z}^i(0) = \mathbf{z}^i(t) - \mathbf{z}^0(t),$$

откуда получаем столбцы матрицы $U(t, 0)$ и ее саму в следующей форме:

$$U(t, 0) = \begin{pmatrix} z_{11} - z_{01} & z_{21} - z_{01} & z_{31} - z_{01} & \cdots \\ z_{12} - z_{02} & z_{22} - z_{02} & z_{32} - z_{02} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ z_{1N} - z_{0N} & z_{2N} - z_{0N} & z_{3N} - z_{0N} & \cdots \\ & \cdots & z_{N1} - z_{01} & \\ & \cdots & z_{N2} - z_{02} & \\ & & \vdots & \\ & \cdots & z_{NN} - z_{0N} & \end{pmatrix}$$

или

$$U(t, 0) = \begin{pmatrix} p_{11} - p_{01} & p_{21} - p_{01} & p_{31} - p_{01} & \cdots \\ p_{12} - p_{02} & p_{22} - p_{02} & p_{32} - p_{02} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ p_{1N} - p_{0N} & p_{2N} - p_{0N} & p_{3N} - p_{0N} & \cdots \\ & \cdots & p_{N1} - p_{01} & \\ & \cdots & p_{N2} - p_{02} & \\ & & \vdots & \\ & \cdots & p_{NN} - p_{0N} & \end{pmatrix}.$$

Из сделанного предположения $\|\mathbf{p}^i(1) - \mathbf{p}^0(1)\| \leq \leq \omega < 1$ вытекает, что $\|U(1, 0)\| \leq \omega < 1$. А тогда матрица $I - U(1, 0)$ обратима и из (5) можно в принципе точно выразить начальное состояние, соответствующее периодическому решению $\zeta(t)$:

$$\zeta(0) = (I - U(1, 0))^{-1} \int_0^1 U(1, s)\mathbf{f}(s) ds. \quad (6)$$

Однако реальное применение этой формулы затруднительно. Рассмотрим более простой способ приближенного вычисления $\zeta(0)$.

Для разности двух произвольных решений получаем при любых $t \geq s$

$$\zeta(t) - \mathbf{z}(t) = U(t, s)(\zeta(s) - \mathbf{z}(s)).$$

Обозначив $\hat{\mathbf{z}}(t) = \zeta(t) - \mathbf{z}(t)$, получаем

$$\hat{\mathbf{z}}(t) = U(t, s)\hat{\mathbf{z}}(s).$$

С учетом 1-периодичности получаем для любого натурального n равенство:

$$U(n+1, n) = U(1, 0).$$

Тогда $\|\hat{\mathbf{z}}(n)\| \leq \omega^n \|\hat{\mathbf{z}}(0)\|$, а значит,

$$\lim_{n \rightarrow \infty} \|\hat{\mathbf{z}}(n)\| = 0.$$

Следовательно,

$$\begin{aligned} \hat{\mathbf{z}}(n+2) - \hat{\mathbf{z}}(n+1) &= U(n+2, n+1)\hat{\mathbf{z}}(n+1) - \\ &- U(n+1, n)\hat{\mathbf{z}}(n) = U(n+1, n)(\hat{\mathbf{z}}(n+1) - \hat{\mathbf{z}}(n)). \end{aligned}$$

Отсюда

$$\|\hat{\mathbf{z}}(n+2) - \hat{\mathbf{z}}(n+1)\| \leq \omega \|\hat{\mathbf{z}}(n+1) - \hat{\mathbf{z}}(n)\|$$

и для любого натурального m :

$$\begin{aligned} \|\hat{\mathbf{z}}(n) - \hat{\mathbf{z}}(m)\| &\leq \|\hat{\mathbf{z}}(n) - \hat{\mathbf{z}}(n+1)\| + \\ &+ \|\hat{\mathbf{z}}(n+1) - \hat{\mathbf{z}}(n+2)\| + \cdots + \|\hat{\mathbf{z}}(m-1) - \hat{\mathbf{z}}(m)\| \leq \\ &\leq (1 + \omega + \omega^2 + \cdots) \|\hat{\mathbf{z}}(n) - \hat{\mathbf{z}}(n+1)\| \leq \\ &\leq \frac{1}{1 - \omega} \|\hat{\mathbf{z}}(n) - \hat{\mathbf{z}}(n+1)\|. \end{aligned}$$

Устремив m к бесконечности, получаем

$$\|\hat{\mathbf{z}}(n)\| \leq \frac{1}{1 - \omega} \|\hat{\mathbf{z}}(n) - \hat{\mathbf{z}}(n+1)\|.$$

С учетом

$$\begin{aligned} \|\hat{\mathbf{z}}(n+1) - \hat{\mathbf{z}}(n)\| &= \|\mathbf{z}(n+1) - \zeta(n+1) + \zeta(n) + \mathbf{z}(n)\| = \\ &= \|\mathbf{z}(n+1) - \mathbf{z}(n)\| \end{aligned}$$

имеем:

$$\begin{aligned} \|\mathbf{z}(n) - \zeta(0)\| &\leq \frac{1}{1 - \omega} \|\mathbf{z}(n) - \mathbf{z}(n+1)\| \leq \\ &\leq \frac{\omega}{1 - \omega} \|\mathbf{z}(n-1) - \mathbf{z}(n)\|. \end{aligned}$$

Таким образом, зная $\mathbf{z}(n-1)$ и $\mathbf{z}(n)$, можно оценить точность вычисления $\zeta(0)$.

С учетом неравенства

$$\|p(t) - \pi(s)\| = |p_0(t) - \pi_0(s)| + \|z(t) - \zeta(s)\| \leq \leq 2\|z(t) - \zeta(s)\|$$

получаем следующее утверждение.

Теорема 2. Любая марковская цепь $X(t)$ с 1-периодической матрицей интенсивностей $A(t)$ и хотя бы одним достижимым состоянием слабо эргодична, имеет 1-периодический предельный режим $\pi(t)$ и для нее гарантировано выполнение неравенства:

$$\|p(n) - \pi(0)\| \leq \frac{2\omega}{1 - \omega} \|p(n - 1) - p(n)\|,$$

где $\pi(t)$ — периодическое решение и $p(t)$ — произвольное решение системы (3).

Таким образом, можно указать несколько способов нахождения предельного периодического решения.

1. Зная матрицу Коши $V(1, 0)$ системы (3), можно получить начальное условие периодического решения из системы

$$\begin{aligned} V(1, 0)\pi(0) &= \pi(0); \\ \|\pi(0)\| &= 1. \end{aligned}$$

2. Зная матрицу Коши $U(1, 0)$ системы (4), можно получить начальное условие с помощью формулы (6).
3. Решать систему (3) на отрезке $[0, n]$ до тех пор, пока $(2\omega/(1 - \omega))\|p(n - 1) - p(n)\|$ не будет получено достаточно точно. В качестве вектора, приближенно задающего периодическое решение, можно тогда взять $p(n)$. Здесь $p(t)$ — произвольное решение.
4. Из неравенства (1) можно аналогично предыдущему пункту получить следующее правило. Решать систему (3) на отрезке $[0, n]$ до тех пор, пока $((1 - \alpha)/\alpha)\|p(n - 1) - p(n)\|$ не станет меньше значения требуемой точности. В качестве вектора, приближенно задающего периодическое решение, берем $p(n)$. Здесь $p(t)$ — решение, которое в начальный момент находится в достижимом состоянии.

5. С помощью оценки скорости сходимости найти целое t_* , после которого векторы решений станут практически неразличимы. Решить систему (3) с произвольным начальным условием на отрезке $[0, t_*]$. В качестве вектора, приближенно задающего периодическое решение, взять $p(t_*)$.

4 Пример

Пусть требования поступают на сервер только при наличии подключения к интернету с интенсивностью $\lambda(t)$. Если есть подключение к интернету, требование обслуживается с интенсивностью $\mu(t)$. В случае обрыва соединения с интернетом требование не обслуживается. При возобновлении соединения обслуживание требования начинается заново. Интенсивность обрыва $\eta(t)$. Интенсивность восстановления соединения $\nu(t)$. Сервер обслуживает только одно требование. Требования обслуживаются по очереди. Очередь считаем конечной.

Граф процесса представлен на рис. 1.

Состояния $2k$ соответствуют ситуации нахождения в системе k требований и наличия соединения с интернетом. Состояния $2k + 1$ соответствуют ситуации нахождения в системе k требований и отсутствия соединения с интернетом.

Выпишем матрицу $A(t)$ для данного процесса:

$$\begin{pmatrix} -(\lambda + \eta) & \nu & \mu & 0 & 0 & \dots \\ \eta & -\nu & 0 & 0 & 0 & \dots \\ \lambda & 0 & -(\lambda + \mu + \eta) & \nu & \mu & \dots \\ 0 & 0 & \eta & -\nu & 0 & \dots \\ 0 & 0 & \lambda & 0 & -(\lambda + \mu + \eta) & \dots \\ 0 & 0 & 0 & 0 & \eta & \dots \\ 0 & 0 & 0 & 0 & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Рассмотрим вначале для наглядности простейшую ситуацию, считая $N = 5$.

Пусть интенсивности $\lambda(t) = 2 + \cos(2\pi t)$; $\mu(t) = 3 + \sin(2\pi t)$; $\eta(t) = 1 + \cos(2\pi t)$; $\nu(t) = 2 + \sin(2\pi t)$.

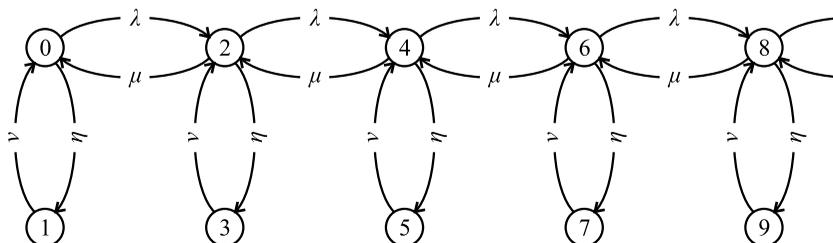


Рис. 1 Граф процесса

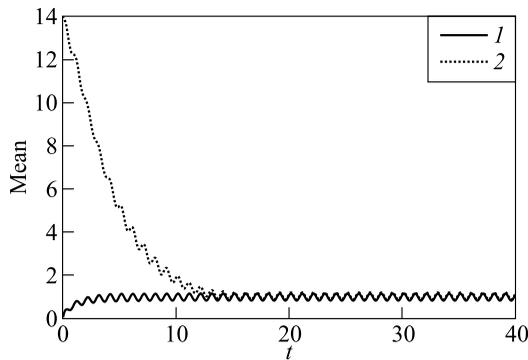


Рис. 2 Математическое ожидание $E(t, k)$ при $t \in [0, 40]$ с начальными условиями $X(0) = 0$ (1) и 29 (2)

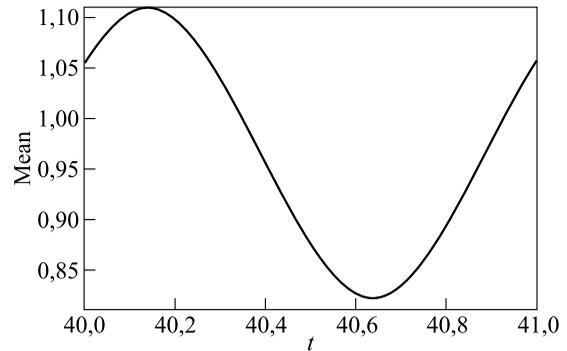


Рис. 3 Аппроксимация предельного математического ожидания $E(t, 0)$ при $t \in [40, 41]$

Вычислим матрицы Коши за период для систем (3) и (4).

$$V(1, 0) = \begin{pmatrix} 0,290760 & 0,297274 & 0,251463 & 0,214199 \\ 0,221605 & 0,317077 & 0,155889 & 0,115851 \\ 0,202991 & 0,178127 & 0,210925 & 0,212677 \\ 0,105084 & 0,077259 & 0,141829 & 0,251418 \\ 0,130241 & 0,098805 & 0,161605 & 0,147329 \\ 0,049318 & 0,031459 & 0,078289 & 0,058525 \\ & 0,212764 & 0,151714 \\ & 0,116136 & 0,072951 \\ & 0,206426 & 0,181198 \\ & 0,116579 & 0,089175 \\ & 0,204803 & 0,241294 \\ & 0,143292 & 0,263669 \end{pmatrix},$$

откуда $\alpha \geq 0,1781$.

Далее,

$$U(1, 0) = \begin{pmatrix} 0,09547 & -0,02486 & -0,02783 \\ -0,06572 & 0,00793 & 0,03675 \\ -0,10575 & 0,00969 & 0,14633 \\ -0,10547 & 0,00343 & 0,01149 \\ -0,14865 & -0,02179 & -0,01591 \\ & -0,03144 & -0,01786 \\ & 0,03136 & 0,02897 \\ & 0,01709 & 0,00921 \\ & 0,07456 & 0,09397 \\ & 0,11105 & 0,21435 \end{pmatrix},$$

откуда $\|U(1, 0)\| = \omega \leq 0,51176$.

Отметим, что при этом получается

$$\frac{1 - \alpha}{\alpha} > \frac{2\omega}{1 - \omega},$$

так что в этой ситуации третий способ приводит к цели быстрее, чем четвертый.

С другой стороны, первый и второй способы позволяют только найти начальное условие, соответствующее периодическому решению, но не дают возможности оценить скорость сходимости к нему.

Отметим, что для получения оценки

$$\|p^*(t) - p^{**}(t)\| \leq \varepsilon = 10^{-3}$$

достаточно, в соответствии с пятым способом, взять $t^* \geq 12$.

Отметим еще, что вектор, дающий начальное условие для предельного периодического режима, имеет вид:

$$\zeta(0) = (0,250499, 0,183664, 0,199890, 0,126187, 0,153903, 0,085858)^T.$$

Рассмотрим теперь ситуацию с большей размерностью. Пусть $N = 49$, а интенсивности $\lambda(t) = 2 + \cos(2\pi t)$; $\mu(t) = 4 + \sin(2\pi t)$; $\eta(t) = 1 + \cos(2\pi t)$; $\nu(t) = 12 + 6 \sin(2\pi t)$.

Тогда, находя матрицы Коши систем (3) и (4) за первый единичный период, получаем оценки: $\alpha < 10^{-5}$, а $\omega \leq 1,5570$, так что в этой ситуации выбрать $n = 1$, как это делалось в начале п. 3, не удастся.

При нахождении же матриц Коши $V(10, 0)$ и $U(10, 0)$ удастся получить приемлемые оценки: $\alpha \geq 0,1439$, и $\omega \leq 0,7578$, что позволяет, пользуясь третьим способом, при $t \geq 40$ получить погрешность для среднего меньше 0,02.

Графики соответствующих средних (сходимость с двумя разными начальными условиями и предельное среднее) приведены на рис. 2 и 3.

Литература

1. А.А. Зейфман, А.В. Коротышева, В.Ю. Королев, Я.А. Сатин. Оценки погрешности аппроксимаций

- неоднородных марковских цепей с непрерывным временем // Теория вероятностей и ее применения, 2017. Т. 61. № 3. С. 563–569. doi: 10.4213/tvp5073.
2. Zeifman A., Razumchik R., Satin Y., Kiseleva K., Korotysheva A., Korolev V. Bounds on the rate of convergence for one class of inhomogeneous Markovian queueing models with possible batch arrivals and services // Int. J. Appl. Math. Comp., 2018. Vol. 28. No. 1. P. 66–72.
 3. Zeifman A., Satin Y., Kiseleva K., Korolev V., Panfilova T. On limiting characteristics for a non-stationary two-processor heterogeneous system // Appl. Math. Comput., 2019. Vol. 351. P. 48–65.
 4. Ammar S. I., Zeifman A., Satin Y., Kiseleva K., Korolev V. On limiting characteristics for a non-stationary two-processor heterogeneous system with catastrophes, server failures and repairs // J. Ind. Manag. Optim., 2020. doi: 10.3934/jimo.2020011.
 5. Zeifman A., Satin Y., Kryukova A., Razumchik R., Kiseleva K., Shilova G. On the three methods for bounding the rate of convergence for some continuous-time Markov chains // Int. J. Appl. Math. Comp., 2020. Vol. 30. P. 1–43. <https://arxiv.org/abs/1911.04086>.
 6. Zeifman A., Korolev V., Satin Y. Two approaches to the construction of perturbation bounds for continuous-time Markov chains // Mathematics, 2020. Vol. 8. Iss. 2. Art. No. 253. 25 p.
 7. Zeifman A. I., Satin Y. A., Kiseleva K. M. On obtaining sharp bounds of the rate of convergence for a class of continuous-time Markov chains // Stat. Probabil. Lett., 2020. Vol. 161. Art. ID: 108730.
 8. Markova E., Sinitcina A., Satin Y., Gudkova I., Samouylov K., Zeifman A. Admission control scheme model described as queueing system with unreliable servers under licensed shared access // Mathematics, 2020 (in press).
 9. Di Crescenzo A., Giorno V., Nobile A. G. Constructing transient birth-death processes by means of suitable transformations // Appl. Math. Comput., 2016. Vol. 281. P. 152–171.
 10. Ammar S. I., Alharbi Y. F. Time-dependent analysis for a two-processor heterogeneous system with time-varying arrival and service rates // Appl. Math. Model., 2018. Vol. 54. P. 743–751.
 11. Chakravarthy S. R. A catastrophic queueing model with delayed action // Appl. Math. Model., 2017. Vol. 46. P. 631–649.
 12. Зейфман А. И., Бенинг В. Е., Соколов И. А. Марковские цепи и модели с непрерывным временем. — М.: ЭЛЕКС-КМ, 2008. 168 с.

Поступила в редакцию 16.03.20

ON APPROACHES TO CONSTRUCTING LIMITING REGIMES FOR SOME QUEUEING MODELS

Ya. A. Satin¹, A. I. Zeifman^{1,2,3}, and G. N. Shilova¹

¹Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

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

³Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation

Abstract: The authors consider nonstationary queueing models, the number of customers in which is described by finite Markov chains with periodic intensities. For many classes of such models, the methods of obtaining upper bounds on the rate of convergence to the limiting regime were developed in previous papers of the authors. Using these methods, one can find the main limiting characteristics of the system, study their stability with respect to small perturbations of the arrival and service intensities, and receive information on how current characteristics of the system differ from the limiting characteristics at each moment of time. In the present paper, the authors study a different situation, namely, it is assumed that explicit estimates of the rate of convergence to the limiting regime cannot be obtained. The methods for constructing the limiting regimes of such systems and for obtaining information on the rate of convergence to them are considered. As an example, the authors consider a simple model of a nonstationary system with a rather slow rate of convergence to the limiting regime.

Keywords: queueing system; finite Markov chains; periodic intensities; limiting characteristics; rate of convergence

DOI: 10.14357/19922264200201

Acknowledgments

The results of Sections 3 and 4 were obtained by Ya. A. Satin and A. I. Zeifman supported by the Russian Science Foundation under grant 19-11-00020.

References

1. Zeifman, A. I., A. V. Korotysheva, V. Y. Korolev, and Ya. A. Satin. 2017. Truncation bounds for approximations of inhomogeneous continuous-time Markov chains. *Theor. Probab. Appl.* 61(3):513–520.
2. Zeifman, A., R. Razumchik, Y. Satin, K. Kiseleva, A. Korotysheva, and V. Korolev. 2018. Bounds on the rate of convergence for one class of inhomogeneous Markovian queuing models with possible batch arrivals and services. *Int. J. Appl. Math. Comp.* 28(1):66–72.
3. Zeifman, A., Ya. Satin, K. Kiseleva, V. Korolev, and T. Panfilova. 2019. On limiting characteristics for a non-stationary two-processor heterogeneous system. *Appl. Math. Comput.* 351:48–65.
4. Ammar, S. I., A. Zeifman, Ya. Satin, K. Kiseleva, and V. Korolev. 2020. On limiting characteristics for a non-stationary two-processor heterogeneous system with catastrophes, server failures and repairs. *J. Ind. Manag. Optim.* doi: 10.3934/jimo.2020011.
5. Zeifman, A., Y. Satin, A. Kryukova, R. Razumchik, K. Kiseleva, and G. Shilova. 2020. On the three methods for bounding the rate of convergence for some continuous-time Markov chains. *Int. J. Appl. Math. Comp.* 30:1–43. Available at: <https://arxiv.org/abs/1911.04086> (accessed April 28, 2020).
6. Zeifman, A., V. Korolev, and Y. Satin. 2020. Two approaches to the construction of perturbation bounds for continuous-time Markov chains. *Mathematics* 8(2):253–25 p.
7. Zeifman, A. I., Y. A. Satin, and K. M. Kiseleva. 2020. On obtaining sharp bounds of the rate of convergence for a class of continuous-time Markov chains. *Stat. Probabil. Lett.* 161:108730.
8. Markova, E., A. Sinitcina, Ya. Satin, I. Gudkova, K. Samouylov, and A. Zeifman. 2020 (in press). Admission control scheme model described as queuing system with unreliable servers under licensed shared access. *Mathematics*.
9. Di Crescenzo, A., V. Giorno, and A. G. Nobile. 2016. Constructing transient birth-death processes by means of suitable transformations. *Appl. Math. Comput.* 281:152–171.
10. Ammar, S. I., and Y. F. Alharbi. 2018. Time-dependent analysis for a two-processor heterogeneous system with time-varying arrival and service rates. *Appl. Math. Model.* 54:743–751.
11. Chakravarthy, S. R. 2017. A catastrophic queueing model with delayed action. *Appl. Math. Model.* 46:631–649.
12. Zeifman, A. I., V. E. Bening, and I. A. Sokolov. 2008. *Markovskie tsepi i modeli s nepreryvnym vremenem* [Markov chains and models with continuous time]. Moscow: ELEKS-KM. 168 p.

Received March 16, 2020

Contributors

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

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of Department, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; principal scientist, Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation; a.zeifman@mail.ru

Shilova Galina N. (b. 1961) — Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; shgn@mail.ru

ЧИСЛЕННЫЕ СХЕМЫ ФИЛЬТРАЦИИ МАРКОВСКИХ СКАЧКООБРАЗНЫХ ПРОЦЕССОВ ПО ДИСКРЕТИЗОВАННЫМ НАБЛЮДЕНИЯМ III: СЛУЧАЙ МУЛЬТИПЛИКАТИВНЫХ ШУМОВ*

А. В. Борисов¹

Аннотация: Статья завершает цикл исследований, начатых в работах «Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям I: характеристики точности» и «Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям II: случай аддитивных шумов». На основании представленных ранее теоретических результатов разработан алгоритм численной реализации задачи фильтрации состояний однородных марковских скачкообразных процессов (МСП) по косвенным непрерывным зашумленным наблюдениям, дискретизованным по времени. Класс систем наблюдения ограничен системами с *мультипликативными* винеровскими шумами: аддитивная полезная составляющая в наблюдениях отсутствует, а интенсивность шумов является функцией оцениваемого состояния. Для вычисления интегралов, присутствующих в оценках, использовался составной вариант численной схемы «средних» прямоугольников порядка точности 3 для вычисления одномерных интегралов, а также формула среднего порядка 4 для интегрирования по треугольнику. В итоге были получены численные схемы порядка точности 1 и 2.

Ключевые слова: марковский скачкообразный процесс; оптимальная фильтрация; мультипликативные шумы в наблюдениях; стохастическое дифференциальное уравнение; аналитическая и численная аппроксимация

DOI: 10.14357/19922264200202

1 Введение

Работа завершает цикл статей [1, 2]. Она посвящена созданию и сравнительному анализу алгоритмов численного решения задач оптимальной фильтрации состояний однородных МСП по дискретизованным косвенным наблюдениям с мультипликативными винеровскими шумами. Наличие шумов такого рода означает, что их интенсивность является случайной и зависит от оцениваемого состояния. Теоретические основы решения такой задачи заложены в [3, 4]. Так как оценка оптимальной фильтрации имеет почти на верное неотрицательные компоненты и удовлетворяет условию нормировки, то представленные в данном цикле алгоритмы численного решения задачи фильтрации сохраняют для своих реализаций указанные свойства и названы за это *устойчивыми*. Точность оценок зависит от шага дискретизации наблюдений и *порядка s аналитической аппроксимации* — числа возможных скачков оцениваемого состояния на интервале дискретизации, учитываемого в алгоритме.

Статья организована следующим образом. Раздел 2 содержит постановку задачи оптимальной

фильтрации по дискретизованным наблюдениям, а также сведения из [1, 2], необходимые для ее численного решения.

В разд. 3 для систем с мультипликативными шумами разработаны численные схемы фильтрации, соответствующие аналитическим аппроксимациям порядка $s = 1$ и 2. Для численного интегрирования при вычислении аппроксимации порядка $s = 1$ выбрана составная схема «средних» прямоугольников, а для $s = 2$ — та же схема прямоугольников для вычисления одномерных интегралов, и составной вариант метода средних — для вычисления двумерных интегралов по треугольникам. Полученные в результате численные аппроксимации имеют порядки 1 и 2.

Заключительные замечания представлены в разд. 4.

2 Оценка фильтрации и ее аппроксимации

На триplete с фильтрацией $(\Omega^X \times \Omega^W, \mathcal{F}^X \times \mathcal{F}^W, \mathcal{P}^X \times \mathcal{P}^W, \{\mathcal{F}_t^X \times \mathcal{F}_t^W\}_{t \geq 0})$ рассматривается система наблюдения

* Работа выполнена при частичной поддержке РФФИ (проект 19-07-00187 А).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, aborISOV@frcscs.ru

$$\left. \begin{aligned} X_t &= X_0 + \int_0^t \Lambda^\top X_s ds + \mu_t; \\ Y_r &= \int_{(r-1)h}^{rh} f X_s ds + \int_{(r-1)h}^{rh} \sum_{n=1}^N X_s^n g_n^{1/2} dW_s, \\ r &\in \mathbb{N}, \end{aligned} \right\} (1)$$

где $X_t \triangleq \text{col}(X_t^1, \dots, X_t^N) \in \mathbb{S}^N$ — ненаблюдаемое состояние системы — однородный МСП с конечным множеством состояний $\mathbb{S}^N \triangleq \{e_1, \dots, e_N\}$ (\mathbb{S}^N — множество единичных векторов евклидова пространства \mathbb{R}^N), матрицей интенсивностей переходов Λ и начальным распределением π ; $\mu_t \triangleq \text{col}(\mu_t^1, \dots, \mu_t^N) \in \mathbb{R}^N$ — \mathcal{F}_t^X -согласованный мартингал; $\{Y_r\}_{r \in \mathbb{N}} : Y_r \triangleq \text{col}(Y_r^1, \dots, Y_r^M) \in \mathbb{R}^M$ — последовательность дискретизованных наблюдений, доступных в известные равноотстоящие моменты времени $\{rh\}_{r \in \mathbb{N}}$; $W_t \triangleq \text{col}(W_t^1, \dots, W_t^M) \in \mathbb{R}^M$ является \mathcal{F}_t^W -согласованным стандартным винеровским процессом; f — $(M \times N)$ -мерная матрица; $\{g_n\}_{n=1, \dots, N}$ — симметричные положительно определенные матрицы; процессы X и W независимы.

Задача оптимальной фильтрации состояния X по дискретизованным наблюдениям Y заключается в нахождении условного математического ожидания

$$\hat{X}_r \triangleq \mathbf{E} \{X_{t_r} | \mathcal{O}_r\},$$

где $\mathcal{O}_r \triangleq \sigma\{Y_\ell : 1 \leq \ell \leq r\}$ — σ -алгебра, порожденная наблюдениями, полученными до момента времени rh включительно; $\mathcal{O}_0 \triangleq \{\emptyset, \Omega\}$.

Пусть $N_r^X(\omega)$ — число скачков процесса X , произошедших на отрезке $[(r-1)h, rh]$, $\tau_r \triangleq \int_0^t X_s ds$ — случайный вектор времени пребывания X в различных состояниях на $[(r-1)h, rh]$, а $\rho^{n,j,m}(\cdot)$ — распределение вектора $\tau_r X_{t_r}^j \mathbf{I}_{\{m\}}(N_r^X)$ при условии $X_{t_{r-1}} = e_k$. Это означает, что для любого $\mathcal{G} \in \mathcal{B}(\mathbb{R}^M)$ верно равенство

$$\mathbf{E} \left\{ \mathbf{I}_{\mathcal{G}}(\tau_r) X_{t_r}^j \mathbf{I}_{\{m\}}(N_r^X) | X_{t_{r-1}} = e_n \right\} = \int_{\mathcal{G}} \rho^{n,j,m}(du).$$

Аналитическая аппроксимация \bar{X}_r порядка s определяется рекурсивной схемой

$$\bar{X}_r = (\mathbf{1} \xi_r^\top \bar{X}_{r-1})^{-1} \xi_r^\top \bar{X}_{r-1}, \quad r \geq 1, \quad \bar{X}_0 = \pi, \quad (2)$$

где $\mathbf{1} = \text{row}(1, \dots, 1)$ — вектор-строка подходящей размерности; $\xi_q \triangleq \|\xi^{ij}(Y_q)\|_{i,j=1, \dots, N}$, $q \in \mathbb{N}$ — $(N \times N)$ -мерные случайные матрицы — функции наблюдений Y_q :

$$\xi^{ij}(y) \triangleq \sum_{m=0}^s \int_{\mathcal{D}} \mathcal{N}\left(y, fu, \sum_{p=1}^N u^p g_p\right) \rho^{i,j,m}(du). \quad (3)$$

В (4) $\mathcal{N}(y, m, K) \triangleq (2\pi)^{-M/2} \det^{-1/2} K \times \exp\{- (1/2) \|y - m\|_{K^{-1}}^2\}$ — это M -мерная плотность гауссовского распределения с математическим ожиданием m и невырожденной ковариационной матрицей K ; $\mathcal{D} \triangleq \{u \in \mathbb{R}_+^M : \mathbf{1}u = h\}$.

Из построения оценки \bar{X}_r (2) следует, что она обладает свойством устойчивости.

Интегралы $\xi^{ij}(y)$ приближенно вычисляются в виде сумм:

$$\xi^{ij}(y) \approx \psi^{ij}(y) \triangleq \sum_{\ell=1}^L \mathcal{N}\left(y, fw_\ell, \sum_{p=1}^N w_\ell^p g_p\right) \varrho_\ell^{ij};$$

$$\psi(y) \triangleq \|\psi^{ij}(y)\|_{i,j=1, \dots, N},$$

определяемых набором пар $\{(w_\ell, \varrho_\ell^{ij})\}_{\ell=1, \dots, L}$. Здесь $\varrho_\ell^{ij} \geq 0$, $\ell = 1, \dots, L$, — веса; $\sum_{\ell=1}^L \varrho_\ell^{ij} \leq 1$; $w_\ell \triangleq \text{col}(w_\ell^1, \dots, w_\ell^N) \in \mathcal{D}$ — точки; $\psi_q \triangleq \|\psi^{ij}(Y_q)\|_{i,j=1, \dots, N}$.

По построению ψ_q^{ij} являются положительными случайными величинами, поэтому численная аппроксимация \tilde{X}_r оценки \bar{X}_r

$$\tilde{X}_r \triangleq (\mathbf{1} \psi_r^\top \tilde{X}_{r-1})^{-1} \psi_r^\top \tilde{X}_{r-1}, \quad r \geq 1, \quad \tilde{X}_0 = \pi,$$

также обладает свойством устойчивости.

Если для схемы численного интегрирования выполнено условие

$$\max_{k=1, \dots, N} \sum_{j=1}^N \int_{\mathbb{R}^M} |\psi^{kj}(y) - \xi^{kj}(y)| dy < \delta,$$

то глобальный показатель точности приближения оптимальной оценки \hat{X}_r ее численной аппроксимацией \tilde{X}_r ограничен сверху

$$\begin{aligned} \sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\hat{X}_r - \tilde{X}_r\|_1 \right\} &\leq \\ &\leq 4 \left[1 - \left(1 - \frac{(\bar{\lambda}h)^{s+1}}{(s+1)!} \right)^r \right] + 2r\delta, \end{aligned} \quad (4)$$

где $\Pi \triangleq \{\pi \in \mathbb{R}_+^N : \mathbf{1}\pi = 1\}$ — вероятностный симплекс; $\bar{\lambda} \triangleq \max_k |\lambda_{kk}|$.

Для фиксированного момента времени T с уменьшением шага дискретизации $h \rightarrow 0$ неравенство (4) принимает асимптотический вид:

$$\begin{aligned} \sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\tilde{X}_{T/h} - \hat{X}_{T/h}\|_1 \right\} &\leq \\ &\leq 2T \left(2\bar{\lambda} \frac{(\bar{\lambda}h)^s}{(s+1)!} + \frac{\delta}{h} \right). \end{aligned} \quad (5)$$

Поэтому для эффективного выбора схемы численного интегрирования, чтобы не понизить порядок точности, обеспечиваемый аналитической аппроксимацией s -го порядка, необходимо подбирать δ так, чтобы $\bar{\lambda}\delta/(\bar{\lambda}h)^{s+1} \rightarrow C \geq 0$ при $h \rightarrow 0$.

3 Приближенное решение задачи фильтрации по наблюдениям с мультипликативными шумами

Для построения численных схем фильтрации состояний однородных МСП по наблюдениям с мультипликативными шумами и исследования их точности без ограничения общности рассмотрим случай «чисто мультипликативных шумов». Это означает, что в (1) $f = 0$, т. е. аддитивный полезный сигнал полностью отсутствует, а информация об оцениваемом состоянии скрыта в интенсивности шумов в наблюдениях. Будем также считать, что все условные интенсивности шумов мультипликативных наблюдений $\{g_n\}_{n=1, \overline{N}}$ различны.

Ниже исследуются аппроксимации порядка $s = 1$ и 2 . Для них с помощью обобщенной формулы полной вероятности легко получить вид интегралов (3), используемых в дальнейшем изложении:

$$\int_{\mathcal{D}} \mathcal{N} \left(y, fu, \sum_{p=1}^N u^p g_p \right) \rho^{k,j,0}(du) = \delta_{kj} e^{\lambda_{kk} h} \mathcal{N}(y, 0, hg_k); \quad (6)$$

$$\int_{\mathcal{D}} \mathcal{N} \left(y, fu, \sum_{p=1}^N u^p g_p \right) \rho^{k,j,1}(du) = (1 - \delta_{kj}) \lambda_{kj} e^{\lambda_{jj} h} \int_0^h Q^{kj}(y, u) du;$$

$$Q^{kj}(y, u) \triangleq e^{(\lambda_{kk} - \lambda_{jj})u} \mathcal{N}(y, 0, ug_k + (h - u)g_j);$$

$$\left. \begin{aligned} & \int_{\mathcal{D}} \mathcal{N} \left(y, fu, \sum_{p=1}^N u^p g_p \right) \rho^{k,j,2}(du) = \\ & = \sum_{\substack{i:i \neq k, \\ i \neq j}} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj} h} \int_0^h \int_0^{h-u} R^{kij}(y, u, v) dv du, \quad (7) \\ & R^{kij}(y, u, v) \triangleq e^{(\lambda_{kk} - \lambda_{ii})u + (\lambda_{ii} - \lambda_{jj})v} \times \\ & \times \mathcal{N}(y, 0, ug_k + vg_i + (h - u - v)g_j). \end{aligned} \right\}$$

За исключением (6), остальные интегралы не имеют явного аналитического представления, и для них будут рассмотрены различные варианты численной реализации.

3.1 Порядок $s = 1$, схема «средних» прямоугольников

Рассмотрим аналитическую аппроксимацию \overline{X}_r порядка $s = 1$. В этом случае

$$\xi^{kj}(y) = \delta_{kj} e^{\lambda_{jj} h} \mathcal{N}(y, 0, hg_j) + (1 - \delta_{kj}) \lambda_{kj} e^{\lambda_{jj} h} \int_0^h Q^{kj}(y, u) du,$$

а для ее приближения попробуем использовать схему «средних» прямоугольников без дополнительно-го дробления отрезка $[0, h]$:

$$\psi^{kj}(y) = \delta_{kj} e^{\lambda_{jj} h} \mathcal{N}(y, 0, hg_j) + (1 - \delta_{kj}) \lambda_{kj} h Q^{kj} \left(y, \frac{h}{2} \right).$$

При этом ошибка численного интегрирования [5] определяется формулой:

$$\begin{aligned} \gamma^{kj}(y) &= (1 - \delta_{kj}) \frac{\lambda_{kj} h^3 e^{\lambda_{jj} h}}{24} \frac{\partial^2}{\partial u^2} Q^{kj}(y, u) \Big|_{u=z} = \\ &= (1 - \delta_{kj}) \frac{\lambda_{kj} h^3 e^{\lambda_{jj} h}}{24} Q^{kj}(y, z) [\zeta_0^2(y, z) - \zeta_1(y, z)], \end{aligned}$$

где $z = z(y) \in [0, h]$ — некоторый параметр, зависящий от y ;

$$\begin{aligned} \zeta_0(y, z) &\triangleq \lambda_{kk} - \lambda_{jj} - \\ &- \frac{d}{dz} |zg_k + (h - z)g_j| / (2 |zg_k + (h - z)g_j|) + \\ &+ \frac{1}{2} y^\top [zg_k + (h - z)g_j]^{-1} (g_k - g_j) \times \\ &\times [zg_k + (h - z)g_j]^{-1} y; \end{aligned}$$

$$\begin{aligned} \zeta_1(y, z) &\triangleq \\ &\triangleq \left(|zg_k + (h - z)g_j| \frac{d^2}{dz^2} |zg_k + (h - z)g_j| - \right. \\ &- \left. \left(\frac{d}{dz} |zg_k + (h - z)g_j| \right)^2 \right) / (2 |zg_k + \\ &+ (h - z)g_j|^2) + y^\top [zg_k + (h - z)g_j]^{-1} \times \\ &\times (g_k - g_j) [zg_k + (h - z)g_j]^{-1} \times \\ &\times (g_k - g_j) [zg_k + (h - z)g_j]^{-1} y. \end{aligned}$$

Непосредственно интегрировать абсолютную величину γ^{kj} проблематично, так как $\int_{\mathbb{R}^M} |\gamma^{kj}(y)| dy = \int_{\mathbb{R}^M} |\gamma^{kj}(y, z^{kj}(y))| dy$, а зависимость $z^{kj}(y)$ в общем случае неизвестна. Поэтому предварительно оценим $|\gamma^{kj}|$ сверху.

Свойства системы (1) гарантируют, что существуют такие симметричные матрицы g и G , что $0 < g \leq g_n \leq G$ для всех $n = 1, \overline{N}$. Поэтому выполняется неравенство

$$Q^{kj}(y, u) \leq K_1 \mathcal{N}(y, 0, hg),$$

где

$$K_1 = \exp \left(\max_{\substack{k,j=1,N:k \neq j \\ u \in [0,h]}} (\lambda_{kk} - \lambda_{jj})u \right) \frac{|G|}{|g|}.$$

Из свойства определителей [6] следует, что

$$\begin{aligned} |zg_k + (h-z)g_j| &= |z(g_k - g_j) + hg_j| = \\ &= \sum_{n=0}^N z^n h^{N-n} G_{kjn}, \end{aligned} \quad (8)$$

где G_{kjn} — сумма всех определителей матриц, полученных из матрицы $u(g_k - g_j)$ путем замены n столбцов соответствующими столбцами матрицы hg_j . Отсюда следует, что

$$\frac{d}{dz} |zg_k + (h-z)g_j| = \sum_{n=1}^N n z^{n-1} h^{N-n} G_{kjn}, \quad (9)$$

поэтому верно неравенство:

$$\begin{aligned} \left| \frac{d}{dz} |zg_k + (h-z)g_j| / (2 |zg_k + (h-z)g_j|) \right| &= \\ &= h^{-1} \left| \frac{\sum_{n=1}^N n (u/h)^{n-1} G_{kjn}}{\sum_{m=0}^N (u/h)^m G_{kjm}} \right| \leq \frac{K_2}{h}, \end{aligned}$$

где

$$K_2 = \max_{\substack{k,j=1,N: \\ k \neq j}} \frac{\sum_{n=1}^N n |G_{kjn}|}{2 \min_{w \in [0,1]} \left| \sum_{m=0}^N G_{kjm} w^m \right|}.$$

Таким образом, для функции $|\zeta_0(y, u)|$ верна следующая оценка сверху:

$$|\zeta_0(y, u)| \leq K_3 + \frac{K_2}{h} + \frac{K_4}{h^2} \|y\|_I^2,$$

где $K_3 = \max_{\substack{k,j=1,N: \\ k \neq j}} |\lambda_{kk} - \lambda_{jj}|$, $K_4 = \|g^{-1}Gg^{-1}\|_2^2$ — квадрат спектральной нормы матрицы. Из этого следует оценка сверху для квадрата $\zeta_0(y, u)$:

$$\begin{aligned} \zeta_0^2(y, u) &= K_3^2 + \frac{K_2^2}{h^2} + \frac{K_4^2}{h^4} \|y\|_I^4 + \\ &+ 2 \frac{K_2 K_3}{h} + 2 \frac{K_3 K_4}{h^2} \|y\|_I^2 + 2 \frac{K_2 K_4}{h^3} \|y\|_I^2. \end{aligned}$$

Используя формулы (8) и (9), можно получить оценку сверху для абсолютного значения первого слагаемого в $\zeta_1(y, u)$:

$$\begin{aligned} &\left| \left(|zg_k + (h-z)g_j| \frac{d^2}{dz^2} |zg_k + (h-z)g_j| - \right. \right. \\ &\quad \left. \left. - \left(\frac{d}{dz} |zg_k + (h-z)g_j| \right)^2 \right) (2 |zg_k + \right. \\ &\quad \left. + (h-z)g_j|^2)^{-1} \right| = \left| \left(\sum_{n=0}^N z^n h^{N-n} G_{kjn} \times \right. \right. \\ &\quad \left. \left. \times \sum_{m=2}^N m(m-1) z^{m-2} h^{N-m} G_{kjm} - \right. \right. \\ &\quad \left. \left. - \left(\sum_{\ell=1}^N \ell z^{\ell-1} h^{N-\ell} G_{kj\ell} \right)^2 \right) \times \right. \\ &\quad \left. \times \left(2 \left(\sum_{s=0}^N z^s h^{N-s} G_{kjs} \right)^2 \right)^{-1} \right| = \frac{1}{h^2} \times \\ &\quad \times \left| \left(\sum_{n=0}^N \left(\frac{z}{h} \right)^n G_{kjn} \sum_{m=2}^N m(m-1) \left(\frac{z}{h} \right)^{m-2} G_{kjm} - \right. \right. \\ &\quad \left. \left. - \left(\sum_{\ell=1}^N \ell \left(\frac{z}{h} \right)^{\ell-1} G_{kj\ell} \right)^2 \right) \times \right. \\ &\quad \left. \times \left(2 \left(\sum_{s=0}^N \left(z/h \right)^s G_{kjs} \right)^2 \right)^{-1} \right| \leq \frac{K_5}{h^2}, \end{aligned}$$

где

$$\begin{aligned} K_5 &= \max_{k,j=1,N:k \neq j} \left(\left(\sum_{n=0}^N |G_{kjn}| \sum_{m=2}^N m(m-1) |G_{kjm}| + \right. \right. \\ &\quad \left. \left. + \left(\sum_{\ell=1}^N \ell |G_{kj\ell}| \right)^2 \right) / \left(2 \min_{w \in [0,1]} \left(\sum_{s=0}^N w^s |G_{kjs}| \right)^2 \right) \right). \end{aligned}$$

Абсолютное значение второго слагаемого в $\zeta_1(y, u)$ также оценивается сверху:

$$\begin{aligned} &y^\top [zg_k + (h-z)g_j]^{-1} (g_k - g_j) \times \\ &\quad \times [zg_k + (h-z)g_j]^{-1} (g_k - g_j) [zg_k + \\ &\quad + (h-z)g_j]^{-1} y \leq \frac{K_6}{h^3} \|y\|_I^2, \end{aligned}$$

где $K_6 = 4 \|g^{-1}Gg^{-1}Gg^{-1}\|_2^2$.

Обозначим

$$K_0 \triangleq \max_{\substack{k,j=1,N: \\ k \neq j}} \frac{\lambda_{kj} e^{\lambda_{jj} h}}{24}$$

и оценим сверху $\int_{\mathbb{R}^M} |\gamma^{kj}(y)| dy$, используя связь моментов 2-го и 4-го порядка гауссовского распределения:

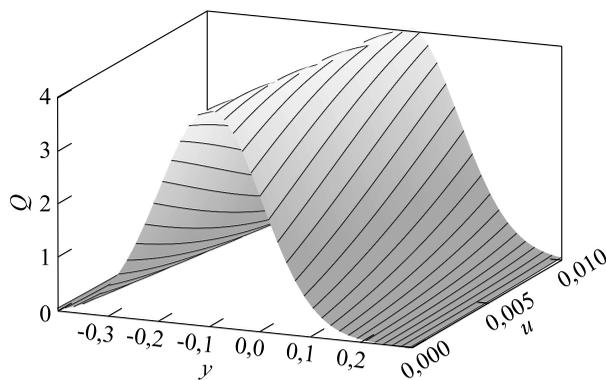


Рис. 1 Общий вид функции $\Xi(y, u)$ в случае аддитивных шумов в наблюдениях

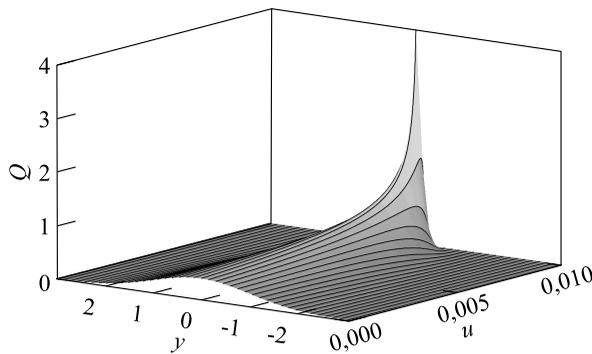


Рис. 2 Общий вид функции $Q(y, u)$ в случае мультипликативных шумов в наблюдениях

$$\begin{aligned} & \int_{\mathbb{R}^M} |\gamma^{kj}(y)| dy \leq \\ & \leq h^3 K_0 K_1 \left[K_3^2 + \frac{2K_2 K_3}{h} + \frac{K_2^2 + K_5}{h^2} \right] + h^3 K_0 K_1 \times \\ & \times \int_{\mathbb{R}^M} \left[\frac{2K_3 K_4}{h^2} + \frac{2K_2 K_4 + K_6}{h^3} \right] \|y\|_T^2 \mathcal{N}(y, 0, hG) dy + \\ & + h^3 K_0 K_1 \frac{K_4^2}{h^4} \int_{\mathbb{R}^M} \|y\|_T^4 \mathcal{N}(y, 0, hG) dy = \\ & = K_7 h + K_8 h^2 + K_9 h^3 \end{aligned}$$

для некоторых положительных констант K_7 , K_8 и K_9 . Это значит, что

$$\int_{\mathbb{R}^M} |\gamma^{kj}(y)| dy = O(h)$$

и согласно (5)

$$\sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\tilde{X}_{T/h} - \hat{X}_{T/h}\|_1 \right\} = O(h^0)$$

и точности несоставного метода «средних» прямоугольников недостаточно для построения численного алгоритма фильтрации любой степени. Какая-либо замена этой схемы на другую несоставную (например, на схему Симпсона, квадратуры Гаусса и пр.) к улучшению не приведет. Причиной этому является связь между порядком производной и степенью h в оценке ошибки интеграла по остатку ряда Тейлора. Дело в том, что функция $Q_{kj}(y, u)$ при малых значениях параметра h и некоторых фиксированных значениях y быстро меняется по аргументу u . Примечательно, что в случае наблюдений с аддитивными шумами [2] аналогичная функция

$$\Xi_{kj}(y, u) = e^{(\lambda_{kk} - \lambda_{jj})u} \mathcal{N}(y, u f^k + (h - u) f^j, hg)$$

такими отрицательными свойствами не обладает. В последней формуле f^k и f^j означают k -й и j -й столбцы матрицы f .

Для иллюстрации данного факта рассмотрим два примера наблюдений одинакового масштаба:

1. Наблюдения с аддитивными шумами:

$$N = 2; M = 1; f = [1 \ 10]; \lambda_{11} = \lambda_{22} = -1; g_1 = g_2 = 1.$$

2. Наблюдения с мультипликативными шумами:

$$N = 2; M = 1; f = [0 \ 0]; \lambda_{11} = \lambda_{22} = -1; g_1 = 1; g_2 = 100.$$

Общие трехмерные графики функций $\Xi(y, u)$ и $Q(y, u)$ представлены на рис. 1 и 3, а отдельные их сечения — на рис. 3 и 4 соответственно. На рис. 4 видно, что в окрестности $y = 0$ функция $Q(y, u)$ меняется быстро и любая простая (несоставная) схема не обеспечит необходимой точности интегрирования.

Используем для приближенного вычисления ξ^{kj} составную схему «средних» прямоугольников, разбив отрезок интегрирования $[0, h]$ с шагом h^2 . В этом случае

$$\gamma^{kj}(y) = (1 - \delta_{kj}) \frac{\lambda_{kj} h^5 e^{\lambda_{jj} h}}{24} \frac{\partial^2 Q_{kj}(y, u)}{\partial u^2} \Big|_{u=z}.$$

Повторяя все выводы этого подраздела для составной схемы «средних» прямоугольников, можно проверить, что она обеспечивает порядок точности 3:

$$\int_{\mathbb{R}^M} |\gamma^{kj}(y)| dy = O(h^3)$$

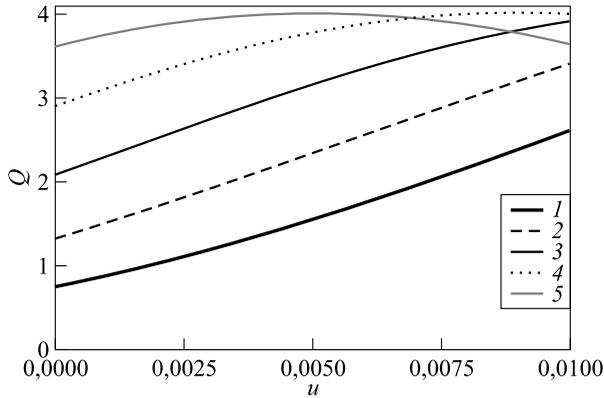


Рис. 3 Сечения $\Omega(y, u)$ для некоторых фиксированных y в случае аддитивных шумов в наблюдениях: 1 — $y = -0,083$; 2 — $-0,0485$; 3 — $-0,014$; 4 — $-0,0205$; 6 — $y = 0,055$

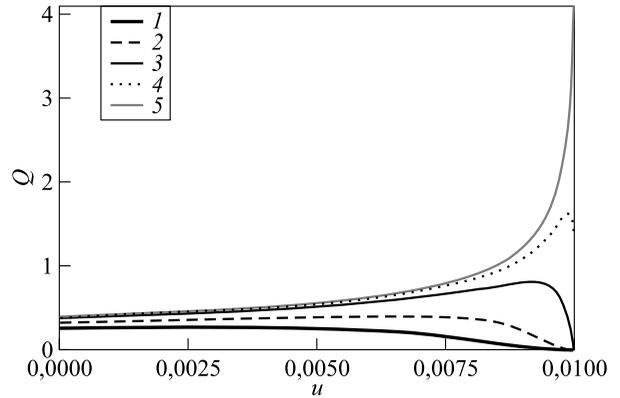


Рис. 4 Сечения $Q(y, u)$ для некоторых фиксированных y в случае мультипликативных шумов в наблюдениях: 1 — $y = -0,9$; 2 — $-0,6$; 3 — $-0,3$; 4 — $-0,15$; 5 — $y = 0$

и согласно (5) глобальный показатель точности имеет порядок 1:

$$\sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\tilde{X}_{T/h} - \hat{X}_{T/h}\|_1 \right\} \leq CTh.$$

Окончательно схема вычисления ψ^{kj} при $s = 1$ примет вид:

$$\begin{aligned} \psi^{kj}(y) &= \delta_{kj} e^{\lambda_{jj}h} \mathcal{N}(y, 0, hg_j) + \\ &+ (1 - \delta_{kj}) \lambda_{kj} h^2 \sum_{i=1}^{[1/h]-1} Q^{kj} \left(y, h^2 \left(i - \frac{1}{2} \right) \right). \end{aligned}$$

3.2 Порядок $s = 2$, схема средних

В случае $s = 2$ элементы ξ^{jk} вычисляются по формуле:

$$\begin{aligned} \xi^{kj}(y) &= \delta_{kj} e^{\lambda_{jj}h} \mathcal{N}(y, 0, hg_j) + \\ &+ (1 - \delta_{kj}) \lambda_{kj} e^{\lambda_{jj}h} \int_0^h Q^{kj}(y, u) du + \\ &+ \sum_{\substack{i:i \neq k, \\ i \neq j}} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} R^{kij}(y, u, v) dv du. \end{aligned} \quad (10)$$

Согласно (5) для сохранения второго порядка точности аналитической аппроксимацией необходимо, чтобы локальная ошибка численного интегрирования на каждом шаге была не более $O(h^3)$. Составная схема «средних» прямоугольников, предложенная в предыдущем подразделе, обеспечивает эту точность для вычисления одномерного интеграла — второго слагаемого в (10).

Выберем подходящую схему вычисления двойных интегралов по треугольнику, входящих в третье слагаемое (10). Прежде всего, определим величину ошибки приближения интеграла в (7) простым методом средних:

$$\begin{aligned} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} R^{kij}(y, u, v) dv du &= \\ &= \frac{h^2}{2} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} R^{kij} \left(y, \frac{h}{3}, \frac{h}{3} \right) + \\ &+ \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} \chi_2^{kij}(y, u, v) dv du, \end{aligned}$$

где функция $\chi_2^{kij}(y, u, v)$ имеет вид:

$$\begin{aligned} \chi_2^{kij}(y, u, v) &\triangleq \frac{1}{2} \left(\left(z - \frac{h}{3} \right) \frac{\partial}{\partial z} + \right. \\ &\left. + \left(w - \frac{h}{3} \right) \frac{\partial}{\partial w} \right)^2 R^{kij}(y, z, w) \Big|_{(z(y,u), w(y,v))}. \end{aligned}$$

Согласно [5], для некоторой положительной константы C_1 верно неравенство:

$$\begin{aligned} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} \chi_2^{kij}(y, u, v) dv du &\leq \\ &\leq h^4 C_1 \max_{\substack{\ell=0,1,2; \\ (z,w) \in \mathcal{D}}} \left| \frac{\partial^2}{\partial z^\ell \partial w^{2-\ell}} \chi_2^{kij}(y, z, w) \right|. \end{aligned}$$

В предыдущем разделе также оценивалась вторая производная, однако от другой функции — Q^{kj} . Она содержала h^2 в знаменателе. Сравнивая Q^{kj}

и R^{kij} , можно заключить, что вторая производная от R^{kij} также будет содержать h^2 в знаменателе, т. е.

$$\lambda_{ki}\lambda_{ij}e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} \chi_2^{kij}(y, u, v) dvdu \leq h^2 C_2$$

для некоторой положительной константы C_2 . Так как требуемый порядок точности — третий, последнее равенство позволяет сделать вывод о том, что простой метод средних в данном случае нужной точности не обеспечивает.

Используем для вычисления двойного интеграла составной метод средних, разбив область интегрирования, прямоугольный треугольник с катетами длины h , на подобные треугольники с катетами h^2 . В этом случае

$$\lambda_{ki}\lambda_{ij}e^{\lambda_{jj}h} \int_0^h \int_0^{h-u} \chi_2^{kij}(y, u, v) dvdu \leq h^4 C_3$$

для некоторой положительной константы C_3 . Тогда согласно (5)

$$\sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\tilde{X}_{T/h} - \hat{X}_{T/h}\|_1 \right\} \leq CT h^2,$$

т. е. для реализации аналитической аппроксимации порядка $s = 2$ правомерно использование составных схем средних при вычислении одномерных и двойных интегралов.

Окончательно схема вычисления ψ^{kj} при $s = 2$ примет вид:

$$\begin{aligned} \psi^{kj}(y) &= \delta_{kj} e^{\lambda_{jj}h} \mathcal{N}(y, 0, hg_j) + \\ &+ (1 - \delta_{kj}) \lambda_{kj} h^2 \sum_{i=1}^{[1/h]-1} Q^{kj} \left(y, h^2 \left(i - \frac{1}{2} \right) \right) + \\ &+ \frac{h^4}{2} \sum_{\substack{i:i \neq k, \\ i \neq j}} \lambda_{ki} \lambda_{ij} e^{\lambda_{jj}h} \times \\ &\times \sum_{n=1}^{[1/h]-1} \sum_{m=1}^{[1/h]-1-n} R^{kij} \left(y, h^2 \left(n - \frac{2}{3} \right), h^2 \left(m - \frac{2}{3} \right) \right). \end{aligned}$$

4 Заключение

Данный цикл работ посвящен разработке алгоритмов численного решения задач фильтрации МСП по косвенным наблюдениям в присутствии винеровских шумов. Перед обработкой непрерывные наблюдения дискретизируются по времени. Таким образом, полученные результаты могут применяться для практических задач оценивания как

по исходным непрерывным наблюдениям, так и по заранее дискретизованным данным. Формула (5) при этом представляется ключевой. Она описывает влияние «входных» характеристик на итоговую точность численного алгоритма фильтрации. К «входным» относятся: характеристика темпа изменения состояний МСП (величина $\bar{\lambda}$), шаг дискретизации по времени h , порядок реализуемой аналитической аппроксимации s и точность используемых схем численного интегрирования δ . Основываясь на этих данных, (5) позволяет оценить величину $\sup_{\pi \in \Pi} \mathbf{E} \left\{ \|\tilde{X}_{T/h} - \hat{X}_{T/h}\|_1 \right\}$ — потерю точности при переходе от оптимальной оценки к некоторой ее численной аппроксимации. Данная формула позволяет решать ряд практических задач системного анализа. Во-первых, по фиксированным $\bar{\lambda}$ и h можно выбирать схему численной реализации оценки с учетом требуемой точности и ресурсоемкости различных методов численного интегрирования. Во-вторых, при проектировании реальных систем наблюдения можно подбирать максимальный шаг дискретизации h , обеспечивающий для выбранной схемы численной реализации требуемую точность.

Важно отметить еще одну особенность полученных результатов. В отличие от классической \mathcal{L}_2 -метрики, используемой для характеристики ошибок численного решения стохастических дифференциальных уравнений [7], в предложенном цикле для этого используется \mathcal{L}_1 -метрика. Дело в том, что решением задачи оптимальной фильтрации является условное распределение состояния МСП по имеющимся наблюдениям. Приближенные решения, предлагаемые в данной работе, обладают свойством устойчивости, т. е. также могут рассматриваться как некоторые распределения. Очевидно, что \mathcal{L}_1 -метрика оказывается более естественной характеристикой, описывающей расстояние между распределениями.

В данных статьях оценено расстояние между оптимальной оценкой фильтрации по дискретизованным наблюдениям и ее численной аппроксимацией, выбранной из предложенного класса. Однако итоговая точность предложенных численных аппроксимаций, т. е. расстояние между ними и истинным текущим состоянием оцениваемого МСП, пока не охарактеризовано. Решение данной задачи представляется одним из направлений дальнейших исследований.

Литература

1. Борисов А. Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным на-

- блюдениям I: характеристики точности // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 68–75. doi: 10.14357/19922264190411.
2. Борисов А. Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям II: случай аддитивных шумов // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 17–23. doi: 10.14357/19922264200103.
 3. Борисов А. Фильтрация Вонэма по наблюдениям с мультипликативными шумами // Автоматика и телемеханика, 2018. № 1. С. 52–65.
 4. Борисов А. Фильтрация состояний марковских скачкообразных процессов по дискретизованным наблюдениям // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 115–121. doi: 10.14357/19922264180316.
 5. Isaacson E., Keller H. Analysis of numerical methods. — New York, NY, USA: Dover Publications, 1994. 541 p.
 6. Magnus J., Neudecker H. Matrix differential calculus with applications in statistics and econometrics. — New York, NY, USA: Wiley, 2019. 504 p.
 7. Kloeden P., Platen E. Numerical solution of stochastic differential equations. — Berlin: Springer, 1992. 636 p.

Поступила в редакцию 11.10.19

NUMERICAL SCHEMES OF MARKOV JUMP PROCESS FILTERING GIVEN DISCRETIZED OBSERVATIONS III: MULTIPLICATIVE NOISES CASE

A. V. Borisov

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

Abstract: The paper presents the final part of investigations initialized in the papers Borisov, A. 2019. Numerical schemes of Markov jump process filtering given discretized observations I: Accuracy characteristics. *Inform. Appl.* 13(4):68–75 and Borisov, A. 2020. Numerical schemes of Markov jump process filtering given discretized observations II: Multiplicative noises case. *Inform. Appl.* 14(1):17–23. Relying on the theoretical results, this paper presents a numerical algorithm of the state filtering of homogeneous Markov jump processes (MJP) given indirect noisy continuous time observations discretized by time. The class of observation systems under consideration is restricted by ones with multiplicative noises: any additive payload component is absent in the observable signal, but the observation noise intensity is a function of the MJP state under estimation. To calculate the integrals in the estimate, the author uses the composite midpoint rule of the precision order 3, along with the composite midpoint rule for triangles of the precision order 4. The constructed numerical algorithms of filtering have the final precision of the orders 1 and 2.

Keywords: Markov jump process; optimal filtering; additive and multiplicative observation noises; stochastic differential equation; analytical and numerical approximation

DOI: 10.14357/19922264200202

Acknowledgments

The work was supported in part by the Russian Foundation for Basic Research (project No. 19-07-00187 A).

References

1. Borisov, A. 2019. Chislennye skhemy fil'tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya I: kharakteristiki tochnosti [Numerical schemes of Markov jump process filtering given discretized observations I: Accuracy characteristics]. *Informatika i ee Primeneniya* — *Inform. Appl.* 13(4):68–75. doi: 10.14357/19922264190411.
2. Borisov, A. 2020. Chislennye skhemy fil'tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya II: sluchay additivnykh shuchmov [Numerical schemes of Markov jump process filtering given discretized observations II: Additive noises case]. *Informatika i ee primeneniya* — *Inform. Appl.* 14(1):17–23. doi: 10.14357/19922264200103.
3. Borisov, A. 2018. Wonham filtering by observations with multiplicative noises. *Automat. Rem. Contr.* 79(1):39–50. doi: 10.1134/S0005117918010046.
4. Borisov, A. 2018. Fil'tratsiya sostoyaniy markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya II: sluchay additivnykh shuchmov [Numerical schemes of Markov jump process filtering given discretized observations II: Additive noises case]. *Informatika i ee primeneniya* — *Inform. Appl.* 14(1):17–23. doi: 10.14357/19922264200103.

- deniyam [Filtering of Markov jump processes by discretized observations]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):115–121. doi: 10.14357/19922264180316.
5. Isaacson, E., and H. Keller. 1994. *Analysis of numerical methods*. New York, NY: Dover Publications. 541 p.
6. Magnus, J., and H. Neudecker. 2019. *Matrix differential calculus with applications in statistics and econometrics*. New York, NY: Wiley. 504 p.
7. Kloeden, P., and E. Platen. 1992. *Numerical solution of stochastic differential equations*. Berlin: Springer. 636 p.

Received October 11, 2019

Contributor

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

УПРАВЛЕНИЕ ВЫХОДОМ СТОХАСТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ СИСТЕМЫ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ.

V. СЛУЧАЙ НЕПОЛНОЙ ИНФОРМАЦИИ О СОСТОЯНИИ*

А. В. Босов¹

Аннотация: Рассматривается обобщение задачи оптимального управления для диффузионного процесса Ито и линейного управляемого выхода с квадратичным критерием качества на случай косвенных наблюдений за состоянием. Имеющееся решение задачи с полной информацией используется для синтеза управления по косвенным наблюдениям на основании принципа разделения. Возможность разделения задач управления и фильтрации состояния обосновывается свойствами используемого квадратичного критерия. Вместо решения возникающей вспомогательной задачи оптимальной фильтрации, описываемого общими уравнениями нелинейной фильтрации на основе обновляющих процессов, предлагается использовать оценку условно-оптимального фильтра (УОФ) В.С. Пугачева. Таким образом, субоптимальное решение рассматриваемой задачи управления получается как результат традиционного подхода к синтезу управления в задаче с неполной информацией, состоящего в формальной замене в решении соответствующей задачи с полной информацией переменной состояния на ее оценку. Окончательно предлагается вариант численной реализации предложенного субоптимального управления на основе метода имитационного компьютерного моделирования, использующий общий пучок моделируемых траекторий как для расчета параметров УОФ, так и для вычисления параметров в исходной задаче управления.

Ключевые слова: стохастическое дифференциальное уравнение; стохастическая дифференциальная система; оптимальное управление; стохастическая фильтрация; условно-оптимальная фильтрация; имитационное компьютерное моделирование

DOI: 10.14357/19922264200203

1 Введение

Задача управления линейным выходом стохастической дифференциальной системы по квадратичному критерию качества, постановка и оптимальное решение, анализ свойств оптимального решения и основанные на нем варианты численной реализации, а также выполненные модельные расчеты представлены в работах [1–4]. В постановке (здесь и далее для простоты рассматриваются скалярные величины) используется состояние y_t стохастической дифференциальной системы, описываемое нелинейным стохастическим дифференциальным уравнением Ито:

$$dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t, \quad y_0 = Y, \quad (1)$$

где v_t — стандартный винеровский процесс; Y — случайная величина с конечным вторым моментом; функции A_t и Σ_t удовлетворяют условиям Ито, обеспечивающим существование единственного решения [5].

С состоянием y_t линейно связан выход z_t :

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z, \quad (2)$$

где w_t — не зависящий от v_t , Y и Z стандартный винеровский процесс; Z — случайная величина с конечным вторым моментом; u_t — допустимое управление. Функции a_t , b_t , c_t и σ_t предполагаются ограниченными, процесс управления — допустимым неупреждающим [5], что обеспечивает существование решения уравнения (2) для любого допустимого управления.

Используется квадратичный целевой функционал следующего вида:

$$J(U_0^T) = \mathbb{E} \left\{ \int_0^T (S_t(s_t y_t - g_t z_t - h_t u_t)^2 + G_t z_t^2 + H_t u_t^2) dt + S_T(s_T y_T - g_T z_T)^2 + G_T z_T^2 \right\},$$

$$U_0^T = \{u_t, 0 \leq t \leq T\}, \quad (3)$$

* Работа выполнена при частичной поддержке РФФИ (проект 19-07-00187-А).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ABosov@frccsc.ru

где S_t , G_t и H_t — неотрицательные ограниченные функции.

Задача была сформулирована в предположении наличия полной информации о состоянии y_t и выходе z_t (соответствующая σ -алгебра, порожденная величинами $y_\tau, z_\tau, 0 \leq \tau \leq t$, обозначается $\mathcal{F}_t^{y,z}$), поэтому допустимое управление искалось в классе $\mathcal{F}_t^{y,z}$ -измеримых неупреждающих функций и оказалось управлением с обратной связью:

$$u_t^* = u_t^*(y_t, z_t) = -\frac{1}{2} (S_t h_t^2 + H_t)^{-1} \times \\ \times (c_t(2\alpha_t z_t + \beta_t(y_t)) + 2S_t(s_t y_t - g_t z_t)h_t). \quad (4)$$

Это решение обеспечено представлением функции Беллмана в виде

$$V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$$

и решением уравнений динамического программирования, приведшим к линейными уравнениями в частных производных второго порядка параболического типа относительно коэффициентов $\beta_t(y)$ и $\gamma_t(y)$ и уравнению Риккати для коэффициента α_t . Подходы к численной реализации полученного оптимального решения обеспечили, во-первых, традиционные сеточные методы решения уравнений параболического типа, во-вторых, альтернативное описание этих коэффициентов, базирующееся на уравнении А. Н. Колмогорова [6] или эквивалентной ему интегральной формуле Фейнмана–Каца [7], обосновавшие применение в расчетах имитационного компьютерного моделирования [4].

Цель данной работы, подводящей окончательный теоретический итог исследованию задачи, рассмотреть случай неполной информации о состоянии, т. е. предположить, что полная информация о состоянии y_t из (1) недоступна, и интерпретировать уравнение (2) как уравнение наблюдений. Формально такая постановка требует определения допустимых управлений в классе \mathcal{F}_t^z -измеримых неупреждающих функций (\mathcal{F}_t^z — σ -алгебра, порожденная $z_\tau, 0 \leq \tau \leq t$; также далее используется обозначение $E\{\cdot | \mathcal{F}_t^z\}$ — оператор условного математического ожидания относительно \mathcal{F}_t^z). Но, даже несмотря на отсутствие управления в уравнении состояния (1), линейность наблюдений (2) и квадратичный критерий (3), получить оптимальное решение не удастся. Обсуждение этого положения и вариант субоптимального управления, опирающегося на принцип разделения, обосновываемый в данной задаче свойствами квадратичного функционала качества, приводятся во втором разделе статьи.

Соображения по принципу разделения и формулировка отдельно задачи фильтрации для системы наблюдения (1), (2) с точки зрения практической реализуемости приводят к выбору в качестве метода оценивания УОФ В. С. Пугачева [8]. При этом структура оценки УОФ такова, что позволяет еще предложить вариант численной реализации на основе метода имитационного компьютерного моделирования, включающий в том числе расчет параметров оптимального управления методом, представленным и апробированным в [4]. Этот алгоритм приведен в разд. 4. Итоги исследования подведены в заключении.

2 Управление на основе принципа разделения

Понятие «принципа разделения» интерпретируется в смысле классического результата о разделении задач управления и фильтрации в линейно-гауссовской стохастической системе [9]. Этот результат базируется на двух положениях. Во-первых, это дифференциальная форма оценки фильтра Калмана, условного математического ожидания состояния в линейно-гауссовской системе наблюдения, имеющая тот же вид, что и уравнение состояния, т. е. то же самое линейное уравнение, отличающееся лишь ковариацией шума, роль которого выполняет обновляющий процесс, являющийся винеровским. Во-вторых, это возможность представления квадратичного критерия двумя слагаемыми, отвечающими за качество управления и качество оценивания состояния.

Оба эти положения имеют свою интерпретацию в рассматриваемой задаче, но вначале следует обсудить важный дополнительный аспект разделения, а именно: отсутствие в задаче управления дуального эффекта [10], т. е. влияния закона управления на качество оценивания состояния. С этой целью видоизменим исходную постановку следующим образом. Выполним замену переменных в (2), избавляясь от слагаемых $b_t z_t dt$ и $c_t u_t dt$, приводя уравнение выхода к некоторой канонической форме (традиционной форме записи наблюдений за состоянием). Для этого введем новую переменную

$$\tilde{z}_t = B_t z_t - \int_0^t B_s c_s u_s ds,$$

где обозначено

$$B_t = \exp \left\{ - \int_0^t b_s ds \right\}.$$

С учетом того, что

$$dB_t = -b_t B_t dt,$$

дифференциал введенной переменной принимает вид:

$$d\tilde{z}_t = B_t (dz_t - b_t z_t dt - c_t u_t dt) = \tilde{a}_t y_t dt + \tilde{\sigma}_t dw_t,$$

где обозначено

$$\tilde{a}_t = a_t B_t; \quad \tilde{\sigma}_t = \sigma_t B_t.$$

Заметим, что проведенное преобразование обеспечивает совпадение σ -алгебр \mathcal{F}_t^z и $\mathcal{F}_t^{\tilde{z}}$ для любого допустимого управления u_t . Действительно, $\mathcal{F}_t^z = \mathcal{F}_t^{\tilde{z}}$, поскольку выполненная замена — линейное, невырожденное ($B_t > 0$) преобразование z_t , а u_t по условию \mathcal{F}_t^z -измеримо. При этом полученное уравнение для \tilde{z}_t показывает, что \tilde{z}_t не зависит от u_t , т.е., используя \tilde{z}_t в качестве наблюдения, можно строить оценку состояния, не зависящую от конкретного закона управления. Таким образом, вместо задачи оценивания состояния по наблюдениям z_t , зависящим от реализуемого закона управления, можно рассматривать эквивалентную задачу оценивания y_t по наблюдениям \tilde{z}_t , не зависящим от u_t .

Проведенная замена не означает разделение задач управления и фильтрации в рассматриваемой постановке, но обосновывает отсутствие в ней дуального эффекта, так что, не ограничивая общности, задачу фильтрации можно решать, используя вместо уравнения выхода (2) уравнение

$$d\tilde{z}_t = \tilde{a}_t y_t dt + \tilde{\sigma}_t dw_t, \quad \tilde{z}_0 = Z. \quad (5)$$

Оптимальную оценку состояния системы (1) по наблюдениям (5), т.е. условное математическое ожидание $E\{y_t | \mathcal{F}_t^{\tilde{z}}\}$, обозначим через \hat{y}_t . Наложены выше на систему (1), (2) ограничения представляют, по-видимому, наиболее жесткий вариант условий, обеспечивающих существование этой оценки. Соотношения для нее основаны на принципиальных результатах, полученных для общей задачи нелинейной фильтрации на основе обновляющихся процессов [11]. Применительно к рассматриваемой задаче оптимальную оценку дает следующее соотношение:

$$d\hat{y}_t = \widehat{A}_t dt + \frac{\widehat{\Sigma}_t}{\widehat{\sigma}_t} d\hat{v}_t, \quad \hat{y}_0 = E\{Y\}, \quad (6)$$

где

$$\begin{aligned} \widehat{A}_t &= E\{A_t(y_t) | \mathcal{F}_t^{\tilde{z}}\}; \\ \widehat{\Sigma}_t &= \tilde{a}_t E\{y_t(y_t - \hat{y}_t) | \mathcal{F}_t^{\tilde{z}}\}; \\ d\hat{v}_t &= \frac{1}{\widehat{\sigma}_t} (d\tilde{z}_t - \tilde{a}_t \hat{y}_t dt). \end{aligned}$$

Кроме того, обновляющий процесс \hat{v}_t является стандартным винеровским относительно $\mathcal{F}_t^{\tilde{z}}$.

Если исходить из того, что (6) используется в качестве уравнения состояния, то наблюдения (5) можно задать следующим уравнением:

$$d\tilde{z}_t = \tilde{a}_t \hat{y}_t dt + \tilde{\sigma}_t d\hat{v}_t, \quad \tilde{z}_0 = Z. \quad (7)$$

Это уравнение действительно описывает линейные наблюдения для состояния \hat{y}_t , при том что эти наблюдения совпадают с (5). Таким образом, полученную пару уравнений (6), (7) можно использовать в качестве эквивалентного описания системы наблюдения.

На этом возможности разделения, обеспечиваемые средствами оптимальной фильтрации, в рассматриваемой постановке исчерпываются: хотя и удалось сохранить структуру наблюдений, уравнение оценки оказалось принципиально иным. Уравнение (6) никоим образом не повторяет форму (1), так что готовое решение задачи управления по полной информации применить не получится. Кроме того, решение уравнения для оптимального фильтра само по себе является существенной задачей. Это, строго говоря, не уравнение Ито, так как его параметры \widehat{A}_t и $\widehat{\Sigma}_t$ зависят от всей траектории \tilde{z}_τ , $0 \leq \tau \leq t$, и вычисление их — отдельная задача, требующая привлечения сложных дополнительных уравнений. Таким образом, все, на что позволяет опираться это уравнение, выражается целесообразностью использовать в рассматриваемой задаче управления оценку фильтрации и предполагать, что результат будет тем успешнее, чем ближе эта оценка к эталонной, определяемой (6).

Поддерживает этот вывод и рассмотрение второго базового положения принципа разделения — разделение квадратичного критерия. Учитывая обозначения

$$\hat{y}_t = E\{y_t | \mathcal{F}_t^{\tilde{z}}\} = E\{y_t | \mathcal{F}_t^z\}$$

и формулу полного математического ожидания, используемый функционал (3) можно представить в виде:

$$\begin{aligned} J(U_0^T) &= E \left(\int_0^T (S_t (s_t \hat{y}_t - g_t z_t - h_t u_t)^2 + G_t z_t^2 + \right. \\ &\quad \left. + H_t u_t^2) dt + S_T (s_T \hat{y}_T - g_T z_T)^2 + G_T z_T^2 \right) + \\ &\quad + E \left\{ \int_0^T S_t s_t^2 (y_t - \hat{y}_t)^2 dt + S_T s_T^2 (y_T - \hat{y}_T)^2 \right\}. \end{aligned}$$

В этом представлении квадратичного критерия выделено второе слагаемое, составляющее плату за ошибку оценивания. С учетом сказанного выше в отношении независимости оценки фильтрации \hat{y}_t от закона управления это слагаемое можно убрать из целевого функционала и, таким образом, обосновать использование в оптимизационной постановке целевой функции вида:

$$J(U_0^T) = E \left\{ \int_0^T (S_t(s_t \hat{y}_t - g_t z_t - h_t u_t)^2 + G_t z_t^2 + H_t u_t^2) dt + S_T(s_T \hat{y}_T - g_T z_T)^2 + G_T z_T^2 \right\}.$$

Сказанное дает основание в качестве результата неформального применения принципа разделения в рассматриваемой задаче предложить субоптимальный вариант управления u_t^S вида

$$u_t^S = u_t^*(\hat{y}_t, z_t), \tag{8}$$

где u_t^* определено соотношением (4) — решением исходной задачи оптимизации (1)–(3); \hat{y}_t — оценка состояния y_t по преобразованным наблюдениям (5), не зависящая от закона управления u_t . Верхний индекс S можно интерпретировать как подчеркивание субоптимальности управления (8) или как первую букву от английского separation (разделение). Нетрудно видеть, что это управление было бы оптимальным для задачи с полной информацией с состоянием y_t вида

$$dy_t = A_t(y_t) dt + \frac{\Sigma_t(y_t)}{\sigma_t} dv_t,$$

выходом (2) и целевым функционалом (3). Такое уравнение состояния можно интерпретировать как нулевое приближение решения задачи оптимальной фильтрации.

Итак, получен вариант (8) субоптимального управления в исходной задаче, выполняющий формально принцип разделения и обладающий потенциалом практической реализуемости при условии приемлемого качества собственно управления. В отношении этого варианта управления придется сделать еще одно упрощение. Поскольку шансов реализовать расчет условного математического ожидания $E\{y_t | \mathcal{F}_t^z\}$ практически нет, то вместо оптимального фильтра (6) в расчетах придется использовать другую оценку фильтрации. Качество управления при этом, видимо, упадет, но можно предполагать, что потери будут тем меньше, чем лучше, т. е. ближе к оптимальной оценке (6), будет используемый фильтр.

3 Управление на основе условно-оптимального фильтра

Итак, имеется необходимость отказаться от использования оптимальной оценки — условного математического ожидания $E\{y_t | \mathcal{F}_t^z\}$ — в пользу другого решения задачи фильтрации, не являющегося оптимальным, но обладающего хорошим качеством в решении задачи оценивания состояния y_t по наблюдениям \tilde{z}_τ , $0 \leq \tau \leq t$. Такую оценку будем обозначать через \tilde{y}_t и определять ее методом условно-оптимальной фильтрации В. С. Пугачева [8].

Условно-оптимальная оценка \tilde{y}_t состояния системы (1) по наблюдениям (2) ищется в виде:

$$d\tilde{y}_t = \tilde{\alpha}_t \xi_t(\tilde{y}_t, \tilde{z}_t) dt + \tilde{\beta}_t \zeta_t(\tilde{y}_t, \tilde{z}_t) d\tilde{z}_t + \tilde{\gamma}_t dt, \tag{9}$$

$$\tilde{y}_0 = E\{Y\},$$

где $\tilde{\alpha}_t$, $\tilde{\beta}_t$ и $\tilde{\gamma}_t$ — ограниченные функции времени, выполняющие роль параметров; функции $\xi_t(y, z)$ и $\zeta_t(y, z)$ — заданные структурные функции, выбираемые из эмпирических соображений.

Для обеспечения условий существования решения дифференциального уравнения (9) будем накладывать на структурные функции те же ограничения, что и на функции сноса и диффузии в (1) с учетом ограничений, наложенных на параметры уравнения (2), т. е.

$$|\xi_t(y, z)| + |\zeta_t(y, z)| \leq C(1 + |y| + |z|) \tag{10}$$

для всех $0 \leq t \leq T$;

$$|\xi_t(y_1, z_1) - \xi_t(y_2, z_2)| + |\zeta_t(y_1, z_1) - \zeta_t(y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|) \tag{11}$$

для всех $0 \leq t \leq T$ и $y_1, y_2, z_1, z_2 \in \mathbb{R}^1$.

В рассматриваемой задаче в качестве эмпирических соображений, обосновывающих выбор структуры УОФ, предлагается использовать записанные выше уравнения оптимальной оценки (6), точнее их структуру, а именно: представление правой части (6) в виде $(\hat{A}_t - (\hat{\Sigma}_t/\hat{\sigma}_t^2)\hat{a}_t\hat{y}_t) dt + (\hat{\Sigma}_t/\hat{\sigma}_t^2) d\tilde{z}_t$ обосновывает предложение использовать в качестве структурных функций

$$\xi_t(y, z) = \xi_t(y) = A_t(y) - \frac{\Sigma_t(y)}{\tilde{\sigma}_t^2} \tilde{a}_t y;$$

$$\zeta_t(y, z) = \zeta_t(y) = \frac{\Sigma_t(y)}{\tilde{\sigma}_t^2},$$

т. е. оценку фильтрации в виде:

$$d\tilde{y}_t = \tilde{\alpha}_t \left(A_t(\tilde{y}_t) - \frac{\Sigma_t(\tilde{y}_t)}{\tilde{\sigma}_t^2} \tilde{a}_t \tilde{y}_t \right) dt + \tilde{\beta}_t \frac{\Sigma_t(\tilde{y}_t)}{\tilde{\sigma}_t^2} d\tilde{z}_t + \tilde{\gamma}_t dt, \quad \tilde{y}_0 = E\{Y\}. \quad (10)$$

Здесь, как и для варианта управления (8), величины $A_t(\tilde{y}_t)$ и $\Sigma_t(\tilde{y}_t)$ можно интерпретировать как нулевые приближения для \widehat{A}_t и $\widehat{\Sigma}_t$, которые уточняются согласно методу условно-оптимальной фильтрации коэффициентами $\tilde{\alpha}_t$, $\tilde{\beta}_t$ и $\tilde{\gamma}_t$, обеспечивающими оценке фильтра (10) несмещенность и гарантированное качество, лучшее на некотором классе допустимых фильтров [8].

Соотношения для коэффициентов УОФ приведены в [8] и представляют собой комбинации моментных характеристик состояния y_t , наблюдений \tilde{z}_t и структурных функций ξ_t и ζ_t , т. е. не зависят от реализуемого закона управления u_t . Приемлемый вариант для расчета этих коэффициентов на основе имитационного компьютерного моделирования приведен в следующем разделе.

4 Численная реализация управления u_t^S

Подведем итог рассуждениям предыдущего раздела, перечислив шаги, выполняемые для реализации варианта управления u_t^S для решения исходной задачи оптимизации, включающей уравнение (1) ненаблюдаемого состояния, уравнение (2) выхода, формирующего косвенные наблюдения, и критерия (3), минимизируемого на классе \mathcal{F}_t^z -измеримых неупреждающих управлений.

Шаг 1. Рассматривается вспомогательная задача условно-оптимальной фильтрации для системы наблюдения и фильтра вида

$$\left. \begin{aligned} dy_t &= A_t(y_t) dt + \Sigma_t(y_t) dv_t, \quad y_0 = Y; \\ d\tilde{z}_t &= \tilde{a}_t y_t dt + \tilde{\sigma}_t dw_t, \quad z_0 = Z; \\ d\tilde{y}_t &= \tilde{\alpha}_t \xi_t(\tilde{y}_t) dt + \tilde{\beta}_t \zeta_t(\tilde{y}_t) d\tilde{z}_t + \tilde{\gamma}_t dt, \\ \tilde{y}_0 &= E\{Y\}, \end{aligned} \right\} \quad (11)$$

где

$$\xi(\tilde{y}_t) = A_t(\tilde{y}_t) - \frac{\Sigma_t(\tilde{y}_t)}{\tilde{\sigma}_t^2} \tilde{a}_y \tilde{y}_t; \quad \zeta_t(\tilde{y}_t) = \frac{\Sigma_t(\tilde{y}_t)}{\tilde{\sigma}_t^2}.$$

Решение (11) — коэффициенты $\tilde{\alpha}_t$, $\tilde{\beta}_t$ и $\tilde{\gamma}_t$ УОФ — определяются следующей системой уравнений (это частный случай решения задачи УОФ для системы наблюдения общего вида [8]):

$$\left. \begin{aligned} \tilde{\gamma}_t &= E\{A_t\} - \tilde{\alpha}_t E\{\xi_t\} - \tilde{\beta}_t \tilde{a}_t E\{y_t\}; \\ \tilde{\beta}_t &= \frac{\tilde{a}_t E\{(y_t - \tilde{y}_t)y_t \zeta_t\}}{\tilde{\sigma}_t^2 E\{\zeta_t^2\}}; \\ \tilde{\alpha}_t &= E\left\{ (y_t - \tilde{y}_t) \left(\frac{\partial \xi_t}{\partial t} + \frac{1}{2} \frac{\partial^2 \xi_t}{\partial (\tilde{y}_t)^2} (\tilde{\beta}_t \zeta_t \tilde{\sigma}_t)^2 \right) + (y_t - \tilde{y}_t) \frac{\partial \xi_t}{\partial \tilde{y}_t} E\{A_t\} + \tilde{\beta}_t \tilde{a}_t \overline{\zeta_t y_t} \right\} - \\ &\quad - \frac{\partial \xi_t}{\partial \tilde{y}_t} (\tilde{\beta}_t \zeta_t \tilde{\sigma}_t)^2 + \xi_t (\overline{A_t} - \tilde{\beta}_t \tilde{a}_t \overline{\zeta_t y_t}) \left\} \times \right. \\ &\quad \left. \times \left(E\{\xi_t \bar{\xi}_t\} - (y_t - \tilde{y}_t) \frac{\partial \xi_t}{\partial \tilde{y}_t} \bar{\xi}_t \right)^{-1}, \right\} \quad (12) \end{aligned}$$

где обозначено:

$$\begin{aligned} \overline{A_t} &= A_t(y_t); \quad \overline{A_t} = A_t - E\{A_t\}; \\ \xi_t &= \xi_t(\tilde{y}_t); \quad \bar{\xi}_t = \xi_t - E\{\xi_t\}; \\ \zeta_t &= \zeta_t(\tilde{y}_t); \quad \overline{\zeta_t y_t} = \zeta_t y_t - E\{\zeta_t y_t\}. \end{aligned}$$

В качестве численной реализации оценки \tilde{y}_t предлагается использовать метод имитационного компьютерного моделирования, а именно: разбивая ось времени на интервалы равной малой длины Δt , для вычисления коэффициентов $\tilde{\alpha}_t$, $\tilde{\beta}_t$ и $\tilde{\gamma}_t$ в правой части (12) использовать вычисленные на предыдущем шаге $\tilde{\alpha}_{t-\Delta t}$ и $\tilde{\beta}_{t-\Delta t}$, а моменты $E\{\cdot\}$ заменять на их оценки статистическим средним, для чего моделировать пучок траекторий процессов y_t , \tilde{z}_t и \tilde{y}_t . Заметим, что аналогичный метод, основанный на формуле Фейнмана–Каца, использован в [4] для расчета коэффициентов α_t , $\beta_t(y)$ и $\gamma_t(y)$ в задаче оптимального управления по полной информации и хорошо себя показал. Кроме того, для обоих расчетов (параметров управления и коэффициентов УОФ) можно воспользоваться одним и тем же ресурсом — смоделированным пучком траекторий.

Шаг 2. Рассматривается вспомогательная задача управления с тем же целевым функционалом (3), линейным выходом (2) и состоянием (1), но в предположении наблюдаемости состояния. Решение — коэффициенты α_t и $\beta_t(y)$, формирующие закон управления $u_t^* = u_t^*(y_t, z_t)$ согласно (4), — находится любым из двух численных методов, предложенных в [2, 4].

Шаг 3. Формируется расширенная (эквивалентная) постановка для исходной задачи, включающая уравнение состояния (1), линейный выход (2) и псевдонаблюдения

$$\tilde{z}_t = B_t z_t - \int_0^t B_s c_s u_s^S ds.$$

Предложенный вариант управления u_t^S вычисляется по формуле (8), где используются параметры управления α_t и $\beta_t(y)$, вычисленные на шаге 2, и оценка УОФ \tilde{y}_t по наблюдениям \tilde{z}_t с коэффициентами $\tilde{\alpha}_t$, $\tilde{\beta}_t$ и $\tilde{\gamma}_t$ вычисленными на шаге 1.

5 Заключение

В статье подведен итог исследованию задачи оптимизации линейного выхода нелинейной дифференциальной системы по квадратичному критерию, выполненному в [1–4]. Оптимальные алгоритмы, полученные для рассматриваемой задачи оптимизации в случае полной информации, т. е. при наличии возможности наблюдать и выход системы, и ее состояние, проанализированы в связи с постановкой задачи при неполной информации, а именно: в случае, когда выход выполняет функцию косвенных наблюдений за неизвестным состоянием системы. Возможности получить и в такой постановке оптимальные решения не выявлены, поэтому предложен вариант субоптимального решения, опирающийся на концепцию разделения задач управления и фильтрации и на метод условно-оптимальной фильтрации состояний стохастических дифференциальных систем наблюдения В. С. Пугачева.

Литература

1. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. I. Оптимальное решение методом динамического программирования // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 99–106.
2. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. II. Численное решение уравнений динамического программирования // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 9–15.
3. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. III. Анализ свойств оптимального управления // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 41–49.
4. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. IV. Альтернативное численное решение // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 24–30.
5. Флеминг У., Ришел Р. Оптимальное управление детерминированными и стохастическими системами / Пер. с англ. — М.: Мир, 1978. 316 с. (Fleming W. H., Rishel R. W. Deterministic and stochastic optimal control. — New York, NY, USA: Springer-Verlag, 1975. 222 p.)
6. Гухман И. И., Скороход А. В. Теория случайных процессов. — М.: Наука, 1975. Т. III. 496 с.
7. Oksendal B. Stochastic differential equations. An introduction with applications. — New York, NY, USA: Springer-Verlag, 2003. 379 p.
8. Пугачев В. С., Синицын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с.
9. Davis M. H. A. Linear estimation and stochastic control. — London: Chapman and Hall, 1977. 224 p.
10. Фельдбаум А. А. Основы теории оптимальных автоматических систем. — 2-е изд., испр. и доп. — М.: Наука, 1966. 624 с.
11. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов (нелинейная фильтрация и смежные вопросы). — М.: Наука, 1974. 696 с.

Поступила в редакцию 30.01.20

STOCHASTIC DIFFERENTIAL SYSTEM OUTPUT CONTROL BY THE QUADRATIC CRITERION. V. CASE OF INCOMPLETE STATE INFORMATION

A. V. Bosov

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

Abstract: A generalization of the optimal control problem for the Ito diffusion process and a linear controlled output with a quadratic quality criterion for the case of indirect observation of the state is considered. The available solution of the problem with full information is used to synthesize control from indirect observations based on the principle of separation. The ability to separate control and state filtering tasks is justified by the properties

of the quadratic criterion used. Instead of solving the arising auxiliary problem of optimal filtering described by the general equations of nonlinear filtering based on innovation processes, it is proposed to use the estimate of the conditionally optimal filter of V. S. Pugachev. Thus, the suboptimal solution of the control problem under consideration is obtained as a result of the traditional approach to control synthesis in the problem with incomplete information, consisting in a formal replacement in solving the corresponding problem with complete information of the state variable by its estimate. Finally, a case of numerical implementation of the obtained suboptimal control is proposed. It is based on the method of computer simulation, using a common beam of simulated paths both for calculating the parameters of a conditionally optimal filter and for calculating the parameters in the original control problem.

Keywords: stochastic differential equation; stochastic differential system; optimal control; stochastic filtering; conditionally-optimal filtering; computer simulations

DOI: 10.14357/19922264200203

Acknowledgments

This work was partially supported by the Russian Foundation for Basic Research (grant 19-07-00187-A).

References

1. Bosov, A. V., and A. I. Stefanovich. 2018. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadrachnomu kriteriyu. I. Optimal'noe reshenie metodom dinamicheskogo programmirovaniya [Stochastic differential system output control by the quadratic criterion. I. Dynamic programming optimal solution]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):99–106.
2. Bosov, A. V., and A. I. Stefanovich. 2019. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadrachnomu kriteriyu. II. Chislennoe reshenie uravneniy dinamicheskogo programmirovaniya [Stochastic differential system output control by the quadratic criterion. II. Dynamic programming equations numerical solution]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):9–15.
3. Bosov, A. V., and A. I. Stefanovich. 2019. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadrachnomu kriteriyu. III. Analiz svoystv optimal'nogo upravleniya [Stochastic differential system output control by the quadratic criterion. III. Optimal control properties analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):41–49.
4. Bosov, A. V., and A. I. Stefanovich. 2020. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadrachnomu kriteriyu. IV. Alternativnoe chislennoe reshenie [Stochastic differential system output control by the quadratic criterion. IV. Alternative numerical decision]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):24–30.
5. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p.
6. Gihman, I. I., and A. V. Skorohod. 2012. *The theory of stochastic processes*. New York, NY: Springer-Verlag. Vol. 3. 388 p.
7. Øksendal, B. 2003. *Stochastic differential equations. An introduction with applications*. New York, NY: Springer-Verlag. 379 p.
8. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, NY: J. Wiley & Sons. 549 p.
9. Davis, M. H. A. 1977. *Linear estimation and stochastic control*. London: Chapman and Hall. 224 p.
10. Feldbaum, A. A. 1966. *Osnovy teorii optimal'nykh avtomaticheskikh sistem* [Foundations of theory of optimal automatic systems]. Moscow: Nauka. 624 p.
11. Liptser, R. S., and A. N. Shiryaev. 2001. *Statistics of random processes II. Applications*. Berlin: Springer-Verlag. 402 p.

Received January 30, 2020

Contributor

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

АСИМПТОТИКА ОЦЕНКИ СРЕДНЕКВАДРАТИЧНОГО РИСКА В ЗАДАЧЕ ОБРАЩЕНИЯ ПРЕОБРАЗОВАНИЯ РАДОНА ПО ПРОЕКЦИЯМ, РЕГИСТРИРУЕМЫМ НА СЛУЧАЙНОЙ СЕТКЕ*

О. В. Шестаков¹

Аннотация: При реконструкции томографических изображений необходимо использовать методы регуляризации, так как задача обращения преобразования Радона, лежащего в основе математических моделей большинства томографических экспериментов, относится к некорректно поставленным. Методы регуляризации, основанные на аппарате вейвлет-анализа, приобрели популярность благодаря адаптации к локальным особенностям изображений и вычислительной эффективности. Анализ погрешностей в томографических изображениях представляет собой важную практическую задачу, поскольку дает возможность оценить качество как самих методов, так и используемого оборудования. Иногда нет возможности регистрировать проекционные данные на равномерной сетке отсчетов. Если точки отсчетов по каждой координате образуют вариационный ряд, построенный по выборке из равномерного распределения, то использование обычных процедур пороговой обработки оказывается адекватным. В данной работе проведен анализ оценки среднеквадратичного риска при обращении преобразования Радона и показано, что если функция изображения равномерно регулярна по Липшицу, то данная оценка является сильно состоятельной и асимптотически нормальной.

Ключевые слова: пороговая обработка; преобразование Радона; случайная сетка; оценка среднеквадратичного риска

DOI: 10.14357/19922264200204

1 Введение

Вычислительная томография используется в медицинской диагностике, астрономии, сейсмологии, электронной микроскопии, диагностике плазмы и во многих других областях. Большинство задач вычислительной томографии основано на обращении преобразования Радона. Если в проекционных данных содержится шум, то необходимо применять методы регуляризации, так как задача обращения преобразования Радона некорректно поставлена.

Большой популярностью пользуются методы подавления шума, основанные на вейвлет-разложении и пороговой обработке. Для решения задач вычислительной томографии эти методы рассматривались в работах [1–3], в которых оценивается порядок среднеквадратичного риска и описываются алгоритмы вычисления оптимальных параметров пороговой обработки.

Также изучены свойства оценки среднеквадратичного риска, позволяющей судить о качестве реконструкции на основе только наблюдаемых дан-

ных. Показано, что при определенных условиях она является сильно состоятельной и асимптотически нормальной [4, 5].

В некоторых ситуациях невозможно (или сложно) регистрировать проекционные данные на равномерной сетке отсчетов [6]. В работе [7] показано, что если точки отсчетов образуют вариационный ряд, построенный по выборке из равномерного распределения на отрезке регистрации данных, то при использовании обычной пороговой обработки вейвлет-коэффициентов порядок среднеквадратичного риска остается с точностью до логарифмического множителя равным оптимальному порядку в классе функций, регулярных по Липшицу.

В данной работе рассматривается статистическая оценка среднеквадратичного риска пороговой обработки коэффициентов при обращении преобразования Радона и показывается, что статистические свойства этой оценки также не меняются при переходе от фиксированной равномерной сетки отсчетов к случайной. Оценка сохраняет асимптотическую регулярность, т. е. остается сильно состоятельной и асимптотически нормальной.

* Работа выполнена при финансовой поддержке РФФИ (проект 18-07-00252).

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

2 Обращение преобразования Радона

Математическая модель задачи реконструкции томографических изображений основана на так называемом преобразовании Радона. Обозначим через $\text{Lip}(\gamma)$ класс функций, равномерно регулярных по Липшицу с показателем $\gamma > 0$. Пусть изображение описывается функцией $f(x, y) \in \text{Lip}(\gamma)$ с компактным носителем (без ограничения общности будем считать, что это круг единичного радиуса с центром в начале координат). Преобразованием Радона функции f называется преобразование вида:

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

где

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

Одним из возможных методов обращения преобразования Rf является метод вейвлет–вейглет-разложения [2].

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

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

Семейство $\{\psi_{j, k_1, k_2}^{[\lambda]}\}$ образует ортонормированный базис в $L^2(\mathbb{R}^2)$. Индекс j в (1) называется масштабom, а индексы k_1 и k_2 — сдвигами. В данной работе предполагается, что используются вейвлеты Мейера [8], имеющие M непрерывных производных ($M \geq \gamma$).

Вейвлет-разложение функции f имеет вид:

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

В силу условий на f и вейвлет-функции коэффициенты разложения в (2) удовлетворяют неравенству:

$$\left| \langle f, \psi_{j, k_1, k_2}^{[\lambda]} \rangle \right| \leq \frac{A}{2^{j(\gamma+1)}},$$

где A — некоторая положительная константа [8].

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

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

где I^α — потенциал Рисса, определяемый в пространстве Фурье по формуле $\widehat{I^\alpha g}(\omega) = |\omega|^{-\alpha} \widehat{g}(\omega)$.

Эти функции называются «вейглетами» [2]. Для них справедливо соотношение:

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

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

$$f = \sum_{\lambda, j, k_1, k_2} 2^{j/2} \langle Rf, \xi_{j, k_1, k_2}^{[\lambda]} \rangle \psi_{j, k_1, k_2}^{[\lambda]}. \quad (3)$$

В разложении (3) используются только проекционные данные, и оно служит основой метода реконструкции.

3 Модель данных и метод подавления шума

В практических задачах проекционные данные описываются дискретными отсчетами некоторой функции. Предположим, что отсчеты регистрируются в случайных точках и содержат аддитивный белый гауссов шум, т.е. рассмотрим следующую модель данных:

$$Y_{i, j} = Rf(s_i, \theta_i) + e_{i, j} \quad i = 1, \dots, N, \quad j = 1, \dots, N,$$

где $N = 2^J$ для некоторого $J > 0$; s_1, \dots, s_N — независимая выборка из равномерного распределения на $[-1, 1]$; $\theta_1, \dots, \theta_N$ — независимая выборка из равномерного распределения на $[0, \pi]$ (выборки из s_i и θ_i также независимы); $e_{i, j}$ независимы между собой и от x_i и y_i и имеют нормальное распределение с нулевым средним и дисперсией σ^2 . Пусть $0 \leq s_{(1)} < \dots < s_{(N)} \leq 1$ — вариационный ряд, построенный по выборке $s_i, i = 1, \dots, N$, а $0 \leq \theta_{(1)} < \dots < \theta_{(N)} \leq 1$ — вариационный ряд, построенный по выборке $\theta_i, i = 1, \dots, N$. Тогда, перенумеровав $Y_{i, j}$ и $e_{i, j}$, получаем модель:

$$Y_{i, j} = Rf(s_{(i)}, \theta_{(i)}) + e_{i, j}, \quad i = 1, \dots, N, \quad j = 1, \dots, N. \quad (4)$$

Наряду с (4) рассмотрим выборку с равными расстояниями между отсчетами:

$$Z_{i, j} = Rf\left(-1 + \frac{2i}{N}, \frac{j\pi}{N}\right) + e_{i, j}, \quad i = 1, \dots, N, \quad j = 1, \dots, N. \quad (5)$$

Применяя к (5) дискретный аналог вейглет-преобразования [9], можно перейти к моделям дискретных коэффициентов:

$$W_{j,k_1,k_2}^{[\lambda]} = \mu_{j,k_1,k_2}^{[\lambda]} + e_{j,k_1,k_2}^{[\lambda]}, \quad (6)$$

где

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

$e_{j,k_1,k_2}^{[\lambda]}$ имеют нормальное распределение с нулевым средним и дисперсией σ_λ^2 , но уже не являются независимыми. Значение σ_λ^2 зависит от выбранного вейглет-базиса и λ , но не зависит от k_1, k_2 и j .

Популярным методом подавления шума служит пороговая обработка эмпирических коэффициентов. К коэффициентам в (6) применяется функция жесткой пороговой обработки $\rho_H(x, T) = y\mathbf{1}(|x| > T)$ или мягкой пороговой обработки $\rho_S(x, T) = \text{sgn}(x)(|x| - T)_+$ с порогом T . Смысл пороговой обработки заключается в удалении достаточно малых коэффициентов, которые считаются шумом.

Если применить дискретный аналог вейглет-преобразования к (4), то получится набор эмпирических коэффициентов:

$$V_{j,k_1,k_2}^{[\lambda]} = \nu_{j,k_1,k_2}^{[\lambda]} + e_{j,k_1,k_2}^{[\lambda]}.$$

В общем случае $V_{j,k_1,k_2}^{[\lambda]}$ не равны $W_{j,k_1,k_2}^{[\lambda]}$, а $\nu_{j,k_1,k_2}^{[\lambda]}$ не равны $\mu_{j,k_1,k_2}^{[\lambda]}$. Однако к $V_{j,k_1,k_2}^{[\lambda]}$ можно применить ту же процедуру, что и к коэффициентам $W_{j,k_1,k_2}^{[\lambda]}$, и получить оценки $\widehat{V}_{j,k_1,k_2}^{[\lambda]}$. В следующем разделе обсуждаются свойства полученных оценок.

Среднеквадратичный риск пороговой обработки для выборки со случайными точками отсчетов определим как

$$R_\nu(f) = \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \mathbb{E} \left(\widehat{V}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right)^2.$$

Также определим среднеквадратичный риск для выборки с равными расстояниями между отсчетами:

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

Используя результаты работы [2], можно убедиться, что справедливо следующее утверждение о порядке $R_\mu(f)$.

Теорема 1. При выборе асимптотически оптимального порога для жесткой и мягкой пороговой обработки справедливо соотношение

$$R_\mu(f) \leq C 2^{6J/(2\gamma+3)} J^{(2\gamma+6)/(2\gamma+3)},$$

где C — некоторая положительная константа.

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

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

Также, повторяя рассуждения работы [7], можно показать, что аналогичное утверждение справедливо для $R_\nu(f)$. Таким образом, замена равноотстоящих точек отсчета на случайные не оказывает влияния на оценку порядка среднеквадратичного риска.

4 Свойства оценки среднеквадратичного риска

На практике вычислить значение среднеквадратичного риска нельзя, поскольку оно зависит от ненаблюдаемых «чистых» коэффициентов. Однако можно построить его оценку, используя только наблюдаемые данные. Эта оценка определяется выражением:

$$\widehat{R}_\nu(f) = \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j F \left[V_{j,k_1,k_2}^{[\lambda]} \right], \quad (7)$$

где

$$F[V_{j,k_1,k_2}^{[\lambda]}] = \left(\left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 - \sigma^2 \right) \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| \leq T \right) + \sigma^2 \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| > T \right)$$

в случае жесткой пороговой обработки и

$$F[V_{j,k_1,k_2}^{[\lambda]}] = \left(\left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 - \sigma^2 \right) \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| \leq T \right) + (\sigma^2 + T^2) \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| > T \right)$$

в случае мягкой пороговой обработки [8].

Выражение (7) дает возможность получить представление о погрешности, с которой оценивается функция f . Докажем утверждение о ее асимптотической нормальности.

Теорема 2. Пусть $f \in \text{Lip}(\gamma)$ с $\gamma > 1/2$ и вейглет-функция удовлетворяет перечисленным выше условиям. Пусть $T = T_\gamma$. Тогда при жесткой или мягкой пороговой обработке

$$P \left(\frac{\widehat{R}_\nu(f) - R_\nu(f)}{D_J} < x \right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty.$$

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

Доказательство. Докажем теорему для случая жесткой пороговой обработки. В случае мягкой пороговой обработки доказательство аналогично.

Наряду с $\widehat{R}_\nu(f)$ рассмотрим

$$\widehat{R}_\mu(f) = \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j F \left[W_{j,k_1,k_2}^{[\lambda]} \right]$$

и запишем разность $\widehat{R}_\nu(f) - R_\nu(f)$ в виде:

$$\widehat{R}_\nu(f) - R_\nu(f) = \widehat{R}_\mu(f) - R_\mu(f) + \widetilde{R},$$

где

$$\widetilde{R} = \widehat{R}_\nu(f) - \widehat{R}_\mu(f) - (R_\nu(f) - R_\mu(f)).$$

В [4] показано, что

$$P \left(\frac{\widehat{R}_\mu(f) - R_\mu(f)}{D_J} < x \right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty.$$

Следовательно, для доказательства теоремы достаточно показать, что

$$\frac{\widetilde{R}}{2^{2J}} \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

В силу теоремы 1 и аналогичного утверждения для $R_\nu(f, T)$

$$\frac{R_\nu(f) - R_\mu(f)}{2^{2J}} \rightarrow 0 \text{ при } J \rightarrow \infty.$$

Далее пусть

$$j_0 \approx \frac{2J}{2\gamma + 3} + \frac{1}{2\gamma + 3} \log_2 J.$$

Представим $\widehat{R}_\nu(f, T_\gamma) - \widehat{R}_\mu(f, T_\gamma)$ в виде:

$$\widehat{R}_\nu(f, T_\gamma) - \widehat{R}_\mu(f, T_\gamma) = S_1 + S_2,$$

где

$$S_1 = \sum_{j=0}^{J_0-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(F[V_{j,k_1,k_2}^{[\lambda]}] - F[W_{j,k_1,k_2}^{[\lambda]}] \right);$$

$$S_2 = \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(F[V_{j,k_1,k_2}^{[\lambda]}] - F[W_{j,k_1,k_2}^{[\lambda]}] \right).$$

Поскольку как в случае жесткой, так и в случае мягкой пороговой обработки

$$\left. \begin{aligned} \left| F \left[V_{j,k_1,k_2}^{[\lambda]} \right] \right| &\leq T_\gamma^2 + \sigma_\lambda^2; \\ \left| F \left[W_{j,k_1,k_2}^{[\lambda]} \right] \right| &\leq T_\gamma^2 + \sigma_\lambda^2 \end{aligned} \right\} \text{ п. в., } \quad (8)$$

то

$$\frac{S_1}{2^{2J}} \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

Далее

$$\begin{aligned} S_2 &= \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(F \left[V_{j,k_1,k_2}^{[\lambda]} \right] - \right. \\ &\quad \left. - F \left[W_{j,k_1,k_2}^{[\lambda]} \right] \right) = \\ &= \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[W_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) + \\ &\quad + \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[W_{j,k_1,k_2}^{[\lambda]} \right]^2 - 2\sigma^2 \right) \times \\ &\quad \times \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| \leq T_\gamma, \left| W_{j,k_1,k_2}^{[\lambda]} \right| > T_\gamma \right) + \\ &\quad + \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(2\sigma^2 - \left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) \times \\ &\quad \times \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| > T_\gamma, \left| W_{j,k_1,k_2}^{[\lambda]} \right| \leq T_\gamma \right) + \\ &\quad + \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[W_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) \times \\ &\quad \times \mathbf{1} \left(\left| V_{j,k_1,k_2}^{[\lambda]} \right| > T_\gamma, \left| W_{j,k_1,k_2}^{[\lambda]} \right| > T_\gamma \right). \quad (9) \end{aligned}$$

Рассмотрим сумму:

$$\begin{aligned} &\sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[V_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[W_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) = \\ &= \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[\nu_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[\mu_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) + \\ &+ 2 \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j e_{j,k_1,k_2}^{[\lambda]} \left(\nu_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right). \end{aligned}$$

Учитывая результаты работ [7, 10], можно показать, что условное распределение этой суммы при фиксированных s_i и θ_i нормально с математическим ожиданием $\sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[\nu_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[\mu_{j,k_1,k_2}^{[\lambda]} \right]^2 \right)$ и дисперсией, не превосходящей

$C_R \sigma^2 \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^{2j} \left(\nu_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right)^2$, где константа C_R зависит от выбранного базиса.

Так как $f \in \text{Lip}(\gamma)$, то, повторяя рассуждения работы [10], можно показать, что

$$\frac{1}{2^{2J}} \mathbb{E}_{s\theta} \left| \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[\nu_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[\mu_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) \right| \rightarrow 0;$$

$$\frac{1}{2^{2J}} \mathbb{E}_{s\theta} \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^{2j} \left(\nu_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right)^2 \rightarrow 0$$

при $J \rightarrow \infty$. Следовательно, применяя неравенство Маркова, получаем, что

$$\frac{1}{2^{2J}} \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[\nu_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[\mu_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) \xrightarrow{P} 0;$$

$$\frac{1}{2^{2J}} \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^{2j} \left(\nu_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right)^2 \xrightarrow{P} 0$$

при $J \rightarrow \infty$. Таким образом,

$$\frac{1}{2^{2J}} \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j \left(\left[\nu_{j,k_1,k_2}^{[\lambda]} \right]^2 - \left[\mu_{j,k_1,k_2}^{[\lambda]} \right]^2 \right) \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

В оставшихся суммах в (9) содержатся индикаторы, в которых либо $|V_{j,k_1,k_2}^{[\lambda]}| > T_\gamma$, либо $|W_{j,k_1,k_2}^{[\lambda]}| > T_\gamma$. Повторяя рассуждения из работы [10], можно показать, что эти суммы при делении на 2^{2J} также сходятся к нулю по вероятности. Теорема доказана.

Помимо асимптотической нормальности оценка (7) также обладает свойством сильной состоятельности.

Теорема 3. Пусть выполнены условия теоремы 2. Тогда при жесткой или мягкой пороговой обработке для любого $\alpha > 2$

$$\frac{\widehat{R}_\nu(f) - R_\nu(f)}{2^{\alpha J}} \rightarrow 0 \text{ п. в. при } J \rightarrow \infty.$$

Поскольку выполнено (8), при фиксированных x_i , к (7) применимо неравенство Боска [11]. Следовательно, доказательство этого утверждения практически полностью повторяет доказательство соответствующего свойства оценки риска в работе [5].

Литература

1. *Kolaczyk E. D.* Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. — Stanford, CA, USA: Stanford University, 1994. PhD Diss. 152 p.
2. *Donoho D.* Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition // Appl. Comput. Harmon. A., 1995. Vol. 2. P. 101–126.
3. *Lee N.* Wavelet-vaguelette decompositions and homogeneous equations. — West Lafayette, IN, USA: Purdue University, 1997. PhD Diss. 103 p.
4. *Шестаков О. В.* Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. — М.: АРГАМАК-МЕДИА, 2016. 200 с.
5. *Шестаков О. В.* Усиленный закон больших чисел для оценки риска в задаче реконструкции томографических изображений из проекций с коррелированным шумом // Информатика и её применения, 2016. Т. 10. Вып. 3. С. 41–45.
6. *Cai T., Brown L.* Wavelet shrinkage for nonequispaced samples // Ann. Stat., 1998. Vol. 26. No. 5. P. 1783–1799.
7. *Cai T., Brown L.* Wavelet estimation for samples with random uniform design // Stat. Probabil. Lett., 1999. Vol. 42. P. 313–321.
8. *Mallat S.* A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 с.
9. *Ерошенко А. А., Шестаков О. В.* Асимптотические свойства оценки риска в задаче восстановления изображения с коррелированным шумом при обращении преобразования Радона // Информатика и её применения, 2014. Т. 8. Вып. 4. С. 32–40.
10. *Шестаков О. В.* Свойства вейвлет-оценок сигналов, регистрируемых в случайные моменты времени // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 16–21.
11. *Johnstone I. M., Silverman B. W.* Wavelet threshold estimates for data with correlated noise // J. Roy. Stat. Soc. B, 1997. Vol. 59. P. 319–351.

Поступила в редакцию 07.04.20

ASYMPTOTICS OF THE MEAN-SQUARE RISK ESTIMATE IN THE PROBLEM OF INVERTING THE RADON TRANSFORM FROM PROJECTIONS REGISTERED ON A RANDOM GRID

O. V. Shestakov^{1,2}

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

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

Abstract: When reconstructing tomographic images, it is necessary to use regularization methods, since the problem of inverting the Radon transform, which is the basis of mathematical models of most tomographic experiments, is ill-posed. Regularization methods based on wavelet analysis have become popular due to their adaptation to local image features and computational efficiency. The analysis of errors in tomographic images is an important practical task, since it makes it possible to evaluate the quality of both the methods themselves and the equipment used. Sometimes, it is not possible to register projection data on a uniform grid of samples. If sample points for each coordinate form a variation series based on a sample from a uniform distribution, then the use of the usual threshold processing procedure is adequate. In this paper, the author analyzes the estimate of the mean-square risk in the Radon transform inversion problem and demonstrates that if the image function is uniformly Lipschitz-regular, then this estimate is strongly consistent and asymptotically normal.

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

DOI: 10.14357/19922264200204

Acknowledgments

This research is supported by the Russian Foundation for Basic Research (project No. 18-07-00252).

References

- Kolaczyk, E. D. 1994. Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. Stanford, CA: Stanford University. PhD Diss. 152 p.
- Donoho, D. 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. A.* 2:101–126.
- Lee, N. 1997. Wavelet-vaguelette decompositions and homogeneous equations. West Lafayette, IN: Purdue University. PhD Diss. 103 p.
- Shestakov, O. V. 2016. *Veroyatnostno-statisticheskie metody analiza i obrabotki signalov na osnove veyvlet-algoritmov* [Probabilistic-statistical methods of signal analysis and processing based on wavelet algorithms]. Moscow: ARGAMAK-MEDIA. 200 p.
- Shestakov, O. V. 2016. Usilennyi zakon bol'shikh chisel dlya otsenki riska v zadache rekonstruktsii tomograficheskikh izobrazheniy iz proektsiy s korrelirovannym shumom [The strong law of large numbers for the risk estimate in the problem of tomographic image reconstruction from projections with a correlated noise]. *Informatika i ee Primeneniya — Inform. Appl.* 10(3):41–45.
- Cai, T., and L. Brown. 1998. Wavelet shrinkage for nonequidistant samples. *Ann. Stat.* 26(5):1783–1799.
- Cai, T., and L. Brown. 1999. Wavelet estimation for samples with random uniform design. *Stat. Probabil. Lett.* 42:313–321.
- Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
- Eroshenko, A. A., and O. V. Shestakov. 2014. Asimptoticheskie svoystva otsenki riska v zadache vosstanovleniya izobrazheniya s korrelirovannym shumom pri obrashchenii preobrazovaniya Radona [Asymptotic properties of risk estimate in the problem of reconstructing images with correlated noise by inverting the Radon transform]. *Informatika i ee Primeneniya — Inform. Appl.* 8(4):32–40.
- Shestakov, O. V. 2019. Svoystva veyvlet-otsenok signalov registriruemyykh v sluchaynye momenty vremeni [Properties of wavelet estimates of signals recorded at random time points]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):16–21.
- Johnstone, I. M., and B. W. Silverman. 1997. Wavelet threshold estimates for data with correlated noise. *J. Roy. Stat. Soc. B* 59:319–351.

Received April 7, 2020

Contributor

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

МЕТОДЫ НАХОЖДЕНИЯ ПРИЧИН СБОЕВ В ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ С ПОМОЩЬЮ МЕТАДАННЫХ*

Н. А. Грушо¹, А. А. Грушо², М. И. Забейло³, Е. Е. Тимонина⁴

Аннотация: Статья посвящена вопросам удаленного выявления и локализации сбоя в информационных системах. Определены информационные ресурсы для решения перечисленных выше задач и исследованы модели использования этих информационных ресурсов. В данной статье метаданные описываются естественными для моделей бизнес-процессов и информационных технологий (ИТ) ориентированными ациклическими графами (DAG — directed acyclic graph). Рассматриваемая задача состоит в следующем. При сбое или ошибке необходимо быстро выйти на содержащий причину сбоя блок, который уже не настолько дорог, чтобы его беречь, и он был бы заменяем. Если блок содержит очень дорогие компоненты, например сервер, то его замена может быть экономически недоступна организации. Если говорить о программных приложениях, то они могут быть легко переустановлены, и цена такой замены невелика. Следовательно, причину надо искать в направлении сверху вниз по уровням иерархической декомпозиции (ИД) DAG информационной технологии. Проведено исследование по анализу и выявлению сбойных данных в ИТ при условии правильности работы всех блоков информационной системы.

Ключевые слова: модели информационных технологий; метаданные; ориентированные ациклические графы; причины сбоев и ошибок в информационных технологиях

DOI: 10.14357/19922264200205

1 Введение

В задачи, решаемые системными администраторами, входит выявление ошибок и сбоев ИТ в распределенных информационно-вычислительных системах (РИВС). Сложность этих задач привела к тому, что наблюдается дефицит высококвалифицированных системных администраторов. В связи с расширением числа информационных систем поддержки цифровой экономики высококвалифицированные системные администраторы все чаще переходят на удаленную работу по обслуживанию информационных систем малого и среднего бизнеса. При этом необходимо удаленно решать следующие задачи:

- как узнать о сбое;
- какую информацию необходимо получить о деталях сбоя;
- как определить причину сбоя;

- как дистанционно осуществить восстановление ИТ;
- как обеспечить безопасность дистанционной работы системного администратора.

В работе не рассматриваются задачи безопасной связи и сбор информации мониторинга в контролируемых информационных системах [1]. В статье рассмотрены вопросы удаленного выявления и локализации сбоя в информационных системах. Определены информационные ресурсы для решения перечисленных выше задач и исследованы модели использования этих информационных ресурсов.

В работах [2–4] определены метаданные для управления соединениями в РИВС. Метаданные представляют информацию о необходимых и допустимых взаимодействиях хостов, на которых реализуется ИТ. Метаданные отражают логику бизнес-процессов и позволяют дистанционно контролировать ИТ, связанные с их выполнением. В [5]

* Работа частично поддержана РФФИ (проекты 18-29-03081 и 18-07-00274).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, info@itake.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

³ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

показано, что метаданные можно также использовать для поиска сбоев и ошибок. Поскольку в работе не рассматриваются сетевые вопросы, то поиск причин сбоев и ошибок в SDN (software-defined networking) можно найти в работе [6].

В отличие от предыдущих работ в данной статье метаданные описываются естественным для моделей бизнес-процессов и ИТ способом — в виде DAG.

Причина A свойства P в данных определяется как детерминированное появление свойства P в данных при появлении в них причины A [7]. В этом определении свойство P называется следствием причины A . Исследования формализованных описаний причин, следствий и связанных с этими понятиями исследований можно найти в работах В. К. Финна (см., например, [8]).

Следует отметить, что причина свойства P не всегда является единственной. Например, предположим, что некоторая система выполняет свои функции последовательно на устройствах A , B и C . Сбой работы системы выявляется на устройстве C (свойство P) и может быть связан с выходом из строя устройства A или устройства B . Каждое из этих событий порождает в системе свойство P . Тогда сбой A и B являются причинами свойства P . Но отсюда не следует, что в случае появления свойства P на устройстве C причина однозначно определена. Вместе с тем уточнение того, какая из причин повлияла на сбой в системе, в этом примере имеет практическое значение. Если можно заменить любое из устройств A или B , то экономически целесообразно сначала узнать, какое из них вышло из строя.

2 Модели графов для описания информационных технологий

В работе рассмотрены ИТ [9], представимые в виде DAG. Прописными латинскими буквами A, B, \dots будем обозначать *данные* (объекты), служащие входными или выходными данными преобразований информации в ИТ, на рисунках данные будем представлять окружностями. Преобразования будем называть *блоками*, обозначать строчными латинскими буквами и на рисунках обозначать прямоугольниками.

Каждый блок соответствует преобразованию информации и решает одну или несколько задач, необходимых для реализации ИТ. Дуги DAG соответствуют передаче данных блокам от предыдущих блоков, т. е. дуга выходит из выходных данных выполненного преобразования и направляется во входные данные для следующего преобразования.

В ИТ могут быть ориентированные циклы преобразований. Очевидно, что входные и выходные данные каждого цикла отличаются. Следовательно, в графе ИТ ориентированные циклы отсутствуют. Если выходные данные преобразования полностью определяют входные данные одного или нескольких следующих преобразований, то из выходных данных такого преобразования выходят несколько дуг, при этом часть выходных данных в следующем блоке может не использоваться.

При рассмотрении ИД ИТ может возникнуть необходимость рассматривать преобразования на уровне операционной системы (ОС) и аппаратной платформы. Тогда преобразование верхнего блока можно представить в виде суперпозиции части преобразования этого блока, блока, связанного с преобразованием информации на уровне ОС и аппаратной платформы, и оставшейся части преобразования верхнего блока. Повтор преобразований допустим, если они имеют разные входные данные.

Простейший DAG, описывающий произвольную ИТ, имеет вид, представленный на рис. 1, где A — входные данные ИТ; B — выходные данные ИТ; f — преобразование, реализующее ИТ.

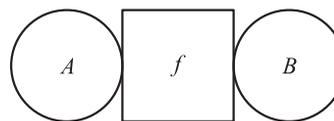


Рис. 1 Простейший DAG

Если выходные данные A какого-то преобразования или выходные данные B другого какого-то преобразования используются для преобразования f , то входные данные для преобразования f представляют собой вектор (A, B) (рис. 2).

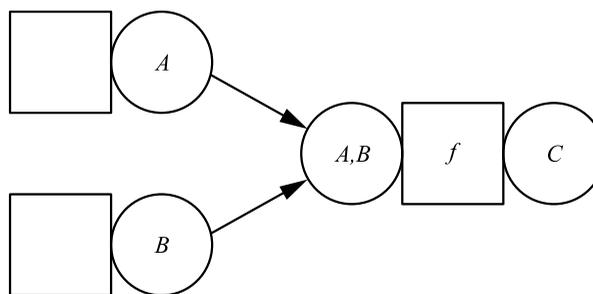


Рис. 2 Векторные входные данные

Далее потребуется ИД DAG. Иерархическая декомпозиция определяется итерационно с помощью двух операций.

1. *Операция деления блока.* Пусть преобразование данных f в текущем DAG имеет вид, показанный на рис. 1. Предположим, что f может

быть представлено в виде суперпозиции двух преобразований:

$$B = f(A) = f_2(f_1(A)); \quad f_1(A) = C.$$

Тогда фрагмент DAG на рис. 1 может быть графически представлен в виде, отраженном на рис. 3. Такое преобразование сохраняет ациклическость исходного графа.



Рис. 3 Операция деления блока

2. *Операция детализации блоков.* Пусть преобразование f зависит от исходных данных (A, B) и может быть представлено в виде $C = f(A, B)$. Предположим, что существуют функции f_1 и f_2 такие, что $f(A, B) = f_2(f_1(A), B)$. Тогда результат этой детализации имеет вид, показанный на рис. 4.

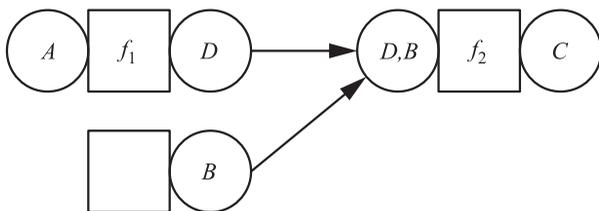


Рис. 4 Операция детализации блока

Очевидно, что операция детализации блоков не порождает ориентированных циклов в порождаемом графе.

Применяя операции деления и детализации, можно из общего представления ИТ (см. рис. 1) получить DAG, в котором блоки соответствуют используемым в ИТ приложениям. Выделим этот этап ИД в связи с тем, что создатели приложений, как правило, готовят для них хорошие средства тестирования и описания входных и выходных данных.

3 Использование иерархической декомпозиции ориентированных ациклических графов для анализа сбоев и ошибок

Рассмотрим этот вопрос на уровне DAG блоков приложений, реализующихся в ИТ. Задача состоит

в следующем. При сбое или ошибке необходимо быстро выйти на содержащий причину сбоя блок, который уже не настолько дорог, чтобы его беречь, и он был бы заменяем. Если блок содержит очень дорогие компоненты, например сервер, то его замена может быть экономически недоступна организации. Если говорить о программных приложениях, то они могут быть легко реинсталлированы, и цена такой замены невелика. Следовательно, причину надо искать в направлении сверху вниз по уровням иерархической декомпозиции DAG ИТ. При этом если выявлены блоки, ставшие причиной выхода из строя ИТ, и их удастся заменить или «отремонтировать», то апостериорная вероятность выхода из строя таких же блоков в дальнейшем увеличивается. Тогда необходимо готовиться к тому, что эти блоки, возможно, будут выходить из строя и в других системах.

Если все приложения (включая ОС и программно-аппаратные платформы) работают исправно, то причину сбоя следует искать в данных. Этот вопрос рассмотрен в следующем разделе.

4 Поиск причин сбоя в данных

Сбои могут возникать даже при правильной работе всех приложений и технических систем. Эти сбои связаны с неправильными входными данными. В ИТ, представленной на рис. 1, сбой данных B при правильно работающем преобразовании f связан с неправильными исходными данными A .

Более детальный анализ причины сбоя в этой ситуации потребует введения новых обозначений и доказательства ряда утверждений.

Пусть A, B и f — параметры блока на рис. 1. Обозначим через $D(f)$ множество возможных исходных данных, которые могут быть поданы на вход преобразования f . Если на данных $A \in D(f)$ преобразование f не может быть вычислено, то $f(A) = 0$. Если сбой произошел на предыдущих блоках ИТ, то формально положим $0 \in D(f)$ и $f(0) = 0$. Таким образом, значение 0 — это очевидно идентифицируемый сбой.

Пусть $J(f)$ — это множество выходных данных преобразования f при входных данных, принадлежащих $D(f)$, $0 \in J(f)$. Предположим, что определена функция $\varphi(B)$, принимающая значение 1, когда B можно признать допустимым результатом преобразования f , и значение 0 в противном случае. Тогда $J(f)$ разбивается на два непересекающихся подмножества $J^+(f)$ и $J^-(f)$, $J^+(f) \cap J^-(f) = \emptyset$, соответственно допустимых и недопустимых значений. Недопустимое значение означают сбой или

ошибку в ИТ по результатам выполнения преобразования f .

Разбиение $J(f) = J^+(f) \cup J^-(f)$ индуцирует разбиение множества $D(f) = D^+(f) \cup D^-(f)$, где $D^-(f)$ — множество тех исходных данных A , которые преобразуются в $J^-(f)$. Пусть при выполнении ИТ встретился фрагмент DAG, изображенный на рис. 3, когда последовательно выполняются преобразования f_1 и f_2 , а выходные данные преобразования f_1 служат выходными данными преобразования f_2 . Тогда из недопустимости выходных данных B следует, что входные данные преобразования f_2 принадлежат $D^-(f_2)$. Но тогда выходные данные преобразования f_1 должны удовлетворять условию $J^-(f_1) \subseteq D^-(f_2)$. Иначе данные B — допустимы. Это означает, что недопустимость данных B (результата блока f_2) при правильности работы блоков f_1 и f_2 требует, чтобы данные $A \in D^-(f_1)$.

Предположим, что в ИТ выделен фрагмент, изображенный на рис. 2. При правильности работы f и предыдущих блоков (обозначим их f_1 и f_2), если данные $C \in J^-(f)$, то вектор данных $(A, B) \in D^-(f)$. При этом важно, что данные A и B могут быть допустимыми для преобразований f_1 и f_2 , т.е. помимо недопустимости данных A или B ошибку может инициировать сочетание (A, B) . Прообраз данных (A, B) находится в произведении $D(f_1) \times D(f_2) = D((f_1, f_2))$. Отсюда $(A, B) \in D^-((f_1, f_2))$. Справедливо следующее утверждение.

Утверждение 1. *Если все блоки работают правильно, но результат данного блока является ошибочным, то такой результат возможен, когда исходные данные этого блока представимы в виде вектора, несовместимого с допустимыми данными ИТ.*

Продолжая последовательность приведенных выше рассуждений (возможно, с увеличением размерности данных и преобразований), получим следующее утверждение.

Утверждение 2. *В случае недопустимости выходных данных ИТ их причинами могут быть либо сбои и ошибки в блоках, либо недопустимость исходных данных ИТ.*

Смысл этого утверждения состоит в том, что недопустимость выходных данных при правильной работе блоков не может возникнуть из-за недопустимости данных промежуточных блоков, а является следствием недопустимости входных данных ИТ.

Следствие. Для поиска причин ошибок и сбоев необходимо иметь хорошие тесты для блоков и две бинарные функции:

(1) φ для определения допустимости выходных данных ИТ;

(2) ψ для проверки допустимости входных данных.

Проведенный анализ позволяет определить источник ошибок в данных.

5 Примеры быстрого поиска сбоев блоков и выявление ошибок в данных

Поиск причин скрытых ошибок и сбоев — важная функция системного администратора информационной системы. Для сложных ИТ или для большого количества ИТ метаданные в виде DAG позволяют выделить сбойные фрагменты информационной системы или отдельные блоки, которые необходимо проверить на наличие сбоев. Однако определение множеств $D^+(f)$, $D^-(f)$, $J^+(f)$ и $J^-(f)$ — труднореализуемая задача. Помимо определения этих множеств необходимо строить эффективно вычисляемые функции принадлежности к этим множествам (например, φ и ψ). Поэтому можно воспользоваться методом контейнерной виртуализации [10].

Пусть выделен хотя бы один сервер, на котором можно быстро устанавливать виртуальные контейнеры, эмулирующие проверяемые блоки (приложение и ОС).

Предположим, что входные и выходные данные приложений запоминаются (например, в соседних блоках). Тогда можно использовать известный метод контроля программ для проверки работы блока. Для этого на эмуляторе запускается преобразование контролируемого блока и сравниваются результаты этого преобразования с полученными ранее результатами. Если они совпадают, то можно считать, что блок работает правильно. Если результаты не совпадают, то причина этого может быть или в сбое преобразования, или в сбое ОС, или в сбое аппаратной платформы. Для каждого из этих компонентов должны быть построены свои тесты. При этом целесообразно построить отдельный DAG для преобразований в блоке, позволяющий рассматривать отдельные процессы ИТ.

Хранение контейнеров возможно в базе данных ресурсов системного администратора, что позволит ему быстро на основании метаданных выявлять сбои и заменять сбойный блок временно работающим виртуальным эмулятором (с замыканием на него соответствующих сетевых соединений).

Метаданные можно также использовать для контроля сетевых соединений [5]. Аналогичный подход был разработан в [11] для контроля критических систем.

6 Заключение

В работе рассмотрена задача поиска и локализации причин сбоев и ошибок в ИТ, представимых ИД DAG. Введенные модели используют нетрадиционное определение DAG в том смысле, что дуги графа характеризуют связи между входными и выходными данными различных преобразований в ИТ.

Построена ИД таких DAG, позволяющая решать задачу поиска наименьшей глубины абстракции для экономически осуществимой замены программного обеспечения или элементов аппаратной платформы информационной системы.

Проведено исследование по анализу и выявлению сбойных данных в ИТ при условии правильности работы всех блоков информационной системы.

Литература

1. Грушо Н. А. Метод интеграции многоагентного поиска информации со средствами анализа защищенности и информационными сервисами в цифровых инфраструктурах // Проблемы информационной безопасности. Компьютерные системы, 2019. № 2. С. 45–55.
2. Грушо А. А., Забейхайло М. И., Грушо Н. А., Тимонина Е. Е. Информационная безопасность на основе метаданных в компонентно-интеграционных архитектурах информационных систем // Системы и средства информатики, 2018. Т. 28. № 2. С. 34–41.
3. Грушо А. А., Тимонина Е. Е., Шоргин С. Я. Иерархический метод порождения метаданных для управления сетевыми соединениями // Информатика и её применения, 2018. Т. 12. Вып. 2. С. 44–49.
4. Грушо А. А., Забейхайло М. И., Грушо Н. А., Тимонина Е. Е. Поиск эмпирических причин сбоев и ошибок в компьютерных системах и сетях с использованием метаданных // Системы и средства информатики, 2019. Т. 29. № 4. С. 28–38.
5. Grusho A., Grusho N., Timonina E. Information flow control on the basis of meta data // Distributed Computer and Communication Networks, 22-nd International Conference, DCCN 2019, Revised Selected Papers / Eds. V. M. Vishnevskiy, K. E. Samouylov, D. V. Kozurev. — Lecture notes in computer science ser. — Springer, 2019. Vol. 11965. P. 548–562.
6. Grusho A., Grusho N., Zabezhailo M., Zatsarinny A., Timonina E. Information security of SDN on the basis of metadata // Computer network security / Eds. J. Rak, J. Bay, I. V. Kotenko, et al. — Lecture notes in computer science ser. — Springer, 2017. Vol. 10446. P. 339–347.
7. Милль Дж. С. Система логики силлогической и индуктивной: Изложение принципов доказательства в связи с методами научного исследования / Пер. с англ. — 5-е изд., испр. и доп. — Сер. «Из наследия мировой философской мысли: Логика». — М.: Ленанд, 2011. 832 с. (Mill J. S. A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and the methods of scientific investigation. — Cambridge library collection — philosophy ser. — Cambridge University Press. Vol. 1. 602 p.)
8. Финн В. К. Искусственный интеллект: Методология, применения, философия. — М.: Красанд, 2011. 448 с.
9. Самуйлов К. Е., Чукарин А. В., Яркина Н. В. Бизнес-процессы и информационные технологии в управлении телекоммуникационными компаниями. — М.: Альпина Паблишерс, 2009. 442 с.
10. Кондрашин М. Контейнерная виртуализация: преимущества и проблемы безопасности // ИнформКурьер-Связь, 2019. № 4. С. 92–95.
11. Шубинский И. Б., Николаев В. И., Колганов С. К., Заляц А. М. Активная защита от отказов управляющих модульных вычислительных систем. — СПб.: Наука, 1993. 284 с.

Поступила в редакцию 08.04.20

METHODS OF FINDING THE CAUSES OF INFORMATION TECHNOLOGY FAILURES BY MEANS OF METADATA

N. A. Grusho¹, A. A. Grusho¹, M. I. Zabezhailo², and E. E. Timonina¹

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The work is devoted to remote detection and localization of failures in information systems. Information resources for these tasks have been identified and models for the use of these information resources have been investigated. This article describes metadata by directed acyclic graphs (DAG) which are used for business processes and information technology descriptions. The task in question is as follows. In case of a failure or an error, it is necessary to find quickly the block containing the cause of failure, which is not so expensive that it could

not be replaced, and it would be replaced. If a unit contains very expensive components, such as a server, replacing it may not be cost-effective for an organization. In terms of software applications, they can be easily reinstalled and the cost of such a replacement is low. Therefore, the cause should be sought in the top-down direction along the levels of hierarchical decomposition of DAG of information technology. A study has been carried out on analysis and detection of failed data in information technology, provided that all blocks of an information system operate correctly.

Keywords: models of information technologies; metadata; directed acyclic graphs; causes of information technology failures and errors

DOI: 10.14357/19922264200205

Acknowledgments

The paper was partially supported by the Russian Foundation for Basic Research (projects 18-29-03081 and 18-07-00274).

References

1. Grusho, N. 2019. Metod integratsii mnogoagentnogo poiska informatsii so sredstvami analiza zashchishchenosti i informatsionnymi servisami v tsifrovyykh infrastrukturakh [Method of integration of multi-agent information search with security analysis tools and information services for digital infrastructures]. *Problemy Informatsionnoy Bezopasnosti. Komp'yuternye Sistemy* [Information Security Problems. Computer Systems] 2:45–55.
2. Grusho, A. A., M. I. Zabezhailo, N. A. Grusho, and E. E. Timonina. 2018. Informatsionnaya bezopasnost' na osnove metadannykh v komponentno-integratsionnykh arkhitekturakh informatsionnykh sistem [Information security on the basis of meta data in enterprise application integration architecture of information systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):34–41.
3. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2018. Ierarkhicheskiy metod porozhdeniya metadannykh dlya upravleniya setevymi soedineniyami [Hierarchical method of meta data generation for control of network connections]. *Informatika i ee Primeneniya — Inform. Appl.* 12(2):44–49.
4. Grusho, A. A., M. I. Zabezhailo, N. A. Grusho, and E. E. Timonina. 2019. Poisk empiricheskikh prichin sboev i oshibok v komp'yuternykh sistemakh i setyakh s ispol'zovaniem metadannykh [Search of empirical causes of failures and errors in computer systems and networks using metadata]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(4):28–38.
5. Grusho, A. A., N. A. Grusho, and E. E. Timonina. 2019. Information flow control on the basis of meta data. *Distributed computer and communication networks*. Eds. V. M. Vishnevskiy, K. E. Samouylov, and D. V. Kozyrev. Lecture notes in computer science ser. Springer. 11965:548–562.
6. Grusho, A., N. Grusho, M. Zabezhailo, A. Zatsarinny, and E. Timonina. 2017. Information security of SDN on the basis of metadata. *Computer network security*. Eds. J. Rak, J. Bay, I. V. Kotenko, et al. Lecture notes in computer science ser. Springer. 10446:339–347.
7. Mill, J. S. 2011. *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and the methods of scientific investigation*. Cambridge library collection — philosophy ser. Cambridge University Press. Vol. 1. 602 p.
8. Finn, V. K. 2011. *Iskusstvennyy intellekt: Metodologiya, primeneniya, filosofiya* [Artificial intelligence: Methodology, applications, philosophy]. Moscow: Krasand. 448 p.
9. Samuylov, K. E., A. V. Chukarin, and N. V. Yarkina. 2009. *Biznes-protsessy i informatsionnye tekhnologii v upravlenii telekommunikatsionnymi kompaniyami* [Business processes and information technologies in management of the telecommunication companies]. Moscow: Alpina Publ. 442 p.
10. Kondrashin, M. 2019. Konteynernaya virtualizatsiya: preimushchestva i problemy bezopasnosti [Container virtualization: advantages and security problems]. *InformKur'er-Svyaz'* [InformCourier-Communications] 4:92–95.
11. Shubinsky, I. B., V. I. Nikolaev, S. K. Kolganov, and A. M. Zayats. 1993. *Aktivnaya zashchita ot otkazov upravlyayushchikh modul'nykh vychislitel'nykh sistem* [Active protection against failures of control modular computing systems]. St. Petersburg: Nauka. 284 p.

Received April 8, 2020

Contributors

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

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; grusho@yandex.ru

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

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

СОВМЕСТНАЯ ОЦЕНКА ПРЕДСКАЗУЕМОСТИ ДАННЫХ И КАЧЕСТВА ПРЕДИКТОРОВ*

С. Л. Френкель¹, В. Н. Захаров²

Аннотация: Предлагается и анализируется новый подход к выбору предикторов, необходимых для предсказания будущих значений в последовательностях данных в конкретный временной период. Цель — недорогие реализуемые техники, обеспечивающие выбор приемлемого предиктора для текущего сеанса предсказания или же принятия решения о невозможности выполнить надежный прогноз, если обнаружится, что данный участок последовательности не обладает свойством предсказуемости. Для этого предсказуемость последовательности определяется как максимальная по множеству доступных предикторов условная вероятность правильного предсказания при данном наборе наблюдаемых значений. Выбор (отсевание) предикторов выполняется как по величине оценки данной вероятности, так и по степени отличия конкретного предиктора от предиктора, оптимального для предсказаний следующего исхода последовательности испытаний Бернулли.

Ключевые слова: предсказание случайных последовательностей; предикторы; анализ данных

DOI: 10.14357/19922264200206

1 Введение

В настоящее время значительное число практических задач в области экономического и финансового управления, медицины, управления вычислительными и коммуникационными сетями решается с использованием различных математических методов прогнозирования.

Примером служит задача онлайн-прогнозирования, выполняемого на основе больших наборов ранее накопленных данных, используемых в алгоритмах *машинного обучения* (МО, *англ.* ML — Machine Learning), с помощью готовых программных инструментов предсказания (далее — *предикторов*), например для предсказания объемов рассматриваемого трафика в определенные моменты в будущем или возможных вредоносных атак [1]. Целью может быть прогноз для администратора сети, чтобы тот успел начать действовать еще до того, как возникнет проблема.

Часто онлайн-работа предикторов требует настройки их параметров, что влияет на суммарные вычислительные затраты всей схемы управления сетью. Разработчик подсистемы прогнозирования системы управления трафиком и безопасности должен предусмотреть возможность оценки ожидаемого качества и возможных затрат на прогнозирование в каждый требуемый момент, когда работает подсистема прогнозирования, чтобы избежать лишних затрат. Для этого необходим инструмент оценки

эффективности прогнозирования на данном временном интервале. Качество прогноза в некоторый момент времени зависит от того, насколько правильно используемый предиктором алгоритм на основе данных в прошлом отражает ситуацию в будущем. Поэтому важными становятся задачи разработки моделей предсказуемости данных, в которых «предсказуемость» (predictability) выступает как одна из характеристик набора данных, и задачи оценки прогностической способности предиктора для набора данных с теми или иными характеристиками, поддающимися измерению (оценке) в требуемый период времени. В настоящее время можно выделить два основных подхода к решению этих задач: вероятностно-статистический и логико-семантический.

Предметом исследований предсказания поведения систем с помощью семантических моделей выступают в значительной степени языки описания предметной области, для которой будут построены модели предсказания, и теории, на основе которых выполняются предсказания [2].

Будем рассматривать вероятностно-статистические подходы как наиболее распространенные в современных инструментальных средствах прогнозирования. В рамках этого подхода наборы данных (например, записи трафика в тех или иных терминах) моделируются как случайные последовательности или процессы.

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00669, 18-29-03100 и 18-07-01434).

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, fsergei51@gmail.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vzakharov@ipiran.ru

Итак, рассматривается решение следующей проблемы:

- имеется множество последовательно рассматриваемых данных;
- имеется доступный набор предикторов, реализующих различные алгоритмы предсказания для этой последовательности.

Как осуществить выбор предиктора для предсказания следующего значения элемента последовательности, используя только результаты предсказаний на предыдущих элементах, и какие свойства индивидуальных последовательностей могут позволить решить эту задачу?

Прежде всего укажем, что данная проблема связана с известными результатами относительно разрешимости задач предсказания. В самом общем виде они формулируются в так называемой теореме об отсутствии «бесплатного обеда» (The No Free Lunch Theorem [3]) в задачах МО, которая гласит, что не существует единственного алгоритма, который мог бы быть лучшим для решения любых проблем МО, и необходимо пробовать разные алгоритмы, принимая во внимание размер и структуру набора данных, чтобы определить, какой из них окажется лучшим для решения конкретной проблемы. Известно также о невозможности построения единственного предиктора с нулевой средней ошибкой предсказания для семейств дискретных последовательностей с неизвестным вероятностным распределением генерирующего их источника [4].

Подчеркнем, что рассматривается ситуация, когда разработчик системы, в которую должна входить подсистема прогноза, не является экспертом в МО настолько, чтобы разрабатывать новые алгоритмы, а намерен использовать лишь доступные готовые предикторы и выбор возможен лишь по наборам данных, формирующимся в результате работы этих предикторов.

В свое время в работах [5, 6] были сформулированы и предложены подходы к решению задачи оценки двоичных последовательностей относительно некоторого множества предикторов, однако получены в основном результаты асимптотического характера и их использование в практических случаях затруднительно. В последние годы появилось много работ по статистическим оценкам работы предикторов из известных облачных ресурсов (MS AWS Azure Machine Learning, Google Cloud Machine Learning и т. д.), однако оценки получены относительно так называемой средней доли ошибок предсказаний, что, как будет показано в разд. 3, требует существенного уточнения.

В данной работе предлагается использовать эвристические методы априорного выбора предик-

торов и/или подмножества двоичных наборов данных для данного предиктора, позволяющих улучшить качество прогноза. Данные рассматриваются как двоичные последовательности, поскольку многие практические задачи прогнозирования (предсказания финансовых индексов, вредоносных атак на сетевые ресурсы и т. д.) требуют предсказания не самих абсолютных значений, а знаков (направления) их изменения, когда задача может быть сведена к предсказанию двоичных последовательностей [7].

В качестве эвристической меры эффективности использования данного предиктора для предсказания двоичной последовательности в следующий после последнего наблюдения момент времени предлагается некоторая мера «близости» характеристик качества предсказания рассматриваемой последовательности данным предиктором, к характеристикам качества предсказания последовательности независимых испытаний (испытаний Бернулли) некоторым тестовым предиктором, который обеспечивает минимальные потери предсказания, при функции потерь (loss function [8]) Хэмминга. В дальнейшем будем называть такой тестовый предиктор «предиктором Бернулли» (ПБ). Данный критерий качества обобщенно учитывает как структуру множества ошибок предсказаний оцениваемым предиктором, так и статистическую оценку условной вероятности правильного предсказания элемента последовательности в момент t по ранее наблюдаемым ее значениям, и чем ближе эти характеристики к характеристикам ПБ для данной последовательности, тем менее пригодным считается предиктор. Заметим, что в разговоре о близости не предполагалось, чтобы она формально отвечала требованиям метрики расстояния, неравенству треугольника прежде всего. При этом выполняется указанное выше требование обеспечить реализацию предлагаемого эвристического подхода без существенных дополнительных затрат на его программную имплементацию.

2 Задача оценки предсказуемости данных

Определим задачу прогнозирования последовательности. Наблюдатель последовательно наблюдает значения x_1, x_2, \dots, x_t известных типов, например вещественные значения (тогда речь идет о временных рядах) или символьные в некотором алфавите A . В момент времени $t - 1$, получив значения x_1, \dots, x_{t-1} , наблюдатель *прогнозирует* следующее значение x_t , т. е. предсказывает, что x_t примет значение b_t . Проблема состоит в том, какое прави-

ло должен использовать наблюдатель для принятия решения о значении x_t .

В рамках широко используемой теории принятия решений [8] для решения этой задачи предполагается наличие некоторой числовой функции потерь $l : \{b_t, x_t\} \rightarrow R$, где R — некоторое измеримое множество, на котором задается мера качества прогноза. Эта функция может быть, например, расстоянием Хэмминга (для двоичных последовательностей) или квадратом ошибки (для числовых). Функция потерь $l(p, x)$ при значениях $p = b_t, x = x_t$ — это величина потери, которая связана с результатом прогноза (принятием решения о значении b_t), если истинным оказалось значение x_t , т. е. потеря, убыток, штраф от неправильного прогноза.

Соответственно, в самом общем виде задача предсказания состоит в построении алгоритма назначения (выбора) значения b_t , обеспечивающего минимум потерь (максимум качества прогноза) относительно принятой функции потерь.

Предсказуемость последовательности данных служит характеристикой эффективности решения задачи предсказания (прогнозирования) значения данных в будущем, или появления событий в будущем, по ранее наблюдаемым данным или событиям.

Концепция предсказуемости (называемая, в том числе и в русскоязычной литературе, predictability) широко изучалась в многочисленных работах по теории случайных процессов [9]. Однако предлагаемый подход к оценке предсказуемости основывается на работах [4–6, 10], относящихся к задачам построения так называемых «универсальных» алгоритмов предсказания «индивидуальных» последовательностей, т. е. таких методов и таких случайных последовательностей, для которых невозможно сделать какие-либо предположения о том, к каким классам вероятностных распределений они принадлежат [11].

В рамках данного подхода сформулированное в работах [10, 11] понимание предсказуемости двоичной последовательности предиктором, реализующим алгоритм A , позволяет оценивать предсказуемость степенью отличия вероятности предсказания этим предиктором следующего значения от $1/2$, т. е. от вероятности правильного предсказания, которое реализует предиктор, эквивалентный подбрасыванию честной монеты.

Решению описанной во введении общей задачи в большей степени отвечает подход, сформулированный в [5]. Дадим его краткое изложение (в более подходящих для рассматриваемой задачи терминах).

Пусть $x_1, x_2, \dots, x_n, x_t \in \{0, 1\}$, — двоичная последовательность независимых испытаний длины n

(последовательность Бернулли) с параметром (вероятностью единицы) $\theta = \Pr \{x_t = 1\}$. Предиктор $f = (f_1, f_2, \dots)$ — это последовательность функций, принадлежащих некоторому семейству. В момент t используется функция предсказания $f_t : \{0, 1\}^t \rightarrow \{0, 1\}$, т. е. она вычисляет по наблюдаемой двоичной последовательности x_1, x_2, \dots, x_t длины t следующее значение: $x'_{t+1} = f_t(x_1, x_2, \dots, x_t)$. Отметим, что x'_{t+1} означает именно оценку истинного значения x_{t+1} и может не совпадать с ним. Обозначим число неправильных предсказаний после применения предиктора f для всей указанной последовательности длины n как $n_e(f) = \sum_{t=1, n} 1\{x'_t \neq x_t\}$, где $1\{\cdot\}$ — функция-индикатор, равная единице при выполнении условия в скобках и нулю в противном случае.

Задача состоит в минимизации

$$\pi_\theta(f) = \limsup_{n \rightarrow \infty} \frac{E_\theta n_e(f)}{n},$$

где E_θ — математическое ожидание случайной величины $n_e(f)$ для последовательности Бернулли x_1, x_2, \dots, x_n с параметром θ . Предсказуемость последовательности определяется как значение

$$\pi_\theta = \inf_f (\pi_\theta(f)) = \inf_f \left(\limsup_{n \rightarrow \infty} \frac{E_\theta n_e(f)}{n} \right). \quad (1)$$

Точная нижняя грань \inf_f берется по всем возможным предикторам f .

Рассмотрим возможность практического вычисления (1) для последовательности Бернулли. Известно [5], что \inf_f для последовательности Бернулли с параметром θ достигается на предикторе, реализующем следующую функцию предсказания:

$$x'_{t+1} = \begin{cases} 1, & \text{если } \theta > \frac{1}{2}; \\ 0, & \text{если } \theta \leq \frac{1}{2}. \end{cases}$$

Однако, если θ точно не известно, то неясно, какой предиктор следует использовать. Интуитивно предиктор

$$x^*_{t+1} = f^*(x_1, x_2, \dots, x_t) = \begin{cases} 0, & \text{если } \theta'(t) < \frac{1}{2}; \\ 1, & \text{если } \theta'(t) > \frac{1}{2}; \\ 0 \text{ или } 1 & \text{в зависимости от результата} \\ & \text{подбрасывания «честной» монеты,} \\ & \text{если } \theta = \frac{1}{2}, \end{cases}$$

где $\theta'(t) = t^{-1} \sum_{r=1,t}(x_r)$ — текущая статистическая оценка θ , представляется в некотором смысле лучшим, если значение θ неизвестно. В дальнейшем будем называть такой предиктор «предиктором Бернулли».

В [5] было доказано, что

- (а) для каждого предиктора f и для любого $\theta \neq 1/2$ либо

$$E_{\theta}n_e(f) \geq n\pi_{\theta} + c_0(\theta) - o(1),$$

либо

$$E_{\underline{\theta}}n_e(f) \geq n\pi_{\underline{\theta}} + c_0(\underline{\theta}) - o(1),$$

где $\underline{\theta} = 1 - \theta$, $c_0(\theta) = [2(1 - 2\pi_{\theta})]^{-1}$;

- (б) предиктор f^* удовлетворяет обоим условиям:

$$E_{\theta}n_e(f^*) \leq n\pi_{\theta} + c_0(\theta);$$

$$E_{\underline{\theta}}n_e(f^*) \leq n\pi_{\underline{\theta}} + c_0(\underline{\theta}).$$

Тем самым было доказано, что предиктор f^* оказывается в определенном смысле лучшим, если значение θ неизвестно.

Однако, учитывая, что в используемых на практике предикторах основным показателем качества служит доля правильных предсказаний (success rate — SR), оцениваемая по достаточно большой выборке (аналогично рассмотренной выше n -последовательности), перейдем к неравенствам именно для этого показателя, связанного с долей неправильных предсказаний, как $1 - n_e(f)/n$ и перепишем указанные в (б) неравенства как

$$\left. \begin{aligned} E_{\theta}SR(f^*) &\geq 1 - \pi_{\theta} - \frac{c_0(\theta)}{n}; \\ E_{\underline{\theta}}SR(f^*) &\geq 1 - \pi_{\underline{\theta}} - \frac{c_0(\underline{\theta})}{n}. \end{aligned} \right\} \quad (2)$$

Итак, для последовательности независимых испытаний Бернулли известен оптимальный предиктор и можно оценить ошибки предсказания с учетом только статистически оценочного знания параметра Бернулли. Нетрудно показать, что описанный ПБ обеспечивает оптимальное предсказание также по критерию минимизации потерь при использовании расстояния Хэмминга в качестве функции потерь, а именно: присваивает функции значения:

$$l(x_t, x'_t) = \begin{cases} 0 & \text{при } x_t = x'_t; \\ 1 & \text{в противном случае,} \end{cases}$$

т. е. «потерял все», если предсказал неправильно, и ничего не потерял, если правильно (совершенно разумный принцип при оценивании, например,

прогнозирования DDoS (distributed denial of service) атак в сети). Этот предиктор, разумеется, довольно примитивный, но при этом все-таки обеспечивает большую вероятность правильного прогноза, чем простое угадывание (подбрасывание «честной» монеты).

Однако, как указывалось, предсказуемость в понимании (1) зависит только от свойств последовательности, а не от предикторов. Поэтому ставится задача: на основании (1) сформулировать количественное определение прогнозируемости случайной последовательности x в момент t относительно заданного множества предикторов f , т. е. как наименьшее значение доли ошибочных предсказаний, которые могут обеспечить предикторы из данного набора доступных предикторов при их независимом использовании на достаточно длинных отрезках последовательности.

Очевидно, что любой предиктор, обеспечивающий лучшую долю предсказаний, чем БП, не должен повторять структуру множества ошибочных решений БП. Учитывая это обстоятельство, будем оценивать предсказуемость таких индивидуальных последовательностей для конкретных предикторов, сравнивая их с ПБ.

3 Вычисление количественной меры предсказуемости случайной последовательности

Из сопоставления формул (1) и (2) нетрудно видеть, что доля правильных предсказаний $SR(\cdot)$ представляет собой статистическую оценку условной вероятности:

$$\pi(x, t) = \text{Prob}(b_t = x_t | x_{t-k}, \dots, x_{t-1}),$$

где x_t — истинное значение члена последовательности; k — длина наблюдаемой последовательности; $b_t = f_p(x_{t-m}, \dots, x_{t-1})$ — предсказанное предиктором f_p значение случайной последовательности в момент t , $m \leq k$ — число членов последовательности, предшествующих t , которое используется в данном предикторе p для получения прогноза.

Соответственно, в определении предсказуемости согласно (1) можно от минимизации математического ожидания доли неправильных предсказаний по данному распределению вероятностей нулей и единиц последовательности длиной n перейти к максимизации вероятности (доли, в практическом аспекте) правильных предсказаний, понимая под количественной мерой предсказуемости статистическую оценку величины $\pi(x, t)$.

Заметим, что отличием длины наблюдаемой последовательности k и числа элементов последовательности m , используемой предиктором, учитывается то, что выбор некоторой обучающей выборки, по которой в конечном итоге вычисляется прогнозируемое значение, путем ли автоматической установки параметров используемой модели предиктора (например, нейронной сетью некоторого типа) или вручную пользователем программы, которая реализует алгоритм, является неотъемлемым компонентом алгоритма предсказания.

4 Вычисление $\pi(x, t)$

Приведем порядок вычисления $\pi(x, t)$ как эмпирической функции распределения. Пусть x_{t-M}, \dots, x_t (рис. 1) — участок двоичной последовательности $x(t)$, рассматриваемый в пределах временного окна W_M размером M в предположении, что используемый предиктор f_p выполняет предсказание по m прошлым наблюдаемым значениям последовательности.

Первое предсказание предиктором f_p в данном окне выполняется для члена последовательности в момент $(t - M + m)$, зная значения членов последовательности в моменты $(t - M)$, $(t - M + 1)$, \dots , $(t - M + m - 1)$. Следующее предсказание делается для момента $(t - M + m + 1)$ по данным, начиная со следующего за $(t - M)$ момента $(t - M + 1)$ до момента $(t - M + m)$, вычисляя $f_p(x_{t-M+1}, \dots, x_{t-M+m})$, и так далее до последнего в данном окне момента t , в котором вычисляют $f_p(x_{t-m}, x_{t-m+1}, \dots, x_{t-1})$.

Тогда, рассматривая поток предсказаний как последовательность Бернулли с единичными исходами (общим числом N_1) для правильных предсказаний и нулевыми для неправильных, в качестве оценки $SR(W_M, p, t)$ данной условной вероятности можно взять долю правильных предсказаний на временном окне W_M :

$$\begin{aligned} SR(W_M, p, t) &= \frac{N_1}{M} = \\ &= M^{-1} \sum_{i=1, M-m+1} 1(f_p(x_{t-M+i-1}, \dots, x_{t-M+i+m-2}) = \\ &= x_{t-M+i+m-1}), \end{aligned}$$

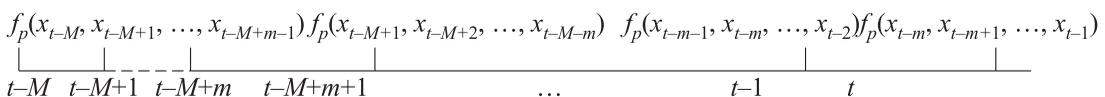


Рис. 1 Интервал прогнозирования, $(t - M + m)$ — момент первого предсказания предиктором с m входными (наблюдёнными) значениями последовательности

где $x_{t-M+i-1}, \dots, x_{t-M+i+m-2}$ — значения членов последовательности в m моментов времени, предшествующих моменту $(t - M + i + m - 1)$, $i = 1, \dots, M - m + 1$, которые предиктор использует в качестве входных (в соответствии с его алгоритмом).

Очевидно, что величина $SR(\cdot)$, часто называемая в документации по использованию программных реализаций алгоритмов предсказаний «коэффициентом успешных предсказаний», служит естественной мерой успешных предсказаний к моменту t .

Вероятность единицы p_1 (параметр Бернулли, аналогичный θ в предыдущем разделе) точно не известна, а известна ее статистическая оценка, позволяющая судить о ее величине в пределах доверительного интервала. Однако (2) и [6] показывают, что ширина этих интервалов имеет асимптотическую оценку, в зависимости от предположений о моделях предикторов, между $O(1/\sqrt{M})$ и $O(1/M)$, т.е. не шире, чем доверительные интервалы стандартных статистических оценок, что допускает практическое использование оценок $SR(\cdot)$.

5 Использование оценок прогнозирования для выбора предиктора или отказа от выполнения прогноза в данный момент

Предположим, что предсказываемая последовательность — это последовательность независимых испытаний с неизвестным параметром распределения Бернулли. Пусть к некоторому моменту t выполнено M предсказаний с предиктором f_p и вычислено $SR(\cdot)$. Определим, разумно ли использовать этот предиктор для дальнейших предсказаний или надо выбрать другой предиктор из доступного списка предикторов. Для оценки качества предиктора используем оценку $SR(\cdot)$ и ее сравнение с оптимальным ПБ. Поясним смысл предлагаемого сравнения на примере нескольких хорошо известных предикторов XGD (eXtreme Gradient Boosting), SGD (stochastic gradient descent) и SVR (Support

Vector Regression) из библиотек Python (из-за ограниченности размера статьи их описание не приводится. Читатель без труда найдет многочисленные описания в Сети, например, в [12]).

Пусть использован один из таких предикторов, например XGD, и вычислена оценка $SR(\cdot)$. Очевидно, что логика использования $SR(\cdot)$ как характеристики предсказуемости состоит в том, что некоторые явные или скрытые характеристики последовательности, значимые для алгоритма работы данного предиктора, не изменятся в следующий момент и можно считать, что ожидаемая доля успешных предсказаний будет примерно такой же, как ранее вычисленная величина $SR(\cdot)$.

Опишем кратко, что это за характеристики последовательности, для XGBoost.

В основе XGBoost лежит алгоритм градиентного бустинга деревьев решений, который строит модель предсказания в форме ансамбля предсказывающих моделей, обычно деревьев решений. Обучение ансамбля проводится последовательно на входных данных предсказываемой последовательности, в которой выделяются паттерны определенной длины (например, «0110»), за которыми должны следовать предсказываемые «0» или «1».

Пусть рассматривается упоминавшаяся последовательность в окне W_M (т.е. $(t-M, t)$) просто как последовательность Бернулли и используется для предсказания также оптимальный предсказатель Бернулли. Поэтому можно говорить об отличии некоторого предиктора двоичной последовательности f_p от такого простого предиктора (ПБ) как по оцененной величине $SR(\cdot)$, так и по структуре ошибок, т.е. по соотношению ошибок «0 вместо 1» («0 → 1») и «1 вместо 0» («1 → 0»). Очевидно, что если сколько-нибудь существенного отличия нет, то можно говорить о неэффективности предиктора f_p , поскольку аналогичное решение дает существенно более простой ПБ.

Критерий перехода к другому предиктору для прогнозирования в момент времени $(t-M+1)$ формулируется следующим образом (считая, что величина M достаточно велика, чтобы обеспечить необходимый узкий доверительный интервал оценок вероятностей нулей, единиц и $SR(\cdot)$).

Предположим, что на интервале длиной M , на котором предиктор f_p делает M предсказаний значений x_{t-M+i} из предыдущих значений, частота единиц равна p_1 , частота правильных прогнозов нулевых значений относительно всех нулевых значений — $p_{V=1}(W_M, f_p = 0)$, где $(V=1)$ — знак правильного предсказания предиктором f_p внутри окна.

Тогда прогноз этого предиктора f_p принимается и для следующего интервала, если при $p_1 > 1/2$

$$\left. \begin{aligned} p_1 < SR(W_M, p, t); \\ p_{V=1}(W_M, f_p = 0) > \frac{1-p_1}{\tau}, \end{aligned} \right\} \quad (3)$$

а при $p_1 < 1/2$

$$\left. \begin{aligned} (1-p_1) < SR(W_M, p, t); \\ p_{V=1}(W_M, f_p = 0) > \frac{p_1}{\tau}, \end{aligned} \right\} \quad (4)$$

где $\tau = 2-5$ выбирается в зависимости от получаемых $p_{V=1}$ и, очевидно, отражает уменьшение дисбаланса между ошибками «1 → 0» и «0 → 1», присущими ПБ.

Если эти условия не выполнены, то выбирается следующий предиктор из списка доступных. Действительно, p_1 — это доля единиц в данном окне, а $(1-p_1)$ — доля нулей. Оптимальным ПБ в данном окне все биты последовательности предсказываются как единицы, если единицы более частые ($p_1 > 1/2$), или как нули, если нули более частые в данном окне. Поэтому первое условие означает, что предиктор f_p обеспечивает более высокую вероятность правильного предсказания до последнего наблюдаемого момента, чем ПБ, а вторая — что определенная часть предсказанных нулей (или единиц) является результатом работы алгоритма предсказания предиктора f_p , а не механического обращения нулей в единицы (или единиц в нули), как это делает ПБ.

Был выполнен ряд экспериментов для двоичных последовательностей, соответствующих знакам изменения значений временных рядов («1» — возрастание, «-1» — убывание), полученных как из различных литературных источников, посвященных исследованию изменения объема трафика в сетях, так и из измерений с помощью инструмента IBM QRadar NETflow числа так называемых «потоков событий» для трафика в сети, реализующей географически распределенный гибридный высокопроизводительный вычислительный кластер РАН в разное время суток [13] (рис. 2).

Для проверки предсказаний предиктором, реализующим технику XGB, а также доступными предикторами SGD и SVR (например класс `sklearn.svm.SVR` в Python) были получены массивы данных в окне размером $M = 100$ мин.

Исследовалась следующая полученная знаковая последовательность длиной 100, соответствующая знакам приращения трафика (алфавит $Z = \{-1, 1\}$ используется вместо двоичного $B = \{0, 1\}$ ввиду формата используемых программ):

```
1 1 1 1 -1 1 1 -1 1 -1 1 1 1 -1 1 1 1 -1 -1 -1 1 -1 1 1
1 -1 1 1 1 1 1 1 1 1 -1 1 1 1 1 1 -1 -1 -1 1 -1 -1 -1
1 -1 -1 -1 1 -1 -1 1 1 -1 1 1 1 1 1 -1 -1 1 -1 1 1 -1 1
```


- IEEE Commun. Surv. Tut., 2016. Vol. 18. Iss. 2. P. 1153–1176.
2. Rooba R., Vallimayil V. Semantic aware future page prediction based on domain // Int. J. Pure Appl. Math., 2018. Vol. 118. Iss. 9. P. 911–919.
 3. Wolpert D., Macready W. No free lunch theorems for optimization // IEEE T. Evolut. Comput., 1997. Vol. 1. Iss. 1. P. 67–83.
 4. Ryabko D., Ryabko B. Predicting the outcomes of every process for which an asymptotically accurate stationary predictor exists is impossible // IEEE Symposium (International) on Information Theory. — IEEE, 2015. P. 1204–1206.
 5. Merhav N., Feder M., Gutman M. Some properties of sequential predictors for binary Markov sources // IEEE T. Inform. Theory, 1993. Vol. 39. Iss. 3. P. 887–893.
 6. Feder M., Merhav N. Universal prediction // IEEE T. Inform. Theory, 1998. Vol. 44. Iss. 6. P. 2124–2147.
 7. Hodge V. J., Krishnan R., Austin J., Polak J., Jackson T. Short-term prediction of traffic flow using a binary neural network // Neural Comput. Appl., 2014. Vol. 25. P. 1639–1655. doi: 10.1007/s00521-014-1646-5.
 8. Even-Dar E., Kearns M., Mansour Y., Wortman J. Regret to the best vs. regret to the average // Mach. Learn., 2008. Vol. 72. P. 21–37.
 9. Bass D. Stochastic processes. — Cambridge, U.K.: Cambridge University Press, 2011. 392 p.
 10. Lavasani A., Eghlidos T. Practical next bit test for evaluating pseudorandom sequences // J. Sci. Technol., 2009. Vol. 16. Iss. 1. P. 19–33.
 11. Nobel A. Some stochastic properties of memoryless individual sequences // IEEE T. Inform. Theory, 2004. Vol. 50. P. 1497–1505.
 12. Chen T., Guestrin C. XGBoost: A scalable tree boosting system // arXiv:1603.02754v3 [cs.LG], 10 Jun 2016.
 13. Volovich K., Denisov S., Shabanov A., Malkovsky S. Aspects of the assessment of the quality of loading hybrid high-performance computing cluster // CEUR Workshop Proceedings, 2019. Vol. 2426. P. 7–11.
 14. Frenkel S. On a priori estimation of random sequences predictability // 6th Workshop on Computational Data Analysis and Numerical Methods Book of Abstracts. — Covilhã, Portugal, 2019. P. 109–111. <http://www.wcdanm-ubi19.uevora.pt/wp-content/uploads/2019/09/Book-of-Abstracts.pdf>.

Поступила в редакцию 15.04.20

JOINT ASSESSMENT OF DATA PREDICTABILITY AND QUALITY PREDICTORS

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 paper proposes and analyzes a new approach to the selection of predictors necessary for predicting future values in data sequences in a specific time period. Our goal is low-cost implemented techniques that ensure the selection of an acceptable predictor for the current prediction session, or the decision about the impossibility of making a reliable forecast if one finds that this section of the sequence does not have the predictability property. For this, the predictability of this sequence is defined as the maximum conditional probability of the correct prediction in the set of available predictors for a given set of observed values. The selection of predictors is performed by both the magnitude of the conditional probability estimation and the degree of difference between a specific predictor and a predictor that is optimal for predicting the next outcome of the Bernoulli trials sequence.

Keywords: random sequences prediction; predictors; data analysis

DOI: 10.14357/19922264200206

Acknowledgments

This research was partially supported by the Russian Foundation for Basic Research (projects 18-07-00669, 18-29-03100, and 18-07-01434).

References

1. Buczak, L., and E. Guven. 2016. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun. Surv. Tut.* 18(2):1153–1176.
2. Rooba, R., and V. Vallimayil.— 2018. Semantic aware future page prediction based on domain. *Int. J. Pure Appl. Math.* 118(9):911–919.
3. Wolpert, D., and W. Macready. 1997. No free lunch theorems for optimization. *IEEE T. Evolut. Comput.* 1(1):67–83.

4. Ryabko, D., and B. Ryabko. 2015. Predicting the outcomes of every process for which an asymptotically accurate stationary predictor exists is impossible. *IEEE Symposium (International) on Information Theory*. IEEE. 1204–1206.
5. Merhav, N., M. Feder, and M. Gutman. 1993. Some properties of sequential predictors for binary Markov sources. *IEEE T. Inform. Theory* 39(3):887–893.
6. Feder, M., and N. Merhav. 1998. Universal prediction. *IEEE T. Inform. Theory* 44(6):2124–2147.
7. Hodge, V., R. Krishnan, J. Austin, J. Polak, and T. Jackson. 2014. Short-term prediction of traffic flow using a binary neural network. *Neural Comput. Appl.* 25:1639–1655.
8. Even-Dar, E., M. Kearns, Y. Mansour, and J. Wortman. 2008. Regret to the best vs. regret to the average. *Mach. Learn.* 72:21–37.
9. Bass, D. 2011. *Stochastic processes*. Cambridge, U.K., Cambridge University Press. 392 p.
10. Lavasani, A., and T. Eghlidos. 2009. Bit test for evaluating pseudorandom sequences. *J. Sci. Technol.* 16(1):19–33.
11. Nobel, A. 2004. Some stochastic properties of memoryless individual sequences. *IEEE T. Inform. Theory* 50:1497–1505.
12. Chen, T., and C. Guestrin. 2016. XGboost: A scalable tree boosting system. *arXiv.org*. Available at: <https://arxiv.org/pdf/1603.02754.pdf> (accessed on April 15, 2020).
13. Volovich, K., S. Denisov, A. Shabanov, and S. Malkovsky. 2019. Aspects of the assessment of the quality of loading hybrid high-performance computing cluster. *CEUR Workshop Proceedings* 2426:7–11.
14. Frenkel, S. 2019. On *a priori* estimation of random sequences predictability. *6th Workshop on Computational Data Analysis and Numerical Methods Book of Abstracts*. Covilhã, Portugal. 109–111. Available at: <http://www.wcdanm-ubi19.uevora.pt/wp-content/uploads/2019/09/Book-of-Abstracts.pdf> (accessed June 22, 2020).

Received April 15, 2020

Contributors

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44–2 Vavilov Str., Moscow 119333, Russian Federation; 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

АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ РАЗМЕРОВ ЧАСТИЦ ЛУННОГО РЕГОЛИТА НА ОСНОВЕ МЕТОДА СТАТИСТИЧЕСКОЙ СИМУЛЯЦИИ ВЫБОРОК*

А. К. Горшенин¹, В. Ю. Королев²

Аннотация: Рассмотрена задача моделирования распределения размеров пылевых частиц лунного реголита. В качестве теоретической модели использованы конечные смеси логнормальных законов. Они позволяют учесть стохастический характер интенсивностей процессов дробления/спекания при формировании ансамблей пылевых частиц в результате различных воздействий (бомбардировка метеоритами, излучение). Разработан метод статистической аппроксимации неизвестных распределений на основе симуляции выборок, который демонстрирует высокое согласие с реальными данными, содержащимися в каталоге NASA (все 317 проб, доставленные миссиями «Луна-24» и «Аполлон-11, 12, 14–17»).

Ключевые слова: конечные логнормальные смеси; бутстреп; EM-алгоритм; статистические методы

DOI: 10.14357/19922264200207

1 Введение

В статье рассмотрена задача моделирования распределений размеров пылевых частиц лунного реголита, возникающих в результате различных воздействий, например при бомбардировках поверхности Луны метеоритами. При таких воздействиях развиваются как взрывные процессы разлета частиц с их дроблением, так и спекание частиц в экзотермических плазмохимических реакциях синтеза [1].

Изучение закономерностей, определяющих размеры частиц, образующих лунные и другие планетарные реголиты, имеет очень большое значение при планировании автоматических и пилотируемых миссий для изучения космических тел (Луны, астероидов, планет и их спутников). В условиях малой гравитации и отсутствия плотной атмосферы пылевые структуры приобретают качества, не типичные для земных условий, представляя собой облака заряженных частиц, обладающих высокими абразивными свойствами. Они осаждаются на элементах аппаратуры (например, солнечных батареях) и скафандрах, что быстро выводит их из строя.

Таким образом, решение рассматриваемой задачи может оказаться весьма полезным при подго-

товке новых космических миссий для повышения уровня безопасности и общей успешности подобных проектов.

Изначально в исследованиях на эту тему рассматривались процессы чистого дробления частиц. В работе [2] указано много случаев, в которых логарифмы размеров частиц (золотин в золотоносных россыпях, частиц горных пород при их дроблении и т. п.) имеют примерно нормальное распределение. На эту работу обратил внимание А. Н. Колмогоров, который предложил математическую модель процесса дробления частиц, аналитически объясняющую возникновение логнормального распределения размеров частиц при дроблении, а также содержания минералов в отдельных пробах [3]. Данная модель базируется на изучении изменения во времени числа частиц, размер которых не превосходит заданный порог.

Результат Колмогорова справедлив при довольно сильных предположениях, в частности для логнормальности распределения частиц при дроблении необходимо, чтобы скорость дробления была постоянной, т. е. не зависела от размеров самих частиц. Однако необходимо учитывать тот факт, что с уменьшением размеров частицы интенсивность ее соударений с другими частицами может изменяться, например уменьшаться в силу

* Работа выполнена при поддержке Российского научного фонда (проект 18-11-00155). Для обработки данных использовался гибридный высокопроизводительный вычислительный комплекс ЦКП «Информатика» ФИЦ ИУ РАН: <http://ckp.frccsc.ru>.

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, agorshenin@frccsc.ru

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

того, что вероятность столкновения в определенном смысле пропорциональна размерам частицы.

Известны эмпирические свидетельства того, что в некоторых случаях и логнормальная модель для распределения размера частиц при дроблении не адекватна. Так, в книге [4] отмечено, что логарифм функции плотности распределения для логарифма размера частицы в естественных запасах песка больше похож на гиперболу, чем на параболу. Это означает, что распределение размера не логнормально, а скорее имеет экспоненциально уменьшающиеся хвосты.

Последнее обстоятельство побудило некоторых исследователей обратить внимание на модели типа лог-несимметричного распределения Лапласа, так называемого двойного (двустороннего) Парето-логнормального распределения, а также лог-гауссовского//обратного гауссовского распределения. В частности, в работах [5, 6] рассмотрены модели формирования распределения размера дробящихся частиц, в основе которых лежит предположение о том, что частица при перемещении из одного места в другое может разделиться на несколько меньших частиц вследствие соударения или другого воздействия, что обуславливает случайность масс частиц после разделения. Не все частицы одновременно добираются до места назначения. Некоторые могут поменять направление и застрять надолго, тогда как другие могут, не задев других, пройти весь путь намного быстрее. Очевидно, состояние частицы зависит от того, на каком расстоянии от конечного пункта она находится. Это учитывается в упомянутых моделях Рида–Йоргенсена и Соренсена за счет случайности времени, в течение которого наблюдается частица. По сути, в работах [5, 6] в модель Колмогорова введена рандомизация.

Данные модели демонстрируют хорошее согласие с экспериментальными данными, отражающими размеры частиц в природных залежах. Однако они не учитывают отмеченное Колмогоровым обстоятельство: интенсивность процесса дробления остается постоянной. В работе [7] рассмотрена более общая модель, в которой данная интенсивность может быть переменной и даже случайной. Показано, что случайность интенсивности процесса дробления может привести к тому, что распределение размеров частиц будет иметь вид произвольной смеси логнормальных распределений.

В докладе [8] была отмечена возможность расширения выводов работы [7] на процессы дробления/спекания. Эти результаты используются в данной статье в качестве теоретической базы

для создания метода статистической обработки всех образцов лунного реголита, доставленных миссиями «Аполлон-11, 12, 14–17» и «Луна 24». Данные 317 проб взяты из каталога NASA [9], в котором размеры частиц представлены в так называемой ϕ -шкале [10], а для характеристики эмпирического распределения размеров частиц использованы величины долей частиц того или иного размера в образце, полученные при просеивании. Для всех представленных данных будет продемонстрировано высокое соответствие модели типа смеси конечных логнормальных законов с представленными в каталоге NASA распределениями размеров ансамблей пылевых частиц, возникающих в лунном реголите.

2 Конечные логнормальные смеси как аппроксимации для эмпирических распределений размеров частиц лунного реголита

В соответствии с результатами работ [7, 8] в качестве модели распределения логарифма размера частиц следует рассматривать асимптотическую аппроксимацию вида, вообще говоря, произвольной смеси логнормальных законов.

На практике разумно использовать дискретную аппроксимацию для смешивающего закона, т.е. изначально предполагать, что распределение логарифмов размеров частиц является конечной смесью нормальных законов, что позволяет искать оценки параметров смеси (веса нормальных компонент, их математические ожидания и дисперсии) с помощью хорошо известных статистических процедур, традиционно используемых для этих целей, например с использованием EM (expectation-maximization) алгоритма или его модификаций. Подобные модели допускают удобную интерпретацию, связанную с описанием типичных видов частиц (породы или конкреций) и/или типов плазмохимических реакций, влияющих на формирование реголита.

Итак, в качестве модели распределения размеров частиц в ϕ -шкале (т.е., фактически, логарифмов исходных размеров) будем использовать конечную смесь нормальных законов вида ($k \in \mathbb{N}$, $a_i \in \mathbb{R}$, $\sigma_i > 0$, $p_i \geq 0$, $i = 1, \dots, k$, $p_1 + \dots + p_k = 1$):

$$\mathbb{P}(Z < x) = \sum_{i=1}^k p_i \Phi\left(\frac{\log x - a_i}{\sigma_i}\right). \quad (1)$$

3 Метод на основе статистической симуляции выборок

В каталогах NASA [9] данные представлены в табличной форме в виде пар «размер частицы — доля (в процентах) частиц такого размера в просеиваемых образцах». Доступны сведения только о нескольких (как правило, не более десяти) точках роста (выбранных, вообще говоря, бессистемно) эмпирической функции распределения, но не о ее поведении между ними. В работе [9] описано применение интерполяции Стайнмана [11] для их соединения, благодаря которому с использованием макросов пакета Excel возможно построение соответствующих гистограмм.

В данной статье в аналогичных целях применяются кусочные кубические полиномы Эрмита (см., например, [12]) — они позволили получить наиболее близкие кривые по сравнению с представленными в каталоге NASA. Для симуляции выборок была апробирована и кусочно-линейная интерполяция. Итоговые результаты в этом случае получаются похожими, но менее наглядными, поэтому для сравнения в данной статье не приводятся.

Предложенная аппроксимация полиномами Эрмита позволила получить непрерывную эмпирическую функцию распределения (ECDF, empirical cumulative distribution function), а значит, стало возможным использовать метод обратных функций для генерации тестовых выборок. Объем выборки для оценивания параметров составлял 10 000 наблюдений; кроме того, для проверки соответствия приближающей смеси и исходной эмпирической функции распределения генерировалась еще одна независимая выборка объемом 2500 наблюдения (для проверки гипотез с помощью критерия согласия Колмогорова). Данная процедура в целом близка к такому статистическому подходу, как бут-

стреп. Ее описание в виде псевдокода приведено в алгоритме 1.

Полученные тестовые выборки используются для получения оценок максимального правдоподобия параметров аппроксимирующего смешанного распределения (1) с помощью EM-алгоритма для нормальных законов [13]. Примеры применения данной бутстреп-процедуры к реальным пробам лунного реголита представлены на рис. 1 и 2.

На графиках слева представлены исходные данные из таблиц каталога NASA (1), полученные в процессе просеивания, их интерполяция с помощью полиномов Эрмита (2) и аппроксимирующая смесь (3). Дополнительная вставка на рис. 1 демонстрирует увеличенную область [5, 10]. Видно, что обе кривые практически всюду совпадают. Такая ситуация повторяется для абсолютного большинства анализируемых проб. Отметим, что для числа компонент k в формуле (1) проверялись разные значения. Эмпирически было установлено, что необходимый баланс между качеством аппроксимации и вычислительной сложностью достигается при значении $k = 4$, которое и использовано в данной статье при обработке всех 317 выборок.

На графиках справа на рис. 1 и 2 приведены гистограммы для имитационных выборок и просеянных данных, построенных по ECDF, для конкретной пробы, а также их приближение плотностью смешанного распределения. Необходимо отметить, что точность аппроксимации определялась именно по сравнению с эмпирической функцией распределения, а данные графики служат лишь для более подробной иллюстрации работы метода.

Очевидно, что как функции распределения, так и гистограммы приближаются визуально очень хорошо, даже с учетом различных особенностей в них. Ясно, что форма распределений в каждом из случаев далека от стандартного гауссовского вида.

На рис. 3 приведены результаты проверки с помощью критерия однородности Колмогоро-

Алгоритм 1 Имитационное моделирование выборок для аппроксимации и статистического теста

```
function [SAMPLE, TESTSAMPLE] ← GENERATINGSAMPLES (ECDF)
    // SampleSize — объем выборки для оценивания параметров
    // TestSampleSize — объем выборки для критерия Колмогорова
    // Случайные векторы для метода обратных функций
    r ← RAND(SampleSize);
    rTest ← RAND(TestSampleSize);
    // Имитационное моделирование
    for i = 1:SampleSize do
        Sample(i) ← FSOLVE (ECDF, r(i)); // Метод обратных функций
    if i ≤ TestSampleSize then
        TestSample(i) ← FSOLVE (ECDF, rTest(i));
```

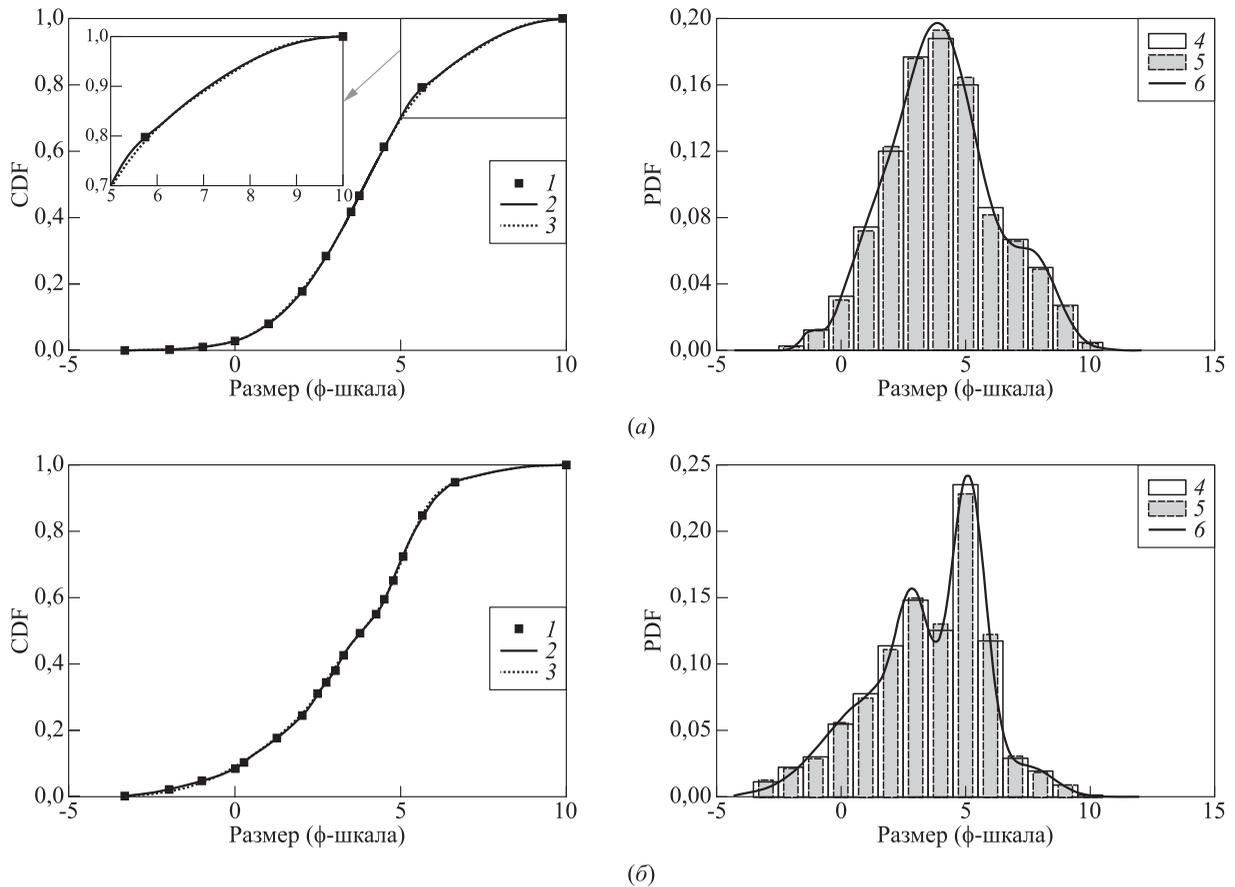


Рис. 1 Пробы лунного грунта 12042,24 (миссия «Аполлон-12») (а) и 14156,22 (миссия «Аполлон-14») (б): 1 — исходные данные из таблиц каталога NASA; 2 — интерполяция с помощью полиномов Эрмита; 3 — аппроксимирующая смесь; 4 — гистограммы для имитационных выборок; 5 — гистограммы для просеянных данных по ECDF; 6 — приближение плотностью смешанного распределения

ва и дополнительно симулированных выборок. Для большей наглядности на графиках обозначены стандартные критические уровни 0,01 и 0,05. Первый из них превышен P -значениями для 84,5% выборок (268 из 317), а второй — для 70,7% наборов (224 из 317), т. е. для абсолютного большинства проб с помощью бутстреп-метода получены достаточно хорошие результаты аппроксимации.

Отметим, что описанная процедура обработки данных была реализована на языке программирования MATLAB. Для расчетов использовались ресурсы гибридного высокопроизводительного вычислительного кластера архитектуры Intel x86_64: сервер Huawei XH 622 V3 (два процессора Intel Xeon E5-2683V4 с тактовой частотой 2,1 ГГц (16 ядер), 512 ГБ оперативной памяти и 2 видеокарты NVIDIA Tesla K80). Это позволило повысить скорость вычислений в среднем в 3,8 раза по сравнению со стандартными настольными решениями.

4 Кластеризация параметров аппроксимирующих смесей

В данном разделе приводятся результаты анализа параметров сразу всех аппроксимирующих конечных нормальных смесей. Кроме того, используется тривиальное обратное преобразование для перехода от ϕ -шкалы для размеров к метрической (в микрометрах). На рис. 4 иллюстрируется взаимозависимость математического ожидания и среднего квадратического отклонения для ϕ -шкалы и стандартных единиц измерения. Верхние графики на рис. 4, а и 4, б демонстрируют вид этой зависимости, при этом размер и интенсивность цвета точек соответствуют их весам (см. формулу (1) и цветовую шкалу вверх). Эти рисунки существенно уточняют предложенную в статье [14] линейную аппроксимацию для данной зависимости, которая, как видно из рис. 4, оказывается чрезмерным загромождением.

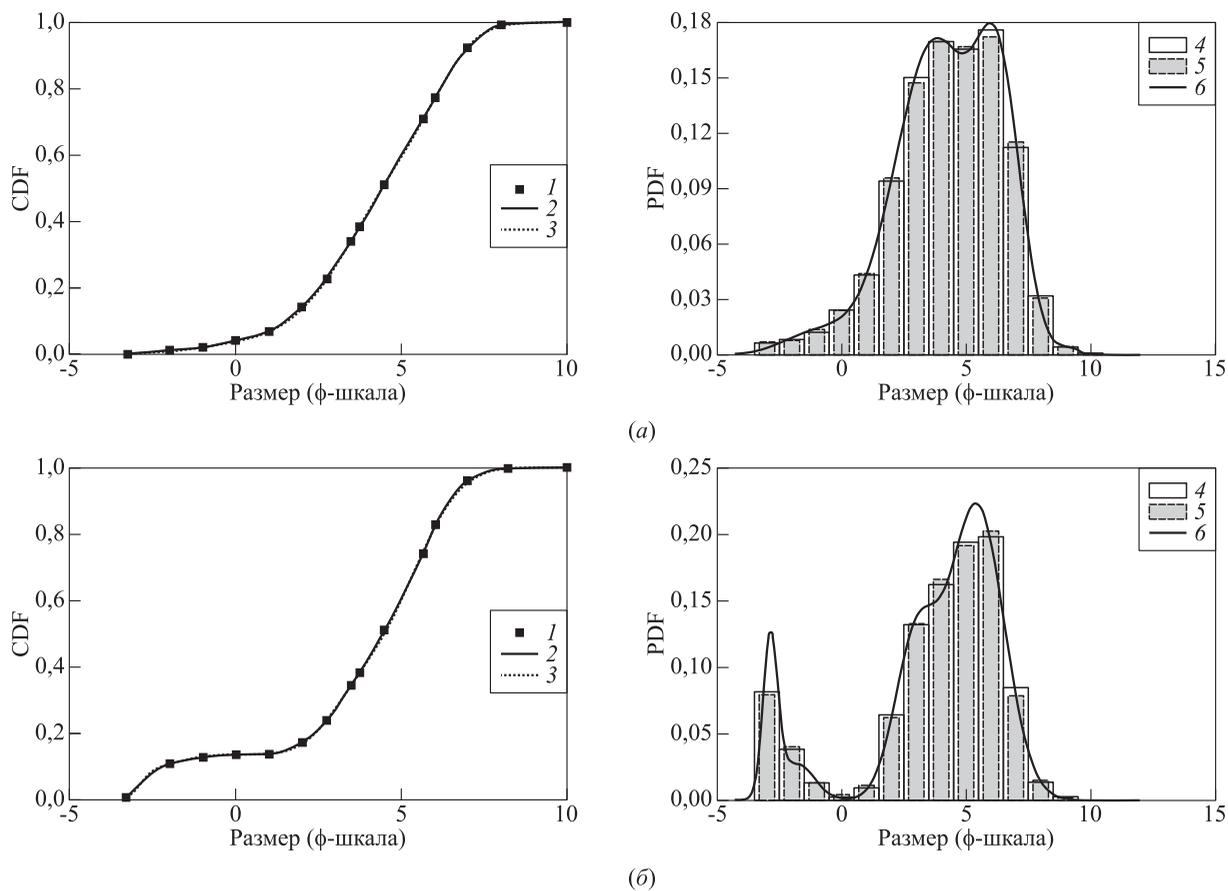


Рис. 2 Пробы лунного грунта 72141,15 (а) и 74221,82 (б) (миссия «Аполлон-17»): 1 — исходные данные из таблиц каталога NASA; 2 — интерполяция с помощью полиномов Эрмита; 3 — аппроксимирующая смесь; 4 — гистограммы для имитационных выборок; 5 — гистограммы для просеянных данных по ECDF; 6 — приближение плотностью смешанного распределения

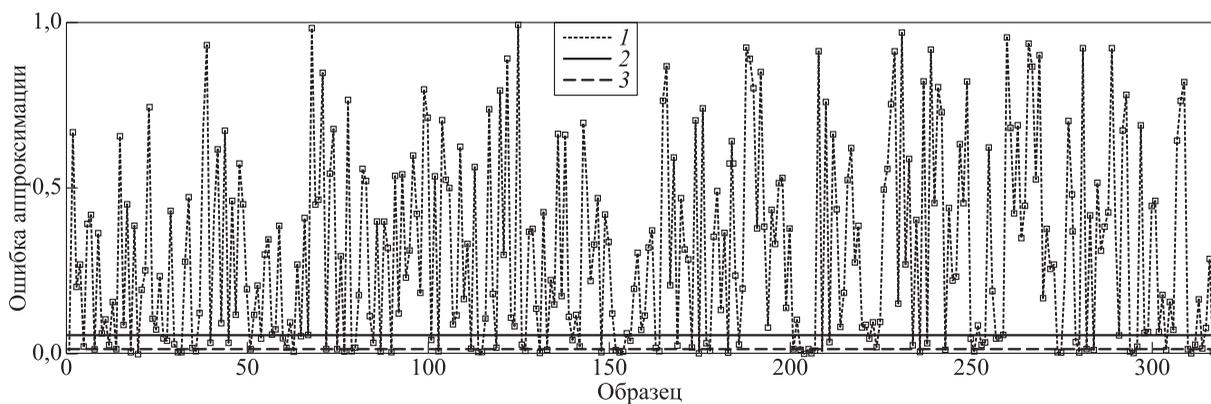


Рис. 3 Ошибки аппроксимации (критерий Колмогорова): 1 — P -значения; 2 и 3 — критические уровни $\alpha = 0,05$ и $0,01$ соответственно

Два нижних графика на каждом из рис. 4, а и 4, б демонстрируют разбиение параметрического пространства на 5 групп методами k -медоид (см. рис. 4, а) и нечеткой кластеризации s -средних (см. рис. 4, б) (в этом случае в качестве окончательного решения выбирается кластер, для которого дости-

гается максимальное значение величины вероятности членства для данного элемента среди всех возможных. Очевидно, что решения обоих методов в каждом из случаев оказываются достаточно близкими. Представленные на рис. 4 данные позволяют выделить некоторые типичные кластеры, которые

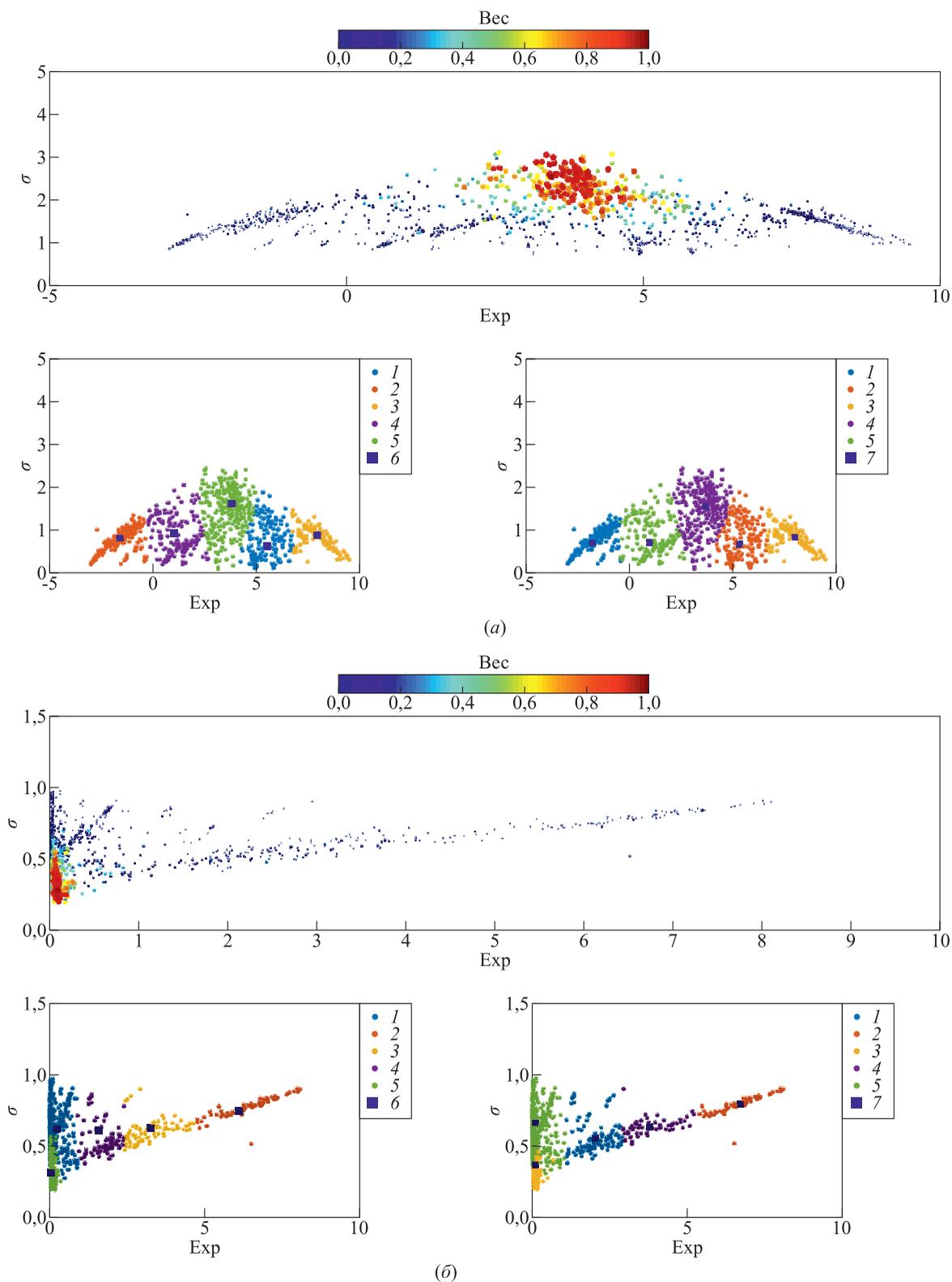


Рис. 4 Кластеризация параметров аппроксимирующих смесей: (а) ϕ -шкала; (б) метрическая шкала (мкм); 1–5 — кластеры; 6 — медиоиды кластеров; 7 — центры кластеров по методу c -средних

могут быть использованы, например, для соотнесения с химическим составом проб или иными характеристиками реголита. На рис. 4 нумерация кластеров произведена в произвольном порядке и не связана с какими-либо их характеристиками, например весом.

5 Заключение

В работе рассмотрена методология аппроксимации распределений размеров частиц лунного реголита с помощью логнормальных смешанных моделей. Продемонстрировано высокое согласие получаемых таким образом вероятностных распределений и данных просеивания образцов лунного реголита. Дальнейшие исследования будут ориентированы на отказ от необходимости интерполяции точек — исходных данных в каталоге NASA. При этом их приближение также возможно функциями, имеющими вид конечной смеси логнормальных законов, однако подобная процедура не требует статистической симуляции выборок. Это может повысить точность аппроксимации, а также уменьшить затрачиваемое на расчеты время.

Литература

1. Попель С. И., Голубь А. П., Захаров А. В. и др. Формирование плазменно-пылевых облаков при ударе метеороида о поверхность Луны // Письма в Журнал экспериментальной и теоретической физики, 2018. Т. 108. Вып. 6. С. 379–387.
2. Разумовский Н. К. Характер распределения содержания металлов в рудных месторождениях // Докл. Акад. наук СССР, 1940. Т. 28. Вып. 9. С. 815–817.
3. Колмогоров А. Н. О логарифмически нормальном законе распределения размеров частиц при дроблении // Докл. Акад. наук СССР, 1941. Т. 31. Вып. 2. С. 99–101.
4. Bagnold R. A. The physics of blown sand and desert dunes. — London: Methuen, 1954. 265 p.
5. Reed W. J., Jorgensen M. The double Pareto-Lognormal distribution — a new parametric model for size distribution // Commun. Stat. Theor. M., 2004. Vol. 33. Iss. 8. P. 1733–1753.
6. Sørensen M. On the size distribution of sand. — Copenhagen: Department of Applied Mathematics and Statistics, University of Copenhagen, 2006. Working paper. 11 p.
7. Королёв В. Ю. О распределении размеров частиц при дроблении // Информатика и её применения, 2009. Т. 3. Вып. 3. С. 60–68.
8. Korolev V. Yu., Skvortsova N. N. Size distribution of dust grains created in plasma-chemical reactions initiated by the microwave radiation of a gyrotron in regolith // Complex Systems of Charged Particles and Their Interactions with Electromagnetic Radiation: 17th Workshop (International) Proceedings. — Moscow: Prokhorov General Physics Institute of the Russian Academy of Sciences, 2019. P. 25.
9. Graf J. C. Lunar soils grain size catalog. — NASA, 1993. NASA Reference Publication 1265. 484 p. <https://ntrs.nasa.gov/search.jsp?R=19930012474/>.
10. Donoghue J. F. Phi scale // Encyclopedia of estuaries / Ed. M. J. Kennish. — Encyclopedia of Earth sciences ser. — Dordrecht: Springer, 2016. 789 p. doi: 10.1007/978-94-017-8801-4.
11. Stineman R. W. A consistently well behaved method of interpolation // Creative Comput., 1980. Vol. 6. Iss. 7. P. 54–57.
12. Fritsch F. N., Carlson R. E. Monotone piecewise cubic interpolation // SIAM J. Numer. Anal., 1980. Vol. 17. P. 238–246.
13. Королёв В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 512 с.
14. Слюта Е. Н. Физико-механические свойства лунного грунта (обзор) // Астрономический вестник, 2014. Т. 48. Вып. 5. С. 358–382.

Поступила в редакцию 15.04.20

APPROXIMATION OF PARTICLE SIZE DISTRIBUTIONS OF LUNAR REGOLITH BASED ON THE RESAMPLING

A. K. Gorshenin^{1,2} and V. Yu. Korolev^{1,2}

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

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

Abstract: The paper considers the problem of modeling the size distribution of dust particles of lunar regolith based on approximating with finite lognormal mixtures. These models make it possible to take into account the stochastic nature of the intensities of splitting/baking processes during the formation of ensembles of dust particles as a result

of various influences (bombardment by meteorites, radiation). A method for statistical approximation of unknown distributions based on simulation of samples was developed. It is demonstrated that the model distributions fit very well to the real observations of lunar regolith gathered by missions “Apollo 11, 12, 14–17” and “Luna-24” that had been presented in the NASA’s grain size catalog (317 samples).

Keywords: finite lognormal mixtures; bootstrap; EM algorithm; statistical methods

DOI: 10.14357/19922264200207

Acknowledgments

The research is supported by the Russian Science Foundation (project 18-11-00155). The calculations were performed using Hybrid high-performance computing cluster of FRC CSC RAS (<http://ckp.frccsc.ru/>).

References

1. Popel, S. I., A. P. Golub, A. V. Zakharov, *et al.* 2018. Formation of dusty plasma clouds at meteoroid impact on the surface of the Moon. *JETP Lett.* 108(6):356–363.
2. Razumovskii, N. K. 1940. Kharakter raspredeleniya sodержaniya metallov v rudnykh mestorozhdeniyakh [The nature of the distribution of metal content in ore deposits]. *Dokl. Akad. Nauk SSSR* 28(9):815–817.
3. Kolmogorov, A. N. 1941. O logarifmicheski normal’nom zakone raspredeleniya razmerov chastits pri droblenii [On the log-normal law of particle size distribution during fracturing]. *Dokl. Akad. Nauk SSSR* 31(2):99–101.
4. Bagnold, R. A. 1954. *The physics of blown sand and desert dunes*. London: Methuen. 265 p.
5. Reed, W. J., and M. Jorgensen. 2004. The double Pareto-Lognormal distribution — a new parametric model for size distribution. *Commun. Stat. Theor. M.* 33(8):1733–1753.
6. Sørensen, M. 2006. On the size distribution of sand. Copenhagen: Department of Applied Mathematics and Statistics, University of Copenhagen. Working paper. 11 p.
7. Korolev, V. Yu. 2009. O raspredelenii razmerov chastits pri droblenii [On the distribution of particle size under fracturing]. *Informatika i ee Primeneniya — Inform. Appl.* 3(3):60–68.
8. Korolev, V. Yu., and N. N. Skvortsova. 2019. Size distribution of dust grains created in plasma-chemical reactions initiated by the microwave radiation of a gyrotron in regolith. *17th Workshop (International) “Complex Systems of Charged Particles and Their Interactions with Electromagnetic Radiation” Proceedings*. Moscow: Prokhorov General Physics Institute of the Russian Academy of Sciences. 25.
9. Graf, J. C. 1993. Lunar soils grain size catalog. NASA Reference Publication 1265. Available at: <https://ntrs.nasa.gov/search.jsp?R=19930012474/> (accessed May 25, 2020).
10. Donoghue, J. F. 2016. Phi scale. *Encyclopedia of estuaries*. Ed. M. J. Kennish. Encyclopedia of Earth sciences ser. Dordrecht: Springer. 789 p. doi: 10.1007/978-94-017-8801-4.
11. Stineman, R. W. 1980. A consistently well behaved method of interpolation. *Creative Comput.* 6(7):54–57.
12. Fritsch, F. N., and R. E. Carlson. 1980. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.* 17:238–246.
13. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil’nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Izd-vo Moskovskogo un-ta. 512 p.
14. Slyuta, E. N. 2014. Physical and mechanical properties of the lunar soil (a review). *Solar Syst. Res.* 48(5):330–353.

Received April 15, 2020

Contributors

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

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

ВВЕДЕНИЕ ОТНОШЕНИЯ ПОРЯДКА НА МНОЖЕСТВЕ ПАРАМЕТРОВ АППРОКСИМИРУЮЩИХ МОДЕЛЕЙ*

А. В. Грабовой¹, О. Ю. Бахтеев², В. В. Стрижов³

Аннотация: Исследуется проблема введения отношения порядка на множестве параметров сложных аппроксимирующих моделей. В качестве параметрических моделей исследуются линейные и нейросетевые модели. Порядок на множестве параметров задается при помощи ковариационной матрицы градиентов функции ошибки по параметрам модели. Предлагается использовать заданный порядок для фиксации параметров модели во время решения оптимизационной задачи. Предполагается, что после небольшого числа итераций алгоритма оптимизации некоторые параметры модели можно зафиксировать без значимой потери качества модели. Это позволит существенно понизить размерность задачи оптимизации. В вычислительном эксперименте сравниваются модели, в которых параметры фиксируются в соответствии с предложенным порядком, с моделями, в которых параметры фиксируются произвольным образом.

Ключевые слова: аппроксимация выборки; линейная модель; нейросеть; выбор модели; функция ошибки

DOI: 10.14357/19922264200208

1 Введение

Оптимизация глубоких нейронных сетей относится к задачам высокой сложности и требует больших вычислительных мощностей [1]. При этом алгоритм оптимизации сходится по большинству параметров сети уже после небольшого числа итераций [2]. Своевременное определение начала сходимости параметров позволит существенно снизить вычислительные затраты на обучение моделей с большим числом параметров.

Примером задания порядка на множестве параметров служит l_1 -регуляризация [3] и регуляризация ElasticNet [4] для линейных моделей. Порядок, заданный на множестве значений коэффициентов регуляризации, индуцирует порядок на множестве признаков описаний и указывает на важность признаков. В случае нейросетей для регуляризации параметров используется метод исключения параметров [5, 6]. Данный метод также задает порядок на множестве параметров модели.

В [7] вводится понятие релевантности параметров нейросетевой модели. Оно задает естественный порядок на множестве параметров модели от наименее релевантных до наиболее релевантных. В [8] предложен метод определения релевантности параметров аппроксимирующих моделей при помощи метода Белсли. Релевантность параметров в ра-

боте [8] определяется на основе ковариационной матрицы параметров модели.

В данной работе предлагается метод введения отношения порядка на множестве параметров сложных параметрических моделей, таких как нейросеть. Рассматривается порядок, заданный при помощи ковариационной матрицы градиентов функции ошибки по параметрам модели [9]. В работе [2] предложен итерационный метод для поиска ковариационной матрицы градиентов. Данный итерационный метод интегрируется в градиентный метод оптимизации Adam [10].

Множество параметров упорядочивается по возрастанию дисперсии: от параметра с минимальной дисперсией до параметра с максимальной дисперсией градиента функции ошибки по соответствующему параметру модели. Предполагается, что малая дисперсия градиента указывает на то, что соответствующий параметр можно зафиксировать.

Для задания порядка на множестве параметров при помощи ковариационной матрицы вводится предположение о том, что фиксация параметров происходит в момент, когда все параметры модели находятся в некоторой окрестности локального минимума функции ошибки. Данное условие накладывается для корректного использования итерационного метода поиска ковариационной матрицы градиентов.

* Работа выполнена при поддержке РФФИ (проекты 19-07-01155 и 19-07-00875) и НТИ (проект 13/1251/2018).

¹ Московский физико-технический институт, grabovoy.av@phystech.edu

² Московский физико-технический институт, bakhteev@phystech.edu

³ Вычислительный центр имени А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский физико-технический институт, strijov@phystech.edu

Заданный порядок на множестве параметров модели используется для фиксации тех параметров модели, которые оказываются предстоящими с точки зрения заданного порядка. Сначала фиксируются те параметры, которые имеют минимальную дисперсию градиента в окрестности локального минимума функции ошибки.

Для анализа свойств предложенного метода задания порядка на множестве параметров проводится вычислительный эксперимент. В качестве моделей рассматриваются модели различной структурной сложности: линейные модели, нейросетевые модели. Предложенный метод задания порядка сравнивается с методом, в котором порядок задан произвольным образом.

2 Постановка задачи

Задана выборка:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m, \quad \mathbf{x}_i \in \mathbb{X} = \mathbb{R}^n, \quad y_i \in \mathbb{Y},$$

где n — размерность признакового пространства; m — число объектов в выборке. Пространство ответов

$$\mathbb{Y} = \begin{cases} \mathbb{R} & \text{в случае задачи регрессии;} \\ \{1, \dots, K\} & \text{в случае задачи классификации,} \end{cases}$$

где K — число классов.

Задано семейство моделей, параметрических функций с наперед заданной структурой:

$$\left. \begin{aligned} \mathfrak{F} &= \{f(\mathbf{w}) : \mathbb{X} \rightarrow \mathbb{Y} \mid \mathbf{w} \in \mathbb{R}^p\}; \\ \mathbf{h}(\mathbf{w}, \mathbf{x}) &= \mathbf{W}_1 \sigma(\mathbf{W}_2 \sigma(\dots \sigma(\mathbf{W}_r \mathbf{x}) \dots)); \\ f_{\text{cl}}(\mathbf{w}, \mathbf{x}) &= \arg \max_{j \in \{1, \dots, K\}} \text{softmax}(\mathbf{h}(\mathbf{w}, \mathbf{x}))_j; \\ f_{\text{reg}}(\mathbf{w}, \mathbf{x}) &= \mathbf{h}(\mathbf{w}, \mathbf{x}), \end{aligned} \right\} (1)$$

где p — размерность пространства параметров; r — число слоев нейросети; $\mathbf{w} = \text{vec}[\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_r]$; σ — функция активации. В случае задачи регрессии структура модели имеет вид f_{reg} , а в случае классификации имеет вид f_{cl} . Задана функция потерь:

$$\left. \begin{aligned} \mathcal{L}(\mathbf{w}, \mathcal{D}) &= \frac{1}{m} \sum_{i=1}^m l(\mathbf{x}_i, y_i, \mathbf{w}); \\ l_{\text{reg}}(\mathbf{x}, y, \mathbf{w}) &= (y - f(\mathbf{w}, \mathbf{x}))^2; \\ l_{\text{cl}}(\mathbf{x}, y, \mathbf{w}) &= \\ &= - \sum_{j=1}^K \left([y = j] \ln \text{softmax}(\mathbf{h}(\mathbf{w}, \mathbf{x}))_j \right), \end{aligned} \right\} (2)$$

где l_{reg} — это функция ошибки на одном элементе для задачи регрессии; l_{cl} — для задачи классифика-

ции. Оптимальный вектор параметров $\hat{\mathbf{w}}$ получим минимизацией функции потерь:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^p} \mathcal{L}(\mathbf{w}, \mathcal{D}).$$

2.1 Задание отношения порядка на множестве параметров

Для поиска оптимальных параметров модели используется градиентный метод оптимизации:

$$\left. \begin{aligned} \mathbf{w}_t &= \mathbf{w}_{t-1} + \Delta \mathbf{w}(\mathbf{g}_{S,t}, \mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots); \\ \mathbf{g}_{S,t} &= \frac{\partial \mathcal{L}(\mathbf{w}_t, \mathbf{X}_S, \mathbf{Y}_S)}{\partial \mathbf{w}}, \end{aligned} \right\} (3)$$

где t — номер итерации; $\mathbf{g}_{S,t}$ — значение градиента на подвыборке размера S ; $\Delta \mathbf{w}$ — приращение вектора параметров.

Порядок на множестве параметров модели задается при помощи ковариационной матрицы \mathbf{C} градиентов функции ошибки \mathcal{L} по параметрам модели \mathbf{w} . Для вычисления ковариационной матрицы \mathbf{C} используется итерационная формула [2], которая вычисляется на каждой итерации (3) градиентного метода оптимизации параметров:

$$\mathbf{C}_t = (1 - \kappa_t) \mathbf{C}_{t-1} + \kappa_t (\mathbf{g}_{1,t} - \mathbf{g}_{S,t}) (\mathbf{g}_{1,t} - \mathbf{g}_{S,t})^T,$$

где t — номер итерации; $\mathbf{g}_{S,t}$ — значение градиента на подвыборке размера S ; $\mathbf{g}_{1,t}$ — значение градиента на первом элементе подвыборки; $\kappa_t = 1/t$ — параметр сглаживания; \mathbf{C}_0 инициализируются из равномерного распределения.

Пусть известно t_0 — число итераций, после которого все параметры находятся в некоторой локальной окрестности минимума, тогда, как показано в работе [2], матрица \mathbf{C}_{t_0} аппроксимирует истинную ковариационную матрицу \mathbf{C} . Ковариационная матрица \mathbf{C}_{t_0} используется для упорядочения параметров модели \mathbf{w}_{t_0} .

Пусть \mathcal{I} — упорядоченный вектор индексов $[1, 2, \dots, p]$. Обозначим $\mathcal{I}_{\mathbf{w}_{t_0}}$ вектор индексов, порядок которого задан при помощи ковариационной матрицы \mathbf{C}_{t_0} .

Например, если ковариационная матрица \mathbf{C}_{t_0} имеет вид:

$$\begin{bmatrix} 0,3 & 0 & 0 \\ 0 & 0,2 & 0 \\ 0 & 0 & 0,25 \end{bmatrix},$$

то вектор индексов $\mathcal{I}_{\mathbf{w}_{t_0}} = [3, 1, 2]$.

2.2 Фиксация параметров

Для фиксации параметров \mathbf{w}_{t_0} при помощи вектора индексов $\mathcal{I}_{\mathbf{w}_{t_0}}$ используется бинарный вектор $\alpha(k)$:

$$\alpha_i(k) = \begin{cases} 1, & \text{если } \mathcal{I}_{\mathbf{w}_{t_0}}[j] \leq k; \\ 0 & \text{иначе,} \end{cases} \quad (4)$$

где k — число фиксирующих параметров.

Учитывая (4), уравнение (3) приводится к виду:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \alpha(k) \cdot \Delta \mathbf{w}(\mathbf{g}_{S,t}, \mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots). \quad (5)$$

После умножения на бинарный вектор α часть параметров не оптимизируется, что приводит к фиксации параметров.

3 Вычислительный эксперимент

Для анализа результатов, полученных предложенным алгоритмом, проводится вычислительный эксперимент. В качестве данных используются синтетические и реальные данные, которые описаны в таблице. Выборки MNIST [11] и Boston Housing [12] рассматриваются в качестве реальных данных, для которых решается задача классификации и регрессии соответственно. Синтетические выборки задаются следующим образом:

$$\begin{aligned} \mathcal{D}_{\text{reg}} &= \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \\ & \quad y_i \sim \mathcal{N}(\mathbf{w}^\top \mathbf{x}_i, \mathbf{I}_n), \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)\}; \\ \mathcal{D}_{\text{cl}} &= \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \\ & \quad y_i \sim \text{Be}(\mathbf{w}^\top \mathbf{x}_i), \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)\}. \end{aligned}$$

В качестве аппроксимирующих моделей рассматриваются линейные и нейросетевые модели (1). В качестве функции ошибки для задачи регрессии рассматривается MSELoss, а для задачи классификации — CrossEntropyLoss (2).

Предварительно для каждой модели и выборки определяется число t_0 — номер итерации, после которой все параметры модели находятся в некоторой окрестности локального минимума. Параметр t_0 устанавливается экспериментальным путем

для каждой модели и выборки отдельно из условия, что качество модели меняется незначительно при числе итераций $t > t_0$.

После t_0 шагов алгоритма оптимизации часть параметров модели фиксируется в соответствии с формулами (4) и (5). Результат работы получается усреднением по 25 независимым запускам оптимизации модели. Значение функции ошибки \mathcal{L} усредняется по разным запускам алгоритма оптимизации. В ходе эксперимента проводится анализ вектора α , который также усредняется по разным запускам алгоритма оптимизации. Усредненное значение бинарного вектора α обозначим $\hat{\alpha}$.

Выборка Synthetic 1. Эксперимент проводился на синтетически построенных данных. В качестве модели использовалась двухслойная нейросеть — перцептрон. На рис. 1, *a* показаны графики зависимости функции потерь \mathcal{L} от числа фиксируемых параметров. В случае фиксации параметров предложенным методом функция потерь \mathcal{L} растет медленней, чем в случае фиксации параметров произвольным образом.

На рис. 2 показана зависимость векторов $\hat{\alpha}(k)$ от числа фиксируемых параметров. Каждый столбец соответствует одному вектору $\hat{\alpha}(k)$. На рис. 2, *a* и 2, *b* видно, что $\hat{\alpha}(k)$ имеет большое число компонент вектора, близких к 1. Так как $\hat{\alpha}(k)$ является усреднением вектора с компонентами 0 или 1, то предложенный порядок задает некоторый устойчивый порядок на множестве параметров модели. На рис. 2, *b* и 2, *g* видно, что в случае произвольной фиксации параметров компоненты вектора $\hat{\alpha}(k)$ имеют одинаковые значения; следовательно, никакого порядка на множестве параметров нет.

Выборка Boston Housing. Эксперимент проводился на реальных данных. На рис. 1, *b* показаны графики зависимости функции потерь \mathcal{L} от числа фиксируемых параметров. В случае фиксации параметров предложенным методом функция потерь \mathcal{L} растет так же, как и в случае фиксации параметров произвольным образом. Данный результат следует из того, что нейросеть оказалась избыточно слож-

Описание выборок, используемых в эксперименте

Выборка \mathcal{D}	Тип	Число признаков n	Модель	Число параметров p
Boston Housing	Регрессия	13	Нейросеть	301
MNIST	Классификация	784	Нейросеть	7960
Synthetic 3	Регрессия	200	Линейная	200
Synthetic 2	Классификация	200	Линейная	200
Synthetic 1	Регрессия	200	Нейросеть	4041

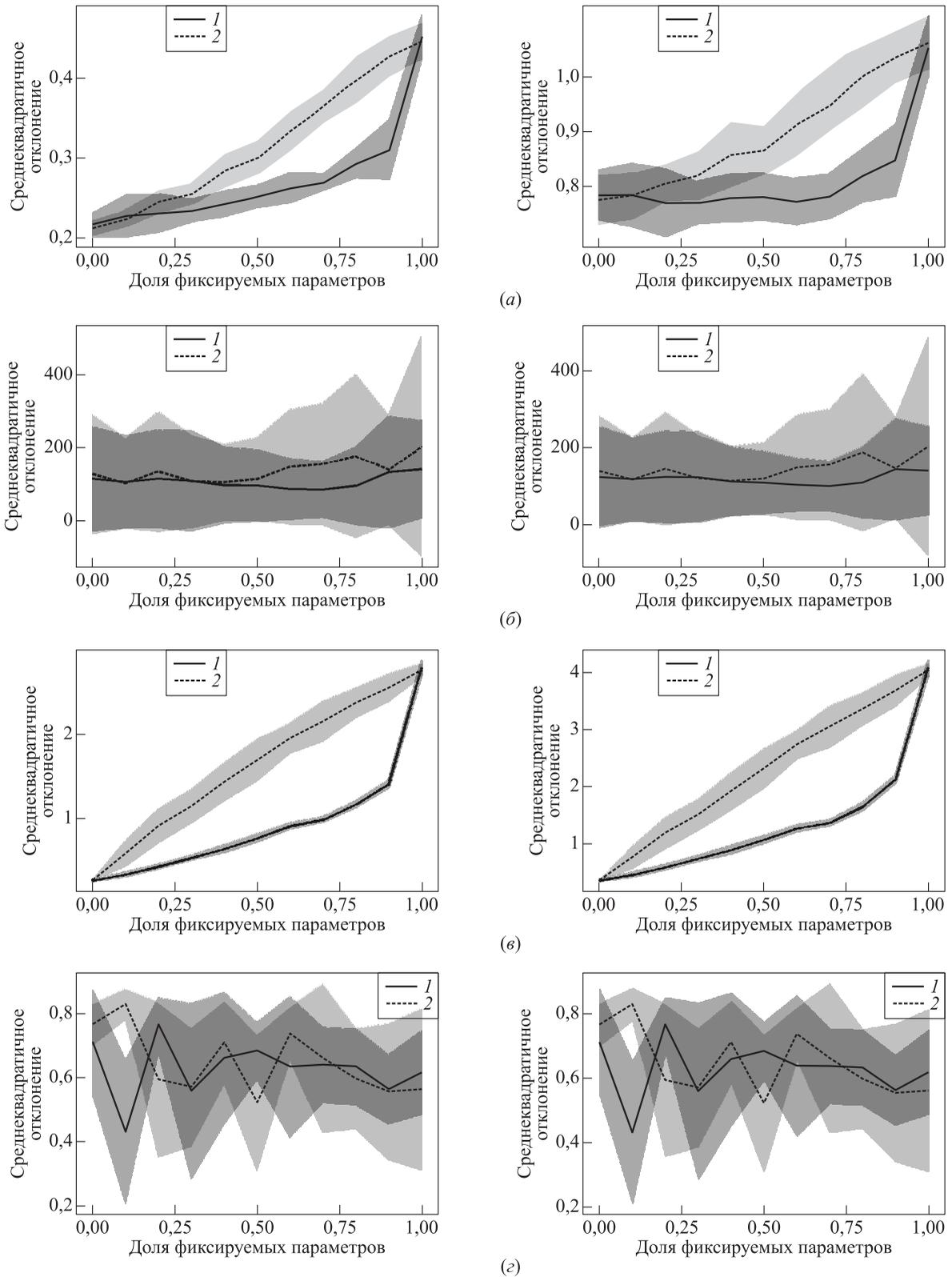


Рис. 1 Зависимость качества модели от числа зафиксированных параметров: (а) Synthetic 1; (б) Boston Housing; (в) Synthetic 3; (г) MNIST; левый столбец — на обучающей выборке; правый столбец — на тестовой выборке; 1 — предложенный метод; 2 — произвольная фиксация

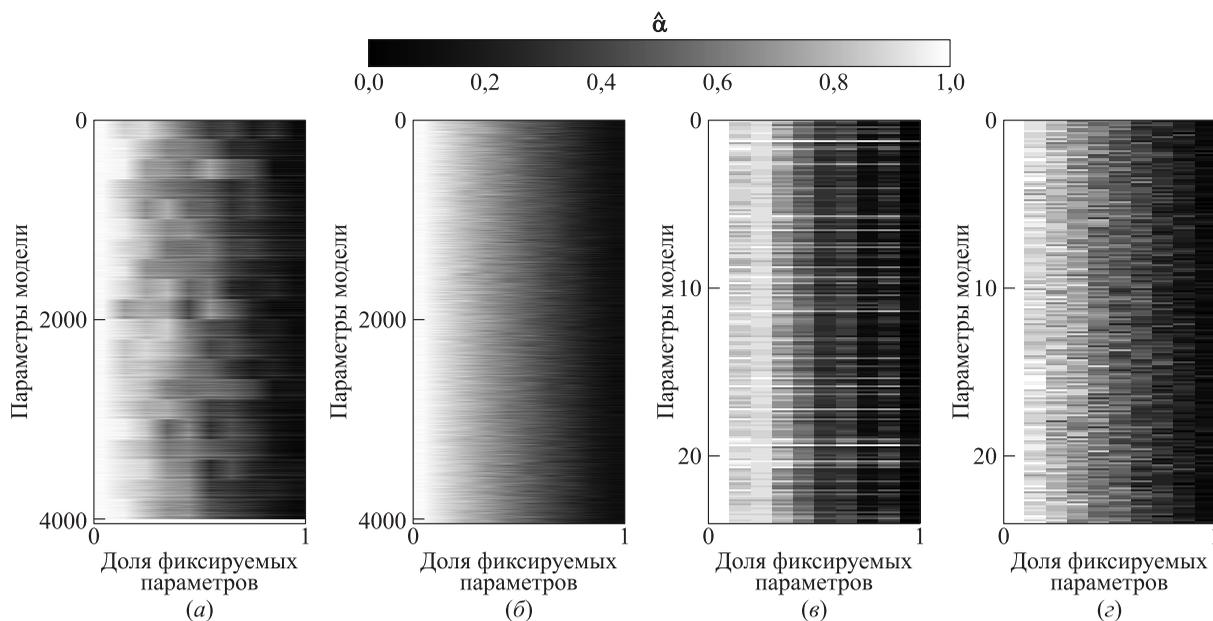


Рис. 2 Визуализация векторов $\hat{\alpha}(k)$ в зависимости от числа фиксируемых параметров (выборка Synthetic 1): (а) все параметры модели упорядочены предложенным методом; (б) все параметры модели упорядочены произвольным образом; (в) часть параметров модели упорядочена предложенным методом; (г) часть параметров модели упорядочена произвольным образом

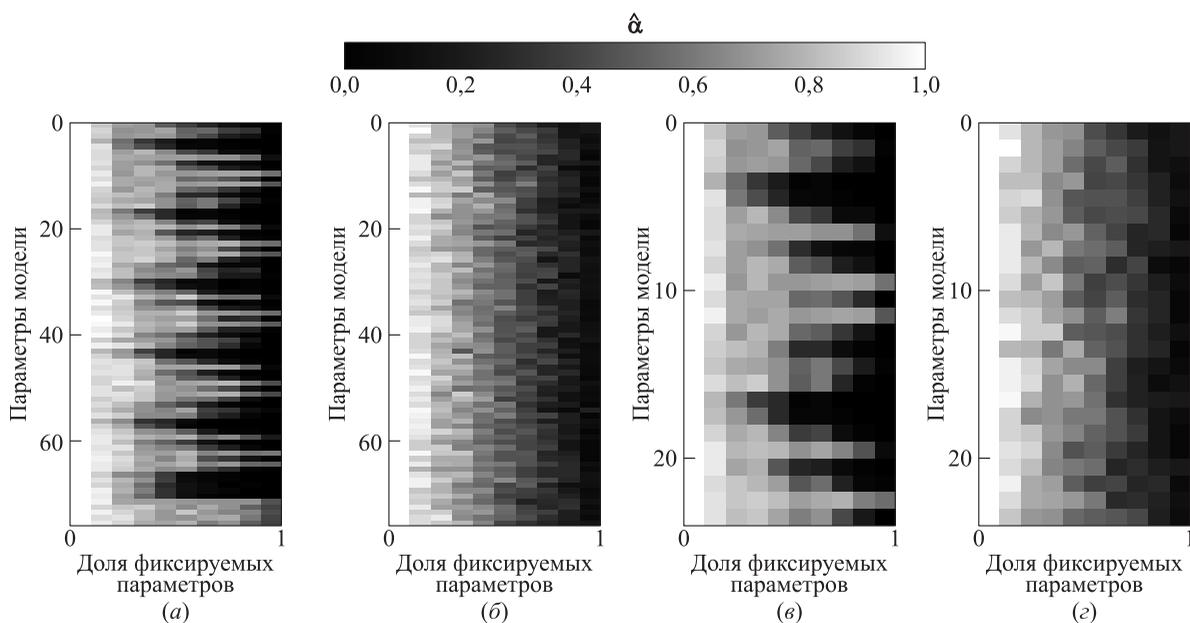


Рис. 3 Визуализация векторов $\hat{\alpha}(k)$ в зависимости от числа фиксируемых параметров (выборка Boston Housing): (а) все параметры модели упорядочены предложенным методом; (б) все параметры модели упорядочены произвольным образом; (в) часть параметров модели упорядочена предложенным методом; (г) часть параметров модели упорядочена произвольным образом

ной моделью с большим числом параметров. После фиксации значимого числа параметров у модели оставалась большое число параметров для дообучения.

На рис. 3 показана зависимость векторов $\hat{\alpha}(k)$ от числа фиксируемых параметров. На рис. 3, а и 3, в видно, что $\hat{\alpha}(k)$ меняется незначительно от запуска к запуску алгоритма. Следовательно, пред-

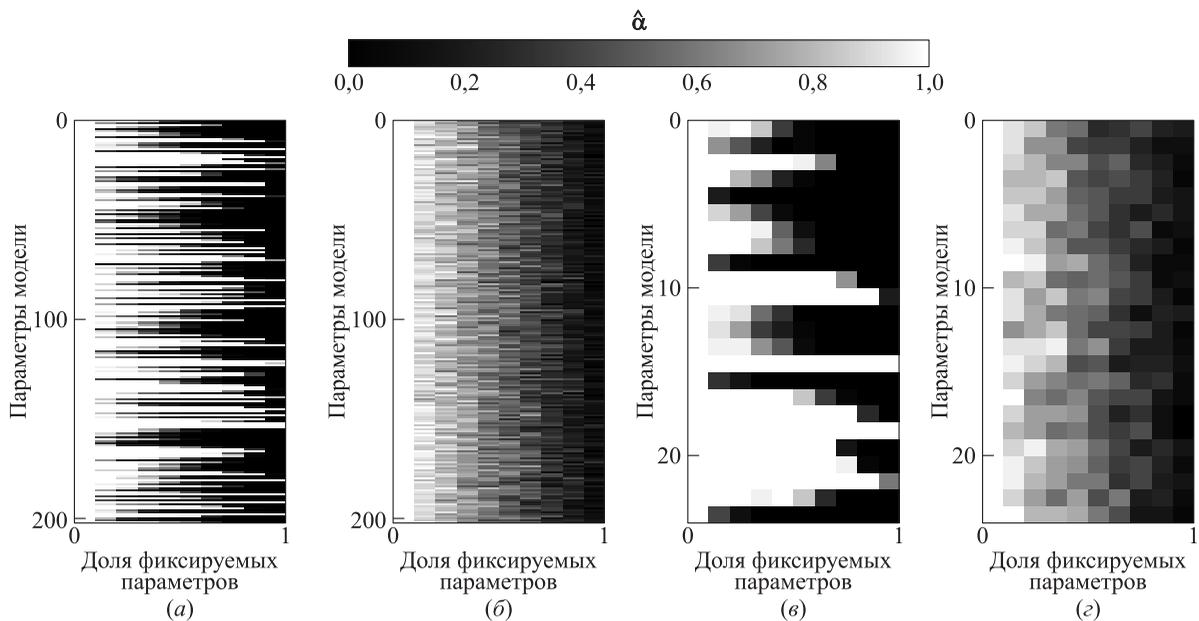


Рис. 4 Визуализация векторов $\hat{\alpha}(k)$ в зависимости от числа фиксируемых параметров (выборка Synthetic 3): (а) все параметры модели упорядочены предложенным методом; (б) все параметры модели упорядочена произвольным образом; (в) часть параметров модели упорядочена предложенным методом; (г) часть параметров модели упорядочена произвольным образом

ложенный порядок задает устойчивый к разным запускам порядок на множестве параметров модели. На рис. 3, б и 3, г видно, что в случае произвольной фиксации параметров вектор $\hat{\alpha}(k)$ является произвольным и никакого порядка на множестве параметров нет.

Выборка Synthetic 3. Эксперимент проводился на синтетически построенных данных. В качестве модели использовалась линейная модель регрессии. На рис. 1, в показаны графики зависимости функции потерь \mathcal{L} от числа фиксируемых параметров. В случае фиксации параметров предложенным методом функция потерь \mathcal{L} растет значительно медленней в сравнении со случаем фиксации параметров произвольным образом. Дисперсия функции ошибки также значительно меньше в случае фиксации параметров предложенным методом.

На рис. 4 показано, что вектор $\hat{\alpha}(k)$ не меняется от запуска к запуску. Так как данная модель линейная, то порядок на параметрах модели индуцирует некоторый порядок на множестве признаков.

Выборка MNIST. В эксперименте рассматривался двухслойный перцептрон для классификации изображений. В качестве входных данных рассматривались изображения размера 28×28 , на которых изображены цифры.

На рис. 1, г показано, что графики функции ошибки похожи в случае фиксации параметров предложенным методом и в случае произвольной фиксации. Данный результат есть следствие того факта, что нейросеть является заведомо переусложненной моделью. После фиксации большого числа параметров у нейросети все еще остается значимое число параметров модели для дообучения.

На рис. 5 показано, что в случае модели со значимым числом оптимизационных параметров предложенный метод упорядочения параметров устойчив от запуска к запуску.

4 Заключение

В данной работе рассмотрена проблема задания порядка на множестве параметров сложных аппроксимирующих моделей. Исследован метод задания порядка на основе анализа стохастических свойств градиента функции ошибки \mathcal{L} по параметрам модели. Для задания порядка использовалась ковариационная матрица градиентов параметров \mathbf{C}_{n_0} , которая рассчитывалась итеративно, в течение t_0 итераций градиентного метода параллельно оптимизации. Число итераций t_0 выбиралось заранее экспериментально.

Предложенный метод был проанализирован в вычислительном эксперименте. Было показано, что порядок, заданный при помощи ковариацион-

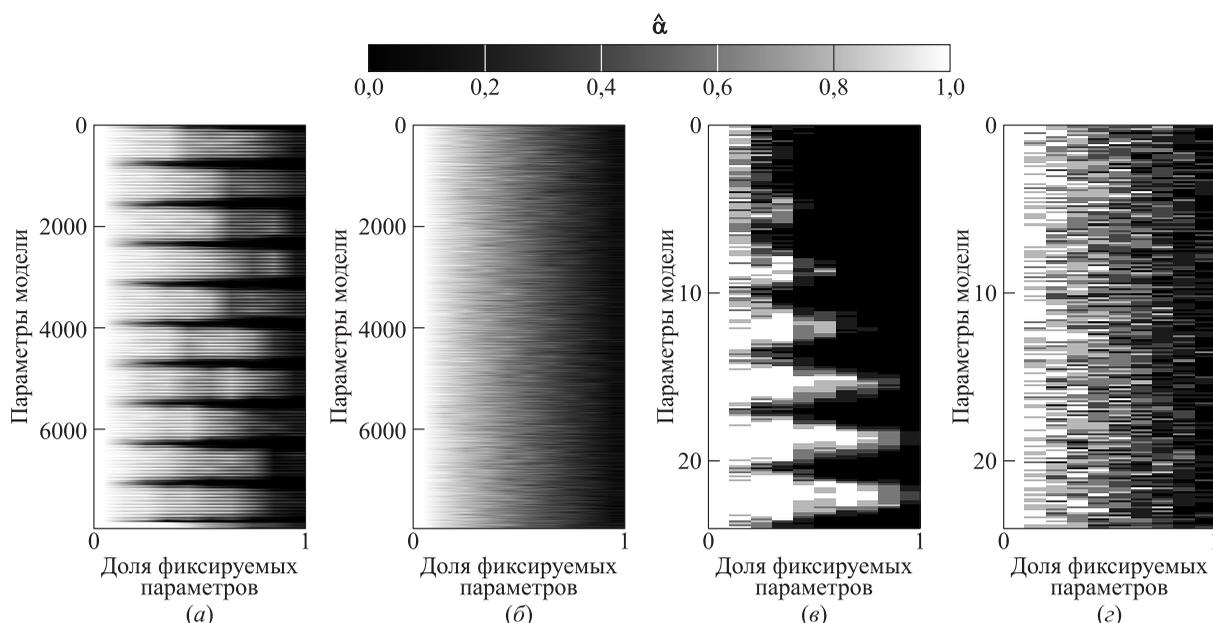


Рис. 5 Визуализация векторов $\hat{\alpha}(k)$ в зависимости от числа фиксируемых параметров (выборка MNIST): (а) все параметры модели упорядочены предложенным методом; (б) все параметры модели упорядочены произвольным образом; (в) часть параметров модели упорядочена предложенным методом; (г) часть параметров модели упорядочена произвольным образом

ной матрицы C_{η_0} , является адекватным, так как фиксация параметров в заданном порядке позволяет зафиксировать значимое число параметров без заметной потери качества.

Отдельно стоит заметить, что параметры упорядочиваются в процессе оптимизации параметров модели. Как было показано в эксперименте, данный порядок устойчив и не меняется от запуска к запуску метода оптимизации.

Литература

1. Sutskever I., Vinyals O., Le Q. Sequence to sequence learning with neural networks // Adv. Neur. Inf., 2014. Vol. 2. P. 3104–3112.
2. Li C., Chen C., Carlson D., Carin L. Preconditioned stochastic gradient Langevin dynamics for deep neural networks // 13th AAAI Conference on Artificial Intelligence. — Phoenix, AZ, USA, 2016. P. 1788–1794.
3. Tibshirani R. Regression shrinkage and selection via the Lasso // J. R. Stat. Soc., 1996. Vol. 58. P. 267–288.
4. Zou H., Hastie T. Regularization and variable selection via the Elastic Net // J. R. Stat. Soc., 2005. Vol. 67. P. 301–320.
5. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting // J. Mach. Learn. Res., 2014. Vol. 15. P. 1929–1958.
6. Molchanov D., Ashukha A., Vetrov D. Variational dropout sparsifies deep neural networks // 34th Conference (International) on Machine Learning. — Sydney, Australia, 2017. Vol. 70. P. 2498–2507.
7. LeCun Y., Denker J., Solla S. Optimal brain damage // Adv. Neur. Inf., 1989. Vol. 2. P. 598–605.
8. Грабовой А. В., Бахтеев О. Ю., Стрижов В. В. Определение релевантности параметров нейросети // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 62–70.
9. Mandt S., Hoffman M., Blei D. Stochastic gradient descent as approximate Bayesian inference // J. Mach. Learn. Res., 2017. Vol. 18. P. 1–35.
10. Kingma D., Ba L. Adam: A method for stochastic optimization // 3rd Conference (International) on Learning Representations. — San Diego, CA, USA, 2015. <https://hdl.handle.net/11245/1.505367>.
11. LeCun Y., Cortes C., Burges C. The MNIST dataset of handwritten digits, 1998. <http://yann.lecun.com/exdb/mnist/index.html>.
12. Harrison D., Rubinfeld D. Hedonic prices and the demand for clean air // J. Environ. Econ. Manag., 1991. Vol. 5. P. 81–102.

Поступила в редакцию 07.10.19

ORDERING THE SET OF NEURAL NETWORK PARAMETERS

A. V. Grabovoy¹, O. Yu. Bakhteev¹, and V. V. Strijov^{1,2}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

²A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: This paper investigates a method for setting order on a set of the model parameters. It considers linear models and neural networks. The set is ordered by the covariance matrix of the gradients. It is proposed to use a given order to freeze the model parameters during the optimization procedure. It is assumed that, after few iterations of the optimization algorithm, most of the model parameters can be frozen without significant loss of the model quality. It reduces the dimensionality of the optimization problem. This method is analyzed in the computational experiment on the real data. The proposed order is compared with the random order on the set of the model parameters.

Keywords: sample approximation; linear model; neural network; model selection; error function

DOI: 10.14357/19922264200208

Acknowledgments

This research was supported by the Russian Foundation for Basic Research (projects 19-07-01155 and 19-07-00875) and NTI (project 13/1251/2018).

References

1. Sutskever, I., O. Vinyals, and Q. Le. 2014. Sequence to sequence learning with neural networks. *Adv. Neur. Inf.* 2:3104–3112.
2. Li, C., C. Chen, D. Carlson, and L. Carin. 2016. Pre-conditioned stochastic gradient Langevin dynamics for deep neural networks. *13th AAAI Conference on Artificial Intelligence Proceedings*. Phoenix, AZ. 1788–1794.
3. Tibshirani, R. 1998. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc.* 58:267–288.
4. Zou, H., and T. Hastie. 2005. Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc.* 67:301–320.
5. Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15:1929–1958.
6. Molchanov, D., A. Ashukha, and D. Vetrov. 2017. Variational dropout sparsifies deep neural networks. *34th Conference (International) on Machine Learning* 70:2498–2507.
7. LeCun, Y., J. Denker, and S. Solla. 1989. Optimal brain damage. *Adv. Neur. Inf.* 2:598–605.
8. Grabovoy, A. V., O. Yu. Bakhteev, and V. V. Strijov. 2019. Opređenje relevantnosti parametara neyroseti [Estimation of the relevance of the neural network parameters]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):62–70.
9. Mandt, S., M. Homan, and D. Blei. 2017. Stochastic gradient descent as approximate Bayesian inference. *J. Mach. Learn. Res.* 18:1–35.
10. Kingma, D., and L. Ba. 2015. Adam: A method for stochastic optimization. *3rd Conference (International) on Learning Representations*. Available at: <https://hdl.handle.net/11245/1.505367> (accessed May 26, 2020).
11. LeCun, Y., C. Cortes, and C. Burges. 1998. The MNIST dataset of handwritten digits. Available at: <http://yann.lecun.com/exdb/mnist/index.html> (accessed May 26, 2020).
12. Harrison, D., and D. Rubinfeld. 1991. Hedonic prices and the demand for clean air. *J. Environ. Econ. Manag.* 5:81–102.

Received October 7, 2019

Contributors

Grabovoy Andrey V. (b. 1997) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; grabovoy.av@phystech.edu

Bakhteev Oleg Yu. (b. 1993) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region, 141701, Russian Federation; bakhteev@phystech.edu

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; strijov@ccas.ru

СТАЦИОНАРНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ $M/G/2/\infty$ С ОДНИМ ЧАСТНЫМ СЛУЧАЕМ ДИСЦИПЛИНЫ ИНВЕРСИОННОГО ПОРЯДКА ОБСЛУЖИВАНИЯ С ОБОБЩЕННЫМ ВЕРОЯТНОСТНЫМ ПРИОРИТЕТОМ*

Л. А. Мейханаджян¹, Р. В. Разумчик²

Аннотация: Рассматривается система массового обслуживания (СМО) $M/G/2/\infty$ с идентичными приборами. Предполагается, что в любой момент времени известна остаточная длина каждой заявки в системе. В системе реализован частный случай дисциплины инверсионного порядка обслуживания с обобщенным вероятностным приоритетом, заключающийся в следующем. В момент поступления новой заявки в систему все заявки на приборах приобретают новую случайную длину. При этом, если есть хотя бы один свободный прибор, новая заявка занимает его; иначе она становится на первое место в очереди. Длины заявок представляют собой независимые одинаково распределенные случайные величины с произвольным абсолютно непрерывным распределением. В предположении, что стационарный режим существует, найдены основные стационарные характеристики функционирования, включая совместное стационарное распределение общего числа заявок в системе и остаточных длин заявок на приборах.

Ключевые слова: многолинейная система; инверсионный порядок обслуживания; вероятностный приоритет

DOI: 10.14357/19922264200209

1 Введение

В работе [1] были изучены стационарные характеристики СМО $M/G/1/\infty$ с одним из частных случаев дисциплины инверсионного порядка обслуживания с обобщенным вероятностным приоритетом [2], названной там дисциплиной ресамплинга. Она заключается в следующем. Предполагается, что в момент поступления новой заявки становится известно ее остаточное время обслуживания. В дальнейшем удобно называть случайную величину остаточного времени обслуживания остаточной длиной, подразумевая, что длина измеряется в единицах времени. Каждая поступающая в непустую систему заявка назначает новую остаточную длину заявке на приборе и становится на первое место в очереди. Когда остаточная длина заявки на приборе становится равной нулю, заявка навсегда покидает систему и на обслуживание выбирается первая заявка из очереди. Также в [1–5] было показано, что описанная выше СМО оказывается полезной в задачах оценки среднего времени

пребывания в однолинейных системах с неточной априорной информацией о временах обслуживания.

Представляет несомненный интерес изучение обобщения полученных в [1] результатов на случай многолинейных систем³. В этой работе исследуется одно такое обобщение: находятся основные стационарные характеристики в системе $M/G/2/\infty$ с дисциплиной ресамплинга и идентичными приборами.

2 Описание системы

Рассмотрим систему с двумя идентичными приборами и очередью неограниченной емкости, на вход которой поступает пуассоновский поток заявок интенсивности λ .

Определим дисциплину обслуживания следующим образом. В момент прихода очередной заявки в систему становится известно ее остаточное время обслуживания (далее — остаточная длина,

* Работа выполнена при поддержке РФФИ (проект 18-37-00283).

¹ Финансовый университет при Правительстве РФ, lamejkanadzhyan@fa.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grazumchik@ipiran.ru

³ Отметим, что вопросы анализа многолинейных СМО $M/G/n/\infty$ остаются по большей части открытыми и связаны с серьезными трудностями. Одно из немногих исключений — СМО $M/G/2/\infty$, для которой известен ряд аналитических результатов (см., например, [6–8]).

или просто длина) и прерывается обслуживание заявок на всех приборах. Каждой заявке, обслуживание которой было прервано, независимо от всей предыстории функционирования системы назначается новая остаточная длина. Затем их обслуживание возобновляется. Новая заявка становится на свободный прибор, если такой имеется; в противном случае она занимает первое место в очереди. Когда остаточная длина заявки на приборе становится равной нулю, она покидает систему и на обслуживание выбирается заявка с первого места в очереди. Остаточные длины заявок представляют собой независимые одинаково распределенные случайные величины с произвольной функцией распределения $B(x)$. Для упрощения выкладок будем предполагать существование у распределения $B(x)$ непрерывной плотности $b(x) = B'(x)$. Также всюду в дальнейшем будем предполагать, что выполняются условия существования стационарного распределения.

3 Вспомогательные функции

Пусть в некоторый момент в систему поступила новая заявка и сразу после ее поступления в системе оказалось $n \geq 3$ заявок с длинами¹ y_1, \dots, y_n . Обозначим через $f_n(s; x_1, x_2, y_4, \dots, y_n | y_1, \dots, y_n)$ преобразование Лапласа–Стилтьеса (ПЛС) времени до момента, когда в системе впервые останется $(n - 1)$ заявка и плотность вероятности того, что в тот же момент длины оставшихся в системе заявок будут равны $x_1, x_2, y_4, \dots, y_n$. Из описания системы и свойств дисциплины обслуживания следует, что функции f_n симметричны на паре переменных (x_1, x_2) , не зависят от y_4, y_5, \dots, y_n и совпадают при $n \geq 3$. Воспользовавшись формулой полной вероятности, получаем уравнение для расчета условной плотности $f = f_n, n \geq 3$:

$$\begin{aligned}
 f(s; x_1, x_2 | y_1, y_2, y_3) = & \\
 = \mathbf{1}_{(y_1 \leq y_2)} e^{-(\lambda+s)y_1} \delta(y_2 - (y_1 + x_2)) \delta(y_3 - x_1) + & \\
 + \mathbf{1}_{(y_1 > y_2)} e^{-(\lambda+s)y_2} \delta(y_1 - (y_2 + x_1)) \delta(y_3 - x_2) + & \\
 + \frac{\lambda}{\lambda + s} \left(1 - e^{-(\lambda+s)\min(y_1, y_2)} \right) \times & \\
 \times \int_0^\infty \int_0^\infty f(s; u_1, u_2) f(s; x_1, x_2 | u_1, u_2, y_3) du_1 du_2, & \quad (1)
 \end{aligned}$$

где

¹Предполагается, что на первом приборе заявка длины y_1 , на втором — y_2 , в очереди на первом месте заявка длины y_3 , на втором — y_4 и т.д.

²Это означает, что $\theta(x) = 1$ при $x \geq 0$ и $\theta(x) = 0$ иначе.

³Например, при $s = 0$ в случае $B(x) = 1 - e^{-\mu x}$ решение (2) есть $g(0; x) = (1/2)\mu e^{-\mu x}$.

$$\begin{aligned}
 f(s; x_1, x_2) = \int_0^\infty \int_0^\infty \int_0^\infty f(s; x_1, x_2 | u_1, u_2, u_3) \times & \\
 \times b(u_1)b(u_2)b(u_3) du_1 du_2 du_3; &
 \end{aligned}$$

δ — дельта-функция Дирака; $\mathbf{1}_{(A)}$ — индикатор множества A . Решая (1), получаем, что для $f(s; x_1, x_2)$ справедлива формула

$$f(s; x_1, x_2) = b(x_1)g(s; x_2) + b(x_2)g(s; x_1),$$

в которой неизвестная функция g есть решение интегрального уравнения:

$$g(s; x) = y(s; x) + \gamma(s) \int_0^\infty K(s; u, x)g(u) du, \quad (2)$$

где

$$y(s; x) = \int_0^\infty e^{-(\lambda+s)u} b(u)b(u+x) du;$$

$$\begin{aligned}
 K(s; u, x) = \theta(u-x)e^{-(\lambda+s)(u-x)}b(u-x) + & \\
 + e^{-(\lambda+s)u}b(u+x); &
 \end{aligned}$$

θ — функция Хевисайда²;

$$\begin{aligned}
 \gamma(s) = & \\
 = \left[\frac{2\lambda}{\lambda + s} \int_0^\infty \int_0^u b(u)b(v) \left(1 - e^{-(\lambda+s)v} \right) dv du \right] \times & \\
 \times \left[1 - \frac{2\lambda}{\lambda + s} \int_0^\infty \int_0^u \left(1 - e^{-(\lambda+s)v} \right) (b(u)g(s; v) + & \right. \\
 \left. + b(v)g(s; u)) dv du \right]^{-1}. &
 \end{aligned}$$

Отметим, что значение $\gamma(s)$ зависит от неизвестной функции g , и уравнение (2), по-видимому, не обладает хорошими особенностями, кроме одной: свободный член и ядро являются неотрицательными функциями. В некоторых частных случаях³ решение (2) может быть выписано в явном виде. В общем же случае его приходится искать численно. Хорошие результаты дает итерационный метод, причем в качестве начальной итерации необходимо брать нулевое приближение. Тогда итерации будут возрастать, что позволит контролировать сходимость к точному решению. При $s = 0$ для

контроля точности можно пользоваться условием нормировки, из которого следует, что

$$\int_0^{\infty} g(0; x) dx = \frac{1}{2}.$$

Пусть теперь в некоторый момент в систему поступила новая заявка и сразу после ее поступления в системе оказалось две заявки с остаточными длинами y_1 и y_2 . Обозначим через $f_2(s; x|y_1, y_2)$ ПЛС времени до момента, когда в системе впервые останется одна заявка, и плотность вероятности того, что в тот же момент ее остаточная длина будет равна x . Выписывая для $f_2(s; x|y_1, y_2)$ уравнение, аналогичное (1), и решая его, получаем, что $f_2(s; x) = 2g(s; x)$.

4 Стационарные вероятности состояний

Рассмотрим процесс $(\xi_1(t), \dots, \xi_{\nu(t)}(t))$. Здесь $\nu(t) = n$, когда в момент t в системе находится n заявок. Координаты $\xi_1(t)$ и $\xi_2(t)$ — это остаточные времена обслуживания заявок на приборах, $\xi_3(t)$ — длина первой заявки в очереди, а $\xi_{\nu(t)}(t)$ — последней. В том случае, когда в системе отсутствуют заявки, координаты $\xi_i(t)$ не определяются. Наконец, при $\nu(t) = 1$ координата $\xi_1(t)$ хранит остаточное время обслуживания находящейся на приборе единственной заявки. Процесс $(\xi_1(t), \dots, \xi_{\nu(t)}(t))$ является марковским и описывает характеристики очереди в момент t .

Обозначим через

$$P_0 = \lim_{t \rightarrow \infty} P_0(t),$$

$$P_n(x_1, \dots, x_n) = \lim_{t \rightarrow \infty} P_n(t; x_1, \dots, x_n), \quad n \geq 1,$$

стационарное распределение процесса $(\xi_1(t), \dots, \xi_{\nu(t)}(t))$, где $P_0(t) = P(\nu(t) = 0)$, $P_n(t; x_1, \dots, x_n) = P(\nu(t) = n, \xi_1(t) < x_1, \dots, \xi_n(t) < x_n)$. Для плотностей введенных вероятностей (в предположении, что они существуют) можно выписать интегродифференциальные уравнения, на основе которых теоретически можно производить вычисления. Однако на практике они не реализуемы уже при совсем небольших значениях n , поскольку размерность уравнений растет пропорционально n . Поэтому далее речь пойдет только о маргинальных стационарных вероятностях $P_1(x), P_n(x_1, x_2) = P_n(x_1, x_2, \infty, \dots, \infty)$, $n \geq 2$, для которых предполагается существование плотностей:

$$p_1(x) = P_1'(x); \quad p_n(x_1, x_2) = \frac{\partial^2 P_n(x_1, x_2)}{\partial x_1 \partial x_2}, \quad n \geq 2.$$

Принятая дисциплина обслуживания обладает известным свойством, которое в данном случае позволяет рекуррентно вычислять стационарные плотности (подробнее см. [9]) и которое заключается в следующем. Пусть $n > 1$ — произвольное целое число. Выделим для процесса $\nu(t)$ те интервалы времени, когда число заявок в системе будет больше n , т. е. $\nu(t) > n$. Тогда, в силу дисциплины обслуживания, с того момента, как в системе впервые появится $(n + 1)$ -я заявка, и до того момента, как в системе снова будет n заявок, последние $(n - 1)$ компонент процесса $(\xi_1(t), \dots, \xi_{\nu(t)}(t))$ не меняются. Следовательно, если для процесса выкинуть все те интервалы времени, когда $\nu(t) > n$, и оставшиеся куски склеить, то вероятностные характеристики получившегося после склейки процесса будут одинаковыми для всех n и плотности $p_1(x), p_k(x_1, x_2)$, $2 \leq k \leq n$, будут совпадать с точностью до постоянного множителя, не зависящего от k .

Положим

$$P_1 = \int_0^{\infty} p_1(u) du; \quad P_n = \int_0^{\infty} \int_0^{\infty} p_n(u, v) dudv, \quad n \geq 2.$$

С учетом описанного выше свойства дисциплины обслуживания, уравнения Колмогорова—Чепмена для стационарных плотностей будут иметь вид:

$$-p_1'(x) = -\lambda p_1(x) + \lambda b(x)P_0 + \lambda f_2(0; x)P_1; \quad (3)$$

$$\lambda \text{fr} \partial p_n(x_1, x_2) \partial x_1 - \frac{\partial p_n(x_1, x_2)}{\partial x_2} = -\lambda p_n(x_1, x_2) + \lambda b(x_1)b(x_2)P_{n-1} + \lambda f_{n+1}(0; x_1, x_2)P_n, \quad n \geq 2, \quad (4)$$

с начальными условиями

$$\lim_{x \rightarrow \infty} p_1(x) = 0;$$

$$\lim_{x_1 \rightarrow \infty} p_n(x_1, x_2) = 0;$$

$$\lim_{x_2 \rightarrow \infty} p_n(x_1, x_2) = 0, \quad n \geq 2.$$

Систему (3)–(4) можно решить следующим образом. Заметим, что (4) есть уравнение в частных производных первого порядка. Воспользовавшись методом характеристик, находим выражение для плотности $p_n(x_1, x_2)$:

$$p_n(x_1, x_2) = e^{\lambda x_1} \int_{x_1}^{\infty} e^{-\lambda u} (\lambda b(u)b(x_2 - x_1 + u) P_{n-1} + \lambda f(0; u, x_2 - x_1 + u) P_n) du, \quad n \geq 2, \quad (5)$$

где

$$f(0; x_1, x_2) = b(x_1)g(0; x_2) + b(x_2)g(0; x_1);$$

функция g есть решение уравнения (2) при $s = 0$. Для нахождения неизвестных в (5) вероятностей P_{n-1} и P_n проинтегрируем (5) по всем значениям x_1 и x_2 . С помощью обычных преобразований, вводя для сокращения записи обозначения

$$\hat{\beta}(\lambda) = \lambda \int_0^\infty e^{-\lambda u} (1 - B(u))^2 du;$$

$$\bar{\beta}(\lambda) = 2\lambda \int_0^\infty \int_0^\infty e^{-\lambda z} (1 - B(z))g(0; z + x) dx dz,$$

получим соотношение

$$P_n = \hat{\beta}(\lambda)P_{n-1} + \bar{\beta}(\lambda)P_n,$$

из которого следует, что

$$P_n = \left(\frac{\hat{\beta}(\lambda)}{1 - \bar{\beta}(\lambda)} \right)^{n-1} P_1, \quad n \geq 1.$$

Поступая аналогичным образом с решением уравнения (3), которое имеет вид:

$$p_1(x) = e^{\lambda x} \int_x^\infty e^{-\lambda u} (\lambda b(u)P_0 + 2\lambda g(0; u)P_1) du,$$

находим

$$P_1 = P_0 \frac{1 - \beta(\lambda)}{1 - \bar{\beta}(\lambda)},$$

где $\beta(\lambda)$ — ПЛС $B(x)$;

$$\bar{\beta}(\lambda) = 2\lambda \int_0^\infty \int_0^\infty e^{-\lambda z} g(0; z + x) dx dz.$$

Оставшаяся неизвестной вероятность P_0 , как обычно, находится из условия нормировки $\sum_{i=0}^\infty P_i = 1$, т.е.

$$P_0 = \left(1 + \frac{1 - \beta(\lambda)}{1 - \bar{\beta}(\lambda)} \frac{1 - \bar{\beta}(\lambda)}{1 - \bar{\beta}(\lambda) - \hat{\beta}(\lambda)} \right)^{-1}. \quad (6)$$

Таким образом, стационарное распределение общего числа заявок в системе образует начиная с P_1 геометрическую прогрессию, что позволяет легко находить моменты числа заявок в системе. В частности, среднее число N заявок в системе в стационарном режиме равно¹ $N = (1 - P_0)^2 / P_1$.

В заключение раздела отметим, что из (6) и того, что $P_0 > 0$, следует необходимое условие существования стационарного распределения.

¹Примечательно, что такая же формула для N через вероятности состояний системы имеет место и в случае одноканальной системы [1, Remark 3].

5 Стационарное распределение времени пребывания

Время пребывания в системе заявки представляет собой сумму двух независимых частей: времени ожидания начала обслуживания и собственно времени нахождения заявки на приборе. Обозначим соответствующие ПЛС через $\chi(s)$, $\psi(s)$ и $\omega(s)$. Поскольку с вероятностью $P_0 + P_1$ поступающая заявка попадает сразу на прибор, а с дополнительной вероятностью — в очередь, то ПЛС $\chi(s)$ времени пребывания заявки в системе имеет вид:

$$\chi(s) = (P_0 + P_1)\psi(s) + (1 - P_0 - P_1)\psi(s)\omega(s). \quad (7)$$

Из результатов разд. 3 немедленно следует, что

$$\omega(s) = \int_0^\infty f_2(s; u) du.$$

Нахождение же ПЛС $\psi(s)$ времени пребывания заявки на приборе (с учетом всех возможных прерываний, т.е. ее истинного времени обслуживания) ничем не отличается от нахождения этого ПЛС в случае однолинейной системы. Поэтому, согласно [1, формула (13)], имеем:

$$\psi(s) = \frac{\beta(\lambda + s)(\lambda + s)}{s + \lambda\beta(\lambda + s)}.$$

Остановимся на нахождении стационарного распределения периода занятости системы (ПЗ). Будем считать, что он начинается в момент поступления заявки в пустую систему и заканчивается в тот момент, когда система впервые оказалась свободной от заявок. Обозначим ПЛС ПЗ, открываемого заявкой длины x , через $\vartheta(s; x)$. Воспользовавшись формулой полной вероятности, получим интегральное уравнение для $\vartheta(s; x)$, решение которого имеет вид:

$$\vartheta(s; x) = e^{-(s+\lambda)x} + \left(1 - e^{-(\lambda+s)x} \right) \times$$

$$\times \left[\frac{2\lambda}{\lambda + s} \int_0^\infty e^{-(s+\lambda)u} g(s; u) dv \right] \times$$

$$\times \left[1 - \frac{2\lambda}{\lambda + s} \int_0^\infty \left(1 - e^{-(\lambda+s)u} \right) g(s; u) du \right]^{-1}.$$

Безусловное ПЛС $\vartheta(s)$ ПЗ получается усреднением $\vartheta(s; x)$ по длине заявки, т. е.

$$\vartheta(s) = \int_0^{\infty} \vartheta(s; u)b(u) du.$$

К сожалению, найденные выражения для стационарных распределений времени пребывания и ПЗ, по-видимому, не могут привести к явным выражениям для моментов и позволяют их находить лишь численно, причем при расчетах основная сложность связана с решением уравнения (2).

6 Заключение

В [1] доказано, что для аналогичной, но одноканальной системы справедлива формула Литтла. Предложенный в этой работе метод анализа, по-видимому, не позволяет установить ее справедливость для двухканальной системы. Однако численные эксперименты показывают, что формула Литтла имеет место и в этом случае. Тогда для расчета стационарного среднего времени v пребывания заявки в системе нет необходимости дифференцировать (7) и можно пользоваться формулой

$$v = \frac{N}{\lambda} = \frac{(1 - P_0)^2}{\lambda P_1}.$$

Всюду выше предполагалось существование стационарного распределения, но критерий не был найден; получено лишь необходимое условие. Результаты, полученные для одноканальной системы в [1], подсказывают, что для двухканальной системы необходимое и достаточное условие существования стационарного режима, по-видимому, не должно зависеть от моментов длины заявки какого-либо порядка, т. е. для любого распределения длины заявки при достаточно малой интенсивности λ существует стационарное распределение. Как показывают численные эксперименты, средняя длина ПЗ рассматриваемой системы равна¹ $3/2v$. Отсюда получаем (при условии справедливости формулы Литтла), что $P_0 > 0$ является не только необходимым, но и достаточным условием существования стационарного распределения.

Отметим наконец, что полученные формулы не позволяют производить расчеты в случае, когда длины заявок принимают только конечное число значений. Этот случай требует специального исследования.

Литература

1. *Horváth I., Razumchik R., Telek M.* The resampling $M/G/1$ non-preemptive LIFO queue and its application to systems with uncertain service time // *Perform. Evaluation*, 2019. Vol. 134. Art. ID: 102000.
2. *Мейханаджян Л. А., Милованова Т. А., Печинкин А. В., Разумчик Р. В.* Стационарные вероятности состояний в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом // *Информатика и её применения*, 2014. Т. 8. Вып. 3. С. 28–38.
3. *Meykhanadzhyan L., Razumchik R.* New scheduling policy for estimation of stationary performance characteristics in single server queues with inaccurate job size information // *30th European Conference on Modelling and Simulation Proceedings*. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2016. P. 710–716.
4. *Milovanova T. A., Meykhanadzhyan L. A., Razumchik R. V.* Bounding moments of Sojourn time in $M/G/1$ FCFS queue with inaccurate job size information and additive error: Some observations from numerical experiments // *CEUR Workshop Proceedings*, 2018. Vol. 2236. P. 24–30.
5. *Мейханаджян Л. А., Разумчик Р. В.* Система массового обслуживания $Geo/G/1/\infty$ с инверсионным порядком обслуживания и ресамплингом в дискретном времени // *Информатика и её применения*, 2019. Т. 13. Вып. 4. С. 60–67.
6. *Hokstad P.* On the steady-state solution of the $M/G/2$ queue // *Adv. Appl. Probab.*, 1979. Vol. 11. Iss. 1. P. 240–255.
7. *Wiens D. P.* On the busy period distribution of the $M/G/2$ queueing system // *J. Appl. Probab.*, 1989. Vol. 26. Iss. 4. P. 858–865.
8. *Knessl C., Matkowsky B. J., Schuss Z., Tier C.* An integral equation approach to the $M/G/2$ queue // *Oper. Res.*, 1990. Vol. 38. Iss. 3. P. 506–518.
9. *Печинкин А. В.* Об одной инвариантной системе массового обслуживания // *Math. Operationsforsch. Statist. Ser. Optimization*, 1983. Vol. 14. Iss. 3. P. 433–444.

Поступила в редакцию 14.04.20

¹ В отличие от случая одноканальной системы, в которой стационарное среднее время пребывания заявки в системе совпадает со средней длиной ПЗ [1, Corollary 3].

STATIONARY CHARACTERISTICS OF $M/G/2/\infty$ QUEUE WITH IDENTICAL SERVERS, LIFO SERVICE, AND RESAMPLING POLICY

L. A. Meykhanadzhyan¹ and R. V. Razumchik²

¹Financial University under the Government of the Russian Federation, 49 Leningradsky Prospekt, Moscow 125993, Russian Federation

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

Abstract: Consideration is given to the $M/G/2/\infty$ queue with identical servers, LIFO (last in, first out) service discipline and one special case of the generalized probabilistic priority policy called resampling. The latter implies that a customer arriving to the nonidle system assigns independently new remaining service time to each customer currently in service. The new customer itself either enters a free server, if there is any, or occupies a place in the queue. Remaining service times are assumed to be independent identically distributed random variables with the known general absolute continuous distribution. Under the assumption that the stationary regime exists, the main performance characteristics of the system, including the joint stationary distribution of the total number of customers in the system and the remaining service times of customers in service, are derived.

Keywords: multiserver system; inverse service order; probabilistic priority; resampling

DOI: 10.14357/19922264200209

Acknowledgments

The reported study was funded by the Russian Foundation for Basic Research according to the research project 18-37-00283.

References

1. Horváth I., R. Razumchik, and M. Telek. 2019. The resampling $M/G/1$ non-preemptive LIFO queue and its application to systems with uncertain service time. *Perform. Evaluation* 134:102000.
2. Meykhanadzhyan, L. A., T. A. Milovanova, A. V. Pechinkin, and R. V. Razumchik. 2014. Stacionarnye veroyatnosti sostoyaniy v sisteme obsluzhivaniya s inverzionnym poryadkom obsluzhivaniya i obobshchennym veroyatnostnym prioritetom [Stationary distribution in a queueing system with inverse service order and generalized probabilistic priority]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):28–38.
3. Meykhanadzhyan, L., and R. Razumchik. 2016. New scheduling policy for estimation of stationary performance characteristics in single server queues with inaccurate job size information. *30th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 710–716.
4. Milovanova, T. A., L. A. Meykhanadzhyan, and R. V. Razumchik. 2018. Bounding moments of sojourn time in $M/G/1$ FCFS queue with inaccurate job size information and additive error: Some observations from numerical experiments. *CEUR Workshop Proceedings* 2236:24–30.
5. Meykhanadzhyan, L. A., and R. V. Razumchik. 2019. Sistema massovogo obsluzhivaniya $Geo/G/1/\infty$ s inversionnym poryadkom obsluzhivaniya i resamplingom v diskretnom vremeni [Discrete-time $Geo/G/1/\infty$ LIFO queue with resampling policy]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):60–67.
6. Hokstad, P. 1979. On the steady-state solution of the $M/G/2$ queue. *Adv. Appl. Probab.* 11(1):240–255.
7. Wiens, D. P. 1989. On the busy period distribution of the $M/G/2$ queueing system. *J. Appl. Probab.* 26(4):858–865.
8. Knessl, C., B. J. Matkowsky, Z. Schuss, and C. Tier. 1990. An integral equation approach to the $M/G/2$ queue. *Oper. Res.* 38(3):506–518.
9. Pechinkin, A. V. 1983. Ob odnoy invariantnoy sisteme massovogo obsluzhivaniya [On an invariant queueing system]. *Math. Operationsforsch. Statist. Ser. Optimization* 14(3):433–444. doi: 10.1080/02331938308842876.

Received April 14, 2020

Contributors

Meykhanadzhyan Lusine A. (b. 1990) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Data Analysis, Decision-Making and Financial Technology, Financial University under the Government of the Russian Federation, 49 Leningradsky Prospekt, Moscow 125993, Russian Federation; lamejkhanadzhyan@fa.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; rrazumchik@ipiran.ru

ОПТИМИЗАЦИЯ ЕМКОСТИ ОСНОВНОГО НАКОПИТЕЛЯ В СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ ТИПА $G/M/1/K$ С ДОПОЛНИТЕЛЬНЫМ НАКОПИТЕЛЕМ*

Я. М. Агаларов¹

Аннотация: Для системы массового обслуживания (СМО) типа $G/M/1/K$ с дополнительным накопителем сформулирована задача оптимизации емкости основного накопителя при стоимостной целевой функции, учитывающей затраты системы, связанные с потерей заявок, хранением заявок в накопителях, техническим обслуживанием накопителей и прибора, простоем прибора. Заявка, поступившая в систему, принимается в основной накопитель, если есть свободное место, иначе согласно заданному вероятностному распределению либо покидает систему (теряется), либо поступает в дополнительный накопитель, если там есть свободное место. Если оба накопителя переполнены, заявка покидает систему (теряется). Если в основном накопителе освобождается место, то одна из заявок из очереди в дополнительном накопителе (если он не пуст) сразу поступает в очередь в основной накопитель. Доказана унимодальность целевой функции и предложена процедура гарантированного поиска оптимальной емкости основного накопителя.

Ключевые слова: система массового обслуживания; оптимизация; накопитель; емкость накопителя

DOI: 10.14357/19922264200210

1 Введение

В качестве математических моделей многих производственных, информационных и транспортных систем, в которых генерируются случайные потоки заявок на обслуживание, как правило, используют СМО с ограниченными накопителями [1]. Одним из параметров, от значения которого существенным образом может зависеть экономическая эффективность такой системы, служит объем накопителя. При этом кроме стоимости самого накопителя объем накопителя влияет на эффективность системы через зависящие от него другие характеристики, такие как потери заявок, задержки заявок в накопителе, время простоя обслуживающего устройства, загруженность обслуживающего устройства, интенсивность повторных поступлений заявок и т. д.

Вопросам оптимизации СМО с ограниченными накопителями посвящено большое число работ, в которых можно найти множество различных постановок оптимизационных задач. Приведем краткий обзор некоторых работ [2–9], опубликованных в последние годы и наиболее близких к рассматриваемой в данной статье тематике оптимизации емкости накопителя при стоимостной целевой функции.

В [2] рассматривается $M/M/1$ с управляемой нагрузкой, в которой можно отклонять заявки по мере

прибытия и корректировать скорость обслуживания в каждый момент принятия решения. Цель состоит в минимизации предельных (долгосрочных) средних затрат, которые включают затраты на эксплуатацию мест хранения, затраты за единицу времени обслуживания заявки при заданной скорости обслуживания, штраф за отклонение задания. Разработан алгоритм вычисления оптимальной политики, который позволяет на каждом шаге вычислять достигнутую точность.

В [3] решается задача выбора оптимальной емкости буферной памяти узла сети массового обслуживания в условиях интенсивного трафика, где в качестве модели узла используется СМО типа $M/M/1$ и в качестве целевой функции — функционал затрат, учитывающий три типа затрат: «стоимость перегрузки», зависящую от числа клиентов в системе, «стоимость контроля», связанную с динамически контролируемой скоростью обслуживания, и «стоимость штрафа» за отклонение клиентов. Предложено приближенное решение, полученное для аппроксимирующей диффузионной модели как решение одномерной броуновской задачи управления (Brownian control problem, BCP).

В работе [4] ставится задача оптимального управления очередью системы типа $GI/GI/1$ с нетерпеливыми заявками (заявка после поступления в очередь через экспоненциально распреде-

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00692, 19-07-00739 и 20-07-00804).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, agglar@yandex.ru

ленный интервал времени, если она за это время не обслужилась, покидает систему) в условиях интенсивного трафика и в качестве оптимальной политики ограничения длины очереди в смысле минимизации затрат предлагается использовать решение диффузионной аппроксимирующей задачи. Система при поступлении заявки в очередь получает платеж, при отклонении заявки и уходе необслуженной заявки из очереди платит соответствующие штрафы.

В работе [5] рассмотрена проблема управления доступом заявок в систему $M/M/N/M$, когда существуют затраты, связанные с отказом от клиентов, простоем прибора и отказывающимися заявками. Сформулирован соответствующий марковский процесс принятия решений (MDP, Markov decision process), и показано, что оптимальная политика — пороговая. Предлагается итерационный алгоритм, который при определенных условиях обеспечивает точное оптимальное решение, когда он останавливается; в противном случае алгоритм останавливается, когда достигается предусмотренная заранее точность приближения. Решена также соответствующая аппроксимирующая задача управления диффузией (DCP, diffusion control problem) и проведен численный анализ минимальных затрат системы для MDP и DCP.

В работе [6] рассматривается последовательность СМО типа $M/M/1$ с возможностью ограничения длины очереди и корректировки скорости обслуживания с целью минимизации предельных средних затрат при интенсивном трафике. Функция затрат включает штраф за каждую отвергнутую заявку, затраты, связанные с корректировкой скорости обслуживания, и штраф за каждую отказывающуюся заявку. Для построения оптимального управления для СМО в условиях интенсивного трафика (выбора оптимальной емкости буфера и оптимальной скорости обслуживания) предлагается использовать оптимальную стратегию для предельной задачи управления диффузией (задача броуновского управления, или BCP — Brownian control problems).

В работе [7] рассмотрена задача выбора оптимальной емкости накопителя системы $G/M/1/K$ с учетом штрафов из-за задержки заявок в очереди, потерь заявок и простоя прибора, доказана унимодальность целевой функции.

Задача выбора оптимальной емкости накопителя системы $M/M/s/K$ при целевой функции, учитывающей штрафы за пребывание заявки в очереди дольше заданного времени (дедлайна) и потери заявки исследована в работе [9]. Доказано, что функция дохода — унимодальна.

В [8] рассмотрена СМО типа $G/G/1/K$, в которой прибывающие заявки, когда число заявок в буфере превышает значение K , отклоняются и с фиксированной вероятностью p покидают систему. В противном случае они поступают на орбиту и повторяют попытку позже через экспоненциально распределенный интервал времени с параметром $\mu > 0$ для всех заявок. Повторные заявки рассматриваются системой как первичные. В работе ставится задача оптимизации емкости накопителя независимо от емкости орбиты для стоимостной целевой функции, учитывающей потери и задержки заявок. Получен асимптотически оптимальный размер буфера для диффузионной модели СМО для случаев $\mu \rightarrow 0$ и $\mu \rightarrow \infty$.

Задача, которая рассматривается ниже, близка к задаче из работы [8] с точки зрения модели и структуры затрат для случая $\mu \rightarrow \infty$.

2 Описание задачи

Рассматривается одноканальная СМО с двумя накопителями (основной и дополнительный) ограниченной емкости и заданной функцией распределения интервалов времени между поступлениями заявок $A(t)$ и экспоненциальной функцией распределения времени обслуживания заявок с заданным параметром $\gamma > 0$ (рис. 1). Поступившая извне заявка принимается в основной накопитель системы (занимает любое свободное место в этом накопителе), если в момент ее поступления число занятых в нем мест меньше его емкости $K > 0$. Если заявка

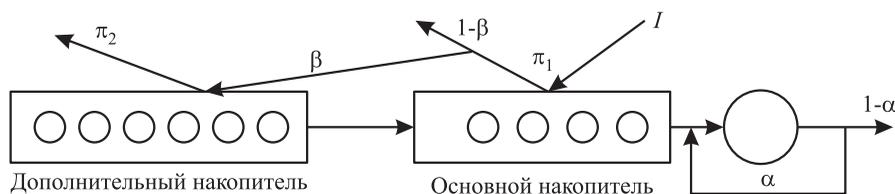


Рис. 1 Схема системы: I — источник потока первичных заявок; π_1 — вероятность переполнения основного накопителя; π_2 — вероятность переполнения дополнительного накопителя; α — вероятность повтора обслуживания; β — вероятность поступления в дополнительный накопитель заявки, не допущенной в основной

допущена в накопитель, она занимает любое свободное место в нем и, если прибор свободен, немедленно начинает обслуживаться прибором, в противном случае становится к нему в очередь. На время обслуживания заявки место в накопителе сохраняется за ней. Заявка после завершения обслуживания с фиксированной вероятностью $(1 - \alpha) > 0$ покидает систему (успешно обслуживается), освободив одновременно прибор и накопитель, а с вероятностью $\alpha \geq 0$ освобождает прибор, но продолжает занимать место в накопителе и повторно становится в очередь к прибору. Освободившийся прибор сразу приступает к обслуживанию очередной заявки, если очередь не пуста, иначе простаивает до момента поступления в основной накопитель новой заявки. Заявка, заставшая основной накопитель системы полностью занятым, с вероятностью $(1 - \beta) \geq 0$ теряется, а с вероятностью $\beta \geq 0$ поступает на дополнительный накопитель ограниченной емкости $M \geq 0$, если там есть свободное место, иначе теряется. В момент освобождения места в основном накопителе одна из заявок перемещается в него из дополнительного накопителя, освободив там место. Время перемещения в рассматриваемой модели считается равным нулю.

Система получает доход, включающий в себя следующие составляющие:

- d_0 — плата, получаемая системой, если поступившая заявка успешно обслужена системой ($0 \leq d_0 < \infty$);
- d_1 — величина штрафа, который платит система, если поступившая заявка потеряна ($0 \leq d_1 < \infty$);
- d_{21} — затраты системы в единицу времени на хранение заявки в основном накопителе ($0 \leq d_{21} < \infty$);
- d_{22} — затраты системы в единицу времени на хранение в дополнительном накопителе ($0 \leq d_{22} < \infty$);
- d_3 — величина штрафа за единицу времени простоя прибора ($0 \leq d_3 < \infty$);
- d_4 — затраты системы в единицу времени на обслуживание заявки прибором ($0 \leq d_4 < \infty$);
- d_5 и d_6 — затраты системы в единицу времени на техническое обслуживание одного места в основном и дополнительном накопителе соответственно ($0 \leq d_5 < \infty, 0 \leq d_6 < \infty$).

Обозначим через $D(K)$ предельный (стационарный) доход, получаемый системой в единицу времени, и рассмотрим его как функцию переменной K при фиксированных значениях остальных параметров системы.

Ставится задача оптимизации объема основного накопителя при фиксированных остальных параметрах системы в смысле максимизации функции дохода $D(K)$ по $K > 0$. В следующих разделах приводятся результаты исследования данной задачи.

3 Решение задачи

Получим явное аналитическое выражение для функции $D(K)$. Введем обозначения:

$\bar{v} = \int_0^\infty v dA(v) > 0$ — среднее время между моментами поступления заявок извне;

$W_1(K)$ — среднее время пребывания заявки в основном накопителе;

$W_2(K)$ — среднее время пребывания заявки в дополнительном накопителе;

$\bar{T}(K)$ — среднее время простоя прибора.

Как следует из описания модели, среднее число попыток обслуживания для каждой заявки равно $(1 - \alpha)^{-1}$ и функция распределения суммарного времени обслуживания одной заявки имеет вид:

$$B(t) = 1 - \exp(-\mu t), \mu = \gamma(1 - \alpha), t \geq 0.$$

Как известно, процесс обслуживания заявок в данной системе описывается цепью Маркова [1], где переходы цепи определяются моментами поступления заявок, а состояние цепи — суммарное число заявок, находящихся в двух накопителях в момент поступления. Для вероятностей переходов p_{ij} вложенной цепи Маркова справедливы формулы:

$$p_{ij}(K) = \begin{cases} r_{i+1-j}, & 0 \leq i \leq K-1, 1 \leq j \leq i+1; \\ \beta r_{i+1-j} + (1-\beta)r_{i-j}, & K \leq i \leq K+M-1, 1 \leq j \leq i+1; \\ r_{K+M-j}, & i = K+M, 1 \leq j \leq K+M; \\ 1 - \sum_{m=0}^i r_m, & i \leq K-1, j=0; \\ 1 - \beta \sum_{m=0}^i r_m - (1-\beta) \sum_{m=0}^{i-1} r_m, & K \leq i \leq K+M-1, j=0; \\ 1 - \sum_{m=0}^{K+M-1} r_m, & i = K+M, j=0; \\ 0 & \text{в остальных случаях,} \end{cases} \quad (1)$$

где

$$r_m = \int_0^\infty \frac{(\mu t)^m}{m!} e^{-\mu t} dB(t) \text{ при } m \geq 0.$$

Для данной цепи система уравнений равновесия в матричном виде задается равенством:

$$\bar{\pi}(K) = \bar{\pi}^T(K)P(K), \quad (2)$$

где $\bar{\pi}(K) = (\pi_0(K), \dots, \pi_{K+M}(K))$ — столбец стационарных вероятностей состояний системы; $\bar{\pi}^T$ — строка стационарных вероятностей состояний системы; $P(K) = (p_{i,j}(K))$ — треугольная матрица переходных вероятностей с элементами $p_{i,j}(K)$ вида (1). Из (2) следует, что стационарные вероятности состояний вычисляются по формуле:

$$\pi_j(K) = \frac{R_j^H}{\sum_{i=0}^{h_2} R_i^H}, \quad 0 \leq j \leq K + M,$$

где

$$R_j(K) = \begin{cases} \frac{R_{j+1}(K)(1-r_1) - R_{K+M}(K)r_{K+M-1-j}}{r_0} - \frac{\sum_{i=j+2}^{K-1} R_i(K)r_{i-j}}{r_0} - \frac{\sum_{i=K}^{K+M-1} R_i(K)[\beta r_{i-j} + (1-\beta)r_{i-j-1}]}{r_0}, & 0 \leq j \leq K-2; \\ \frac{R_K(K)[1-\beta r_1 - (1-\beta)r_0] - R_{K+M}(K)r_M}{r_0} - \frac{\sum_{i=K+1}^{K+M-1} R_i(K)[\beta r_{i+1-K} + (1-\beta)r_{i-K}]}{r_0}, & j = K-1; \\ 1, & j = K+M, \beta > 0; \\ \frac{1-r_0}{\beta r_0}, & j = K+M-1, \beta > 0; \\ \frac{R_{j+1}(K)[1-\beta r_1 - (1-\beta)r_0]}{\beta r_0} - \frac{R_{K+M}(K)r_{K+M-1-j}}{\beta r_0} - \frac{\sum_{i=j+2}^{K+M-1} R_i(K)[\beta r_{i-j} + (1-\beta)r_{i-j-1}]}{\beta r_0}, & K \leq j \leq K+M-2, \beta > 0; \\ 0, & K+1 \leq j \leq K+M, \beta = 0; \\ 1, & j = K, \beta = 0. \end{cases}$$

Заметим, что

$$\pi_{j+1}(K+1) = [1 - \pi_0(K+1)] \pi_j(K), \quad 0 \leq j \leq K+M. \quad (3)$$

Обозначим через $q_j(K)$ средний доход, получаемый системой в состоянии j . Тогда

$$D(K) = \frac{1}{v} \sum_{j=0}^{K+M} \pi_j(K) q_j(K). \quad (4)$$

Справедливо следующее утверждение.

Утверждение 1. $D(K)$ — унимодальная функция по переменной $K > 0$. При этом

- (1) если $\inf_{K>0} G(K) < \sup_{K>0} D(K)$ и $d_{21} > 0$, то существует $K^* < \infty$, иначе, если $D(1) < G(1)$, то $K^* = \infty$;
- (2) если $D(1) \geq G(1)$, то $K^* = 1$.

Доказательство. Верны равенства:

$$q_j(K) = \sum_{m=0}^{\infty} \int_0^{\infty} r_m(v) q_j(K, m, v) dv, \quad j = 0, \dots, K+M, \quad (5)$$

где $q_j(K, m, v)$ — средний доход, получаемый системой в состоянии j при условии, что время пребывания в этом состоянии равно v и за время завершится полное обслуживание ровно m заявок (в дальнейшем условии $B_{m,v}$).

Ниже, чтобы получить явные выражения для $q_j(K, m, v)$, найдем соответствующие выражения для следующих величин:

- $W_{i,1}(m, v)$ — среднее суммарного времени занятия мест в основном накопителе;
- $W_{i,2}(m, v)$ — среднее суммарного времени занятия мест в дополнительном накопителе;
- $W_{i,пp}(m, v)$ — среднее время простоя прибора при условии $B_{m,v}$ и при условии, что число заявок в системе в момент поступления с учетом поступившей равно i .

Как известно [7], в СМО $G/M/1$ среднее время ожидания l -й по очереди заявки при условии $B_{m,v}$ равно $lv/(m+1)$. После несложных преобразований для заданного i получим равенства:

$$W_{i,1}(m, v) = \begin{cases} Kv, & 0 \leq m \leq i-K, i \geq K; \\ iv - \frac{mv}{2} - \frac{(i-K)(i-K+1)v}{2(m+1)}, & i \geq m > i-K, i \geq K; \\ \frac{i(i+1)v}{2(m+1)} - \frac{(i-K)(i-K+1)v}{2(m+1)}, & i \geq K, m > i; \\ iv - \frac{mv}{2}, & i < K, 0 \leq m < i; \\ \frac{i(i+1)v}{2(m+1)}, & i < K, m \geq i, \end{cases} \quad (6)$$

$$W_{i,2}(m, v) = \begin{cases} (i - K)v - \frac{mv}{2}, & 0 \leq m \leq i - K, i \geq K; \\ \frac{(i - K)(i - K + 1)v}{2(m + 1)}, & m > i - K, i \geq K; \\ 0, & i \leq K, \end{cases} \quad (7)$$

$$W_{i,sp}(m, v) = \begin{cases} \frac{(m - i + 1)v}{m + 1}, & m \geq i; \\ 0, & 0 \leq m < i. \end{cases} \quad (8)$$

Отсюда находим

$$q_j(K, m, v) = \begin{cases} d_0 - \tilde{q}_{j+1}(K, m, v), & j \leq K - 1; \\ \beta[d_0 - \tilde{q}_{j+1}(K, m, v)] - \\ - (1 - \beta)[d_1 + \tilde{q}_j(K, m, v)], & K \leq j \leq K + M - 1; \\ -\tilde{q}_{K+M-1}(K, m, v) - d_1, & j = K + M, \end{cases} \quad (9)$$

где

$$\tilde{q}_i(K, m, v) = d_{21}W_{i,1}(m, v) + d_{22}W_{i,2}(m, v) + d_3W_{i,sp}(m, v) + d_4v + d_5Kv + d_6Mv, \quad 1 \leq i \leq K + M.$$

По формуле (5), воспользовавшись (1), (6)–(9), получим:

$$q_j(K) = \begin{cases} d_0 + \tilde{q}_{j+1}(K), & 0 \leq j \leq K - 1; \\ \beta[d_0 + \tilde{q}_{j+1}(K)] - (1 - \beta)[d_1 - \tilde{q}_j(K)], & K \leq j \leq K + M - 1; \\ \tilde{q}_{K+M}(K), & j = K + M, \end{cases} \quad (10)$$

где

$$\begin{aligned} \tilde{q}_i(K) = & \frac{d_{21}}{\mu} \left[\frac{1}{2} \sum_{m=2}^{i+1} (m - 1)mr_m - \right. \\ & \left. - i \sum_{m=1}^{i+1} mr_m - \frac{1}{2} i(i + 1) \sum_{m=i+2}^{\infty} r_m \right] + \\ & + \theta_{(i-K)} \frac{d_{22} - d_{21}}{\mu} \left[\frac{1}{2} \sum_{m=2}^{i-K+1} (m - 1)mr_m - \right. \\ & \left. - (i - K) \sum_{m=1}^{i-K+1} mr_m - \right. \\ & \left. - \frac{(i - K)(i - K + 1)}{2} \sum_{m=i-K+2}^{\infty} r_m \right] - \\ & - \frac{d_3}{\mu} \sum_{m=i+1}^{\infty} (m - i)r_m - d_4\bar{v} - d_5K\bar{v} - d_6M\bar{v}; \quad (11) \end{aligned}$$

$\theta_{(i-K)}$ — функция Хевисайда переменной $(i - K)$, $1 \leq i \leq K + M - 1$; $\tilde{q}_{K+M}(K) = \tilde{q}_{K+M-1}(K) - d_1$.

Из (10) и (11) для величин $\Delta_j(K) = q_j(K + 1) - q_{j-1}(K)$, $1 \leq j \leq K + M - 1$, и $\tilde{\Delta}_i(K) = \tilde{q}_i(K + 1) - \tilde{q}_{i-1}(K)$, $2 \leq i \leq K + M$, следуют равенства:

$$\Delta_j(K) = \begin{cases} \tilde{\Delta}_{j+1}(K), & 1 \leq j \leq K; \\ \beta\tilde{\Delta}_{j+1}(K) + (1 - \beta)\tilde{\Delta}_j(K - 1), & K \leq j \leq K + M - 1; \\ \tilde{\Delta}_{K+M}, & j = K + M. \end{cases}$$

$$\begin{aligned} \tilde{\Delta}_i(K) = & \frac{d_3}{\mu} \sum_{m=i+2}^{\infty} r_m - \frac{d_{21}}{\mu} \sum_{m=1}^{i+1} mr_m - \\ & - \frac{d_{21}(i + 1)}{\mu} \sum_{m=i+2}^{\infty} r_m - d_5\bar{v}, \\ & K > 0, i = 1, \dots, K + M. \quad (12) \end{aligned}$$

Из (2) и (3) следует:

$$\begin{aligned} r_0\pi_0(K + 1) = & [1 - \pi_0(K + 1)] \left\{ \sum_{j=0}^{K-1} \pi_j(K) \sum_{m=j+2}^{\infty} r_m + \right. \\ & + \sum_{j=K}^{K+M-1} \pi_j(K) \left[\beta \sum_{m=j+2}^{\infty} r_m + (1 - \beta) \sum_{m=j+1}^{\infty} r_m \right] + \\ & \left. + \pi_{K+M}(K) \sum_{m=K+M+1}^{\infty} r_m \right\}. \quad (13) \end{aligned}$$

Введем обозначения $\Psi(K)$ и $H(K)$:

$$\begin{aligned} \Psi(K) = & \bar{v} - \frac{1}{\mu} \left\{ \sum_{j=0}^{K-1} \pi_j(K) \sum_{m=j+2}^{\infty} (m - j - 1)r_m + \right. \\ & + \sum_{j=K}^{K+M-1} \pi_j(K) \left[\beta \sum_{m=j+2}^{\infty} (m - j - 1)r_m + \right. \\ & \left. + (1 - \beta) \sum_{m=j+1}^{\infty} (m - j)r_m \right] + \\ & \left. + \pi_{K+M}(K) \sum_{m=K+M+1}^{\infty} (m - K - M)r_m \right\}. \end{aligned}$$

$$\begin{aligned} H(K) = & \sum_{j=0}^{K-1} \pi_j(K) \sum_{m=j+2}^{\infty} r_m + \\ & + \sum_{j=K}^{K+M-1} \pi_j(K) \left[\beta \sum_{m=j+2}^{\infty} r_m + (1 - \beta) \sum_{m=j+1}^{\infty} r_m \right] + \\ & + \pi_{K+M}(K) \sum_{m=K+M+1}^{\infty} r_m. \end{aligned}$$

Рассмотрим функцию:

$$G(K) = \frac{d_3 r_0}{\mu} - \frac{d_5 r_0 \bar{v}}{H(K)} - d_{21} r_0 \frac{\Psi(K)}{H(K)} + q_0(K). \quad (14)$$

Используя (3), (10)–(13), из (4) после несложных преобразований получим равенство:

$$D(K) - D(K + 1) = \pi_0(K + 1)[D(K) - G(K)]. \quad (15)$$

Отметим, что $G(K)$ — убывающая функция от переменной $K > 0$. Действительно, так как $\Psi(K)$ — среднее время занятости прибора в произвольно взятом состоянии системы (следует из определения $\Psi(K)$ и формулы для $W_{i,пр}(m, v)$) (7) и $\pi_0(K)$ убывает по $K > 0$ (доказательство аналогично доказательству этого свойства $\pi_0(K)$ в работе [7]), то $\Psi(K)$ возрастает, а $H(K) = r_0 \pi_0(K + 1) / (1 - \pi_0(K + 1))$ убывает по $K > 0$ и, как следует из (14), $G(K)$ убывает по $K > 0$.

Далее, так как функция $D(K)$ удовлетворяет условиям теоремы из [10] (т. е. для $D(K)$ имеет место равенство (15), $0 < \pi_0(K) < 1$, $G(K)$ — убывающая функция по $K > 0$), то непосредственно получаем доказательство унимодальности функции $D(K)$. Доказательство условий 1 и 2 утверждения 1 следует из свойств функций $D(K)$ и $G(K)$, а именно: если $D(K) < G(K)$, то $D(K) < D(K + 1)$ возрастает по $K > 0$, иначе $D(K) > D(K + 1)$ (см. доказательство теоремы в [10]).

Как следствие утверждения 1 получим простое правило вычисления оптимальной емкости основного накопителя: если $D(1) \geq D(2)$, то оптимальная емкость основного накопителя $K^* = 1$, иначе минимальное значение K^* , удовлетворяющее условию $D(K^*) \geq D(K^* + 1)$, является оптимальным.

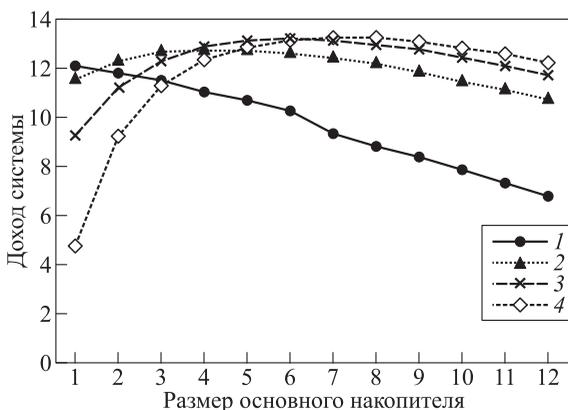


Рис. 2 Зависимость дохода системы от емкости основного накопителя

4 Пример

Здесь приводятся результаты вычислительных экспериментов, проведенных для СМО $M/H_n/1$ с функцией распределения времени обслуживания $A(t) = \sum_{i=1}^n f_i(1 - e^{-\delta_i t})$ при $n = 2$; $f_1 = 0,3$; $f_2 = 0,7$; $\delta_1 = 2$; $\delta_2 = 1$; $\gamma = 1$; $d_0 = 20$; $d_1 = 10$; $d_{21} = d_{22} = 0,5$; $d_3 = d_4 = 0,01$; $d_5 = 0,1$; $d_6 = 0$.

Кривые 1, 2, 3 и 4 на рис. 2 иллюстрируют соответствующие зависимости предельного дохода системы для четырех значений β : 1,0; 0,7; 0,5 и 0,1. Соответствующие оптимальные значения для K и $D(K)$ равны: 1, 4, 6, 7 и 11,95, 12,82, 13,16, 13,2.

На рис. 3 показаны графики, иллюстрирующие зависимость предельного максимального дохода системы от значения параметра M , для трех значений параметра β (соответствующие кривые 1, 2 и 3): 1,0, 0,5 и 0,1.

5 Заключение

В рассмотренной модели емкость дополнительного накопителя считается ограниченной, и согласно утверждению 1 функция $D(K)$ унимодальна для любого конечного значения $M \geq 0$. Так как сумма (4) сходится при $M \rightarrow \infty$ (ряд $(1/\bar{v}) \sum_{j=0}^K \pi_j(K) q_j(K) + (1/\bar{v}) \sum_{j=K+1}^{\infty} \pi_j(K) q_j(K)$ сходится) при условиях, что $d_6 = 0$ и что средняя длина очереди в дополнительном накопителе ограничена, то есть основание считать, что функция $D(K)$ унимодальна при выполнении этих условий и в случае $M = \infty$. Вычислительные эксперименты показывают, что при $d_6 = 0$ оптимальная емкость основного накопителя начиная с некоторого M остается постоянной (соответствующее максимальное значение дохода по K слабо зависит

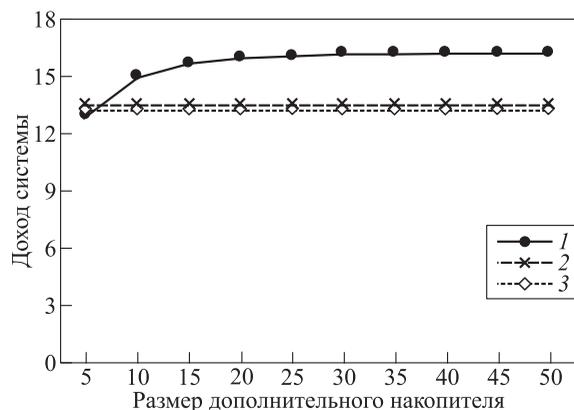


Рис. 3 Зависимость максимального дохода от емкости дополнительного накопителя

от M , когда $M \gg 0$) (см. рис. 3) и правило вычисления оптимального значения K , предложенное в конце разд. 3, может найти применение и в случае $M = \infty$.

Результаты данной работы могут также найти применение при исследовании более сложных моделей систем с повторными заявками (например, в качестве предельной модели для систем с ненулевым временем τ перехода из дополнительного накопителя в основной при $\tau \rightarrow 0$).

Литература

1. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
2. Ata B., Shneorson S. Dynamic control of an $M/M/1$ service system with adjustable arrival and service rates // Manage. Sci., 2006. Vol. 52. Iss. 11. P. 1778–1791.
3. Ghosh A. P., Weerasinghe A. P. Optimal buffer size for a stochastic processing network in heavy traffic // Queueing Syst., 2007. Vol. 55. Iss. 3. P. 147–159.
4. Ward A., Kumar S. Asymptotically optimal admission control of a queue with impatient customers // Math. Oper. Res., 2008. Vol. 33. Iss. 1. P. 167–202.
5. Koçağa Y. L., Ward A. R. Admission control for a multi-server queue with abandonment // Queueing Syst., 2010. Vol. 65. Iss. 3. P. 275–323.
6. Ghosh A. P., Weerasinghe A. P. Optimal buffer size and dynamic rate control for a queueing system with impatient customers in heavy traffic // Stoch. Proc. Appl., 2010. Vol. 120. Iss. 11. P. 2103–2141.
7. Агаларов Я. М., Агаларов М. Я., Шоргин В. С. Об оптимальном пороговом значении длины очереди в одной задаче максимизации дохода системы массового обслуживания типа $M/G/1$ // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 70–79.
8. Atar R., Lev-Ari A. Optimizing buffer size for the retrial queue: Two state space collapse results in heavy traffic // Queueing Syst., 2018. Vol. 90. Iss. 3–4. P. 225–255.
9. Агаларов Я. М., Коновалов М. Г. Доказательство унимодальности целевой функции в задаче порогового управления нагрузкой на сервер // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 2–6.
10. Агаларов Я. М. Признак унимодальности целочисленной функции одной переменной // Обзорение прикладной и промышленной математики, 2019. Т. 26. Вып. 1. С. 65–66.

Поступила в редакцию 15.04.20

OPTIMIZATION OF THE CAPACITY OF THE MAIN STORAGE IN $G/M/1/K$ QUEUEING SYSTEM WITH AN ADDITIONAL STORAGE DEVICE

Ya. M. Agalarov

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

Abstract: The problem of optimizing the capacity of the main storage device of a queueing system of the type $G/M/1/K$ with an additional storage device with the cost objective function is formulated taking into account the costs of the system associated with the loss of requests, storage of requests, maintenance of storage devices, and device downtime. A request arriving to the system is accepted into the main storage device if there is a free space; otherwise, according to the given probability distribution, it goes to the additional device if there is a free space. A request leaves the system (is lost) if both storage devices are full. If space is freed up in the main storage device, then one of the requests from the queue in the additional device immediately enters the queue in the main device. The unimodality of the objective function is proved and the procedure for finding the optimal capacity of the main storage device is proposed.

Keywords: queueing system; optimization; storage device; storage capacity

DOI: 10.14357/19922264200210

Acknowledgments

The reported study was partly funded by the Russian Foundation for Basic Research according to the research projects Nos. 18-07-00692, 19-07-00739, and 20-07-00804.

References

1. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
2. Ata, B., and S. Shneerson. 2006. Dynamic control of an $M/M/1$ service system with adjustable arrival and service rates. *Manage. Sci.* 52(11):1778–1791.
3. Ghosh, A. P., and A. P. Weerasinghe. 2007. Optimal buffer size for a stochastic processing network in heavy traffic. *Queueing Syst.* 55(3):1572–9443.
4. Ward, A., and S. Kumar. 2008. Asymptotically optimal admission control of a queue with impatient customers. *Math. Oper. Res.* 33(1):167–202.
5. Koçağa, Y. L., and A. R. Ward. 2010. Admission control for a multi-server queue with abandonment. *Queueing Syst.* 65(3):275–323.
6. Ghosh, A. P., and A. P. Weerasinghe. 2010. Optimal buffer size and dynamic rate control for a queueing system with impatient customers in heavy traffic. *Stoch. Proc. Appl.* 120(11):2103–2141.
7. Agalarov, Ya. M., M. Ya. Agalarov, and V. S. Shorgin. 2016. Ob optimal'nom porogovom znachenii dliny ocheredi v odnoy zadache maksimizatsii dokhoda sistemy massovogo obsluzhivaniya tipa $M/G/1$ [About the optimal threshold of queue length in particular problem of profit maximization in $M/G/1$ queueing system]. *Informatika i ee Primeneniya — Inform. Appl.* 10(2):70–79.
8. Atar, R., and A. Lev-Ari. 2018. Optimizing buffer size for the retrial queue: Two state space collapse results in heavy traffic. *Queueing Syst.* 90(3-4):225–255.
9. Agalarov, Ya. M., and M. G. Kononov. 2019. Dokazatel'stvo unimodal'nosti tselevoy funktsii v zadache porogovogo upravleniya nagruzkoy na server [Proof of the unimodality of the objective function in $M/M/N$ queue with threshold-based congestion control]. *Informatika i ee primeneniya — Inform. Appl.* 13(2):2–6.
10. Agalarov, Ya. M. 2019. Priznak unimodal'nosti tselochislennoy funktsii odnoy peremennoy [A sign of unimodality of an integer function of one variable]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Surveys Applied and Industrial Mathematics] 26(1):65–66.

Received April 15, 2020

Contributor

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

СТАТИСТИЧЕСКИЕ СВОЙСТВА ДВОИЧНЫХ НЕАВТОНОМНЫХ РЕГИСТРОВ СДВИГА С ВНУТРЕННИМ СУММИРОВАНИЕМ*

С. Ю. Мельников¹, К. Е. Самуйлов²

Аннотация: Проводится сравнение статистических и алгебраических свойств двоичных неавтономных регистров сдвига и регистров сдвига с внутренним суммированием, при шаге которых вектор состояния суммируется со своим сдвигом на один шаг. Доказан изоморфизм графов переходов этих автоматов. Показано, что при бернуллиевском случайном входе стационарное распределение на состояниях регистра с внутренним суммированием равномерное. Получен вид вероятностной функции этих регистров. Показано, что при определенных ограничениях на функцию выходов регистры с внутренним суммированием не чезарово-наследственные. Предъявлены входные последовательности, которые обладают свойством устойчивости относительных частот произвольных мультиграмм, в то время как выходные последовательности таким свойством не обладают.

Ключевые слова: автомат со случайным входом; регистр сдвига; граф де Брейна; вероятностная функция

DOI: 10.14357/19922264200211

1 Введение

При конструировании генераторов случайных последовательностей широко используются как линейные, так и нелинейные регистры сдвига с теми или иными элементами усложнения или обратной связи [1]. Это во многом обусловлено совокупностью «хороших» комбинаторных и структурных свойств графов де Брейна, описывающих преобразования информации в таких регистрах.

Пусть $f(x_1, x_2, \dots, x_n)$ — булева функция от n двоичных переменных, $n = 1, 2, \dots$. Будем рассматривать два автомата Мура, множество состояний каждого из которых есть $V_n = \{0, 1\}^n$, входной и выходной алфавиты — множества $\{0, 1\}$, выходом служит значение функции f от текущего состояния.

Автомат (регистр сдвига), который обозначим A_f , под действием входного символа $a_0 \in \{0, 1\}$ из состояния (a_1, a_2, \dots, a_n) переходит в состояние $(a_0, a_1, \dots, a_{n-1})$.

Регистром сдвига с внутренним суммированием назовем автомат, который под действием входного символа $a_0 \in \{0, 1\}$ из состояния (a_1, a_2, \dots, a_n) переходит в состояние $(a_0 \oplus a_1, a_1 \oplus a_2, \dots, a_{n-1} \oplus a_n)$, где \oplus — суммирование по модулю 2. Такой автомат будем обозначать A_f^\oplus .

Алгебраические и статистические свойства обычных регистров хорошо изучены. Теоретико-автоматные свойства регистров, аналогичных A_f^\oplus , в автономном случае рассматривались в [2], а во-

просы их аппаратной реализации — в [3]. В настоящей работе доказывается изоморфизм графов переходов и проводится сравнение статистических свойств автоматов A_f и A_f^\oplus .

2 Связь графов переходов автоматов A_f и A_f^\oplus

Графом переходов автомата A_f , как нетрудно видеть, служит ориентированный граф с множеством вершин V_n , с дугами, выходящими из вершин (a_1, a_2, \dots, a_n) и заходящими в вершины $(a_0, a_1, a_2, \dots, a_{n-1})$, $a_i \in \{0, 1\}$. Это хорошо известный граф де Брейна, который будем обозначать G_n .

Граф переходов регистра сдвига с внутренним суммированием с накопителем размера $n = 1, 2, \dots$ — это ориентированный граф с множеством вершин V_n , содержащий дуги, выходящие из вершин (a_1, a_2, \dots, a_n) , и заходящие в вершины $(a_0 \oplus a_1, a_1 \oplus a_2, \dots, a_{n-1} \oplus a_n)$, $a_i \in \{0, 1\}$. Такой граф будем обозначать Γ_n . Графы G_3 и Γ_3 представлены на рисунке.

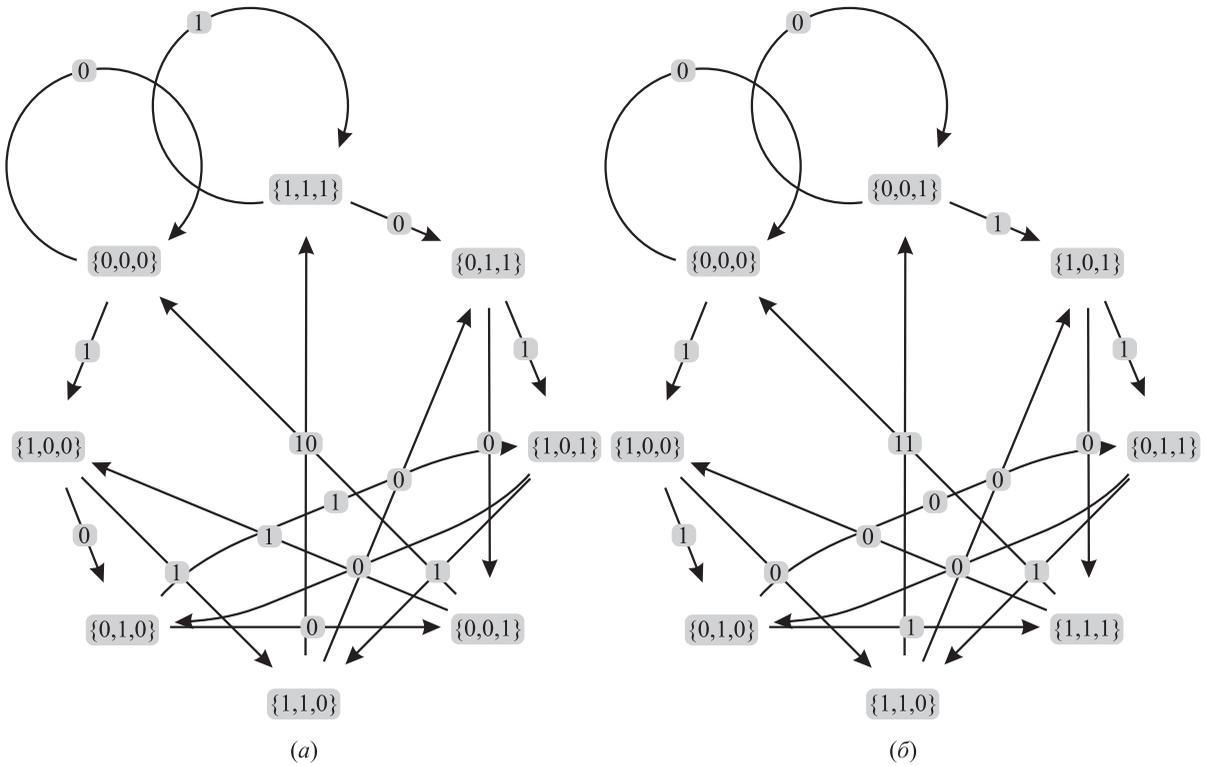
Изоморфизм графов G_n и Γ_n может быть задан с помощью комбинаторных формул обращения [4].

Утверждение 1. Пусть (a_1, a_2, \dots, a_n) , $a_i \in \{0, 1\}$, — вершина графа G_n , (b_1, b_2, \dots, b_n) , $b_i \in \{0, 1\}$, — вер-

* Публикация подготовлена при поддержке Программы РУДН «5-100» (К. Е. Самуйлов, постановка задачи) и при финансовой поддержке РФФИ (проекты 18-00-01555, 18-00-01685 и 19-07-00933).

¹ Российский университет дружбы народов, melnikov@linfotech.ru

² Российский университет дружбы народов, ksam@sci.pfu.edu.ru



Графы G_3 (а) и Γ_3 (б)

шина графа Γ_n . Отображение $\varphi_n : V_n \rightarrow V_n$, определяемое формулой $\varphi_n(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)$, где

$$b_t = \sum_{j=t}^n \binom{n-t}{n-j} a_j \pmod{2}, \quad t = 1, \dots, n,$$

является изоморфизмом графов G_n и Γ_n .

Доказательство. Отображение φ_n — это линейное преобразование пространства V_n с треугольной матрицей, составленной из биномиальных коэффициентов:

$$A = \begin{pmatrix} \binom{n-1}{n-1} & \binom{n-1}{n-2} & \dots & \binom{n-1}{0} \\ 0 & \binom{n-2}{n-2} & \dots & \binom{n-2}{0} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \binom{0}{0} \end{pmatrix} \pmod{2}.$$

Поскольку $\text{Det}(A) = 1$, отображение φ_n биективно.

Осталось показать, что данное отображение соседние вершины графа G_n переводит в соседние вершины графа Γ_n . Пусть вершины (a_1, a_2, \dots, a_n) и $(a_0, a_1, \dots, a_{n-1})$ соединены дугой в графе G_n . Покажем, что их образы $\varphi_n(a_1, a_2, \dots, a_n)$ и $\varphi_n(a_0, a_1, \dots, a_{n-1})$ также соединены дугой в графе Γ_n .

Рассмотрим k -ю и $(k+1)$ -ю координаты первого вектора и $(k+1)$ -ю координату второго, $k = 1, 2, \dots, n-1$:

$$[\varphi_n(a_1, a_2, \dots, a_n)]_k = \sum_{i=k}^n \binom{n-k}{n-i} a_i;$$

$$[\varphi_n(a_1, a_2, \dots, a_n)]_{k+1} = \sum_{i=k+1}^n \binom{n-k-1}{n-i} a_i;$$

$$[\varphi_n(a_0, a_1, \dots, a_{n-1})]_{k+1} = \sum_{i=k+1}^n \binom{n-k-1}{n-i} a_{i-1},$$

Покажем, что для $k = 1, 2, \dots, n-1$ выполняется соотношение:

$$[\varphi_n(a_1, a_2, \dots, a_n)]_k + [\varphi_n(a_1, a_2, \dots, a_n)]_{k+1} = [\varphi_n(a_0, a_1, \dots, a_{n-1})]_{k+1} \pmod{2}.$$

В самом деле:

$$\begin{aligned} & [\varphi_n(a_1, a_2, \dots, a_n)]_k + [\varphi_n(a_1, a_2, \dots, a_n)]_{k+1} = \\ &= \sum_{i=k}^n \binom{n-k}{n-i} a_i \pmod{2} + \\ &+ \sum_{i=k+1}^n \binom{n-k-1}{n-i} a_i \pmod{2} = \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=k}^n \left(\binom{n-k}{n-i} + \binom{n-k-1}{n-i} \right) a_i \pmod{2} = \\
 &= \sum_{i=k}^n \binom{n-k-1}{n-i-1} a_i \pmod{2} = \\
 &= \sum_{j=k+1}^n \binom{n-k-1}{n-j} a_{j-1} \pmod{2} = \\
 &= [\varphi_n(a_0, a_1, \dots, a_{n-1})]_{k+1}.
 \end{aligned}$$

Здесь использовалось соотношение:

$$\binom{a}{b} + \binom{a-1}{b} = \binom{a-1}{b-1} \pmod{2},$$

вытекающее из известного комбинаторного тождества

$$\binom{a}{b} = \binom{a-1}{b} + \binom{a-1}{b-1}.$$

Пример. $n = 4$. Изоморфизм $G_4 \cong \Gamma_4$ описывается следующим образом:

$$A = \begin{pmatrix} 1 & 3 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \pmod{2}; \quad \varphi_4 = \begin{pmatrix} a_1 \oplus a_2 \oplus a_3 \oplus a_4 \\ a_2 \oplus a_4 \\ a_3 \oplus a_4 \\ a_4 \end{pmatrix}.$$

Утверждение 2. Справедливо равенство $\varphi_n^{-1} = \varphi_n$, $n = 1, 2, \dots$

Доказательство. В [5, п. 2.2] показано, что матрица, обратная к

$$A = \left(\binom{n-1-i}{n-1-j} \right)_{i,j=0}^{n-1},$$

имеет вид:

$$A^{-1} = \left((-1)^{i+j} \binom{n-1-i}{n-1-j} \right)_{i,j=0}^{n-1}.$$

По модулю два эти матрицы равны.

В [6] доказано, что у графа G_n существуют ровно два автоморфизма, тождественный и соответствующий инверсии координат вершин.

Утверждение 3. Пусть ϕ и η — преобразования линейного пространства V_n вида:

$$\begin{aligned}
 \phi(x_1, x_2, \dots, x_n) &= (\overline{x_1}, \overline{x_2}, \dots, \overline{x_n}); \\
 \eta(x_1, x_2, \dots, x_n) &= (x_1, x_2, \dots, x_{n-1}, \overline{x_n}).
 \end{aligned}$$

Тогда

$$\begin{aligned}
 \phi: G_n &\rightarrow G_n \text{ — автоморфизм графа } G_n; \\
 \eta: \Gamma_n &\rightarrow \Gamma_n \text{ — автоморфизм графа } \Gamma_n,
 \end{aligned}$$

диаграмма

$$\begin{array}{ccc}
 G_n & \cong_{\varphi_n} & \Gamma_n \\
 \downarrow \phi & & \downarrow \eta \\
 G_n & \cong_{\varphi_n} & \Gamma_n
 \end{array}$$

коммутативна.

Доказательство. Очевидно, что инверсия координат вершин задает автоморфизм графа де Брейна. Для доказательства коммутативности диаграммы и того, что отображение $\eta: \Gamma_n \rightarrow \Gamma_n$ есть автоморфизм графа Γ_n , равенство $\phi\varphi_n = \varphi_n\eta$ проверяется непосредственно.

3 Вероятностная функция двоичных регистров сдвига с внутренним суммированием

Вероятностная функция [7] конечного связного автомата определяется как действительная функция, заданная на множестве стохастических векторов, соответствующих возможным полиномиальным распределениям на входном алфавите. Значение функции есть предел относительной частоты встречаемости знака в выходной последовательности автомата в предположении, что на его вход поступает последовательность независимых одинаково распределенных по заданной полиномиальной схеме случайных величин. В двоичном случае входная последовательность бернуллиевская с параметром p , $0 < p < 1$. В качестве знака выходной последовательности рассматривается единица. Для вывода формулы вероятностной функции автомата A_f^{\oplus} получим стационарное распределение на состояниях автомата. Определим случайное блуждание на графе Γ_n следующим образом. Начальная вершина выбирается в соответствии с некоторой полиномиальной схемой на множестве V_n . Шаг блуждания проходит по одной из исходящих из нее дуг. Предположим, что переходы из вершины в вершину независимы и вероятность шага блуждания из вершины (a_1, a_2, \dots, a_n) в вершину $(1 \oplus a_1, a_1 \oplus a_2, \dots, a_{n-1} \oplus a_n)$ равна p , а в вершину $(a_1, a_1 \oplus a_2, \dots, a_{n-1} \oplus a_n)$ равна $1 - p$, $0 < p < 1$.

Утверждение 4. Стационарное распределение описанного случайного блуждания имеет вид:

$$P(a_1, a_2, \dots, a_n) = \frac{1}{2^n}.$$

Доказательство. Покажем, что матрица переходных вероятностей марковской цепи, описывающей рассматриваемое случайное блуждание, дважды стохастическая. Ориентированный

граф Γ_n — регулярный степени 2, т. е. из каждой вершины выходят две дуги и в каждую вершину заходят две дуги, что вытекает из доказанного выше изоморфизма Γ_n и G_n . Рассмотрим вершину (a_1, a_2, \dots, a_n) . В нее можно попасть из двух разных вершин, которые обозначим (b_1, b_2, \dots, b_n) (по дуге, которая помечена символом b_0) и (c_1, c_2, \dots, c_n) (по дуге, которая помечена символом c_0). Тогда справедливо соотношение:

$$(b_1, b_2, \dots, b_n) \oplus (b_0, b_1, \dots, b_{n-1}) = (c_1, c_2, \dots, c_n) \oplus (c_0, c_1, \dots, c_{n-1}).$$

Поскольку $(b_1, b_2, \dots, b_n) \neq (c_1, c_2, \dots, c_n)$, то из этого соотношения нетрудно вывести, что $b_0 \neq c_0$. Это означает, что одна из дуг, заходящих в вершину (a_1, a_2, \dots, a_n) , помечена символом 1, а другая — символом 0. Следовательно, в матрице переходных вероятностей в столбце, соответствующем вершине (a_1, a_2, \dots, a_n) , расположено ровно два ненулевых элемента: p и $1-p$. Поэтому данная матрица — дважды стохастическая, что означает равномерность стационарного распределения.

Утверждение 5. Вероятностная функция регистра сдвига с внутренним суммированием с выходной функцией f имеет вид:

$$P_{A_f^\oplus}(p) = \frac{\|f\|}{2^n},$$

где $\|f\|$ — вес функции f :

$$\|f\| = \sum_{(x_1, x_2, \dots, x_n) \in V_n} f(x_1, x_2, \dots, x_n).$$

4 Чезарово-наследственность регистра сдвига с внутренним суммированием

В [8] дано следующее определение: слово называется чезаровским для бесконечной последовательности, если определен предел относительной частоты его встречаемости в растущих начальных отрезках этой последовательности. Последовательность над некоторым алфавитом называется чезаровской, если произвольное слово над этим алфавитом является для нее чезаровским. Конечный автомат называют чезарово-наследственным, если из любого начального состояния чезаровские последовательности во входном алфавите он перерабатывает в чезаровские последовательности в выходном алфавите. В [8] показано, что при любой функции f автомат A_f чезарово-наследственный.

Утверждение 6. Если для булевой функции $f(x_1, x_2, \dots, x_n)$ выполнено условие

$$f(0, 0, \dots, 0) \neq f(0, 0, \dots, 0, 1),$$

то A_f^\oplus не чезарово-наследственный автомат.

Доказательство. В графе автомата A_f , как нетрудно видеть, существует ровно один цикл, движение по которому происходит при подаче на вход последовательности, состоящей только из нулей, это петля в нулевой вершине. Для автомата A_f^\oplus таких циклов существует несколько. В самом деле, поскольку при нулевом символе на входе автомата состояние (a_1, a_2, \dots, a_n) переходит в состояние $(a_1, a_1 \oplus a_2, \dots, a_{n-1} \oplus a_n)$, условие наличия цикла длины (петли) выглядит так: $a_2 = a_1 \oplus a_2, a_3 = a_2 \oplus a_3, \dots, a_n = a_{n-1} \oplus a_n$. Отсюда получаем равенства: $a_i = 0, i = 1, 2, \dots, n-1, a_n = 0, 1$. Поэтому в графе автомата A_f^\oplus имеются две петли, движение по которым происходит при подаче на вход: петля в вершине $(0, 0, \dots, 0)$ и петля в вершине $(0, 0, \dots, 0, 1)$ (см. рисунок).

Пусть теперь выполнено условие утверждения. Из него следует, что при движении по одной из указанных выше петель выходная последовательность автомата A_f^\oplus состоит из единиц, а при движении по другой эта последовательность состоит из нулей. Поскольку диаметр графа Γ_n равен n , из состояния $(0, 0, \dots, 0)$ в состояние $(0, 0, \dots, 0, 1)$ и наоборот можно перейти за не более чем n шагов. Через ξ_{01} и ξ_{10} обозначим входные последовательности, которые обеспечивают эти переходы. Рассмотрим теперь бесконечную входную последовательность

$$\chi = \xi_0 0^{k_1} \xi_{01} 0^{k_2} \xi_{10} 0^{k_3} \xi_{01} 0^{k_4} \xi_{10} \dots,$$

где ξ_0 — входная последовательность, переводящая автомат A_f^\oplus из заданного состояния в состояние $(0, 0, \dots, 0)$; 0^{k_i} — последовательность, состоящая из k_i нулей, k_i — целые числа, $i = 1, 2, \dots$. Выходной последовательностью, очевидно, будет

$$\gamma = \zeta_0 0^{k_1} \zeta_{01} 1^{k_2} \zeta_{10} 0^{k_3} \zeta_{01} 1^{k_4} \zeta_{10} \dots,$$

где ζ_0, ζ_{01} и ζ_{10} — некоторые двоичные последовательности, длина каждой из которых не превосходит n ; 1^{k_j} — последовательность, состоящая из k_j единиц.

Нетрудно убедиться в том, что при $k_i = 2^{2^i}$ последовательность ξ чезаровская, а для последовательности γ не существует предела относительной частоты встречаемости единицы в растущих начальных отрезках. Поэтому она не чезаровская и автомат A_f^\oplus не чезарово-наследственный.

5 Заключение

Регистры сдвига различных типов широко применяются в качестве узлов генераторов случайных последовательностей. В работе исследованы алгебраические и статистические свойства семейства двоичных регистров сдвига с внутренним суммированием. Доказан изоморфизм графов переходов регистров сдвига с внутренним суммированием и обычных регистров. Доказано, что, в отличие от обычного регистра сдвига, при случайном бернуллиевском входе вероятностное распределение на состояниях регистра с внутренним суммированием равномерное. Показано, что рассматриваемые регистры не обладают свойством цезарово-наследственности, которым обладают обычные регистры сдвига.

Литература

1. Грушо А. А., Применко Э. А., Тимонина Е. Е. Теоретические основы компьютерной безопасности. — М.: Академия, 2009. 267 с.
2. Golomb S. W. Shift register sequences. — Laguna Hills, CA, USA: Aegean Park Press, 1981. 247 p.
3. The VLSI handbook / Ed. W.-K. Chen. — 2nd ed. — Chicago, IL, USA: CRC Press, 2006. 2320 p.
4. Сачков В. Н. Курс комбинаторного анализа. — Ижевск: НИЦ РХД, 2013. 336 с.
5. Риордан Дж. Комбинаторные тождества / Пер. с англ. — М.: Наука, 1982. 256 с. (Riordan J. 1968. Combinatorial identities. — New York, NY, USA: Wiley. 256 p.)
6. Liu M. Homomorphisms and automorphisms of 2-D de Bruijn–Good graphs // Discrete Math., 1990. Vol. 85. Iss. 1. P. 105–109.
7. Melnikov S. Yu., Samouylov K. E. The recognition of the output function of a finite automaton with random input // Distributed Computer and Communication Networks: 21st Conference (International) Revised Selected Papers / Eds. V. M. Vishnevskiy, D. V. Kozyrev. — Communications in computer and information science ser. — Springer, 2018. Vol. 919. P. 525–531. doi: 10.1007/978-3-319-99447-5_45.
8. Мельников С. Ю. О переработке конечными автоматами цезаровских последовательностей // Лесной вестник, 2004. № 1(32). С. 169–174.

Поступила в редакцию 14.04.20

STATISTICAL PROPERTIES OF BINARY NONAUTONOMOUS SHIFT REGISTERS WITH INTERNAL XOR

S. Yu. Melnikov and K. E. Samouylov

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

Abstract: The statistical and algebraic properties of binary nonautonomous shift registers and shift registers with internal XOR are compared, during which the state vector is summed with its one-step shift. The isomorphism of transition graphs of these automata is proved. It is shown that, with a Bernoulli random input, the stationary distribution of the register states with internal XOR is uniform. The form of the probability function of these registers is obtained. It is shown that, under certain conditions on the output function, registers with internal XOR are not Cesaro-hereditary. The authors show input sequences that possess the property of stability of the relative frequencies of arbitrary multigrams, while output sequences do not have this property.

Keywords: random input automata; shift register; de Bruijn graph

DOI: 10.14357/19922264200211

Acknowledgments

The publication has been prepared with the support of the “RUDN University Program 5-100” (recipient K. E. Samouylov, problem statement). The reported study was partially funded by the Russian Foundation for Basic Research, projects 18-00-01555, 18-00-01685, and 19-07-00933.

References

1. Grusho, A. A., E. A. Primenko, and E. E. Timonina. 2009. *Teoreticheskie osnovy komp'yuternoy bezopasnosti* [Theoretical foundations of computer security]. Moscow: Akademiya. 267 p.
2. Golomb, S. W. 1981. *Shift register sequences*. Laguna Hills, CA: Aegean Park Press. 247 p.
3. Chen, W.-K., ed. 2006. *The VLSI handbook*. 2nd ed. Chicago, IL: CRC Press. 2320 p.

4. Sachkov, V. N. 2013. *Kurs kombinatornogo analiza* [Combinatorial analysis course]. Izhevsk: NITS RKHD. 336 p.
5. Riordan, J. 1968. *Combinatorial identities*. New York, NY: Wiley. 256 p.
6. Liu, M. 1990. Homomorphisms and automorphisms of 2-D de Bruijn-good graphs. *Discrete Math.* 85(1):105–109.
7. Melnikov, S. Yu., and K. E. Samouylov. 2018. The recognition of the output function of a finite automaton with random input. *Distributed Computer and Communication Networks: 21st Conference (International) Revised Selected Papers*. Eds. V. M. Vishnevskiy and D. V. Kozyrev. Communications in computer and information science ser. Springer. 919:525–531.
8. Melnikov, S. Yu. 2004. O pererabotke konechnymi avtomatami chezarovskikh posledovatel'nostey [On the finite automata with Cesaro sequences input]. *Lesnoy Vestnik* [Forestry Bull.] 1(32):169–174.

Received April 14, 2020

Contributors

Melnikov Sergey Yu. (b. 1963) — Candidate of Science (PhD) in physics and mathematics, doctoral student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; melnikov@linfotech.ru

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ksam@sci.pfu.edu.ru

ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ СЕРИЙ ДАННЫХ НА ОСНОВЕ МНОГОМЕРНЫХ РЕФЕРЕНСНЫХ РЕГИОНОВ

М. П. Кривенко¹

Аннотация: Рассматриваются последовательные процедуры анализа серий данных. Развивается подход, при котором набор многомерных характеристик некоторого объекта, изменяющийся во времени, представляется как единый вектор наблюдаемых значений. За счет увеличения размерности данных удается получить единую картину описания объектов, учесть объективно существующую зависимость отдельных наблюдений, смоделировать изменения во времени. В основу решения задач классификации положено использование многомерных референсных областей. Предлагаются три варианта процедур обработки данных, исследуются их свойства, вырабатываются рекомендации по их применению на практике. Для демонстрации возможностей этих процедур рассматривается задача ранней диагностики рака с использованием биомаркера PSA (prostate-specific antigen — простат-специфический антиген). Указаны особенности применения последовательных методов анализа серий данных, сформированы рекомендации по их эффективному использованию, выявлены преимущества консолидирующего подхода в анализе данных.

Ключевые слова: серии данных; консолидирующий подход; последовательные процедуры; анализ PSA

DOI: 10.14357/19922264200212

1 Введение

В [1] рассматривался метод представления результатов наблюдений отдельных серий данных в виде единой совокупности данных. Подобная консолидация фрагментированных по времени и по характеристикам наблюдений подключает преимущества обработки многомерных данных, а также позволяет избежать потери общей эффективности статистического вывода из-за декомпозиции общей задачи анализа данных при лонгитюдном подходе. Проведенные эксперименты продемонстрировали продуктивность предлагаемого подхода, полученные при этом результаты создали предпосылки для постановки и решения задачи упреждения результатов обследований.

Основная цель исследования серий данных — описать изменения во времени и охарактеризовать влияющие на них факторы. Желание прогнозировать при этом процессы, порождающие серии наблюдений, естественно приводит к постановкам задач последовательного анализа данных. Соответствующая экономия статистического материала приобретает особую жизненную остроту в области медицинской диагностики.

Тема последовательного анализа была инициирована А. Вальдом [2] в ответ на возникшую потребность создания более эффективных процедур

выборочного контроля. Был введен в рассмотрение и проанализирован последовательный критерий отношения вероятностей (sequential probability ratio test — SPRT) для двух простых гипотез $H_0 : f = f_0$ против $H_1 : f = f_1$, использующий последовательность независимых наблюдений X_1, X_2, \dots , имеющих общую плотность распределения f . Критерий прекращает обработку данных на шаге, когда появится первое наблюдаемое значение x_n , при котором $r_n \leq B$ или $r_n \geq A$, где $0 < B < 1 < A$, $r_n = \prod_{i=1}^n [f_1(X_i)/f_0(X_i)]$. По определению $n = \infty$, если $B < r_n < A$ для всех n . Критерий SPRT принимает гипотезу H_0 , если $r_n \leq B$, и H_1 , если $r_n \geq A$. Была обоснована оптимальность этого критерия в смысле ожидаемого размера используемой выборки, даны приближенные формулы для нахождения границ A и B при заданных ошибках 1-го и 2-го рода.

Практически сразу же после появления фундаментальных работ по поводу SPRT было признано, что последовательное тестирование гипотез оказывается полезным инструментом в биомедицинских исследованиях. Впоследствии был предложен производный от SPRT и его вариантов подход — «повторный критерий значимости» (repeated significance test — RST). Основная мотивация при создании RST состояла в том, чтобы в ходе испытаний многократно применять процедуру, эффектив-

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, mkcrivenko@ipiran.ru

ную для отдельного эксперимента. Однако общий уровень значимости α^* , достижимый на некоторой стадии обработки данных, мог оказаться существенно больше, чем α для отдельного эксперимента. Поэтому изменения SPRT в виде RST включали в себя поиск подходящих скорректированных областей принятия решений таким образом, чтобы фактический уровень значимости не превышал α . В виде конкретных результатов речь шла об отдельных распределениях с фиксированными значениями параметров. Историческая справка о зарождении последовательных процедур и их развитии дана, например, в отдельных статьях энциклопедии [3], обширная библиография и представление о современном состоянии в области последовательного анализа данных могут быть почерпнуты из [4, 5].

Объединить потребности обработки серий многомерных данных и сформировавшиеся на данный момент возможности последовательного анализа преимущественно одномерных данных предлагается в данной статье за счет использования многомерных референсных регионов (MRR — multivariate reference region) [6].

Напомним, если плотность распределения референсных значений есть $f(\mathbf{y})$, то MMR есть область $A_t = \{\mathbf{y} \in \mathcal{R}^d | f(\mathbf{y}) \geq t\}$ для некоторого порогового значения t . Для нормального распределения это обычный эллипсоид равной вероятности. Если задается вероятность $(1 - \alpha_0)$ попадания в A_t , то пороговое значение t есть решение уравнения

$$\int_{A_t} f(\mathbf{u}) d\mathbf{u} = 1 - \alpha_0,$$

получить которое аналитически в случае произвольной плотности распределения вряд ли возможно. Здесь присутствуют две проблемы: вычисление многомерного интеграла и зависимость области интегрирования от неизвестного значения. Для решения их предлагается привлечь метод моделирования.

Сгенерируем выборку из $f(\mathbf{y})$, которую обозначим как $\mathbf{Y}^f = \{\mathbf{y}_1^f, \dots, \mathbf{y}_m^f\}$. Для оценки $\int_{A_t} f(\mathbf{u}) d\mathbf{u}$ используем отношение:

$$\begin{aligned} \frac{|\{y_i^f | y_i^f \in A_t\}|}{m} &= \frac{|\{y_i^f | f(y_i^f) \geq t\}|}{m} = \\ &= 1 - \frac{|\{y_i^f | f(y_i^f) < t\}|}{m} = 1 - F_m(t), \end{aligned}$$

где $F_m(t)$ — эмпирическая функция распределения случайной величины $f(\mathbf{y})$, т. е. случайной величины, представляющей собой результат преобразова-

ния с помощью функции $f(\cdot)$ случайной величины, имеющей плотность распределения $f(\mathbf{u})$.

Таким образом, искомая оценка t^* должна удовлетворять уравнению $F_m(t^*) = \alpha_0$ и может быть получена как непараметрическая оценка квантиля порядка α_0 из распределения $F_m(\cdot)$. Заметим, что для такой оценки можно указать доверительный интервал.

Для построения MRR необходимо знать распределение данных. При описании области точек высокой плотности можно обратиться к параметрической модели смеси нормальных распределений, имеющей плотность распределения

$$f(\mathbf{u}) = \sum_{j=1}^k p_j \varphi(\mathbf{u}, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j).$$

Если $\hat{f}(\mathbf{u})$ — оценка смеси, то t^* строится следующим образом: генерация выборки $\{\mathbf{y}_1^f, \dots, \mathbf{y}_m^f\}$ из \hat{f} , подсчет значений $\hat{f}(\mathbf{y}_i^f)$, непараметрическая оценка квантиля порядка α_0 .

Использование MRR высокой плотности для диагностирования сводится к реализации так называемого слабого критерия значимости. Для наблюдаемого элемента серии \mathbf{x} нулевая гипотеза H_0 заключается в том, что $\mathbf{x} \in A_t$; статистика критерия есть $\hat{f}(\mathbf{x})$ и решение о принадлежности критической области A_t принимается при больших значениях $\hat{f}(\mathbf{x})$. Но из-за наличия зависимостей в сериях получающийся для ее элементов набор значений не является, вообще говоря, последовательностью испытаний Бернулли. Поэтому применение SPRT должно осуществляться с осторожностью.

2 Процедуры последовательного анализа

Рассмотрим задачу выбора на основании поступающих данных одного из двух решений (диагнозов). Одно из них соответствует нормальному состоянию объекта исследования (нулевая гипотеза H_0 и диагноз «здоров»), другое — отклонению от нормы (конкурирующая гипотеза H_1). Основой для построения статистического вывода служит слабый критерий значимости. Для задач медицинской диагностики в этом случае речь идет об использовании референсных значений в виде MRR.

В основе построения MRR лежит модель смеси нормальных многомерных распределений, которая может эффективно использоваться как средство аппроксимации реальных данных и при этом быть доступной с точки зрения теоретического анализа. Выбор структурного параметра модели (число

элементов смеси) предлагается искать, комбинируя применение информационных критериев или простого перебора для формирования начальных приближений с последующим уточнением получающихся решений [7]. Принимая во внимание, что статистический вывод строится на основе слабого критерия значимости, качество полученных результатов можно охарактеризовать через оценку по обучающей выборке значения α_0 .

Рассмотрим возможные процедуры последовательного анализа отдельных серий.

Обычная обработка (процедура R). В медицинской диагностике распространено простое правило использования референсных значений: ожидать такое первое наблюдение, которое с этими значениями не согласуется. Таким образом, последовательное тестирование завершается отклонением H_0 , если в пределах серии встретилось событие $x_n \notin A_t$, или ее принятием в противном случае. Недостаток данной процедуры заключается в том, что она приводит к большим ошибкам 1-го рода, чем в случае одного наблюдения. Если это существенно, то можно снизить уровень значимости при построении A_t . Кроме того, при реализации процедуры R целесообразно учесть рекомендации [1]: обрабатывать данные в совокупности, т. е. перейти от последовательности отдельных наблюдений x_1, x_2, \dots, x_s к фрагментированным данным $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{\hat{s}}$, где $\hat{x}_i = (x_{(i-1)g+1}, x_{(i-1)g+2}, \dots, x_{(i-1)g+g})$ для $i = 1, \dots, \hat{s} - 1$ и $\hat{x}_{\hat{s}}$ может оказаться с неполным набором компонент.

Обработка по Вальду (процедура W). Проверки принадлежности элементов серии области MRR приводят к последовательности успехов или неуспехов. Если принять предположение, что при этом формируется последовательность испытаний Бернулли b_1, \dots, b_l с вероятностью появления успеха p то можно обратиться к последовательному анализу Вальда и построить процедуру, обеспечивающую гарантированные ошибки при обработке серий.

В этом случае необходимо проверить нулевую гипотезу $H_0 : p = p_0$ против конкурирующей гипотезы $H_\delta : p = p_0 + \delta$, где $p_0, p_0 + \delta \in (0, 1)$, с вероятностями α_W и β_W ошибок 1-го и 2-го рода соответственно, $\alpha_W, \beta_W < 0,5$. Если через r_m обозначить число успехов (успеху соответствует значение 1) в m испытаниях, т. е. $r_m = \sum_{i=1}^m b_i$, $m = 1, \dots, m_{\max}$, то в соответствии с общими принципами [2] и поправкой на ограниченность длины серии (значение m_{\max} для каждой серии свое) процедура SPRT основывается на отношении правдоподобия и в случае $\delta < 0$ включает следующие шаги.

1. Положить $m = 1$.
2. Если

$$r_m \geq \frac{1}{c_\delta} \ln \frac{\beta_W}{1 - \alpha_W} - \frac{1}{c_\delta} \ln \frac{1 - (p_0 + \delta)}{1 - p_0} m,$$

то принять гипотезу H_0 ; если

$$r_m \leq \frac{1}{c_\delta} \ln \frac{1 - \beta_W}{\alpha_W} - \frac{1}{c_\delta} \ln \frac{1 - (p_0 + \delta)}{1 - p_0} m,$$

то принять гипотезу H_δ ; если принята одна из гипотез, то завершить обработку.

3. Если $m = m_{\max}$, то принять гипотезу H_0 и завершить обработку; в противном случае увеличить m на единицу и перейти к шагу 2, т. е. продолжить обработку серии.

Для c_δ верно представление

$$c_\delta = \frac{\ln((p_0 + \delta)(1 - p_0))}{p_0(1 - (p_0 + \delta))}.$$

Изменение постановки задачи привело к необходимости задания значений дополнительных параметров p_0, δ, α_W и β_W .

Обработка с консолидацией (процедура С). Две описанные выше процедуры основываются на единственном для всех элементов серии MRR. Но в рассматриваемой ситуации, когда регион строится с помощью обучающей выборки (оценивание параметров смеси нормальных распределений и уровня высокой плотности), в принципе, ничто не мешает на каждом шаге последовательной обработки элементов серий опираться на различные MRR. Таким образом, обработка включает шаги: допустить, что для серии длины s выполняется H_0 , далее для $i = 1, \dots, s$ осуществить проверку и, если $(x_1, \dots, x_i) \notin A_t(\alpha_0, i)$, то отклонить H_0 и завершить обработку. Здесь область $A_t(\alpha_0, i)$ построена для заданного α_0 и объединенной совокупности данных из i элементов серии.

3 Эксперименты

Исследовалась возможность сокращения объема данных при обработке серий измерений уровня PSA при раннем диагностировании рака предстательной железы. Для экспериментов использовался набор данных [8]. Результаты его предварительного анализа [1] показали, что целесообразно рассматривать только один маркер — общий PSA (tPSA — total PSA). Для каждого измерения уровня PSA фиксировался промежуток времени до установки окончательного диагноза. Всего имелось 683 наблюдения, каждое из которых включало идентификатор пациента, окончательный диагноз (отсутствие или наличие рака предстательной железы, далее — $D = 0/1$),

промежутков времени до установки окончательного диагноза, уровни tPSA, возраст пациента. Данные распались на серии наблюдений для отдельных пациентов: общее число серий — 70 для $D = 0$ и 71 для $D = 1$; длина серий колебалась от 2 до 9 для $D = 0$ и от 1 до 7 для $D = 1$.

Формирование предварительного диагноза было сведено к задаче обучаемой классификации данных. Для описания данных из класса, соответствующего $D = 0$, принималась вероятностная модель смеси анализаторов главных компонент [7], включая выбор значений параметров числа элементов смеси. Задача выбора эффективной размерности для отдельных элементов смеси не решалась: на начальном этапе исследования возможностей консолидирующих методов анализа серий из-за небольших размерностей вектора признаков в этом не было особой необходимости.

Применение информационных критериев оказалось малоэффективным, так как уже при $g > 3$ оценивание итоговой ошибки 1-го рода с помощью обучающей выборки приводило к большим ошибкам. Поэтому выбор числа элементов смеси k осуществлялся перебором от больших значений к меньшим до удовлетворительных величин разброса оценок ошибки 1-го рода. Прodelьвалось это для ряда необходимых значений α_0 и g . Примером получившихся результатов при $\alpha_0 = 5\%$ служит набор значений: для $g = 1$ число элементов смеси $k^* = 8$, для $g = 2 - k^* = 5$, для $g = 3 - k^* = 6$ и т. д.

При реализации процедуры W необходимо определиться с дополнительными параметрами. Можно предложить следующее: принятые в медицинской диагностике рекомендации приводят к $p_0 = 95\%$; эксперименты с построением различных классификаторов серий измерений PSA показали, что для диагноза «1» достижимы вероятности принятия правильного решения (выход за границы MRR) порядка 60%, т.е. $1 - (p_0 + \delta) \approx 0,6$ и $\delta \approx -55\%$, и, наконец, не противоречат сложившимся традициям статистического анализа значения $\alpha_W = \beta_W = 0,1$.

Применение описанных процедур практически сразу же выявило несоответствие уровней значимости, задаваемых при построении MRR, и итоговых характеристик решающих последовательных правил: с ростом длины серии ошибка 1-го рода увеличивается, что связано с самой природой процедур, для чего надо обратить внимание на точные значения в столбцах табл. 1, озаглавленных как α_R . При этом реальные значения этой характеристики (столбцы α_R^*), в принципе, значимо не отличаясь от теоретических, стабильно ниже, чем теоретические (в столбцах α_E^* приведены критические уровни

Таблица 1 Теоретические и эмпирические вероятности отклонения нулевых гипотез для процедуры R (вероятности α_R , α_R^* и α_E^* заданы в %)

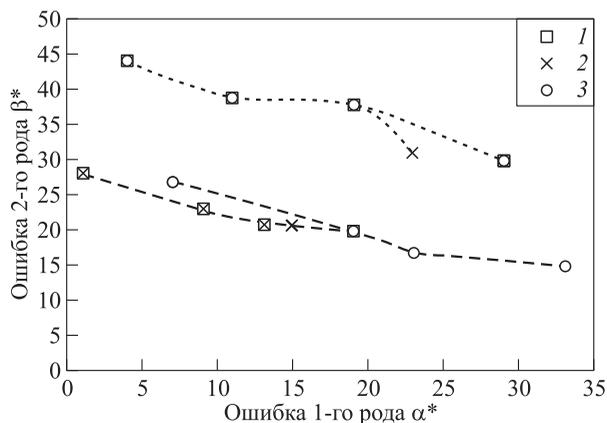
l	$\alpha_0 = 5\%$			$\alpha_0 = 10\%$		
	α_R	α_R^*	α_E^*	α_R	α_R^*	α_E^*
1	5	5	99	10	10	73
2	10	6	6	19	14	4
3	14	10	15	27	18	3
4	18	10	4	34	21	1
5	23	11	6	41	21	0

значимости при сравнении теоретической и эмпирической частот). Причина здесь, скорее всего, в предсказанной зависимости элементов серий [9, разд. 2.5], как теперь оказывается, и в случае $D = 0$. Заметим, что при переходе от $\alpha_0 = 5\%$ к $\alpha_0 = 10\%$ значимость отличий возрастает.

В результате пришлось расширить диапазон значений α_0 , для которых оценивались ошибки классификации, так, чтобы имела возможность сравнивать эффективность различных процедур. Далее при изложении результатов оставлены значения $\alpha_0 = 2\%, 4\%, 6\%, 8\%, 10\%$.

Результаты сравнительного анализа процедур (см. рисунок) позволяют сделать следующие выводы:

- группирование результатов измерений дает ощутимые преимущества;
- управление рассмотренными последовательными процедурами может осуществляться через установление априорных требований к используемому MRR;
- с точки зрения соотношения ошибок классификации более перспективной оказывается процедура С.



Соотношение ошибок 1-го и 2-го рода для последовательных процедур (1 — R; 2 — W; 3 — С) при обработке отдельных ($g = 1$, пунктирные кривые) и группированных наблюдений ($g = 3$, штриховые кривые)

Таблица 2 Доля ожидаемого объема выборки для последовательных процедур

Процедура	g	e
R	1	56
	3	88
W	1	52
	3	87
C	1	56
	3	62

Есть еще одно важное для практики свойство процедур обработки данных — возможность сократить объем данных, необходимых для принятия решений. Для имеющихся данных, когда каждая серия по результатам измерений tPSA и дополнительных исследований завершалась установлением окончательного диагноза, было интересно измерить долю ожидаемого объема выборки по отношению к длине серии — величину e . Она характеризует прогностические свойства применяемой процедуры статистического вывода: чем e меньше, тем выше упреждение при установлении диагноза $D = 1$. Значения этой характеристики в диапазоне значений $\alpha^* \in (1\%, 30\%)$ оказались практически одинаковыми; для исследуемых тестов и вариантов их применения они приведены в табл. 2.

Полученные результаты говорят о том, что:

- с точки зрения минимизации ожидаемого объема выборки предпочтение надо отдать вариантам процедур с наименьшим значением длины обрабатываемой группы;
- в любом случае использования последовательных процедур удается заблаговременно предупредить о возможном значимом отклонении от нормы.

4 Заключение

Последовательные методы анализа данных решают применительно к задачам медицинской диагностики важнейшую задачу скорейшего получения предварительных выводов с гарантированными свойствами. Эффект от подобных действий можно пояснить, переписав фрагмент табл. 2 в других единицах измерения: например, для процедуры R при $g = 1$ значение $e = 56$ эквивалентно периоду времени в 4,1 года, а при $g = 3$ значение $e = 88$ соответствует 1,8 года.

Построение соответствующих процедур осуществляется на основе реальных данных, объем которых, к сожалению, обычно невелик. Как следствие, правила формирования выводов относительно состояния пациента принимают форму рекомендаций (вариантов, сценариев) типа: «если врача удовлетворяют значения ошибок диагностирования и время упреждения при формировании предположения об окончательном диагнозе, то решение на основе данных о пациенте таково» (для подтверждения распространенности подобного подхода см., например, [10]).

Дальнейшее развитие методов анализа серий данных в рамках консолидирующего подхода может пойти по пути построения соответствующего SPRT, ближе всего к которому лежит предложенная в данной работе процедура C.

Литература

1. Кривенко М. П. Байесовская классификация серий многомерных данных // Системы и средства информатики, 2020. Т. 30. № 1. С. 34–45.
2. Wald A. Sequential tests of statistical hypotheses // Ann. Math. Stat., 1945. Vol. 16. No. 2. P. 117–186.
3. Armitage P., Colton T. Encyclopedia of biostatistics: In 8 vols. — 2nd ed. — Hoboken, NJ, USA: Wiley, 2005. 6257 p.
4. Bartroff J., Lai T. L., Shih M.-C. Sequential experimentation in clinical trials. Design and analysis. — New York, NY, USA: Springer, 2013. 237 p.
5. Wassmer G., Brannath W. Group sequential and confirmatory adaptive designs in clinical trials. — Cham, Switzerland: Springer, 2016. 301 p.
6. Кривенко М. П. Многомерный референсный регион высокой плотности // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 59–64.
7. Кривенко М. П. Выбор модели данных в задачах медицинской диагностики // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 27–29.
8. Thornquist M. D., Omenn G. S., Goodman G. E., et al. Statistical design and monitoring of the carotene and retinol efficacy trial (CARET) // Control. Clin. Trials, 1993. Vol. 14. Iss. 4. P. 308–324.
9. Fitzmaurice G. M., Laird N. M., Ware J. H. Applied longitudinal analysis. — 2nd ed. — Hoboken, NJ, USA: Wiley, 2011. 701 p.
10. Etzioni R., Pepe M., Longton G., Hu C., Goodman G. Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer // Med. Decis. Making, 1999. Vol. 19. Iss. 3. P. 242–251.

Поступила в редакцию 05.03.20

SEQUENTIAL ANALYSIS OF SERIAL MEASUREMENTS BASED ON MULTIVARIATE REFERENCE REGIONS

M. P. Krivenko

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

Abstract: Sequential data series analysis procedures are considered. An approach is developed when a set of multivariate features of a certain object, which varies in time, is presented as a single vector of observed values. By increasing the dimensionality of the data, it is possible to obtain a single picture of the description of objects, to take into account the objectively existing dependence of individual observations, and to simulate changes over time. The basis for solving classification problems is the use of multivariate reference regions. Three options for data processing procedures are proposed, their properties are investigated, and recommendations for practical application are developed. To demonstrate the capabilities of these procedures, the task of early diagnosis of cancer using the PSA (prostate-specific antigen) biomarker is considered. Features of the application of sequential methods for analyzing data series are indicated, recommendations for their effective use are formed, and the advantages of the consolidating approach in data analysis are identified.

Keywords: serial measurements; consolidation approach; sequential procedures; analysis of prostate-specific antigen (PSA)

DOI: 10.14357/19922264200212

References

1. Krivenko, M. P. 2020. Bayesovskaya klassifikatsiya seriy mnogomernykh dannykh [Bayesian classification of serial multivariate data]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1): 34–45.
2. Wald, A. 1945. Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16(2):117–186.
3. Armitage, P., and T. Colton. 2005. *Encyclopedia of biostatistics*: In 8 vols. 2nd ed. Hoboken, NJ: Wiley. 6257 p.
4. Bartroff, J., T. L. Lai, and M.-C. Shih. 2013. *Sequential experimentation in clinical trials. Design and analysis*. New York, NY: Springer. 237 p.
5. Wassmer, G., and W. Brannath. 2016. *Group sequential and confirmatory adaptive designs in clinical trials*. Cham, Switzerland: Springer. 301 p.
6. Krivenko, M. P. 2017. Mnogomernyy referensnyy region vysokoy plotnosti [High-density multivariate reference region]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):59–64.
7. Krivenko, M. P. 2019. Vybora modeli dannykh v zadachakh meditsinskoy diagnostiki [Data model selection in medical diagnostic tasks]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):27–29.
8. Thornquist, M. D., G. S. Omenn, G. E. Goodman, et al. 1993. Statistical design and monitoring of the carotene and retinol efficacy trial (CARET). *Control. Clin. Trials* 14(4):308–324.
9. Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2011. *Applied longitudinal analysis*. 2nd ed. Hoboken, NJ: Wiley. 701 p.
10. Etzioni, R., M. Pepe, G. Longton, C. Hu, and G. Goodman. 1999. Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Med. Decis. Making* 19(3):242–251.

Received March 5, 2020

Contributor

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

МНОГОФАКТОРНЫЕ МОДЕЛИ ПОЛНОСВЯЗНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ БЕЗ ОГРАНИЧЕНИЙ НА СООТНОШЕНИЯ ДИСПЕРСИЙ ОШИБОК ПЕРЕМЕННЫХ

М. П. Базилевский¹

Аннотация: Статья посвящена проблеме построения регрессионных моделей с ошибками в объясняющих переменных. В настоящее время такие модели широкого практического применения почти не находят, потому что они не пригодны для прогнозирования и интерпретации, их сложно оценивать, и при этом неизвестны дисперсии ошибок переменных. Для устранения перечисленных недостатков ранее автором были разработаны и исследованы двухфакторные модели полносвязной линейной регрессии. Такие модели легко оцениваются, их можно использовать для прогнозирования, и они лишены эффекта мультиколлинеарности. В данной работе впервые рассмотрены многофакторные модели полносвязной линейной регрессии. Доказано, что в случае снятия ограничений с соотношения дисперсий ошибок переменных существуют единственные оценки полносвязной регрессии, при которых аппроксимационные качества ее вторичного уравнения и классической модели множественной линейной регрессии, оцененной с помощью метода наименьших квадратов, совпадают.

Ключевые слова: EIV-модель; полносвязная регрессия; регрессия Деминга; метод наименьших квадратов

DOI: 10.14357/19922264200213

1 Введение

Одним из методов исследования влияния одной или нескольких объясняющих переменных на объясняемую переменную служит регрессионный анализ [1, 2], применение которого приводит к построению математических моделей статистического типа — регрессионных моделей. Зачастую такие модели оцениваются с помощью метода наименьших квадратов (МНК) в предположении, что объясняющие переменные не содержат случайных ошибок. Однако на практике даже при использовании современных технических средств значения таких переменных часто оказываются не вполне правильно измеренными. Поэтому возникает необходимость в построении моделей с ошибками в объясняющих переменных.

Регрессионные модели с ошибками в объясняющих переменных, известные в настоящее время в зарубежных источниках как «Errors-In-Variables models» (EIV-модели) и «measurement error models», имея внушительный математический аппарат для своего оценивания [3–5], к сожалению, практического применения почти не находят. Исключение составляет регрессия Деминга [6], которая служит прекрасным инструментом в клинических лабораториях для численного сопоставления новых и существующих измерительных методов [7–9].

К недостаткам EIV-моделей относится то, что они не пригодны для прогнозирования, возникают проблемы с их интерпретацией, априори нужно знать дисперсии ошибок переменных, а также они сложны в оценивании.

Для устранения перечисленных недостатков в работах [10, 11] была предложена и исследована двухфакторная модель полносвязной линейной регрессии. Если классическая множественная регрессия строится по принципу «объясняющие переменные влияют на объясняемую», то принцип полносвязной регрессии — «все переменные влияют друг на друга». Установлено, что двухфакторная полносвязная регрессия достаточно просто оценивается, лишена эффекта мультиколлинеарности, имеет гораздо более разнообразную интерпретацию, чем множественная регрессия, и пригодна для прогнозирования.

Целью данной работы ставится обобщение полносвязной регрессии на случай многих переменных.

Стоит отметить, что работа выполнена в рамках логико-алгебраического подхода к обработке статистических данных, при котором предполагается, что никаких априорных сведений об их вероятностной природе нет, поэтому не изучаются традиционные свойства оценок параметров: несмещенность, состоятельность и эффективность.

¹Иркутский государственный университет путей сообщения, кафедра математики, mik2178@yandex.ru

2 Многофакторная модель полносвязной линейной регрессии

Пусть x_{ij} , $i = \overline{1, n}$, $j = \overline{1, m}$, — наблюдаемые значения m объясняющих (входных) переменных x_1, x_2, \dots, x_m . Предположим, что существуют их «истинные» значения $x_{i1}^*, x_{i2}^*, \dots, x_{im}^*$, $i = \overline{1, n}$, которые отличаются от наблюдаемых на случайные отклонения:

$$x_{ij} = x_{ij}^* + \varepsilon_i^{(x_j)}, \quad i = \overline{1, n}, \quad j = \overline{1, m}. \quad (1)$$

Предположим, что между «истинными» переменными $x_1^*, x_2^*, \dots, x_m^*$ имеют место функциональные зависимости

$$x_j^* = a_j + b_j x_m^*, \quad j = \overline{1, m-1}, \quad (2)$$

где a_j и b_j , $j = \overline{1, m-1}$, — неизвестные параметры.

Уравнения (2) означают, что все «истинные» переменные связаны между собой через связующую переменную x_m^* . Тогда совокупность уравнений (1) и (2) будем называть многофакторной (m -факторной) моделью полносвязной линейной регрессии без выходной переменной. Для ее оценивания будем использовать метод наименьших полных квадратов (МНК), состоящий в минимизации функционала

$$S = \lambda_1 \sum_{i=1}^n (x_{i1} - a_1 - b_1 x_{im}^*)^2 + \lambda_2 \sum_{i=1}^n (x_{i2} - a_2 - b_2 x_{im}^*)^2 + \dots + \lambda_{m-1} \sum_{i=1}^n (x_{i,m-1} - a_{m-1} - b_{m-1} x_{im}^*)^2 + \sum_{i=1}^n (x_m - x_{im}^*)^2 \rightarrow \min, \quad (3)$$

где $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ — соотношения дисперсий ошибок переменных:

$$\lambda_1 = \frac{\sigma_{\varepsilon^{(x_m)}}^2}{\sigma_{\varepsilon^{(x_1)}}^2}; \quad \lambda_2 = \frac{\sigma_{\varepsilon^{(x_m)}}^2}{\sigma_{\varepsilon^{(x_2)}}^2}; \quad \dots; \quad \lambda_{m-1} = \frac{\sigma_{\varepsilon^{(x_m)}}^2}{\sigma_{\varepsilon^{(x_{m-1})}}^2}.$$

Если соотношения дисперсий ошибок $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ известны, то, вычислив частные производные функции (3) по переменным x_{im}^* , $i = \overline{1, n}$, и приравняв их к нулю, получим:

$$x_{im}^* = \frac{-\sum_{j=1}^{m-1} \lambda_j a_j b_j + \sum_{j=1}^{m-1} \lambda_j b_j x_{ij} + x_{im}}{1 + \sum_{j=1}^{m-1} \lambda_j b_j^2}, \quad i = \overline{1, n}. \quad (4)$$

Выразив из уравнений $\partial S / \partial a_j = 0$, $j = \overline{1, m-1}$, коэффициенты a_j и подставив их в формулы (4), получим равенство:

$$\overline{x_m^*} = \overline{x_m}. \quad (5)$$

Тогда, применяя равенство (5), запишем уравнения $\partial S / \partial a_j = 0$, $j = \overline{1, m-1}$, в виде

$$a_j = \overline{x_j} - b_j \overline{x_m}, \quad j = \overline{1, m-1}. \quad (6)$$

Используя формулы (4) и (6), представим уравнения $\partial S / \partial b_j = 0$, $j = \overline{1, m-1}$, в виде нелинейной системы:

$$b_p \left(D_{x_m} + \sum_{j=1}^{m-1} \lambda_j^2 b_j^2 D_{x_j} + 2 \sum_{j_1=1}^{m-2} \sum_{j_2=j_1+1}^{m-1} \lambda_{j_1} \lambda_{j_2} b_{j_1} b_{j_2} K_{x_{j_1} x_{j_2}} + 2 \sum_{j=1}^{m-1} \lambda_j b_j K_{x_j x_m} \right) = \left(1 + \sum_{j=1}^{m-1} \lambda_j b_j^2 \right) \left(\sum_{j=1}^{m-1} \lambda_j b_j K_{x_j x_p} + K_{x_m x_p} \right), \quad p = \overline{1, m-1}. \quad (7)$$

Дополним набор входных переменных x_1, x_2, \dots, x_m выходной переменной y и свяжем «истинные» значения, например переменной x_m , со значениями переменной y моделью парной линейной регрессии:

$$y_i = c_0 + c_1 x_{im}^* + \varepsilon_i, \quad i = \overline{1, n}, \quad (8)$$

где c_0 и c_1 — неизвестные параметры, которые находятся с помощью обычного МНК. Уравнение (8) будем называть вторичным уравнением полносвязной регрессии.

Тогда совокупность уравнений (1), (2) и (8) представляет собой многофакторную модель полносвязной линейной регрессии с выходной переменной y .

Если известны соотношения дисперсий ошибок $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$, то оценки полносвязной регрессии (1), (2) и (8) находятся в два этапа.

1. С помощью МНК оценивается полносвязная регрессия без выходной переменной. Для этого численно решается система (7) и находятся оценки $b_1^*, b_2^*, \dots, b_{m-1}^*$, затем по формулам (6) — коэффициенты $a_1^*, a_2^*, \dots, a_{m-1}^*$ и, наконец, по формулам (4) — «истинные» значения переменной x_m^* .

2. С помощью МНК оценивается модель парной линейной регрессии (8) по формулам:

$$c_1^* = \frac{\overline{yx_m^*} - \overline{y}\overline{x_m^*}}{(\overline{x_m^*})^2 - (\overline{x_m^*})^2}, \quad c_0^* = \overline{y} - c_1^*\overline{x_m^*}.$$

Тогда оцененная модель (1), (2) и (8) имеет вид:

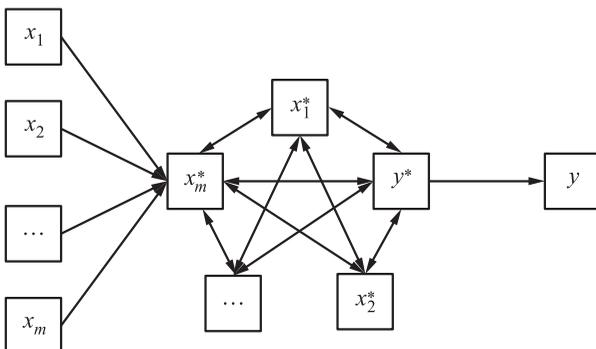
$$\begin{aligned} y^* &= c_0^* + c_1^*x_m^*; \\ x_j^* &= a_j^* + b_j^*x_m^*, \quad j = \overline{1, m-1}; \\ x_m^* &= A_0 + \sum_{j=1}^m A_j x_j, \end{aligned} \quad (9)$$

где

$$\begin{aligned} A_0 &= \frac{-\sum_{j=1}^{m-1} \lambda_j a_j^* b_j^*}{1 + \sum_{j=1}^{m-1} \lambda_j (b_j^*)^2}; \\ A_j &= \frac{\lambda_j b_j^*}{1 + \sum_{j=1}^{m-1} \lambda_j (b_j^*)^2}, \quad j = \overline{1, m-1}; \\ A_m &= \frac{1}{1 + \sum_{j=1}^{m-1} \lambda_j (b_j^*)^2}. \end{aligned}$$

Взаимосвязи между переменными для многофакторной модели полносвязной линейной регрессии представлены на рисунке.

Таким образом, в результате оценивания полносвязной регрессии формируются «истинные» (скрытые, латентные) переменные $x_1^*, x_2^*, \dots, x_m^*$, каждая пара которых связана линейной функциональной зависимостью. При этом каждая из латентных переменных $x_1^*, x_2^*, \dots, x_m^*$ одинаково коррелирует с переменной y , а также представляет собой линейную комбинацию наблюдаемых переменных x_1, x_2, \dots, x_m , поэтому значения переменных $x_1^*, x_2^*, \dots, x_m^*$ легко идентифицируются.



Спецификация многофакторной полносвязной регрессии

Стоит отметить, что на основании предпосылок (2) использовать полносвязные модели стоит в случае сильной корреляции между переменными x_1, x_2, \dots, x_m . В этом случае, согласно (1), латентные переменные $x_1^*, x_2^*, \dots, x_m^*$ практически не будут отличаться от переменных x_1, x_2, \dots, x_m . Следовательно, оценки полносвязной регрессии $b_1^*, b_2^*, \dots, b_{m-1}^*$ будут близки к оценкам угловых коэффициентов соответствующих моделей парной регрессии, знаки которых всегда совпадают со знаками соответствующих коэффициентов корреляции. А это означает, что при сильной корреляции переменных x_1, x_2, \dots, x_m знаки оценок $b_1^*, b_2^*, \dots, b_{m-1}^*$ также будут совпадать со знаками коэффициентов корреляции. То же самое справедливо для оценки c_1^* при сильной корреляции переменной y с каждой из переменных x_1, x_2, \dots, x_m .

3 Связь между моделями множественной и полносвязной линейной регрессии

В дальнейшем будем полагать, что параметры $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ — это не соотношения дисперсий ошибок переменных, а просто некоторые заданные числа. В соответствии с этим не будем ограничивать их условиями положительности.

Используя (9), перепишем вторичную регрессию (8) в виде:

$$y_i = \theta_0 + \sum_{j=1}^m \theta_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}, \quad (10)$$

где $\theta_0 = c_0 + A_0 c_1$; $\theta_j = A_j c_1, j = \overline{1, m}$.

Уравнение (10) есть классическая модель множественной линейной регрессии. На ее основании можно сделать очевидный вывод: для любых $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ сумма квадратов остатков вторичной регрессии (8) не может быть меньше суммы квадратов остатков оцененной с помощью МНК множественной регрессии y от x_1, x_2, \dots, x_m . Возникает вопрос: существуют ли параметры $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$, при которых оценки вторичного уравнения (10) полносвязной регрессии совпадают с МНК-оценками модели множественной линейной регрессии. Ответ на него дает следующая теорема.

Теорема. Пусть $\alpha_0^*, \alpha_1^*, \dots, \alpha_m^*$ — МНК-оценки модели множественной линейной регрессии

$$y_i = \alpha_0 + \sum_{j=1}^m \alpha_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}.$$

Тогда всегда существуют единственные значения параметров $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ полносвязной регрессии (1), (2) и (8), при которых оценки вторичного уравнения (10) равны соответствующим МНК-оценкам множественной модели, т. е. $\theta_j = \alpha_j^*, j = \overline{0, m}$.

Доказательство. Приравняем соответствующие оценки вторичной регрессии (10) к оценкам множественной регрессии $\theta_j = \alpha_j^*, j = \overline{0, m}$. В результате получим систему:

$$\left. \begin{aligned} c_0 + \frac{-\sum_{j=1}^{m-1} \lambda_j a_j b_j}{1 + \sum_{j=1}^{m-1} \lambda_j b_j^2} c_1 &= \alpha_0^*; \\ c_1 \frac{\lambda_p b_p}{1 + \sum_{j=1}^{m-1} \lambda_j b_j^2} &= \alpha_p^*, p = \overline{1, m-1}; \\ c_1 \frac{1}{1 + \sum_{j=1}^{m-1} \lambda_j b_j^2} &= \alpha_m^*. \end{aligned} \right\} (11)$$

Поделив все уравнения системы (11), кроме первого, на последнее, получим:

$$\lambda_p b_p = q_p, p = \overline{1, m-1}, (12)$$

где $q_p = \alpha_p^*/\alpha_m^*, p = \overline{1, m-1}$.

С учетом (12), система (7) примет вид:

$$b_p B = \left(1 + \sum_{j=1}^{m-1} q_j b_j \right) A_p, p = \overline{1, m-1}, (13)$$

где

$$\begin{aligned} B &= D_{x_m} + \sum_{j=1}^{m-1} q_j^2 D_{x_j} + \\ &+ 2 \sum_{j_1=1}^{m-2} \sum_{j_2=j_1+1}^{m-1} q_{j_1} q_{j_2} K_{x_{j_1} x_{j_2}} + 2 \sum_{j=1}^{m-1} q_j K_{x_j x_m}; \\ A_p &= \sum_{j=1}^{m-1} q_j K_{x_j x_p} + K_{x_m x_p}, p = \overline{1, m-1}. \end{aligned}$$

Выражение (13) представляет собой систему линейных алгебраических уравнений относительно переменных b_1, b_2, \dots, b_{m-1} :

$$\begin{pmatrix} A_1 q_1 - B & A_1 q_2 & \dots & A_1 q_{m-1} \\ A_2 q_1 & A_2 q_2 - B & \dots & A_2 q_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m-1} q_1 & A_{m-1} q_2 & \dots & A_{m-1} q_{m-1} - B \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \end{pmatrix} = - \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_{m-1} \end{pmatrix}. (14)$$

Определитель Δ основной матрицы системы (14):

$$\Delta = (-1)^{m-1} B^{m-1} \left(1 + \sum_{j=1}^{m-1} \frac{A_j}{B} q_j \right). (15)$$

Определитель (15) обращается в нуль либо при $B^{m-1} = 0$, либо при $1 - \sum_{j=1}^{m-1} (A_j/B) q_j = 0$. Нетрудно показать, что $B^{m-1} \neq 0$. Так, величина дисперсии связующей переменной x_m^* составляет $B/(1 + \sum_{j=1}^{m-1} q_j b_j)^2$. Из этого следует, что для любых q_j справедливо неравенство $B > 0$, а значит, и неравенство $B^{m-1} > 0$.

Докажем от противного, что

$$1 - \sum_{j=1}^{m-1} \frac{A_j}{B} q_j \neq 0.$$

Пусть

$$1 - \sum_{j=1}^{m-1} \frac{A_j}{B} q_j = 0. (16)$$

Из системы (13) следует, что

$$\frac{A_p}{B} = \frac{b_p}{1 + \sum_{j=1}^{m-1} q_j b_j}, p = \overline{1, m-1}. (17)$$

Подставляя (17) в (16), получим равенство

$$\frac{1}{1 + \sum_{j=1}^{m-1} q_j b_j} = 0,$$

которое никогда не выполняется. Следовательно,

$$1 - \sum_{j=1}^{m-1} \frac{A_j}{B} q_j \neq 0.$$

Таким образом, определитель (15) никогда не обращается в нуль, поэтому система (14) всегда совместна и имеет единственное решение. Теорема доказана.

Решение линейной системы (14) дает оценки $b_1^*, b_2^*, \dots, b_{m-1}^*$, с помощью которых, как следует из (12), определяются параметры $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ по формулам $\lambda_p = q_p/b_p^*, p = \overline{1, m-1}$. Зная эти параметры, нетрудно вычислить оставшиеся оценки полносвязной регрессии в соответствии с описанной выше последовательностью.

Предложенный в рамках данной работы метод построения полносвязных регрессий предназначен для следующего.

1. Для множества наблюдаемых переменных x_1, x_2, \dots, x_m можно легко получить множество соответствующих латентных переменных $x_1^*, x_2^*, \dots, x_m^*$, квадрат коэффициента корреляции для каждой из которых с переменной y равен коэффициенту детерминации множественной модели y от x_1, x_2, \dots, x_m . Иными словами, m переменных можно эквивалентно заменить одной, т.е. провести снижение размерности данных.
2. При сильной корреляции между переменными x_1, x_2, \dots, x_m возникает эффект мультиколлинеарности, который делает невозможным интерпретацию коэффициентов множественной регрессии. Однако, как было отмечено выше, знаки оценок $b_1^*, b_2^*, \dots, b_{m-1}^*, c_1^*$ полносвязной регрессии в этом случае будут совпадать со знаками соответствующих коэффициентов корреляции, поэтому такие оценки можно интерпретировать по принципу: если переменная x_m^* изменится на 1 единицу, то переменная x_1^* изменится на b_1^* , переменная x_2^* на b_2^* и т.д.

Литература

1. Носков С. И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. — Иркутск: Облформпечать, 1996. 321 с.
2. Носков С. И., Базилевский М. П. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования. — Иркутск: ИрГУПС, 2018. 176 с.
3. Кендалл М., Стьюарт А. Статистические выводы и связи / Пер. с англ. — М.: Наука, 1973. 899 с. (Kendall M. G., Stuart A. The advanced theory of statistics: Inference and relationship. — 2nd ed. — Hafner Publishing Co., 1967. Vol. 2. 690 p.)
4. Golub G. H., Van Loan C. F. An analysis of the total least squares problem // SIAM J. Numer. Anal., 1980. Vol. 17. P. 883–893.
5. Демиденко Е. З. Линейная и нелинейная регрессия. — М.: Финансы и статистика, 1981. 304 с.
6. Deming W. E. Statistical adjustment of data. — Wiley, 1943. 273 p.
7. Jensen A. L., Kjelgaard-Hansen M. Method comparison in the clinical laboratory // Vet. Clin. Path., 2006. Vol. 35. Iss. 3. P. 276–286.
8. Dhanoa M. S., Sanderson R., López S., France J. Bivariate relationships incorporating method comparison: A review of linear regression methods // CAB Reviews, 2016. Vol. 11. Art. No. 28. 15 p.
9. Wu C., Yu J. Z. Evaluation of linear regression techniques for atmospheric applications: The importance of appropriate weighting // Atmos. Meas. Tech., 2018. Vol. 11. P. 1233–1250.
10. Базилевский М. П. Синтез модели парной линейной регрессии и простейшей EIV-модели // Моделирование, оптимизация и информационные технологии, 2019. Т. 7. № 1. С. 170–182.
11. Базилевский М. П. Исследование двухфакторной модели полносвязной линейной регрессии // Моделирование, оптимизация и информационные технологии, 2019. Т. 7. № 2. С. 80–96.

Поступила в редакцию 07.09.19

MULTIFACTOR FULLY CONNECTED LINEAR REGRESSION MODELS WITHOUT CONSTRAINTS TO THE RATIOS OF VARIABLES ERRORS VARIANCES

M. P. Bazilevskiy

Irkutsk State Transport University, 15 Chernyshevskogo Str., Irkutsk 664074, Russian Federation

Abstract: The article is devoted to the problem of constructing errors-in-variables regression models. Currently, such models are not widely used because they are not suitable for forecasting and interpretation, they are difficult to estimate, and the variables errors variances are unknown. To eliminate these shortcomings, the author developed and investigated two-factor fully connected linear regression models. Such models are easily estimated, they can be used for forecasting, and they lack the effect of multicollinearity. In this paper, for the first time, multifactor fully connected linear regression models are considered. It is proved that in the case of removing the restrictions, on the ratio of variables errors variances, there are the one estimates of a fully connected regression, in which the approximation qualities of its secondary equation and the classical multiple linear regression model, estimated using the ordinary least squares, coincide.

Keywords: errors-in-variables models; fully connected regression; Deming regression; ordinary least squares

DOI: 10.14357/19922264200213

References

1. Noskov, S. I. 1996. *Tekhnologiya modelirovaniya ob"ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh* [Technology for modeling objects with unstable functioning and data uncertainty]. Irkutsk: Oblinformpechat'. 321 p.
2. Noskov, S. I., and M. P. Bazilevskiy. 2018. *Postroenie regressionnykh modeley s ispol'zovaniem apparata lineynobulevogo programmirovaniya* [Construction of regression models using linear Boolean programming]. Irkutsk: IrGUPS. 176 p.
3. Kendall, M. G., and A. Stuart. 1967. *The advanced theory of statistics: Inference and relationship*. 2nd ed. Hafner Publishing Co. Vol. 2. 690 p.
4. Golub, G. H., and C. F. Van Loan. 1980. An analysis of the total least squares problem. *SIAM J. Numer. Anal.* 17:883–893.
5. Demidenko, E. Z. 1981. *Lineynaya i nelineynaya regressiya* [Linear and nonlinear regression]. Moscow: Finansy i statistika. 304 p.
6. Deming, W. E. 1943. *Statistical adjustment of data*. Wiley. 273 p.
7. Jensen, A. L., and M. Kjelgaard-Hansen. 2006. Method comparison in the clinical laboratory. *Vet. Clin. Path.* 35(3):276–286.
8. Dhanoa, M. S., R. Sanderson, S. López, and J. France. 2016. Bivariate relationships incorporating method comparison: A review of linear regression methods. *CAB Reviews* 11:28. 15 p.
9. Wu, C., and J. Z. Yu. 2018. Evaluation of linear regression techniques for atmospheric applications: The importance of appropriate weighting. *Atmos. Meas. Tech.* 11:1233–1250.
10. Bazilevskiy, M. P. 2019. Sintez modeli parnoy lineynoy regressii i prosteyshy EIV-modeli [Synthesis of linear regression model and EIV-model]. *Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii* [Modeling, Optimization and Information Technology] 7(1):170–182.
11. Bazilevskiy, M. P. 2019. Issledovanie dvukhfaktornoy modeli polnosvyaznoy lineynoy regressii [Investigation of a two-factor fully connected linear regression model]. *Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii* [Modeling, Optimization and Information Technology] 7(2):80–96.

Received September 7, 2019

Contributor

Bazilevskiy Mikhail P. (b. 1987) — Candidate of Science (PhD) in technology, associate professor, Irkutsk State Transport University, 15 Chernyshevskogo Str., Irkutsk 664074, Russian Federation; mik2178@yandex.ru

РЕШЕНИЕ ЗАДАЧИ БЕЗУСЛОВНОГО ЭКСТРЕМУМА ДЛЯ ДРОБНО-ЛИНЕЙНОГО ИНТЕГРАЛЬНОГО ФУНКЦИОНАЛА, ЗАВИСЯЩЕГО ОТ ПАРАМЕТРА

П. В. Шнурков¹, К. А. Адамова²

Аннотация: Работа посвящена исследованию задачи безусловного экстремума дробно-линейного интегрального функционала, заданного на множестве вероятностных распределений. В отличие от результатов, полученных ранее, в рассматриваемой задаче подынтегральные функции интегральных выражений, находящихся в числителе и знаменателе, зависят от некоторого векторного вещественного параметра оптимизации. Таким образом, задача оптимизации исследуется на декартовом произведении множества вероятностных распределений и множества допустимых значений векторного вещественного параметра. Доказаны три утверждения об экстремуме дробно-линейного интегрального функционала. Установлено, что во всех вариантах решение исходной задачи полностью определяется экстремальными свойствами основной функции дробно-линейного интегрального функционала, которая представляет собой отношение подынтегральных функций числителя и знаменателя. Описаны возможные применения полученных результатов в задачах оптимального управления стохастическими системами.

Ключевые слова: дробно-линейный интегральный функционал; задача безусловного экстремума дробно-линейного интегрального функционала; основная функция; задачи оптимального управления марковскими и полумарковскими случайными процессами

DOI: 10.14357/19922264200214

1 Введение

Настоящая работа продолжает цикл исследований проблемы безусловного экстремума для дробно-линейного интегрального функционала, начатый в работах [1, 2].

Данная проблема имеет не только общетеоретическое значение, но и служит основой для решения задач оптимального управления различными классами случайных процессов (регенерирующими, марковскими, полумарковскими). В свою очередь, задачи оптимального управления в указанных классах процессов возникают при анализе многочисленных прикладных моделей в математической теории эффективности и надежности, теории управления запасами и других областях прикладной теории вероятностей.

Сделаем некоторые замечания библиографического характера, относящиеся к общей проблеме безусловного экстремума для дробно-линейного интегрального функционала.

Обычно дробно-линейным программированием называется раздел теории оптимизации, в котором целевой функционал рассматриваемой экстремальной задачи представляет собой отношение двух линейных функционалов, а имеющиеся ограниче-

ния носят линейный характер. В данной области оптимизации имеется обширная научная литература, основная часть которой посвящена исследованию соответствующих задач в конечномерных пространствах.

Современная теория указанного научного направления изложена в фундаментальной монографии [3]. В этой книге не только приводятся теоретические результаты решения соответствующих экстремальных задач, но и описываются численные методы нахождения таких решений. Кроме того, в ней приведена подробная библиография научных исследований в области дробно-линейного программирования. Отметим также некоторые значительные работы последних лет, в которых исследовались теоретические и вычислительные проблемы, связанные с данным научным направлением [4–6].

Специальный раздел дробно-линейного программирования составляют экстремальные задачи, в которых целевой функционал представляет собой отношение двух интегралов. Подынтегральные функции в этих интегралах предполагаются известными, а интегрирование проводится по вероятностной мере, принадлежащей некоторому множеству вероятностных мер, определенных

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

² Научно-производственный центр автоматики и приборостроения им. академика Н. А. Пилюгина, kselya_an@mail.ru

на заданном измеримом пространстве. Решением задачи служит вероятностная мера, доставляющая глобальный экстремум такому функционалу. Функционалы указанного вида можно назвать интегральными дробно-линейными функционалами. Экстремальные задачи для интегральных дробно-линейных целевых функционалов, заданных на множестве вероятностных распределений в конечномерном пространстве, рассматривались в работах [7, 8]. Результаты исследований по теории безусловного экстремума для таких функционалов в наиболее завершённой форме изложены в гл. 10 коллективной монографии [8]. Основное содержание этих результатов заключается в том, что безусловный экстремум функционалов такого вида достигается на вырожденных вероятностных распределениях, имеющих одну точку роста. Однако этот результат был получен в работе [8] при сильных ограничительных условиях, главное из которых — предположение о существовании экстремума целевого функционала, т. е. существовании решения исходной задачи.

В работах [1, 2] было предложено новое решение задачи безусловного экстремума для дробно-линейного интегрального функционала, которое существенно обобщает и усиливает результаты работы [8]. Принципиальное отличие результатов указанных работ от предшествующих заключается в том, что в основном утверждении указываются условия, при выполнении которых экстремум дробно-линейного интегрального функционала существует и достигается на вырожденном распределении, сосредоточенном в одной точке. При этом точка, в которой сосредоточена вся вероятностная мера, — это точка глобального экстремума функции, для которой известно явное аналитическое представление.

Таким образом, исследование экстремальных свойств данной функции позволяет одновременно доказать существование решения исходной экстремальной задачи и найти само это решение, определяемое точкой глобального экстремума.

В работах [1, 2] предполагалось, что подынтегральные функции числителя и знаменателя дробно-линейного интегрального функционала не зависят от вероятностной меры, характеризующей управление. Данное обстоятельство оправдано тем, что во многих задачах прикладного содержания целевой показатель представляет собой стационарный стоимостный функционал, структура которого установлена и обладает указанными особенностями. Однако в ряде задач целевой показатель, представляемый в форме дробно-линейного интегрального функционала, устроен таким образом, что подынтегральные функции его числителя и зна-

менателя зависят от некоторого набора детерминированных параметров управления. В этом случае экстремальная задача меняет свой характер и нуждается в специальном исследовании. Такому исследованию и посвящена данная работа.

2 Постановка основной экстремальной задачи

Обозначим через (U, \mathcal{B}) некоторое измеримое пространство, где U — множество произвольной природы; \mathcal{B} — σ -алгебра подмножеств множества U , включающая в себя все одноточечные множества.

В дальнейшем пространство (U, \mathcal{B}) будет служить множеством допустимых значений стохастических параметров оптимизации или управлений в рассматриваемой экстремальной задаче.

Пусть Γ — некоторое множество вероятностных мер, заданных на σ -алгебре \mathcal{B} , элементы которого будем обозначать $\Psi \in \Gamma$. Некоторые требования к множеству Γ и его элементам, связанные с постановкой основной экстремальной задачи, будут сформулированы в дальнейшем.

Классическая теория измеримых пространств и мер, задаваемых на этих пространствах, изложена в известной работе П. Халмоша [9].

Определение 1. Назовем вероятностную меру Ψ^* , заданную на \mathcal{B} , вырожденной, если существует точка $u^* \in U$ такая, что $\Psi^*({u^*}) = 1$, $\Psi^*(B^*) = 0$, где $u^* = \{u^*\}$ — множество, состоящее из одной точки, B^* — произвольное множество из системы \mathcal{B} , не содержащее точку u^* . Точку u^* будем называть точкой сосредоточения меры Ψ^* и обозначать меру со своей точкой сосредоточения символами $\Psi^* = \Psi_{u^*}^*$.

Обозначим через Γ^* множество всех возможных вырожденных вероятностных мер, заданных на измеримом пространстве (U, \mathcal{B}) . Множество Γ^* находится во взаимно однозначном соответствии с множеством всех точек сосредоточения вырожденных вероятностных мер, т. е. с самим множеством U .

Теперь введем $S \subseteq R^r$ — множество значений векторного параметра $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r) \in S$. В дальнейшем параметр α будет служить дополнительным детерминированным параметром оптимизации (управления) в рассматриваемой экстремальной задаче.

Зададим некоторые измеримые числовые функции:

$$A(\alpha, u) : S \times U \rightarrow R; \quad B(\alpha, u) : S \times U \rightarrow R,$$

где $u \in U$, $\alpha \in S$.

Введем интегральные преобразования, задаваемые функциями $A(\alpha, u)$ и $B(\alpha, u)$:

$$\left. \begin{aligned} I_{1,\alpha}(\Psi) &= \int_U A(\alpha, u) d\Psi(u); \\ I_{2,\alpha}(\Psi) &= \int_U B(\alpha, u) d\Psi(u). \end{aligned} \right\} \quad (1)$$

В своей наиболее общей форме интегральные выражения в соотношениях (1) представляют собой интегралы Лебега по вероятностной мере $\Psi \in \Gamma$. Согласно общей вероятностной модели, подробно изложенной в работе [10, гл. II], эти интегралы имеют смысл математических ожиданий некоторых случайных величин $A(\alpha, u)$ и $B(\alpha, u)$, зависящих от элементарного исхода $u \in U$ случайного эксперимента на вероятностном пространстве (U, \mathcal{B}, Ψ) . Если множество U — конечномерное вещественное пространство, то вероятностная мера Ψ может быть задана функцией распределения одномерной или многомерной случайной величины. Тогда интегралы в соотношениях (1) можно понимать как интегралы Лебега—Стилтьеса по вероятностному распределению Ψ . Заметим, что теория интегралов данного вида в краткой форме изложена в упомянутых работах [9, 10], а в более развернутом виде представлена в книге [11].

Интегральные преобразования (1) также определяют функционалы, заданные на множестве вероятностных мер Γ :

$$I_{1,\alpha}(\Psi) : \Gamma \rightarrow R, \quad I_{2,\alpha}(\Psi) : \Gamma \rightarrow R,$$

которые зависят от параметра $\alpha \in S$.

Теперь введем понятие дробно-линейного интегрального функционала, зависящего от параметра.

Определение 2. Отображение $I_\alpha : \Gamma \rightarrow R$, определяемое соотношением:

$$I_\alpha(\Psi) = \frac{I_{1,\alpha}(\Psi)}{I_{2,\alpha}(\Psi)} = \frac{\int_U A(\alpha, u) d\Psi(u)}{\int_U B(\alpha, u) d\Psi(u)}, \quad (2)$$

будем называть дробно-линейным интегральным функционалом, зависящим от параметра α .

Рассмотрим следующую экстремальную задачу:

$$I_\alpha(\Psi) \rightarrow \text{extr}, \quad \Psi \in \Gamma, \quad \alpha \in S, \quad (3)$$

для дробно-линейного интегрального функционала, зависящего от числового параметра α .

Уточним, что данная экстремальная задача рассматривается на множестве пар (α, Ψ) , которые и представляют собой двумерный параметр оптимизации.

Определение 3. Будем называть функцию $C(\alpha, u) = A(\alpha, u)/B(\alpha, u)$, $u \in U$, $\alpha \in S$, основной функцией дробно-линейного интегрального функционала (2).

Задача безусловного экстремума для дробно-линейного интегрального функционала вида (2), не зависящего от параметра, была исследована в работах [1, 2]. Отличие рассматриваемой задачи от предшествующей заключается в том, что в целевом функционале (2) присутствует зависимость подынтегральных функций числителя и знаменателя от некоторого дополнительного неслучайного параметра оптимизации (управления) $\alpha \in S$.

3 Решение поставленной экстремальной задачи

Введем некоторые предварительные условия для основных объектов, входящих в описание экстремальной задачи (3). Эти условия призваны обеспечить корректность поставленной задачи.

1. Интегральные выражения (1) существуют для всех $\Psi \in \Gamma$, $\alpha \in S$.
2. $\int_U B(\alpha, u) d\Psi(u) \neq 0$, $\Psi \in \Gamma$, $\alpha \in S$.
3. $\Gamma^* \subset \Gamma$.

Замечание 1. Если функция $B(\alpha, u)$ обладает свойством строгого знакопостоянства, т. е. $B(\alpha, u) > 0$, $u \in U$, $\alpha \in S$, или $B(\alpha, u) < 0$, $u \in U$, $\alpha \in S$, то условие 2 из указанной системы предварительных условий выполняется автоматически. В то же время условие строгой положительности функции $B(\alpha, u)$ является естественным для многих задач оптимального управления регенерирующими и полумарковскими случайными процессами (см. соответствующие пояснения в работе [2]). В связи с этим в формулировках всех последующих основных утверждений об экстремуме дробно-линейного интегрального функционала будет предполагаться, что выполнены условия 1, 3 и условие строгого знакопостоянства функции $B(\alpha, u)$.

Теорема 1. *Предположим, что в экстремальной задаче (3) основные объекты удовлетворяют предварительным условиям 1 и 3, а функция $B(\alpha, u)$ строго знакопостоянна (строго положительна или строго отрицательна) для всех $u \in U$, $\alpha \in S$. Предположим также, что основная функция $C(\alpha, u)$ достигает глобального экстремума на всем множестве $(\alpha, u) \in S \times U$ в точке (α^*, u^*) .*

Тогда решение экстремальной задачи (3) существует и достигается на паре $(\alpha^, \Psi_{u^*}^*)$, где $\Psi_{u^*}^*$ — вырожденная вероятностная мера, сосредоточенная в точке u^* , и при этом выполняются соотношения:*

$$\begin{aligned} \max_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi) &= \max_{\alpha \in S} \max_{\Psi^* \in \Gamma^*} I_\alpha(\Psi^*) = \\ &= \max_{(\alpha, u) \in S \times U} \frac{A(\alpha, u)}{B(\alpha, u)} = \frac{A(\alpha^*, u^*)}{B(\alpha^*, u^*)}, \end{aligned}$$

если (α^*, u^*) — точка глобального максимума функции $C(\alpha, u)$;

$$\begin{aligned} \min_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi) &= \min_{\alpha \in S} \min_{\Psi^* \in \Gamma^*} I_\alpha(\Psi^*) = \\ &= \min_{(\alpha, u) \in S \times U} \frac{A(\alpha, u)}{B(\alpha, u)} = \frac{A(\alpha^*, u^*)}{B(\alpha^*, u^*)}, \end{aligned}$$

если (α^*, u^*) — точка глобального минимума функции $C(\alpha, u)$.

Полное доказательство теоремы 1 приведено в приложении к настоящей работе [12], а также в публикации [13].

Теперь исследуем решение экстремальной задачи (3) для вариантов, в которых основная функция дробно-линейного интегрального функционала (2) не достигает глобального экстремума.

Теорема 2. *Предположим, что в экстремальной задаче (3) основные объекты удовлетворяют условиям 1 и 3, а функция $B(\alpha, u)$ строго знакопостоянна (строго положительна или строго отрицательна). Предположим также, что основная функция $C(\alpha, u)$ ограничена (сверху или снизу), но при этом не достигает глобального экстремума (максимума или минимума) на множестве $S \times U$.*

Тогда справедливы следующие утверждения.

1. Если основная функция $C(\alpha, u)$ ограничена сверху и не достигает глобального максимума, то для любого заданного $\varepsilon > 0$ существует пара $\alpha^{(+)}(\varepsilon) \in S$ и $u^{(+)}(\varepsilon) \in U$ такая, что выполняется двойное неравенство:

$$\begin{aligned} \sup_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi) - \varepsilon &< I_{\alpha^{(+)}(\varepsilon)}(\Psi_{u^{(+)}(\varepsilon)}^*) = \\ &= \frac{A(\alpha^{(+)}(\varepsilon), u^{(+)}(\varepsilon))}{B(\alpha^{(+)}(\varepsilon), u^{(+)}(\varepsilon))} < \sup_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi), \end{aligned} \quad (4)$$

где $\Psi_{u^{(+)}(\varepsilon)}^* \in \Gamma^*$ — вырожденная вероятностная мера, сосредоточенная в точке $u^{(+)}(\varepsilon)$.

2. Если основная функция $C(\alpha, u)$ ограничена снизу и не достигает глобального минимума, то для любого заданного $\varepsilon > 0$ существует пара $\alpha^{(-)}(\varepsilon) \in S$ и $u^{(-)}(\varepsilon) \in U$ такая, что выполняется двойное неравенство:

$$\begin{aligned} \inf_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi) &< I_{\alpha^{(-)}(\varepsilon)}(\Psi_{u^{(-)}(\varepsilon)}^*) = \\ &= \frac{A(\alpha^{(-)}(\varepsilon), u^{(-)}(\varepsilon))}{B(\alpha^{(-)}(\varepsilon), u^{(-)}(\varepsilon))} < \inf_{(\alpha, \Psi) \in S \times \Gamma} I_\alpha(\Psi) + \varepsilon, \end{aligned} \quad (5)$$

где $\Psi_{u^{(-)}(\varepsilon)}^* \in \Gamma^*$ — вырожденная вероятностная мера, сосредоточенная в точке $u^{(-)}(\varepsilon)$.

Заметим, что утверждения 1 и 2, т. е. соотношения (4) и (5), могут выполняться как по отдельности для верхней или нижней грани множества значений функционала, так и совместно.

Доказательство теоремы 2 приведено в приложении к настоящей работе [12], а также в публикации [13].

Сущность утверждения теоремы 2 заключается в том, что если основная функция дробно-линейного интегрального функционала ограничена, но не достигает экстремума, то для любого заданного $\varepsilon > 0$ существует так называемая ε -оптимальная детерминированная стратегия управления, задаваемая значениями $\alpha^{(+)}(\varepsilon) \in S$ и $u^{(+)}(\varepsilon) \in U$ для задачи на максимум или значениями $\alpha^{(-)}(\varepsilon) \in S$ и $u^{(-)}(\varepsilon) \in U$ для задачи на минимум.

Перейдем к формулировке еще одного утверждения, связанного с экстремальной проблемой для дробно-линейного интегрального функционала, зависящего от параметра.

Теорема 3. *Предположим, что в экстремальной задаче (3) основные объекты удовлетворяют условиям 1 и 3, а функция $B(\alpha, u)$ знакопостоянна (строго положительна или строго отрицательна). Предположим также, что основная функция $C(\alpha, u)$ не ограничена (сверху или снизу). Тогда соответствующий дробно-линейный интегральный функционал также не ограничен сверху или снизу и справедливы следующие утверждения:*

1. Существует последовательность $(\alpha_n^{(+)}, \Psi_n^{*(+)})$, $\alpha_n^{(+)} \in S$, $\Psi_n^{*(+)} \in \Gamma^*$, $n = 1, 2, \dots$, такая что

$$I(\alpha_n^{(+)}, \Psi_n^{*(+)}) \rightarrow \infty, \quad n \rightarrow \infty, \quad (6)$$

если основная функция не ограничена сверху.

2. Существует последовательность $(\alpha_n^{(-)}, \Psi_n^{*(-)})$, $\alpha_n^{(-)} \in S$, $\Psi_n^{*(-)} \in \Gamma^*$, $n = 1, 2, \dots$, такая что

$$I(\alpha_n^{(-)}, \Psi_n^{*(-)}) \rightarrow -\infty, \quad n \rightarrow \infty, \quad (7)$$

если основная функция не ограничена снизу.

Утверждения 1 и 2, т. е. соотношения (6) и (7), могут выполняться как по отдельности, так и совместно.

Доказательство теоремы 3 приведено в приложении к настоящей работе [12], а также в публикации [13].

Из теоремы 3 непосредственно следует, что если основная функция дробно-линейного интегрального функционала не ограничена сверху или снизу, то и сам этот функционал не ограничен и решения соответствующей экстремальной задачи на максимум или минимум не существует.

4 Заключение

В работе получено исчерпывающее решение задачи безусловного экстремума для дробно-линейного интегрального функционала, зависящего от дополнительного детерминированного параметра оптимизации. Установлено, что решение поставленной экстремальной задачи полностью определяется свойствами основной функции данного функционала, которая предполагается известной. Полученные результаты обобщают соответствующие утверждения о решении экстремальной задачи для дробно-линейного интегрального функционала, у которого основная функция не зависит от дополнительного параметра оптимизации [1, 2]. Данные результаты можно применять для решения различных прикладных задач, в которых параметр управления имеет смешанный характер и состоит из детерминированной и стохастической компонент.

Литература

1. Шнурков П. В. О решении задачи безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. Акад. наук, 2016. Т. 470. № 4. С. 387–392.
2. 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.
3. Bajalinov E. B. Linear-fractional programming. Theory, methods, applications and software. — Boston—Dordrecht—London: Kluwer Academic Publs., 2003. 423 p.
4. Joshi V. D., Singh E., Gupta N. Primal-dual approach to solve linear fractional programming problem // J. Appl. Mathematics Statistics Informatics, 2008. Vol. 4. Iss. 1. P. 61–69.
5. Hasan M. B., Acharjee S. Solving LFP by converting it into a single LP // Int. J. Operations Research, 2011. Vol. 8. Iss. 3. P. 1–14.
6. Borza M., Rambely A. S., Saraj M. Solving linear fractional programming problems with interval coefficients in the objective function. A new approach // Appl. Math. Sciences, 2012. Vol. 6. No. 69-72. P. 3443–3452.
7. Барзилович Е. Ю., Каушанов В. А. Некоторые математические вопросы теории обслуживания сложных систем. — М.: Советское радио, 1971. 272 с.
8. Вопросы математической теории надежности / Под ред. Б. В. Гнеденко. — М.: Радио и связь, 1983. 376 с.
9. Халмош П. Теория меры / Пер. с англ. Д. А. Василькова. — М.: ИИЛ, 1953. 291 с. (Halmos P. Measure theory. — Graduate texts in mathematics ser. — Springer-Verlag New York, 1950. 316 p.)
10. Ширяев А. Н. Вероятность-1. — М.: МЦНМО, 2011. 552 с.
11. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Физматлит, 2004. 572 с.
12. Шнурков П. В., Адамова К. А. Приложение к статье: Решение задачи безусловного экстремума для дробно-линейного интегрального функционала, зависящего от параметра, 2020. 17 с. <http://www.ipiran.ru/publications/publications/supplement.pdf>.
13. Shnurkov P., Adamova K. Solution of the unconditional extremal problem for a linear-fractional integral functional depending on the parameter // arXiv.org, 2019. arXiv:1906.05824 [math.OC]. 14 p. <https://arxiv.org/abs/1906.05824>.

Поступила в редакцию 15.04.20

SOLUTION OF THE UNCONDITIONAL EXTREMAL PROBLEM FOR A LINEAR-FRACTIONAL INTEGRAL FUNCTIONAL DEPENDENT ON THE PARAMETER

P. V. Shnurkov¹ and K. A. Adamova²

¹National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

²Academician Pilyugin Center, 1 Vvedenskogo Str., Moscow, 117342, Russian Federation

Abstract: The paper is devoted to the study of the unconditional extremal problem for a fractional linear integral functional defined on a set of probability distributions. In contrast to results proved earlier, the integrands of the integral expressions in the numerator and the denominator in the problem under consideration depend on a real optimization parameter vector. Thus, the optimization problem is studied on the Cartesian product of a set of probability distributions and a set of admissible values of a real parameter vector. Three statements on the extremum of a fractional linear integral functional are proved. It is established that, in all the variants, the solution of the

original problem is completely determined by the extremal properties of the test function of the linear-fractional integral functional; this function is the ratio of the integrands of the numerator and the denominator. Possible applications of the results obtained to problems of optimal control of stochastic systems are described.

Keywords: linear-fractional integral functional; unconditional extremal problem for a fractional linear integral functional; test function; optimal control problems for Markov and semi-Markov random processes

DOI: 10.14357/19922264200214

References

1. 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.
2. Shnurkov, P.V., A. K. Gorshenin, and V.V. Belousov. 2016. Analiticheskoe reshenie zadachi optimal'nogo upravleniya polumarkovskim protsessom s konechnym mnozhestvom sostoyaniy [An analytic solution of the optimal control problem for a semi-Markov process with a finite set of states]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):72–88.
3. Bajalinov, E. B. 2003. *Linear-fractional programming. Theory, methods, applications and software*. Boston–Dordrecht–London: Kluwer Academic Publs. 423 p.
4. Joshi, V. D., E. Singh, and N. Gupta. 2008. Primal-dual approach to solve linear fractional programming problem. *J. Appl. Mathematics Statistics Informatics* 4(1): 61–69.
5. Hasan, M.B., and S. Acharjee. 2011. Solving LFP by converting it into a single LP. *Int. J. Operations Research* 8(1): 1–14.
6. Borza, M., A. S. Rambely, and M. Saraj. 2012. Solving linear fractional programming problems with interval coefficients in the objective function. A new approach. *Appl. Math. Sciences* 6(69–72):3443–3452.
7. Barzilovich, E.Yu., and V.A. Kashtanov. 1971. *Nekotorye matematicheskie voprosy teorii obsluzhivaniya slozhnykh sistem* [Some mathematical questions in theory of complex systems maintenance]. Moscow: Sovetskoe radio. 272 p.
8. Gnedenko, B. V., ed. 1983. *Voprosy matematicheskoy teorii nadezhnosti* [Questions of mathematics reliability theory]. Moscow: Radio i svyaz'. 376 p.
9. Halmos, P. 1950. *Measure theory*. Graduate texts in mathematics ser. Springer-Verlag New York. 316 p.
10. Shiryaev, A. N. 2011. *Veroyatnost'-1* [Probability-1]. Moscow: MTSNMO. 552 p.
11. Kolmogorov, A. N., and S. V. Fomin. 2004. *Elementy teorii funktsiy i funktsional'nogo analiza* [Elements of function theory and functional analysis]. Moscow: Fizmatlit. 572 p.
12. Shnurkov, P.V., and K. A. Adamova. 2020. Prilozhenie k stat'е: reshenie zadachi bezuslovnogo ekstremuma dlya drobno-lineynogo integral'nogo funktsionala, zavisyashchego ot parametra [Appendix to article: Solution of the unconditional extremal problem for a linear-fractional integral functional depending on the parameter]. Available at: <http://www.ipiran.ru/publications/publications/supplement.pdf> (accessed April 15, 2020).
13. Shnurkov, P., and K. Adamova. 2019. Solution of the unconditional extremal problem for a linear-fractional integral functional depending on the parameter. *arXiv.org*. 14 p. Available at: <https://arxiv.org/abs/1906.05824> (accessed April 5, 2020).

Received April 15, 2020

Contributors

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

Adamova Kseniia A. (b. 1994) — engineer, Academician Pilyugin Center, 1 Vvedenskogo Str., Moscow, 117342, Russian Federation; ksenya_an@mail.ru

ИНТЕГРАЦИОННАЯ ПЛАТФОРМА ДЛЯ МНОГОМАСШТАБНОГО МОДЕЛИРОВАНИЯ НЕЙРОМОРФНЫХ СИСТЕМ*

К. К. Абгарян¹, Е. С. Гаврилов²

Аннотация: Актуальные сегодня многоуровневые элементы резистивной памяти позволяют увеличить плотность интеграции энергонезависимой памяти, а также спроектировать и создать системы с механизмом параллельных вычислений. В основе таких устройств лежат мемристоры, необходимые для разработки основ аналоговых нейроморфных сетей, которые используются для решения задач интеллектуального анализа данных. Однако использование мемристоров в составе нейроморфных устройств сталкивается с рядом проблем, таких как разброс значений параметров переключения (напряжение, окно памяти) от ячейки к ячейке, асимметричность и нелинейные эффекты и др. Такие проблемы диктуют необходимость создания оригинальных имитационных моделей и новых программных инструментов, которые позволят оценить влияние возмущающих факторов на предсказательную точность и процесс обучения сети. В данной работе для решения задачи многомасштабного моделирования нейроморфных систем применяется оригинальная информационная технология построения многомасштабных моделей. Для ее практической реализации построена интеграционная платформа, которая позволяет оценить влияние возмущающих факторов на предсказательную точность и процесс обучения нейроморфной сети, а в дальнейшем сможет обеспечить формирование информации для обоснованного выбора материалов, конфигурации и топологии ячеек памяти компьютеров нового поколения.

Ключевые слова: многомасштабное моделирование; многоуровневые элементы памяти; нейроморфные сети; предсказательное моделирование; мемристор; интеграционная платформа; программный комплекс

DOI: 10.14357/19922264200215

1 Введение

В настоящее время искусственные нейронные сети стали мощным инструментом интеллектуального анализа данных, построения аппроксимационных моделей сложных систем и процессов, распознавания образов, классификации и кластеризации. Актуальные сегодня многоуровневые элементы резистивной памяти позволяют увеличить плотность интеграции энергонезависимой памяти, а также спроектировать и создать системы с механизмом параллельных вычислений, которые необходимы для разработки основ аналоговых нейроморфных сетей. Ожидается, что на основе таких ячеек/матриц резистивной памяти будет создана элементная база компьютеров следующего поколения, работающих на новых физических принципах.

Вместе с тем для решения задач интеллектуального анализа данных использование мемристорных элементов в составе нейроморфных устройств сталкивается с рядом проблем. Среди них — разброс значений параметров переключения (напряжение,

окно памяти) от ячейки к ячейке, асимметричность и нелинейные эффекты, сложность применения градиентных алгоритмов обучения в связи с дискретным характером синаптических весов.

Такие проблемы диктуют необходимость создания оригинальных имитационных моделей и новых программных инструментов, которые позволят оценить влияние возмущающих факторов на предсказательную точность и процесс обучения сети. Кроме того, такие модели обеспечат формирование информации для обоснованного выбора материалов, конфигурации и топологии ячеек памяти, селекторов и электродов, выработки оптимальных схемных и алгоритмических решений.

Сложность создания моделей данного класса связана, в первую очередь, с необходимостью учета широкого спектра пространственно-временных масштабов. В случаях, когда необходимо в рамках одной модели провести исследование многомасштабного физического процесса или явления, возникает проблема взаимосогласованности имеющихся моделей, что требует разработки теоретических основ их объединения.

* Работа выполнена при финансовой поддержке РФФИ (проект 19-29-03051 мк).

¹ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский авиационный институт (национальный исследовательский университет), kristal83@mail.ru

² Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский авиационный институт (национальный исследовательский университет), eugavrilov@gmail.com

Распределение БК по масштабным уровням

№ масштабного уровня	Обозначение и название БК	Название масштабного уровня
0	МС ₀ ¹ «АТОМ А ₀ ⁱ »	Уровень химических элементов
1	МС ₁ ¹ «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» МС ₁ ² «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА»	Квантово-механический уровень
2	МС ₂ ¹ «АТОМНЫЙ КЛАСТЕР – СТАТИКА» МС ₂ ² «АТОМНЫЙ КЛАСТЕР – ДИНАМИКА»	Нано-уровень
3	МС ₃ ¹ «Модель генерации/рекомбинации ионов кислорода и кислородных вакансий (GR-model)» МС ₃ ² «Модель дрейфа-диффузии ионов кислорода (DD-model)» МС ₃ ³ «Модель силового поля (E-model)» МС ₃ ⁴ «Модель переноса электронов (электрического тока) (J-model)» МС ₃ ⁵ «Модель теплопереноса (HT-model)»	Уровень элемента резистивной памяти (мемристора)
4	МС ₄ ¹ «Фитинг моделей. ВАХ» МС ₄ ² «Схемотехническое представление»	Уровень формирования нейроморфной сети
5	МС ₅ ¹ «Нейросетевое моделирование». Обучение по прецедентам	Логический уровень

В настоящей работе для решения задачи многомасштабного моделирования нейроморфных систем применяется оригинальная информационная технология построения многомасштабных моделей, основанная на теоретико-множественном представлении физико-математических моделей и использовании информационных структур, объединяющих данные и методы их обработки [1, 2]. Для ее практической реализации была разработана интеграционная платформа для многомасштабного моделирования, объединяющая информационные потоки на разных масштабных уровнях: на уровне элементов резистивной памяти, на уровне нейроморфной сети, на уровне имитации обучения нейроморфной сети по прецедентам (логический уровень). Для более детального и последовательного построения схемы многомасштабного моделирования можно выделить шесть характерных масштабных уровней: уровень элементов таблицы Менделеева; квантово-механический уровень и наноуровень (их объединение ранее условно называли уровнем атомарной структуры); уровень элемента памяти (мемристора); уровень нейроморфной сети; логический уровень (см. таблицу).

В данной работе представлена оригинальная интеграционная платформа для многомасштабного моделирования нейроморфных систем, размещенная на гибридном кластере ЦКП ФИЦ ИУ РАН, которая позволяет проводить эффективное

параллельное решение задач разных масштабов с учетом постоянного обмена данными. Далее приведено описание используемых теоретических моделей и сценариев работы комплекса, направленных на решение существенных проблем, которые возникают при проектировании элементов мемристора на основе магнитных туннельных переходов (МТП) при последовательной миниатюризации этих устройств. Это придает задаче междисциплинарный характер и обуславливает актуальность применения информационной технологии многомасштабного моделирования в контексте поставленной задачи

2 Принципы формирования информационной поддержки многомасштабного моделирования физических явлений и процессов

В данной работе при построении многомасштабной модели для имитации работы нейроморфных систем используются подходы, разработанные ранее и представленные в [1, 2]. В их основе лежит идея о том, что спецификация информационных объектов предметной области может быть сфор-

мулирована в терминах конечно-множественных представлений [3] которые, в частности, могут быть реализованы посредством доменных моделей [4, 5] — конечных наборов (множеств) данных с поддержкой операций по их обработке. В статье [2] даны понятия и определения, применяемые при описании информационной технологии многомасштабного моделирования. Вводится понятие «базовая модель-композиция» (БК), для ее описания применяется теоретико-множественный аппарат. Используется обозначение базовой модели-композиции: $МС_i^j$, где i — номер масштабного уровня; j — номер данной БК на этом масштабном уровне [1, 2].

Как было показано в работах [1, 2], БК служат информационными аналогами базовых физико-математических моделей, которые применяются в вычислительном процессе для решения конкретных задач на своем масштабном уровне. Их можно представить в виде таблиц данных разного структурного типа (входных/выходных данных, моделей и алгоритмов и др.). Экземпляры БК (таблицы с конкретными данными) представляются объектами классов — наследников БК в объектно-ориентированном языке программирования. Они хранятся в виде документов в документно-ориентированной базе данных [6]. Многомасштабная композиция представляет собой коллекцию, состоящую из сгруппированных документов с иерархической структурой, отражающей последовательность передачи данных в общем вычислительном процессе.

Покажем, как при решении задачи о применении методов многомасштабного моделирования для имитации работы нейроморфной сети с помощью данной технологии из конкретных БК составляются многомасштабные композиции (МК) — информационные аналоги многомасштабных моделей, передающие содержание многомасштабных вычислительных процессов.

3 Интеграционная платформа для моделирования работы нейроморфной сети

Интеграционная платформа для многомасштабного моделирования нейроморфной сети объединяет информационные потоки на разных масштабных уровнях — на квантово-механическом и наноуровнях, на уровне элементов резистивной памяти, на уровне нейроморфной сети, на уровне имитации обучения нейроморфной сети по прецедентам. Для более детального описания модели-

руемого вычислительного уровня введем дополнительные уровни.

Общее представление о распределении БК по масштабным уровням, задействованным в вычислительном процессе, показано в таблице.

Покажем, как организуется вычислительным процесс, имитирующий работу нейроморфной сети, в основе которой лежит мемристивный элемент на базе оксида гафния.

На нулевом масштабном уровне с БК $МС_0^1$ «АТОМ A_0^i » (i — номер элемента в таблице Менделеева) задаются основные данные по конкретным химическим элементам, входящим в состав соединений, участвующих в вычислительном процессе (атомный номер химического элемента, масса атома, заряд ядра, радиус атома, электронная конфигурация, структура решетки и др.). В данном случае формируются экземпляры БК $МС_0^{72}$ «АТОМ Hf_0^{72} » и $МС_0^{16}$ «АТОМ O_0^{16} ».

К первому масштабному уровню отнесены БК $МС_1^1$ «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» и $МС_1^2$ «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». С их помощью данные, полученные с нулевого масштабного уровня из $МС_0^{72}$ «АТОМ Hf_0^{72} » и $МС_0^{16}$ «АТОМ O_0^{16} », передаются на квантово-механический уровень первоначально в $МС_1^1$ «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА», где, используя знания о химическом составе и кристаллографической структуре, с помощью простых физико-математических моделей (ионно-атомных радиусов и др.) определяется кристаллохимическая структура соединения (метрические параметры кристаллической решетки, координаты базисных атомов и др.). Данная БК программно реализована в двух расчетных модулях (модель «Плотная упаковка» и программный комплекс Materials Studio (<https://www.3dsbiovia.com/products/collaborative-science/biovia-materials-studio>)).

Далее полученные в ходе вычислительного процесса данные передаются в БК $МС_1^2$ «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». Данная БК программно реализована в двух расчетных модулях (программный комплекс VASP (<https://www.vasp.at>) и пакет программ с открытым кодом Quantum Espresso (<https://www.quantum-espresso.org>)). Здесь на базе теории функционала электронной плотности решается задача по определению итоговых значений параметров кристаллической решетки, рассчитывается электронная плотность [7], полная энергия для заданной конфигурации базисных атомов. Далее проводятся расчеты констант упругости, поляризация, фононное рассеяние, рассчитываются энергетические барьеры, энергия активации и др.

На втором масштабном уровне, используя данные, полученные на предыдущем уровне, проводится выбор потенциалов межатомного взаимодействия с учетом типа связи моделируемого соединения (в данном случае — оксида гафния). Например, используется модель погруженного атома (embedded atom model, EAM) [8] либо потенциал межатомного взаимодействия MEAM [9]. При помощи БК MC_2^1 «АТОМНЫЙ КЛАСТЕР – СТАТИКА» проводится идентификация параметров потенциала межатомного взаимодействия. Далее потенциал с идентифицированными параметрами используется в ходе молекулярно-динамического моделирования в рамках БК MC_2^2 «АТОМНЫЙ КЛАСТЕР – ДИНАМИКА».

Ранее полученные данные о координатах позиций атомов кислорода, рассчитанные данные об энергии активации, энергии рекомбинации, энергетических барьерах миграции и другие поступают на третий масштабный уровень. Туда же поступают дополнительные данные, такие как значения потенциалов на электродах, коэффициенты, учитывающие вклад поляризации света в энергетические барьеры, и др. (см. рисунок)

На третьем масштабном уровне сформирована математическая модель образования/разрушения проводящих каналов (филаментов) в мемристорных элементах на основе оксидных пленок. Ей соответствует «Композитный модуль расчета мемристора». Необходимо отметить, что в мире весьма активно ведутся работы по исследованию возможностей проектирования и оптимизации ячеек памяти на основе оксидов металлов. Так, в статье [10] представлен электротермический симулятор, позволяющий исследовать физику и потенциал ячеек резистивной памяти на основе SiO_x с произвольным доступом (RRAM, resistive random-access memory). В данной же работе в контексте многомасштабного подхода используется модельное представление, включающее в себя 5 взаимосвязанных подмоделей, каждой из которых поставлены в соответствие БК: MC_3^1 «Модель генерации/рекомбинации ионов кислорода и кислородных вакансий (GR-model)»; MC_3^2 «Модель дрейфа-диффузии ионов кислорода (DD-model)»; MC_3^3 «Модель силового поля (E-model)»; MC_3^4 «Модель переноса электронов (электрического тока) (J-model)»; MC_3^5 «Модель теплопереноса (HT-model)». Из 5 БК формируется композиция для расчета свойств мемристора $K_3^{-3,1;3,2;3,3;3,4;3,5}$. Здесь нижний индекс $i = 3$, так как все БК, образующие композицию, — с третьего уровня; верхние индексы (3,1;3,2;3,3;3,4;3,5) через точку с запятой обозначают, какие именно БК используются.

На четвертом масштабном уровне многомасштабной модели для имитации работы нейроморфной сети представлена БК MC_4^1 «Фитинг моделей. ВАХ» и БК MC_4^2 «Схемотехническая». На вход MC_4^1 подаются сведения (графики) по вольт-амперным характеристикам (ВАХ) моделируемого соединения с нижнего масштабного уровня либо сведения о модели и начальном наборе ее параметров. На выходе получаем уточненные значения параметров модели. Далее полученные данные передаются в БК MC_4^2 «Схемотехническая», туда же поступают данные по параметрам электронной схемы.

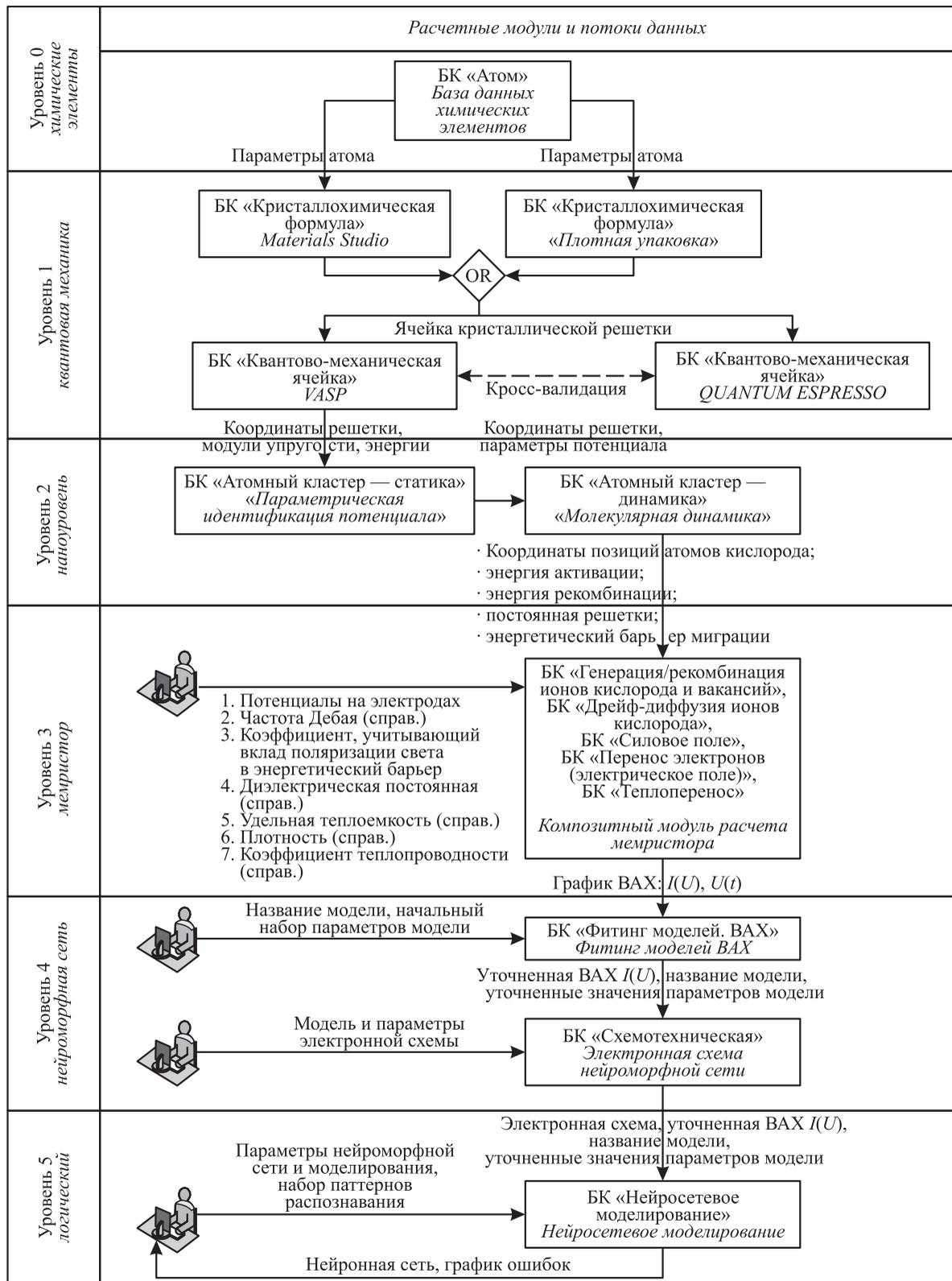
На пятом масштабном уровне проводится моделирование работы аналоговой нейроморфной сети MC_5^1 «Нейросетевое моделирование». Данные, рассчитанные на нижних масштабных уровнях (электронная схема, уточненные ВАХ, уточненные названия параметров модели и др.), а также параметры нейроморфной сети и наборы паттернов распознавания подаются на вход MC_5^1 «Нейросетевое моделирование». С помощью данной модели проводится имитация обучения нейроморфной сети. Более подробно работа модели этого масштабного уровня описана в работе [11].

На рисунке представлена архитектурная схема расчетных модулей и основных потоков данных. В общем случае модуль представлен программным компонентом, состоящим:

- из программы, непосредственно выполняющей расчет в пакетном режиме;
- микросервиса [12], обеспечивающего связь с расчетной программой и предоставляющего программный интерфейс (API, application programming interface) для интеграции в общий сценарий расчета [13];
- интерфейса пользователя для ввода параметров и мониторинга расчета;
- локальной базы данных расчетного модуля для хранения внутренних данных и промежуточных результатов.

На следующем этапе с помощью созданной интеграционной платформы планируется проведение серии вычислительных экспериментов. Будут рассмотрены различные модели многоуровневых устройств памяти с мемристивными системами на основе оксидов металлов и широкозонных полупроводников, таких как MgO , VO_2 , TiO_2 , SiO_2 , ZrO_2 , HfO_2 , TaO_x и др.

В связи с тем, что разработанная интеграционная платформа позволяет оценить влияние возмущающих факторов на предсказательную точность



Интеграционная платформа для многомасштабного моделирования нейроморфных систем

и процесс обучения нейроморфной сети, накопление большого числа данных вычислительных экспериментов в дальнейшем может обеспечить формирование информации для обоснованного выбора материалов, конфигурации и топологии ячеек памяти компьютеров нового поколения.

Литература

1. Абгарян К. К. Многомасштабное моделирование в задачах структурного материаловедения. — М.: МАКС Пресс, 2017. 284 с.
2. Абгарян К. К. Информационная технология построения многомасштабных моделей в задачах вычислительного материаловедения // Системы высокой доступности, 2018. Т. 14. № 2. С. 9–15.
3. Бродский Ю. И. Модельный синтез и модельно-ориентированное программирование. — М.: ВЦ РАН, 2013. 142 с.
4. Evans E. Domain-driven design: Tackling complexity in the heart of software. — Addison-Wesley Professional, 2003. 560 p.
5. Абгарян К. К., Гаврилов Е. С., Марасанов А. М. Информационная поддержка задач многомасштабного моделирования композиционных материалов // Int. J. Open Information Technologies, 2017. № 12. С. 24–29.
6. Fowler M., Sadalage P. J. NoSQL distilled: A brief guide to the emerging world of polyglot persistence. — Addison-Wesley Professional, 2012. 190 p.
7. Kohn W., Sham L. J. Self-consistent equations including exchange and correlation effects // Phys. Rev. A, 1965. Vol. 140. P. 1133–1138.
8. Gao D., Deng H., Heinisch H., Kurtz R. A new Fe–He interatomic potential based on *ab initio* calculations in α -Fe // J. Nucl. Mater., 2011. Vol. 418. Iss. 1. P. 115–120.
9. Lee B.-J., Baskes M. I., Kim H., Cho Ya. K. Second nearest-neighbor modified embedded atom method potentials for bcc transition metals // Phys. Rev. B, 2001. Vol. 64. Art. ID: 184102. doi: 10.1103/PhysRevB.64.184102.
10. Sadi T., Mehonic A., Montesi L., Buckwell M., Kenyon A., Asenov A. Investigation of resistance switching in SiO_x RRAM cells using a 3D multi-scale kinetic Monte Carlo simulator // J. Phys. Condens. Matter, 2018. Vol. 30. Iss. 8. Art. ID: 084005. doi: 10.1088/1361-648X/aaa7C1.
11. Морозов А. Ю., Ревизников Д. Л., Абгарян К. К. Вопросы реализации нейросетевых алгоритмов на мемристорных кроссбарах // Известия высших учебных заведений. Материалы электронной техники, 2019. Т. 22. № 4.
12. Newman S. Building microservices. — Sebastopol, CA, USA: O'Reilly Media, 2015. 282 p.
13. Абгарян К. К., Гаврилов Е. С. Информационная поддержка интеграционной платформы многомасштабного моделирования // Системы и средства информатики, 2019. Т. 29. № 1. С. 53–62.

Поступила в редакцию 15.04.20

INTEGRATION PLATFORM FOR MULTISCALE MODELING OF NEUROMORPHIC SYSTEMS

K. K. Abgaryan^{1,2} and E. S. Gavrilov^{1,2}

¹Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

²Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation

Abstract: The current multilevel resistive memory elements allow increasing the integration density of nonvolatile memory as well as designing and creating systems with a parallel computing mechanism. Such devices are based on memristor elements necessary for developing the foundations of analog neuromorphic networks that are used to solve data mining problems. However, the use of memristors as a part of neuromorphic devices encounters a number of problems such as the scatter of the switching parameters (voltage and memory window) from cell to cell, asymmetry and nonlinear effects, and others. Such problems dictate the need to create original simulation models and new software tools that will allow one to evaluate the influence of disturbing factors on the predictive accuracy and network learning process. In this paper, to solve the problem of multiscale modeling of neuromorphic systems, the authors use the original information technology for constructing multiscale models. For its practical implementation, an integration platform has been built that allows one to evaluate the influence of disturbing factors on the predictive accuracy and learning process of a neuromorphic network and in the future, it will be able to provide information for a reasonable choice of materials, configuration, and topology of memory cells of new-generation computers.

Keywords: multi-scale modeling; multilevel memory elements; neuromorphic networks; predictive modeling; memristor; integration platform; software package

DOI: 10.14357/19922264200215

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 19-29-03051 mk).

References

1. Abgaryan, K. K. 2017. *Mnogomasshtabnoe modelirovaniye v zadachakh strukturnogo materialovedeniya* [Multiscale modeling in material science problems]. Moscow: MAKS Press. 284 p.
2. Abgaryan, K. K. 2018. Informatsionnaya tekhnologiya postroyeniya mnogomasshtabnykh modeley v zadachakh vychislitel'nogo materialovedeniya [Information technology is the construction of multi-scale models in problems of computational materials science]. *Sistemy Vysokoy Dostupnosti* [High Availability Systems] 14(2):9–15.
3. Brodskiy, Yu. I. 2013. *Model'nyy sintez i model'no-oriyentirovannoye programirovaniye* [Model synthesis and model oriented programming]. Moscow: CC RAS. 142 p.
4. Evans, E. 2003. *Domain-driven design: Tackling complexity in the heart of software*. Addison-Wesley Professional. 560 p.
5. Abgaryan, K. K., E. S. Gavrilov, and A. M. Marasanov. 2017. Informatsionnaya podderzhka zadach mnogomasshtabnogo modelirovaniya kompozitsionnykh materialov [Multiscale modeling for composite materials computer simulation support]. *Int. J. Open Information Technologies* 12:24–29.
6. Fowler, M., and P. J. Sadalage. 2012. *NoSQL distilled: A brief guide to the emerging world of polyglot persistence*. Addison-Wesley Professional. 190 p.
7. Kohn, W., and L. J. Sham. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* 140:1133–1138.
8. Gao, D., H. Deng, H. Heinisch, and R. Kurtz. 2011. A new Fe–He interatomic potential based on *ab initio* calculations in α -Fe. *J. Nucl. Mater.* 418(1):115–120.
9. Lee, B.-J., M. I. Baskes, H. Kim, and Ya. K. Cho. 2001. Second nearest-neighbor modified embedded atom method potentials for bcc transition metals. *Phys. Rev. B* 64:184102. doi: 10.1103/PhysRevB.64.184102.
10. Sadi, T., A. Mehonic, L. Montesi, M. Buckwell, A. Kenyon, and A. Asenov. 2018. Investigation of resistance switching in SiOx RRAM cells using a 3D multi-scale kinetic Monte Carlo simulator. *J. Phys. Condens. Matter* 30(8):084005. doi: 10.1088/1361-648X/aaa7C1.
11. Morozov, A. Yu., D. L. Reviznikov, and K. K. Abgaryan. 2019. Voprosy realizatsii neyrosetevykh algoritmov na memristornykh krossbarakh [Implementation of neural network algorithms using memristor crossbars]. *Izvestiya Vysshikh Uchebnykh Zavedeniy. Materialy Elektronnoy Tekhniki* [Proceedings of Higher Educational Institutions. High Materials of Electronics Engineering] 22(4).
12. Newman, S. 2015. *Building microservices*. Sebastopol, CA: O'Reilly Media. 282 p.
13. Abgaryan, K. K., and E. S. Gavrilov. 2019. Informatsionnaya podderzhka integratsionnoy platformy mnogomasshtabnogo modelirovaniya [Informational support of the multiscale modeling integration platform]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):53–62.

Received April 15, 2020

Contributors

Abgaryan Karine K. (b. 1963) — Doctor of Science in physics and mathematics, principal scientist, Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; Head of Department, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation; kristal83@mail.ru

Gavrilov Evgeny S. (b. 1982) — scientist, Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; senior lecturer, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation; eugavrilov@gmail.com

ВЫБОР МОДЕЛЕЙ ОПТИМАЛЬНОЙ СЛОЖНОСТИ МЕТОДАМИ МОНТЕ-КАРЛО (НА ПРИМЕРЕ МОДЕЛЕЙ ПРОИЗВОДСТВЕННЫХ ФУНКЦИЙ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ)*

И. Л. Кирилюк¹, О. В. Сенько²

Аннотация: Описан подход к сравнению альтернативных вариантов линейных регрессионных моделей на временных рядах и к определению целесообразности их усложнения (посредством добавления новых переменных) с использованием нескольких вариантов методов Монте-Карло. Предложенные методы исследования с помощью генерации псевдовыборок позволяют учесть как эффекты, связанные с возможными отличиями распределений в эмпирических данных от распределения Гаусса, так и эффекты, связанные с возможной нестационарностью исследуемых временных рядов. Для этого применяется генерирование псевдовыборок — временных рядов, являющихся гауссовым белым шумом или случайным блужданием, а также перестановочный тест и метод бутстрепа. Достоверность получаемых результатов оценивается при помощи процедур ресэмплинга. Применимость рассматриваемых методов демонстрируется на примере моделей инвестиционных производственных функций регионов Российской Федерации, рассчитываемых на основе данных Федеральной службы государственной статистики.

Ключевые слова: методы Монте-Карло; перестановочные тесты; ложная регрессия; производственные функции; селекция моделей; мезоуровень экономики

DOI: 10.14357/19922264200216

1 Введение

Во многих областях исследований, например в общественных науках, предметная область не описывается надежно верифицированными и строго обоснованными моделями. В этой ситуации приобретает актуальность сравнение альтернативных вариантов регрессионных моделей, проверка оправданности усложнения таких моделей, например посредством включения в них новых переменных. До сих пор в разных областях науки применяются простейшие методы сравнения моделей посредством сравнения коэффициентов детерминации, получаемых при применении этих моделей к единому набору данных. Однако такой подход имеет тот недостаток, что коэффициент детерминации может только расти при добавлении в модель новых переменных.

Дальнейшее развитие математической статистики привело к появлению новых способов сравнения с использованием скорректированных коэффициентов детерминации, критерия Акаике, байесовского критерия Шварца и др. В литературе подобные подходы в ряде публикаций объединены под названием выбор моделей (*англ.* Model selec-

tion) [1]. Однако перечисленные критерии имеют свои ограничения, включая невозможность прямых оценок статистической достоверности превосходства усложненных моделей.

Одно из направлений Model selection — применение методов непараметрической статистики. В отличие от вышеперечисленных критериев они не предполагают конкретную форму распределения ошибок модели, но при этом часто требуют долгих вычислений. Здесь применяются перестановочные тесты [2–4] и метод бутстрепа [5]. Эти методы основаны на генерации множества псевдовыборок, полученных из распределений эмпирических данных. Методом генерации служит взятие последовательности эмпирических данных в произвольном порядке. Для перестановочных тестов в результате получается перемешивание, для бутстрепа же данные берутся «с возвращением» (т. е. какие-то значения могут дублироваться). Используемым альтернативным вариантом метода Монте-Карло служит генерация псевдовыборок с нормальным распределением.

Одной из трудностей, имеющих место при поиске моделей, корректно описывающих данные, является проблема ложных регрессий [6]. Этот

* Исследование выполнено в рамках государственного задания по теме «Феномен мезоуровня в экономическом анализе: новые теории и их практическое применение».

¹ Институт экономики Российской академии наук, igokir@rambler.ru

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук, senkoov@mail.ru

эффект имеет место, когда расчет показывает достоверность регрессионной зависимости (например, высокие значения коэффициентов детерминации), хотя на самом деле такой закономерной зависимости нет. Явление ложной регрессии характерно для стохастически нестационарных процессов, например для случайных блужданий. В таких случаях для верификации зависимостей использовать стандартную статистику (например, t -статистику) неправомерно, в частности из-за несоблюдения требования независимости отдельных наблюдений. Поэтому важную роль в проверке достоверности регрессионной зависимости играет проверка временных рядов на стационарность. Известным способом такой проверки служит тест Дики–Фуллера. Представленное исследование направлено на разработку методов верификации, применимых для регрессионных моделей, включающих нестационарные переменные, т.е. переменные, значения которых являются элементами нестационарного временного ряда. При моделировании учитывается также возможность отличия распределения ошибок от нормального.

Интересным объектом для исследования методами Model selection представляются широко применимые в экономике производственные функции, связывающие объем выпускаемой продукции с факторами производства (классическими факторами выступают труд и капитал). Простейшая спецификация — функция Кобба–Дугласа — посредством логарифмирования сводится к модели линейной регрессии, для которой получено особенно много результатов, в том числе с применением непараметрических методов. За время исследований накопилось множество обобщений этой функции и альтернативных спецификаций. Производственные функции применяются как на микроуровне экономики, так и на мезо- и макроуровне. В данной работе они исследуются на мезоуровне — на уровне российских регионов. Их анализ позволяет лучше понять особенности конкретных регионов, прогнозировать их развитие и управлять ими.

2 Описание исследуемых моделей и данных

Ранее в [7] исследовались различные варианты производственных функций, построенных по временным рядам, характеризующим развитие Российской Федерации в целом. Выводы о качестве моделей делались на основе вычисления обычных и скорректированных коэффициентов детермина-

ции. В данной работе они исследуются для отдельных регионов Российской Федерации (по данным за 1996–2014 гг.). Поскольку удобно исследовать временные ряды фиксированной длины, в расчетах не использовались данные для Чеченской республики, Крыма, Севастополя (они доступны за меньшее число лет). Регионы, входящие в состав других субъектов Российской Федерации, а не подчиняющиеся ей напрямую, также в расчетах не участвовали.

В предыдущей работе [8] для производственных функций регионов проводилась оценка применимости классической модели функции Кобба–Дугласа и определялась роль трендов по времени в обеспечении достоверности зависимости с использованием псевдовыборок в виде гауссова белого шума и случайных блужданий. В данной работе методы Монте–Карло используются для сравнения альтернативных вариантов моделей, анализа достоверности отдельных переменных с привлечением непараметрических методов (перестановочных тестов и бутстрепов).

В литературе ведется дискуссия о целесообразности использования в производственных функциях в качестве фактора производства *инвестиций* вместо *капитала* [9, 10]. Это связано, в частности, с тем, что для инвестиций проще оценивать объективные количественные значения. Кроме того, расчеты показали, что использование инвестиций при аппроксимации исследуемых данных функцией Кобба–Дугласа приводит к более высоким значениям коэффициента детерминации, чем при использовании капитала. Поэтому (в отличие от [7, 8]) в данной работе в качестве исходной модели используется инвестиционная производственная функция.

Также, в отличие от [7, 8], в переменную *труд* производственной функции введена в качестве множителя к числу занятых в экономике заработная плата, что существенно увеличило значимость труда в исследуемых вариантах производственных функций. В литературе инвестиционные производственные функции с трудом, выражаемым через зарплату, были представлены ранее, например в [10].

Основная используемая в статье модель («инвестиционный» аналог функции Кобба–Дугласа) выражается формулой:

$$Y = AI^{\alpha}L^{\beta}, \quad (1)$$

где Y — валовой региональный продукт; I — инвестиции в основной капитал; L — среднегодовая численность занятых в экономике, помноженная на среднемесячную номинальную начисленную заработную плату работающих в экономике; A , α и β — вычисляемые параметры.

Иногда на (1) накладывается условие $\alpha + \beta = 1$, но здесь это условие не используется.

Данные взяты из [11] и аналогичных сборников за 2003, 2005 и 2012 гг. Некоторые данные за одинаковые годы в разных сборниках слегка различаются, что, видимо, обусловлено внесением поправок статистическими органами, в таких случаях использовались данные из более поздних сборников. Было использовано приведение Y , L и I к постоянным ценам с помощью индексов потребительских цен (в [7] использовался индекс-дефлятор валового внутреннего продукта, а в [8] — индекс физического объема валового регионального продукта).

Посредством логарифмирования переменных из (1) получаем линейное соотношение

$$\ln Y = \ln A + \alpha \ln I + \beta \ln L, \quad (2)$$

которое теперь можно исследовать методами линейного регрессионного анализа.

Однако имеет смысл проверять альтернативные гипотезы, например о том, что производственная функция достоверно зависит только от инвестиций или только от труда:

$$\ln Y = \ln A + \alpha \ln I; \quad (3)$$

$$\ln Y = \ln A + \beta \ln L. \quad (4)$$

3 Расчеты по альтернативным моделям производственных функций

Исследуемый в статье набор данных в терминах эконометрики является частным случаем панельных данных (т. е. данных, описывающих совокупность однотипных объектов, каждый из которых характеризуется некоторым числом временных рядов показателей). Для них разработаны особые методы исследования, в частности методы выяснения того, есть ли у объектов закономерные индивидуальные признаки (модель с фиксированными эффектами) или они характеризуются единым набором признаков со случайными отклонениями (модель со случайными эффектами). Этот подход ранее использовался для оценки производственных функций регионов Российской Федерации [12] (на основе иного набора данных) и показал, что они описываются моделью с фиксированными эффектами. Проведенные авторами расчеты также подтвердили справедливость этой модели. Это делает более обоснованными исследования производственных функций регионов независимо друг от друга.

Классические методы оценки достоверности, применяемые в регрессионном анализе (вычисление R^2 , p -значения для модели в целом с использованием F-статистики и для коэффициентов при отдельных переменных с использованием t-статистики Стьюдента), показывают высокую достоверность получаемых зависимостей, однако из литературы известно, что применение стандартных методов верификации к регрессионным моделям приводит к возникновению ложных регрессий — формально весьма достоверных, но фактически бессмысленных.

Важная причина возникновения ложных регрессий кроется в несправедливости для временных рядов предположения о взаимной независимости отдельных наблюдений. Избежать подобного предположения позволяют методы верификации, основанные на искусственной генерации псевдовыборок в соответствии с предполагаемой нулевой гипотезой. В этом случае генерация данных, в частности, будет производиться с использованием предположения о соответствии исследуемого временного ряда процессу случайного блуждания.

Для оценки достоверности (1)–(4) и сравнения моделей были рассмотрены следующие модели генерации данных методами Монте-Карло.

1. Для совокупности используемых показателей генерировалось по 5000 псевдовыборок по формуле

$$X_t = e_t, \quad (5)$$

где e_t — гауссов белый шум. Известно, что он характеризуется средним значением и дисперсией, но значение коэффициентов детерминации для его реализаций конечной длины от них не зависит и определяется числом наблюдений во временных рядах (поэтому среднее и дисперсия могут устанавливаться произвольно, в данном случае они выбирались в соответствии с эмпирическими данными для регионов).

2. Генерировалось по 5000 псевдовыборок по формуле

$$X_t = X_{t-1} + e_t, \quad (6)$$

где e_t — гауссов белый шум с нулевым средним и дисперсией, соответствующей дисперсии эмпирических рядов данных. Процесс (6) называется случайным блужданием и относится к нестационарным процессам.

3. Генерировалось по 5000 псевдовыборок, для получения которых внутри используемых эмпирических временных рядов данных происходило перемешивание случайным образом порядка

их значений во времени (этот подход называется перестановочными тестами).

4. Генерировалось по 5000 псевдовыборок, где внутри используемых эмпирических временных рядов данных происходило перемешивание случайным образом порядка их значений во времени, но при этом (в отличие от предыдущего случая) некоторые значения могли повторяться за счет других данных (выборка с возвращением, может интерпретироваться как бутстреп).
5. Генерировалось по 5000 псевдовыборок, где внутри используемых эмпирических временных рядов данных происходило перемешивание случайным образом порядка приращений во времени, при этом некоторые значения могли повторяться за счет других данных. Дополнительно проводилось исключение влияния прироста значений величин реальных данных за весь период наблюдений.

Перестановочного теста для приращений не проводилось, поскольку его использование оставляет инвариантным суммарное приращение величин за период наблюдений, что вносит искажения в результаты.

Использовалась следующая методология верификации: величины параметра R^2 , характеризующего точность регрессионной модели на искусственно сгенерированных выборках, сравнивались с величинами на истинной выборке. Для полученных каждым из пяти способов псевдовыборок вычислялись коэффициенты детерминации, получающиеся при аппроксимации данных исследуемыми моделями (2)–(4). Затем 5000 полученных коэффициентов детерминации ранжировались по величине и реальное значение R^2 сравнивалось с 250-м по величине значением от максимального среди полученных для псевдовыборок (которое далее будем называть 95%-ным квантилем эмпирического распределения, или просто 95%-ным квантилем). Это способ оценки достоверности полученного результата на уровне 95% (в данном случае того, что значение R^2 получено не случайно, а выражает некоторую закономерность, если реальное R^2 больше 250-го по величине псевдовыборочного).

Можно считать, что для конкретных регионов, для которых значение R^2 превышает соответствующее значение 95%-ного квантиля, выявлена закономерность, достоверная на уровне ниже 0,05. Однако авторам доступна информация о 79 регионах, и это дает дополнительную возможность проверить, не является ли выявленная закономерность на самом деле случайной (такая проб-

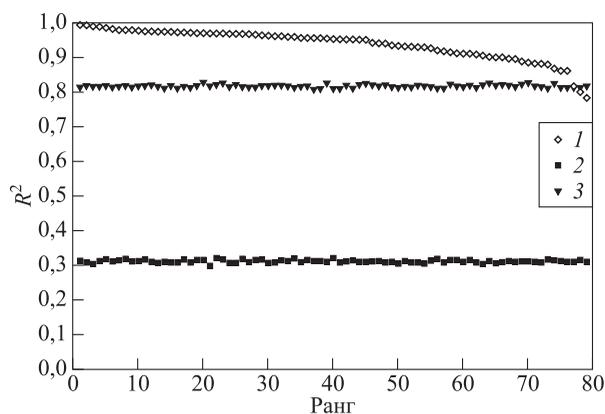


Рис. 1 Ранговое распределение R^2 , рассчитанных по формуле (2), в сравнении с 95%-ными квантилями R^2 для симуляций, рассчитанных по формулам (5) и (6): 1 — реальные значения R^2 ; 2 и 3 — 95%-ные квантили эмпирического распределения R^2 для псевдовыборок, генерируемых процессами (5) и (6) соответственно

лема становится особо актуальной при множественном тестировании). Если, например, лишь для одного региона из 79 для какой-либо модели превзойден 95%-ный квантиль, то достоверность выявленной закономерности значительно меньше, чем когда он превзойден для всех регионов.

Классические оценки по F-критерию соответствия модели (2) реальным данным показывают высокие значения коэффициента детерминации (рис. 1). Они варьируют в диапазоне от 0,994 для Дагестана до 0,783 для Мурманской области. Соответствующие p -значения варьируют в диапазоне $(1,55 \cdot 10^{-18}; 4,96 \cdot 10^{-6})$. При этом 95%-ный квантиль для сгенерированных согласно формуле (5) псевдовыборок равен примерно 0,313, а для псевдовыборок, сгенерированных согласно формуле (6), — 0,813. Только для двух регионов из 79 имеет место соотношение $R^2 < 0,813$.

Описываемая моделью зависимость в данных, таким образом, не может быть полностью объяснена не только чистой случайностью, но и ложной регрессией, обусловленной нестационарностью вида (6) рассматриваемых процессов, из чего можно сделать вывод о достоверности закономерности, выражаемой соотношениями (1)–(2) на уровне меньше 0,05. Результат, полученный с использованием (5) соответствует обычному p -значению для R^2 , вычисляемому в стандартных пакетах в рамках параметрической статистики, а использование (6) позволяет получить более надежные результаты в оценке наличия закономерностей с учетом возможной нестационарности исследуемых процессов. В качестве альтернативы (5) в вычислениях

использовались бутстреппы, но они дали близкие к полученным для псевдовыборок, сгенерированных по моделям (5) и (6), результаты: соответственно $R^2 \approx 0,313$ (как и перестановочный тест) и $R^2 \approx 0,87$ — и не привели к качественно новым выводам о достоверности исследуемых моделей.

Тест Дики–Фуллера в целом формально показывает нестационарность временных рядов и остатков модели (2) по 19 значениям во времени, но для получения достоверных результатов не хватает данных (ряды должны были бы быть хотя бы в несколько раз длиннее), и потому используется сравнение реальных результатов со сгенерированными по формуле (6) в качестве альтернативы этому тесту, пригодной для малых выборок.

Возникает вопрос, не описываются ли данные с настолько приемлемой точностью более простыми соотношениями, например формулами (3) или (4), что усложнение модели до (2) может оказаться нецелесообразным. Дальнейший анализ выявляет подходы к ответу на этот вопрос.

Достоверность отличия от нуля отдельных коэффициентов — $\ln A$, α , β — в формуле (2) по классическому t-критерию Стьюдента следующая: условие $p < 0,05$ выполняется для $\ln A$ у 41 региона, для α — у 37 регионов и для β — у 60 регионов.

Зависимость, определяемая формулой (3), характеризуется для используемых в статье данных коэффициентами детерминации в диапазоне (0,346; 0,986). При этом 95%-ные квантили эмпирического распределения R^2 для псевдовыборок, генерируемых по формулам (5) и (6), равны соответственно 0,209 и 0,707. Для четырех регионов из 79 имеет место соотношение $R^2 < 0,707$.

Зависимость (4) характеризуется R^2 в диапазоне (0,62; 0,99), 95%-ные квантили R^2 для (5) и (6) те же, что для (3). Лишь для одного региона имеет место соотношение $R^2 < 0,707$. Итак, модели (3) и (4) сами по себе с весьма высокой достоверностью описывают данные, хотя достоверность на уровне меньше 0,05 отличия от нуля коэффициентов формулы (2) по t-критерию для значительного числа регионов не имеет места. Это связано с наличием мультиколлинеарности, корреляции между L и I . Существенный вклад в мультиколлинеарность вносит наличие общего тренда по времени, значение которого (для несколько другого набора показателей и с применением другого подхода) обсуждалось в [5]. В следующем разделе описаны способы оценки целесообразности для описания исследуемого набора данных усложнения моделей (3) и (4) до модели (2).

4 Использование метода генерации псевдовыборок для оценки целесообразности усложнения модели

Объединенная модель (2) статистически достоверна, как и более простые модели (3) и (4). Но существует ли при этом комбинированный, синергетический эффект, можно ли доказать, что объединенная модель действительно лучше? Подходы к ответу на этот вопрос будут приведены в данном разделе.

Опишем алгоритм, применявшийся для оценки целесообразности использования для описания динамики Y обоих факторов производства — I и L . Поскольку обычный R^2 может только увеличиваться при введении в модель новой переменной, даже при использовании в качестве всех данных случайно сгенерированных псевдовыборок расчет по модели с дополнительной имитируемой случайным процессом переменной даст некоторый прирост в значении R^2 . Прирост будет и при добавлении такой переменной к реальным данным. Рассчитав эти приросты для достаточно большого числа псевдовыборок и определяя квантили распределения приростов, можно оценить, насколько достоверно усложненный вариант модели в действительности лучше описывает ситуацию по сравнению с исходной моделью, насколько вероятно, что увеличение значения R^2 (будем обозначать его ΔR^2) не является артефактом.

Было проведено сравнение «длинной» модели (2) с «короткими» моделями (3) и (4). Набор из 5000 имитируемых длинных моделей рассчитывался на основе реальных данных короткой модели и случайных псевдовыборок, описываемых формулами (5) и (6) и имитирующих реальные данные той из переменных длинной модели (2), которая отсутствует в сравниваемой короткой модели. Для полученного набора вычислялись значения коэффициентов детерминации. Затем значения R^2 «короткой модели» (3) или (4), посчитанные для реальных данных, вычитались из соответствующих значений R^2 «длинной модели» (2) и определялись 95%-ные квантили полученных наборов ΔR^2 . Они сравнивались с реальными значениями ΔR^2 (т.е. с такими разностями, где и в короткой, и в длинной моделях все данные реальны).

Расчеты по сравнению моделей (2) и (3) показали, что для 61 региона реальные значения ΔR^2 больше 95%-ных квантилей симуляций (5) и для 44 регионов больше 95%-ных квантилей симуляций (6).

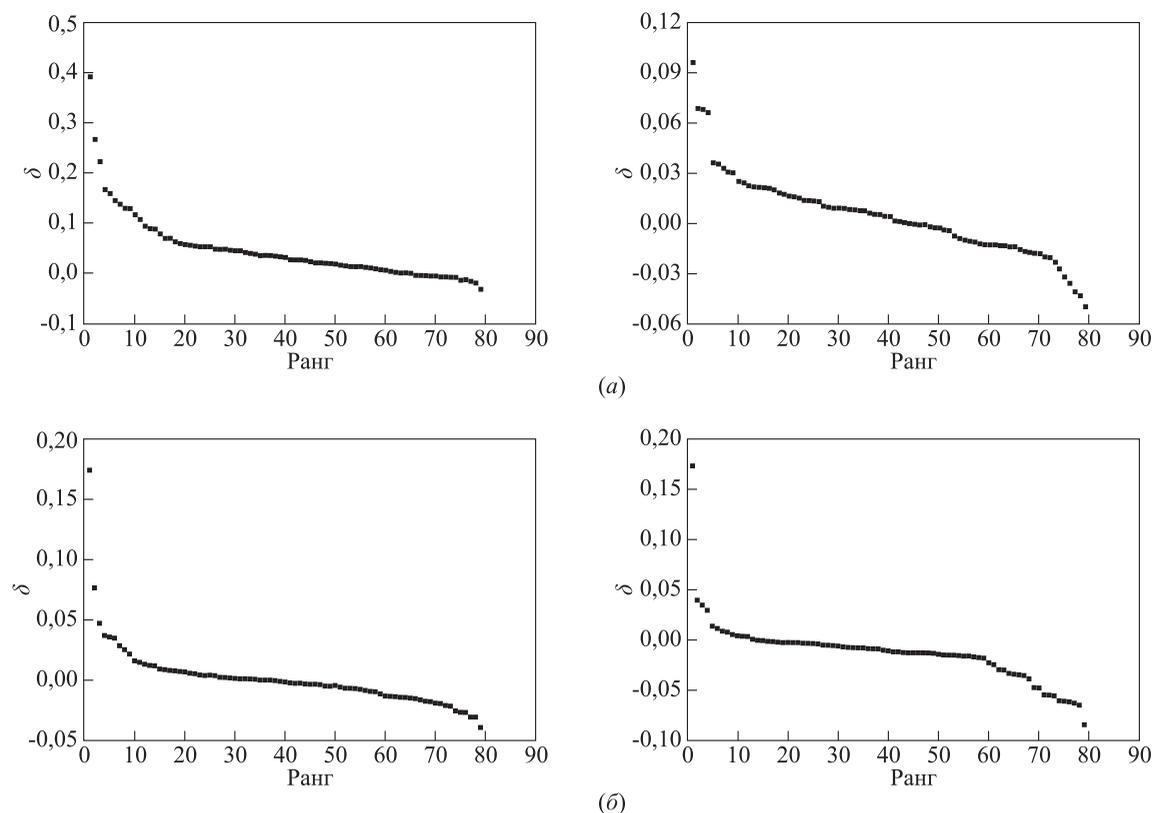


Рис. 2 Ранговое распределение разностей $\delta = (\Delta R^2_{\text{реал}} - \Delta R^2_{250})$ реальных значений ΔR^2 , рассчитанных по формулам (2) и (3) (а) и (2) и (4) (б), и разностей 95%-ных квантилей эмпирических распределений ΔR^2 для симуляций, рассчитанных по формулам (5) (левый столбец) и (6) (правый столбец) для случаев зависимостей от одной и двух переменных

Сравнение моделей (2) и (4) показало, что значения ΔR^2 для 38 регионов больше 95%-ных квантилей симуляций (5) и для 13 регионов больше 95%-ных квантилей симуляций (6).

Как видно из результатов расчетов, представленных на рис. 2, сравнение с симуляциями, основанными на модели (6), не позволяет сделать вывод со значимостью на уровне $p < 0,05$ об информативности переменной I (инвестиции) для 34 из 78 регионов. Симуляции по модели (6) не позволяют сделать вывод со значимостью на уровне $p < 0,05$ об информативности переменной L (труд) уже для 65 из 78 регионов.

В описанных выше расчетах информация о каждом из регионов использовалась независимо от информации о других регионах. Вместе с тем для каждого отдельного региона вывод основывается на анализе весьма короткого временного ряда. Совместное использование информации обо всех регионах может дать более достоверные результаты об информативности переменных I и L в целом безотносительно к каждому конкретному региону. Поэтому были проведены дополнительные расче-

ты. Были вычислены суммы ΔR^2 по всем регионам для пар моделей (2), (3) и (2), (4). Такие же суммы были вычислены в каждом случае для 5000 наборов из 79 псевдовыборок, где переменная, не вошедшая в короткую модель, генерировалась по формулам (5) и (6). Полученные суммы для реальных данных сравнивались с 95%-ными квантилями эмпирического распределения разностей коэффициентов детерминации псевдовыборок. Для обеих пар моделей была показана высокая достоверность исследуемых разностей ΔR^2 . Для случая с короткой моделью (3) получены интервалы сумм ΔR^2 псевдовыборок по формуле (5) — (0,28; 1,24) и по формуле (6) — (0,99; 3,23) при реальном значении 6,02. Для случая с короткой моделью (4) получены интервалы сумм ΔR^2 псевдовыборок по формуле (5) — (0,15; 0,70) и по формуле (6) — (0,42; 1,37), при реальном значении 1,57. Таким образом, в обоих случаях имеет место высокая достоверность, оцениваемая по крайней мере на уровне $p < 0,0002$.

Все использованные в статье расчеты проводились с помощью программ, написанных на языке R.

5 Заключение

Использованные в статье методы оценки целесообразности усложнения моделей имеют преимущества перед классической методологией оценки достоверности регрессии по F- и t-статистикам и могут быть использованы для широкого класса моделей. Преимущество заключается, в частности, в том, что легко учитывается эффект нестационарности, а также существуют непараметрические варианты методов. В исследованном случае перестановочные тесты, бутстрепы и симуляции с использованием нормального распределения дали близкие результаты, но такая ситуация не является общей, непараметрические методы имеют более широкую сферу применимости. Представляет интерес интеграция методов исследования нестационарных процессов непараметрическими методами с существующими подходами к исследованию панельных данных, что предполагается возможным предметом дальнейших исследований.

Литература

1. *Стрижов В. В., Крымова Е. А.* Методы выбора регрессионных моделей. — М.: ВЦ РАН, 2010. 60 с.
2. *Kennedy P. E., Cade B. S.* Randomization tests for multiple regression // *Commun. Stat. Simul. C.*, 1996. Vol. 25. Iss. 4. P. 923–936.
3. *Anderson M. J., Robinson J.* Permutation tests for linear models // *Aust. NZ J. Stat.*, 2001. Vol. 43. Iss. 1. P. 75–88.
4. *Senko O. V., Dzyba D. S., Pigarova E. A., Rozhinskaya L. Ya., Kuznetsova A. V.* A method for evaluating validity of piecewise-linear models // *Conference (International) on Knowledge Discovery and Information Retrieval Short Papers.*— Scitepress, 2014. P. 437–443. doi: 10.5220/0005156904370443.
5. *Skrobotov A.* On bootstrap implementation of likelihood ratio test for a unit root // *Econ. Lett.*, 2018. Vol. 171(C). P. 154–158.
6. *Granger C. J., Newbold P.* Spurious regressions in econometrics // *J. Econometrics*, 1974. Vol. 2. Iss. 2. P. 111–120.
7. *Кирилюк И. Л.* Модели производственных функций для российской экономики // *Компьютерные исследования и моделирование*, 2013. Т. 5. № 2. С. 293–312.
8. *Кирилюк И. Л., Сенько О. В.* Исследования соотношений между нестационарными временными рядами на примере производственных функций // *Машинное обучение и анализ данных*, 2018. Т. 4. № 3. С. 142–151.
9. *Поспелов И. Г., Поспелова И. И., Хохлов М. А., Шипулина Г. Е.* Новые принципы и методы разработки макромоделей экономики и модель современной экономики России. — М.: ВЦ РАН, 2006. 242 с.
10. *Гребнев М. И.* Построение производственных функций регионов России // *ВУЗ. XXI век*, 2015. № 2. С. 50–56.
11. *Регионы России. Социально-экономические показатели.* 2017. — М.: Росстат, 2017. 1402 с. https://gks.ru/bgd/regl/B17_14p/Main.htm.
12. *Бахитова Р. Х., Ахметшина Г. А., Лакман И. А.* Панельное моделирование объема выпуска продукции для регионов России // *Управление большими системами*, 2014. Вып. 50. С. 99–109.

Поступила в редакцию 22.05.19

SELECTION OF OPTIMAL COMPLEXITY MODELS BY METHODS OF NONPARAMETRIC STATISTICS (ON THE EXAMPLE OF PRODUCTION FUNCTION MODELS OF THE REGIONS OF THE RUSSIAN FEDERATION)

I. L. Kirilyuk¹ and O. V. Sen'ko²

¹Institute of Economics of the Russian Academy of Sciences, 32 Nakhimovskiy Pr., Moscow 117218, 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 article describes an approach to comparing alternative variants of linear regression models on time series and determining the appropriateness of complicating them (by adding new variables) using several variants of Monte-Carlo methods. The proposed research methods using pseudosampling generation allow taking into account both the effects associated with possible differences of distributions in empirical data from the Gauss distribution and the effects associated with possible nonstationarity of the time series under study. For this purpose, pseudosampling generation is used — time series, which are Gaussian white noise, random walk generation, as well as the permutation test and the bootstrap method. Reliability of the obtained results is estimated using resampling.

Applicability of the considered methods is demonstrated by the example of models of investment production functions of regions of the Russian Federation, calculated on the basis of data from the Federal State Statistics Service.

Keywords: Monte-Carlo methods; permutation tests; spurious regression; production functions; model selection; meso level of the economy

DOI: 10.14357/19922264200216

Acknowledgments

The research was carried out within the framework of the state task “The phenomenon of mesolevel in economic analysis: New theories and their practical application.”

References

1. Strizhov, V. V., and E. A. Krymova. 2010. *Metody vybora regressionnykh modeley* [Methods for choosing regression models]. Moscow: CC RAS. 60 p.
2. Kennedy, P. E., and B. S. Cade. 1996. Randomization tests for multiple regression. *Commun. Stat. Simul. C.* 25(4):923–936.
3. Anderson, M. J., and J. Robinson. 2001. Permutation tests for linear models. *Aust. NZ J. Stat.* 43(1):75–88.
4. Senko, O. V., D. S. Dzyba, E. A. Pigarova, L. Ya. Rozhinskaya, and A. V. Kuznetsova. 2014. A method for evaluating validity of piecewise-linear models. *Conference (International) on Knowledge Discovery and Information Retrieval Short Papers*. Scitepress. 437–443. doi: 10.5220/0005156904370443.
5. Skrobotov, A. 2018. On bootstrap implementation of likelihood ratio test for a unit root. *Econ. Lett.* 171(C):154–158.
6. Granger, C. J., and P. Newbold. 1974. Spurious regressions in econometrics. *J. Econometrics* 2(2):111–120.
7. Kirilyuk, I. L. 2013. Modeli proizvodstvennykh funktsiy dlya rossiyskoy ekonomiki [Models of production functions for the Russian economy]. *Komp'yuternye issledovaniya i modelirovanie* [Computer Research Modeling] 5(2):293–312.
8. Kirilyuk, I. L., and O. V. Senko. 2018. Issledovaniya sootnosheniy mezhdru nestatsionarnymi vremennymi ryadami na primere proizvodstvennykh funktsiy [Studies of the relationship between nonstationary time series on the example of production functions]. *Mashinnoe Obucheniye i Analiz Dannya* [Machine Learning Data Analysis] 4(3):142–151.
9. Pospelov, I. G., I. I. Pospelova, M. A. Khokhlov, and G. E. Shipulina. 2006. *Novye printsipy i metody razrabotki makromodeley ekonomiki i model' sovremennoy ekonomiki Rossii* [New principles and methods for developing macromodels of the economy and a model of the modern economy of Russia]. Moscow: CC RAS. 242 p.
10. Grebnev, M. I. 2015. Postroenie proizvodstvennykh funktsiy regionov Rossii [Construction of production functions of Russian regions]. *VUZ. XXI vek* [High School. XXI Century] 2:50–56.
11. Regiony Rossii. Sotsial'no-ekonomicheskie pokazateli. 2017 [Regions of Russia. Socio-economic indicators. 2017]. Moscow: Rosstat. 1402 p. Available at: https://gks.ru/bgd/regl/B17_14p/Main.htm (accessed May 1, 2020).
12. Bakhitova, R. Kh., G. A. Akhmetshina, and I. A. Lakman. 2014. Panel'noe modelirovanie ob'ema vypuska produktsii dlya regionov Rossii [Panel modeling of production output in Russian regions]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 50:99–109.

Received May 22, 2019

Contributors

Kirilyuk Igor L. (b. 1974) — scientist, Institute of Economics of the Russian Academy of Sciences, 32 Nakhimovskiy Pr., Moscow 117218, Russian Federation; igokir@rambler.ru

Sen'ko Oleg V. (b. 1957) — Doctor of Science in physics and mathematics, academician of RAS, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; senkoov@mail.ru

РЕДУЦИРОВАНИЕ СПЕКТРА МОДЕЛЕЙ ПЕРЕВОДА В НАДКОРПУСНЫХ БАЗАХ ДАННЫХ

В. А. Нуриев¹, И. М. Зацман²

Аннотация: Описывается подход к редуцированию спектра моделей перевода, которые регистрируются и хранятся в надкорпусных базах данных (НБД). Этот подход ориентирован на применение в профессиональном («человеческом») и машинном переводе. В настоящее время информационные ресурсы НБД обеспечивают решение широкого круга задач, представляющих интерес для информатики, компьютерной лингвистики и языкознания. В данной статье акцент ставится на возможностях использования НБД в интересах переводной деятельности — для редуцирования спектра моделей перевода. Как правило, в каждом конкретном случае переводчик оказывается в ситуации многовариантного выбора: ввиду синонимического потенциала, которым обладают естественные языки, вместо единственно возможного решения при переводе имеется некоторое множество относительно взаимозаменяемых по смыслу эквивалентов, т. е. «веер альтернатив». Выбирая один вариант из целого ряда потенциально возможных, переводчик, чтобы максимально сузить диапазон поиска, руководствуется некоторыми релевантными характеристиками переводимого текста. Цель статьи — описать подход к применению НБД, позволяющий сузить диапазон поиска релевантной модели перевода.

Ключевые слова: параллельные тексты; перевод; модели перевода; надкорпусная база данных; многовариантный выбор

DOI: 10.14357/19922264200217

1 Введение

В последние несколько десятилетий существенно возрос научный и практический интерес к таким информационным ресурсам, как электронные текстовые корпуса и лингвистические базы данных. Эти ресурсы обеспечивают получение нового фундаментального знания, позволяют верифицировать, корректировать и дополнять научные знания, полученные в докорпусную эпоху [1–3]. Они открывают широкие возможности для исследователей-лингвистов, работа которых связана с большими массивами текстов на естественном языке (см., например, [4]). Также представители некоторых профессий — переводчики, писатели, преподаватели (русского языка и литературы, иностранных языков, культуры речи и т. д.) — получают новые инструменты, с помощью которых рабочий процесс можно значительно компьютеризировать и оптимизировать (см., например, [5–7]).

Разновидностью таких информационных ресурсов являются НБД (о НДБ см. [8]), функционирующие на основе текстов параллельных корпусов, входящих в состав Национального корпуса русского языка (НКРЯ). Надкорпусные базы данных раз-

рабатываются с 2012 г. в Институте проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук и представляют собой уникальный инструмент, который имеет ярко выраженную научно-исследовательскую направленность, поскольку используется для извлечения нового лингвистического знания из параллельных текстов. Также он может широко применяться как в профессиональной переводческой практике, так и для формирования лингвистических онтологий, ориентированных на повышение качества машинного перевода, в том числе при отсутствии или недостаточном объеме доступных параллельных текстов [9].

В данной статье рассматриваются возможности применения НБД в интересах переводной деятельности, и основная цель статьи, таким образом, — представление подхода к редуцированию спектра моделей перевода, которые регистрируются и хранятся в НБД. Демонстрация этих возможностей проводится на примере НБД коннекторов³ (см. разд. 3). Однако прежде, чем перейти к описанию указанного подхода, следует дать общее представление о том, как происходит процесс перевода и как НБД регистрирует модели перевода.

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, nurieff.v@gmail.com

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, izatsman@yandex.ru

³ Об НБД коннекторов см. [10–13].

2 Модели перевода

Результат переводной деятельности — это реализация в конечном тексте конкретной переводной модели (устойчивого межъязыкового соответствия) или окказионального варианта. Изучая перевод, исследователь стремится реконструировать переводческое действие, т. е. зафиксировать изменения, которые претерпевает исходный текст (ИТ), оформляясь на переводящем языке (ПЯ).

В некоторых работах по теории и практике перевода отмечается: «Основная особенность... перевода состоит в жестко детерминированных отношениях между системами языка-источника и принимающего языка, которые в большинстве случаев предписывают вполне однозначное решение той или иной конкретной переводческой задачи» [14, с. 47]. В отдельных случаях, возможно, это так, однако обычно при переводе однозначность переводческого решения (в выборе конкретного варианта или модели) не настолько очевидна. Вряд ли было бы корректно «говорить о возможности лишь одного единственно правильного варианта перевода (как некорректно... говорить об одной единственно возможной интерпретации... в обычных условиях общения)» [15, с. 22]. В силу синонимического потенциала, которым обладают естественные языки, вместо единственно возможного решения при переводе, как правило, имеется некоторое множество относительно взаимозаменяемых по смыслу эквивалентов, т. е. «веер альтернатив».

Переводчик интерпретирует ИТ и, руководствуясь своей интерпретацией, занимается отбором в ПЯ соответствующих средств выражения. Если ИТ — это определенным образом упорядоченная информационно нагруженная последовательность знаков, то «интерпретационная сторона знака — это вся совокупность ситуационных факторов понимания (предметных, коммуникативных, социокультурных), которые обуславливают необходимость употребления знака в момент речи» [15, с. 25]. Другими словами, ИТ представляет собой кодированное средствами естественного языка смысловое содержание, объем которого может изменяться в процессе перевода. С точки зрения информатики процессы профессионального (выполняемого человеком-специалистом) перевода суть информационные трансформации объектов ментальной и социоинформационной сред, а процессы машинного перевода — информационные трансформа-

ции объектов социоинформационной и цифровой сред¹. Рисунок иллюстрирует средовую принадлежность информационных трансформаций в профессиональном и машинном переводе на примере перевода книги.

С точки зрения теории перевода переводящий субъект обрабатывает конкретный языковой знак, перебирая возможные его смысловые значения, актуализированные в заданном языковом окружении — контексте, а затем выбирает соответствующие средства, доступные на ПЯ.

Выбор происходит из объектов, представляющих собой переводные соответствия — *модели перевода*, которые уже систематизированы и генерализованы по разным основаниям, а также зафиксированы в контрастивных грамматиках, двуязычных словарях, учебниках по переводу и т. д. Для использования в машинном переводе этим моделям присваиваются формальные признаки, характеризующие сферу применения, что обеспечивает НБД, регистрирующая модели перевода. Регистрация/интеграция информации о моделях перевода и их признаках в НБД осуществляется на основе анализа параллельных текстов (ИТ и его перевода, выполненного человеком-специалистом). Эта информация, в свою очередь, может быть использована для повышения качества машинного перевода путем редуцирования всего спектра моделей на этапе спецификации, т. е. выбора одного конкретного межъязыкового соответствия из этого спектра.

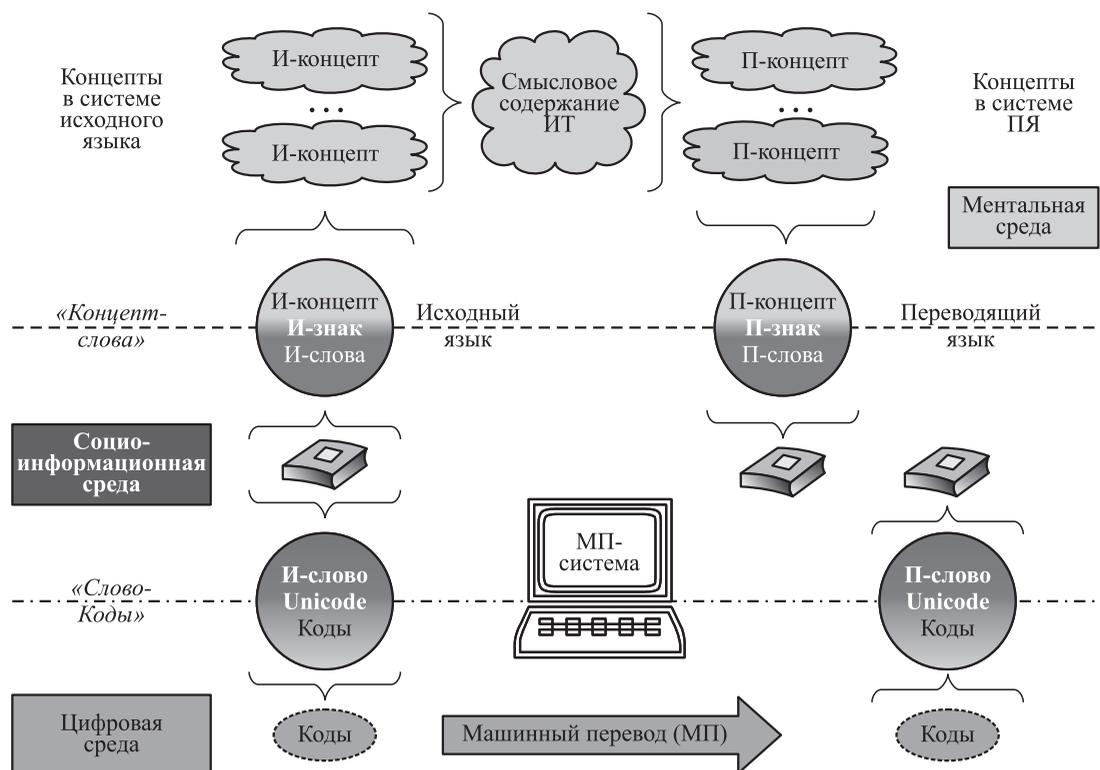
В наборе зафиксированных моделей перевода может не оказаться нужного варианта, и тогда перед переводчиком встает задача выявления смысловых компонентов (= концептов), необходимых для интерпретации обрабатываемой языковой единицы, чтобы впоследствии путем обобщения отнести ее к той или иной семантической общности и обнаружить ее возможные переводные соответствия².

Спектр моделей перевода для разных языковых единиц наряду с двуязычными словарями и контрастивными грамматиками может регистрироваться и аннотироваться (т. е. описываться в соответствии с определенным набором рубрик) также в электронных базах данных, в том числе в НБД. Одной из таких баз данных служит НБД коннекторов³, которая упоминалась выше. Она предоставляет ряд возможностей в отношении динамической (on-the-fly) генерализации примеров и моделей перевода по разным основаниям (логико-семантические от-

¹ Понятие информационной трансформации во многом совпадает с трактовкой, предложенной П. Розенблумом (см. [16, 17]).

² Таким образом, в процессе перевода используются две категории моделей перевода: (1) «внешние», которые описаны в словарях, справочниках, грамматиках и т. д.; (2) «индивидуализированные», которые созданы переводящим субъектом и не являются частью существующего множества «внешних» моделей перевода.

³ Коннектор — языковая единица, функция которой состоит в выражении логико-семантических отношений между двумя соединенными с ее помощью текстовыми фрагментами (см. подробнее [18–20]).



Информационные трансформации в профессиональном и машинном переводе: префикс И (И-концепт, И-слово, И-знак) относится к системе исходного языка; префикс П (П-концепт, П-слово, П-знак) — к системе ПЯ; фигурные скобки указывают направление процессов профессионального и машинного перевода

ношения, устанавливаемые коннектором; структура коннектора; его позиция, статус и т.д.) [21]. В НБД предусмотрено формирование и визуализация спектра моделей перевода каждого коннектора (см. пример в [22, табл. 2]).

Для параллельных текстов пользователь НБД имеет возможность интерактивно осуществить обратный переход от генерализованных моделей перевода к примерам их использования в текстах, т.е. полностью обратить процесс генерализации. Чтобы обеспечить обратимость, в НБД каждый пример перевода связан с соответствующей динамически сформированной моделью перевода, которая имеет гипертекстовую ссылку на список всех примеров ее реализации. Иными словами, наряду с динамической генерализацией моделей перевода НБД обеспечивает и их спецификацию с помощью гипертекстовых ссылок [21].

3 Спектр моделей перевода

Подход к редуцированию спектра переводов, представляемый здесь, носит пошаговый харак-

тер и включает в себя несколько основных этапов. Реализуемость подхода демонстрируется на примере аннотаций НБД коннекторов и текстов французско-русского параллельного подкорпуса¹, входящего в состав НКРЯ. По состоянию на 31.03.2020 объем этих текстов в НБД составлял 649 245 французских словоупотреблений и 509 033 русских, а всего 1 158 278 словоупотреблений (= токенов).

На первом этапе был отобран один из наиболее частотных однокомпонентных французских коннекторов — *c'est-à-dire* (204 аннотации). На основе аннотаций НБД для него был сформирован экспериментальный массив моделей перевода в виде кортежей $\langle c'est-à-dire; \text{вариант перевода} \rangle$.

Например, для наиболее продуктивной модели перевода кортеж принимает следующий вид: $\langle c'est-à-dire; \text{то есть} \rangle$ (см. варианты перевода в табл. 1). Для этого массива генерализованная модель перевода имеет вид $\langle c'est-à-dire; \{ \text{вариант перевода} \} \rangle$, где вариант перевода обозначает весь спектр таких вариантов во втором столбце табл. 1.

¹ Аннотации коннекторов в НБД формируются посредством обработки параллельных текстов, включающих ИТ и их переводы. Для эксперимента, описываемого в данной статье, привлекались профессиональные переводы.

Таблица 1 Варианты перевода французского коннектора *c'est-à-dire*

№	Вариант перевода	Число аннотаций
1	то есть	157
2	zero (нулевой эквивалент)	29
3	иными словами	8
4	иначе говоря	2
5	— (тире)	5
6	a	1
7	итого	1
8	: (двоеточие)	1
Всего аннотаций:		204

Затем было определено число аннотаций с различными сочетаниями рубрик, использованных для аннотирования текстовых фрагментов, где употреблен коннектор *c'est-à-dire* (табл. 2)¹. Эта таблица показывает, какие рубрики характеризуют модель перевода в НБД, а также насколько продуктивна эта модель при том или ином их сочетании.

Экспериментальные данные включают следующие рубрики: порядок следования связываемых коннектором текстовых фрагментов (р CNT q); встраивающий/встроенный коннектор (SubCNT/SuperCNT); маркирует часть предложения без предикации/с предикацией; дистантное/контактное расположение составляющих в неоднословных коннекторах (контакт/дистант); позиция (начальная/неначальная/конечная); логико-семантические отношения, устанавливаемые

коннектором (переформулирование и т.д.); маркирует сверхфразовое единство, самостоятельное повествовательное предложение, сложное предложение и т.д. Релевантность приведенных рубрик определялась в ходе предыдущих корпусно-ориентированных контрастивных исследований, инструментом в которых служила НБД коннекторов. Результаты некоторых исследований см. в [19–25].

Приведенные в табл. 2 рубрики, характеризующие коннектор и его контекст, были проанализированы на предмет их релевантности для выбора переводного варианта². Следовало определить, позволяет ли имеющийся в НБД (исходно заданный исследователями-лингвистами³) набор рубрик сузить область переводческого поиска, т. е. редуцировать спектр моделей перевода.

Планировалось ответить экспериментально на следующий вопрос: «Насколько исходно заданная в НБД система классификационных рубрик дает возможность редуцировать спектр конкретных моделей перевода?» Для получения ответа на этот вопрос был проведен семантический анализ распределения аннотаций как по рубрикам, так и по их сочетаниям.

В табл. 1 показано, что французский коннектор имеет более чем одно переводное соответствие на русском языке: для *c'est-à-dire* в НБД на 31.03.2020 зарегистрировано 8 вариантов перевода, включая 2 с пунктуационными знаками. Чтобы переводчик мог существенно редуцировать этот спектр переводных вариантов, указанных в табл. 2 рубрик недостаточно. По-видимому, в интересах переводной

Таблица 2 Число аннотаций с моделями перевода для *c'est-à-dire* (первые два столбца) при сочетании рубрик (третий столбец)

Оригинал	Вариант перевода	Сочетание рубрик	Число аннотаций
<i>c'est-à-dire</i>	то есть	р CNT q + SubCNT + SuperCNT + без предикации + + контакт + начальная + переформулирование	10
	zero	р CNT q + SubCNT + SuperCNT + без предикации + + контакт + начальная + переформулирование	1
	то есть	р CNT q + SubCNT + SuperCNT + контакт + начальная + + переформулирование + сложное предложение	1
	то есть	р CNT q + SubCNT + без предикации + контакт + начальная + + переформулирование	33
	zero	р CNT q + SubCNT + вставка + контакт + начальная + + переформулирование	1
	—	Другие сочетания рубрик	158
Итого:			204

¹ Исследуемые сочетания рубрик встретились в 46 из 204 аннотаций с коннектором *c'est-à-dire*.

² Анализ проводился одним из соавторов данной статьи В. А. Нуриевым — профессиональным переводчиком, имеющим опубликованные переводы с французского и английского языков.

³ Набор рубрик исходно формировался лингвистами, исходя из определенных задач, напрямую не связанных с переводческой деятельностью.

деятельности число рубрик необходимо увеличить. Этот промежуточный вывод позволили сделать несколько следующих наблюдений:

- 83% случаев употребления *c'est-à-dire* (170 аннотаций из 204) в НБД приходится на сборник научных статей Ж. Женетта «Фигуры»;
- 86% случаев перевода *c'est-à-dire* нулевым эквивалентом (25 аннотаций из 29) также приходится на эту работу, что, по-видимому, мотивировано необходимостью синонимического варьирования в подборе эквивалентов;
- этой же необходимостью мотивировано употребление при переводе *c'est-à-dire* в «Фигурах» Ж. Женетта других переводных соответствий типа *иными словами* (5 вхождений); *иначе говоря* (2 вхождения), *т.е.* (1 вхождение);
- нулевой эквивалент часто требует значительной синтаксической переработки высказывания и влечет за собой изменение его общего смысла;
- реализация модели перевода (*c'est-à-dire; то есть*) больше характерна для научного стиля — 87% случаев употребления *то есть* зафиксировано в переводе «Фигур» Ж. Женетта;
- при переводе текстов художественных жанров наблюдается тенденция к большему синонимическому варьированию, здесь реализуются такие модели перевода *c'est-à-dire*, как *т.е.* (4 вхождения), *иными словами* (3 вхождения), *а* (1 вхождение), *итого* (1 вхождение), *двоеточие* (1 вхождение);
- нулевой эквивалент для текстов художественных жанров менее характерен (14% всех случаев, причем в разных переводах).

На основе этих наблюдений можно предположить, что в НБД необходимо ввести дополнительную рубрику «художественный/нехудожественный текст», которая позволит редуцировать спектр моделей перевода.

4 Заключение

В статье предложен подход к применению НБД для редуцирования спектра моделей перевода, на основе которого планируется вычислить их условные частотности с разными сочетаниями рубрик, проставленными в процессе аннотирования. Предлагаемый подход иллюстрируется на примере анализа моделей перевода французского коннектора

c'est-à-dire. Планируется выполнить проверку реализуемости этого подхода на примере нескольких французских коннекторов, включая *comme*, чье употребление¹ характеризуется большей равномерностью — случаи его употребления относительно равномерно распределяются по целому ряду разножанровых текстов, принадлежащих разным авторам (О. де Бальзак, Ф. Бегбедер, Ж. Верн, Р. Госинни, Ж. Женетт, Ж. Кокто, П. Модигано и др.). Кроме того, для *comme* зафиксировано большее многообразие вариантов перевода (31 вариант на 206 аннотаций в НБД по состоянию на 31.03.2020) и этот коннектор, в отличие от *c'est-à-dire*, устанавливающего только один тип логико-семантических отношений (переформулирования), может сигнализировать о четырех разных типах отношений (пропозициональной причины; аддитивные пропозициональные отношения; отношение аналогии; сравнительные отношения).

Литература

1. Zatsman I., Buntman N., Kruzhhkov M., Nuriev V., Zaliziak Anna A. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery // 15th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2014. Vol.3. P. 1063–1071.
2. Zatsman I., Buntman N. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings // 16th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2015. P. 851–860.
3. Zatsman I., Buntman N., Coldefy-Faucard A., Nuriev V. WEB knowledge base for asynchronous brainstorming // 17th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2016. Vol. 1. P. 976–983.
4. Ляшевская О. Н. Корпусные инструменты в грамматических исследованиях русского языка. — М.: ЯСК: Рукописные памятники Древней Руси, 2016. 520 с.
5. Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate / Eds. A. Beeby, P. Rodríguez Inés, P. Sánchez-Gijón. — Amsterdam—Philadelphia: John Benjamins Publishing, 2009. 149 p.
6. Flowerdew L. Corpora and language education. — New York – Basingstoke: Palgrave Macmillan, 2012. 358 p.
7. Corpora in translator education / Eds. F. Zanettin, S. Bernardini, D. Stewart. — New York, NY, USA: Routledge, 2014. 156 p.
8. Егорова А. Ю., Зацман И. М., Мамонова О. С. Надкорпусные базы данных в лингвистических проектах // Системы и средства информатики, 2019. Т. 29. № 3. С. 77–91.

¹ По данным НБД коннекторов на 31.03.2020.

9. Anthes G. Dead languages come to life // Commun. ACM, 2020. Vol. 63. No. 4. P. 13–15.
10. Зализняк А. А., Зацман И. М., Инькова О. Ю. Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–108.
11. Inkova O., Popkova N. Statistical data as information source for linguistic analysis of Russian connectors // Информатика и её применения, 2017. Т. 11. Вып. 3. С. 123–131.
12. Зацман И. М., Кружков М. Г. Надкорпусная база данных коннекторов: развитие системы терминов проектирования // Системы и средства информатики, 2018. Т. 28. № 4. С. 156–167.
13. Инькова О. Ю. Надкорпусная база данных как инструмент формальной вариативности коннекторов // Компьютерная лингвистика и интеллектуальные технологии: по мат-лам ежегодной Международ. конф. «Диалог». — М.: РГУ, 2018. Вып. 17(24). С. 240–253.
14. Шетинкин В. Е. Пособие по переводу с французского языка на русский. — М.: Просвещение, 1987. 160 с.
15. Иванов Н. В. Интерпретация в знаковой онтологии языка и в переводе. — М.: Международные отношения, 2018. 152 с.
16. Rosenbloom P. On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 308 p.
17. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89.
18. Инькова-Манзотти О. Ю. Коннекторы противопоставления во французском и русском языках. Сопоставительное исследование. — М.: Информэлектро, 2001. 429 с.
19. Семантика коннекторов: контрастивное исследование / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 396 с.
20. Структура коннекторов и методы ее описания / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. 316 с.
21. Зацман И. М., Мамонова О. С., Щурова А. Ю. Обратимость и альтернативность генерализации моделей перевода коннекторов в параллельных текстах // Системы и средства информатики, 2017. Т. 27. № 2. С. 125–142.
22. Зацман И. М. Методология обратимой генерализации в контексте классификации информационных трансформаций // Системы и средства информатики, 2018. Т. 28. № 2. С. 128–144.
23. Амеличева В. Формальное и семантическое варьирование русского коннектора *не только... но и* и его французские эквиваленты // Съпоставительно езикознание (Сопоставительное языкознание / Contrastive Linguistics), 2017. Т. 42. № 4. С. 9–20.
24. Инькова О., Манзотти Э. Tra l'altro, между прочим, entre autres: сходства и различия // Съпоставительно езикознание (Сопоставительное языкознание / Contrastive Linguistics), 2017. Т. 42. № 4. С. 35–47.
25. Нуриев В. Коннектор *раз... то* и модели его перевода на французский язык: между условием, причиной и следствием // Съпоставительно езикознание (Сопоставительное языкознание / Contrastive Linguistics), 2017. Т. 42. № 4. С. 63–76.

Поступила в редакцию 13.04.20

REDUCING THE SPECTRUM OF TRANSLATION MODELS IN SUPRACORPORA DATABASES

V. A. Nuriev and I. M. Zatsman

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

Abstract: The paper describes an approach aimed at reducing the spectrum of translation models that are registered in supracorpora databases (SDB). This approach may be applied to both professional (made by the human) and machine translation. As of today, one can use the information resources of the SDB to research the problems of interest to computing, computer linguistics and linguistic theory. Here, the focus is on examining if and how the SDB can be used in translation practice — for reducing the spectrum of translation models. Very often while translating, a translator finds himself/herself in a multiple-choice situation: due to the synonymic potential, characteristic of natural languages, in translation instead of the only possible solution, there is a set of relatively interchangeable equivalents, i. e., “fan of alternatives.” Choosing from a (sometimes wide) range of variants, a translator, in order to narrow the choice, relies on some specific characteristics of the source text. Hence, the goal of the paper is to describe the approach that would allow one to use the SDB for narrowing the choice set of relevant translation models.

Keywords: parallel texts; translation; translation models; supracorpora database; multiple choice

DOI: 10.14357/19922264200217

References

1. Zatsman, I., N. Buntman, M. Kruzhev, V. Nuriev, and A. Zaluzniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 3:1063–1071.
2. Zatsman, I., and N. Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings. *16th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 851–860.
3. Zatsman I., N. Buntman, A. Coldefy-Faucard, and V. Nuriev. 2016. WEB knowledge base for asynchronous brainstorming. *17th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 1:976–983.
4. Lyashevskaya, O. N. 2016. Korpusnye instrumenty v grammaticheskikh issledovaniyakh russkogo yazyka [Corpus instruments in grammatical studies of Russian]. Moscow: YASK: Rukopisnye Pamyatniki Drevney Rusi. 520 p.
5. Beeby, A., P. Rodríguez Inés, and P. Sánchez-Gijón, eds. 2009. *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate*. Amsterdam–Philadelphia: John Benjamins Publishing. 149 p.
6. Flowerdew, L. 2012. *Corpora and language education*. New York – Basingstoke: Palgrave Macmillan. 358 p.
7. Zanettin, F., S. Bernardini, and D. Stewart, eds. 2014. *Corpora in translator education*. New York, NY: Routledge. 156 p.
8. Egorova, A. Yu., I. M. Zatsman, and O. S. Mamonova. 2019. Nadkorpurnye bazy dannykh v lingvisticheskikh proektakh [Supracorpora databases in linguistic projects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(3):77–91.
9. Anthes, G. 2020. Dead languages come to life. *Commun. ACM* 63(4):13–15.
10. Zaluznyak, A. A., I. M. Zatsman, and O. Yu. In'kova. 2017. Nadkorpurnaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database of connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–108.
11. Inkova, O., and N. Popkova. 2017. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):123–131.
12. Zatsman, I. M., and M. G. Kruzhev. 2018. Nadkorpurnaya baza dannykh konnektorov: razvitie sistemy terminov proektirovaniya [Supracorpora database of connectives: Design-oriented evolution of the term system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):156–167.
13. Inkova, O. Yu. 2018. Nadkorpurnaya baza dannykh kak instrument formal'noy variativnosti konnektorov [Supracorpora database as an instrument of the study of the formal variability of connectives]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog" [Computer Linguistic and Intellectual Technologies: Conference (International) "Dialog" Proceedings]*. Moscow. 17(24):240–253.
14. Schetinkin, V. E. 1987. *Posobie po perevodu s frantsuzskogo yazyka na russkiy* [Manual of translation from French into Russian]. Moscow: Prosveschenie. 160 p.
15. Ivanov, N. V. 2018. *Interpretatsiya v znakovoy ontologii yazyka i v perevode* [Interpretation in the semiotic ontology of the language and in translation]. Moscow: Mezhdunarodnye otnosheniya. 152 p.
16. Rosenbloom, P. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 308 p.
17. 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.
18. In'kova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh. Sopostavitel'noe issledovanie* [Connectors of opposition in French and Russian: A comparative study]. Moscow: Informelektro. 429 p.
19. In'kova, O. Yu., ed. 2018. *Semantika konnektorov: kontrastivnoe issledovanie* [Semantics of connectives: Contrastive study]. Moscow: TORUS PRESS. 396 p.
20. In'kova, O. Yu., ed. 2019. *Struktura konnektorov i metody ee opisaniya* [Structure of connectives and methods of its description]. Moscow: TORUS PRESS. 316 p.
21. Zatsman, I. M., O. S. Mamonova, and A. Yu. Shchurova. 2017. Obratimost' i al'ternativnost' generalizatsii modeley perevoda konnektorov v paralel'nykh tekstakh [Reversibility and alternativeness of generalization of connectives translations models in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):125–142.
22. Zatsman, I. M. 2018. Metodologiya obratimoy generalizatsii v kontekste klassifikatsii informatsionnykh transformatsiy [Methodology of reversible generalization in context of classification of information transformations]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):128–144.
23. Amelicheva, V. 2017. Formal'noe i semanticheskoe var'irovanie russkogo konnektora *ne tol'ko... no i* i ego frantsuzskie ekvivalenty [Formal and semantic variations of the Russian connector *ne tol'ko... no i* and its French equivalents]. *Sopostavitel'noye yazykoznanie* [Contrastive Linguistics] 42(4):9–20.
24. Inkova, O., and E. Manzotti. 2017. Tra l'altro, mezhdu prochim, entre autres: skhodstva i razlichiya [Tra l'altro, mezhdu prochim, entre autres: Similarities and differences]. *Sopostavitel'noye yazykoznanie* [Contrastive Linguistics] 42(4):35–47.
25. Nuriev, V. 2017. Konnektor *raz... to* i modeli ego perevoda na frantsuzskiy yazyk: mezhdu usloviem, prichinoy i sledstviem [The connector *raz... to* and variants of rendering it into French: Between condition, cause, and effect]. *Sopostavitel'noye yazykoznanie* [Contrastive Linguistics] 42(4):63–76.

Received April 13, 2020

Contributors

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

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

Абгарян Каринэ Карленовна (р. 1963) — доктор физико-математических наук, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; заведующая кафедрой Московского авиационного института (национального исследовательского университета)

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Адамова Ксения Андреевна (р. 1994) — инженер Научно-производственного центра автоматизации и приборостроения имени академика Н. А. Пилюгина

Базилевский Михаил Павлович (р. 1987) — кандидат технических наук, доцент Иркутского государственного университета путей сообщения

Бахтеев Олег Юрьевич (р. 1991) — аспирант Московского физико-технического института

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гаврилов Евгений Сергеевич (р. 1982) — научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, старший преподаватель Московского авиационного института (национального исследовательского университета)

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук, ведущий научный сотрудник факультета вычислительной математики и киберне-

тики Московского государственного университета имени М. В. Ломоносова

Грабовой Андрей Валериевич (р. 1997) — студент Московского физико-технического института

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забейхайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой Вологодского государственного университета, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, главный научный сотрудник Вологодского научного центра Российской академии наук

Кирилук Игорь Леонидович (р. 1974) — научный сотрудник Института экономики Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор, заведующий ка-

федрой математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Мейханаджян Лусине Акобовна (р. 1990) — кандидат физико-математических наук, доцент Департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве РФ

Мельников Сергей Юрьевич (р. 1963) — кандидат физико-математических наук, докторант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Нуриев Виталий Александрович (р. 1980) — кандидат филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов

Сатин Яков Александрович (р. 1978) — кандидат физико-математических наук, доцент Вологодского государственного университета

Сенько Олег Валентинович (р. 1957) — доктор физико-математических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Московского физико-технического института

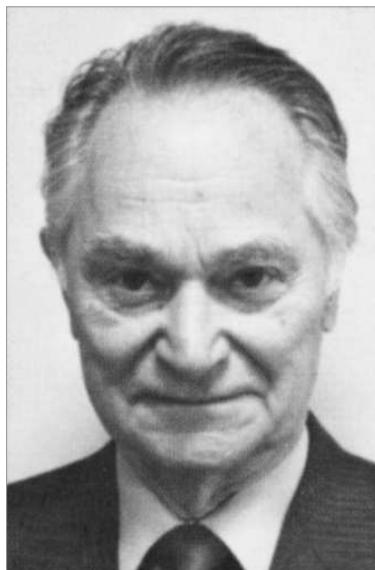
Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, доцент, старший научный сотрудник Института проблем информатики 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 им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шилова Галина Николаевна (р. 1961) — кандидат физико-математических наук, доцент, заведующая кафедрой Вологодского государственного университета

Шнурков Петр Викторович (р. 1953) — кандидат физико-математических наук, доцент Национального исследовательского университета «Высшая школа экономики»



Рустем Бадриевич Сейфуль-Мулюков

1928–2020

Редакция журнала «Информатика и её применения» с глубоким прискорбием сообщают, что 17 марта 2020 г. на 93-м году жизни скончался заведующий редакцией журнала, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук Рустем Бадриевич Сейфуль-Мулюков.

Всю свою жизнь Рустем Бадриевич посвятил служению науке. Закончив в 1956 г. аспирантуру Московского ордена Трудового Красного знамени Нефтяного института им. академика И. М. Губкина, он прошел путь от заведующего отделом Института геологии зарубежных стран Министерства геологии СССР до заместителя директора ВИНТИ АН СССР, доктора геолого-минералогических наук, профессора.

С марта 2002 г. Рустем Бадриевич успешно применял свои знания и организационный талант в ИПИ РАН (в дальнейшем — ФИЦ ИУ РАН), в котором руководил лабораторией и отделом, занимающимися вопросами технологий информационной технической деятельности. Р. Б. Сейфуль-Мулюков, являясь автором значительного количества научных трудов и монографий по геологии, информационным технологиям и теоретической информатике, осуществлял организацию издания монографий ИПИ РАН и ФИЦ ИУ РАН, библиографий научных сотрудников Центра.

Р. Б. Сейфуль-Мулюков являлся заведующим редакцией журналов «Информатика и её применения» и «Системы и средства информатики», членом редколлегии журнала «Системы и средства информатики». Он вложил огромный вклад в становление и развитие этих журналов, организацию их регистрации, функционирования, редактуры и издания. Включение этих журналов в ряд отечественных и зарубежных информационных баз и систем цитирования во многом является его личной заслугой.

На всех занимаемых должностях Рустем Бадриевич отличался высоким профессионализмом, преданностью делу и вниманием к коллегам.

Рустема Бадриевича отличали доброта, отзывчивость, неиссякаемый оптимизм, простота и сердечность.

Коллеги Рустема Бадриевича запомнят его как многогранного в своих увлечениях человека, живописца, эрудита и энциклопедиста, интересующегося историей, литературой и искусством.

Выражаем глубокое соболезнование семье, родственникам, друзьям и коллегам по работе в связи с тяжелой невосполнимой утратой. Светлый образ Рустема Бадриевича навсегда сохранится в нашей памяти.

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp 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.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovaniye plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie 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 Publ. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: rust@ipiran.ru (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. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazoidachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy 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]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform 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: rust@ipiran.ru (to Svetlana Strigina)
<http://www.ipiran.ru/english/journal.asp>