

Информатика и её применения

Том 13 Выпуск 1 Год 2019

СОДЕРЖАНИЕ

Интерполяционное аналитическое моделирование распределений в сложных стохастических системах	
И. Н. Сеницын	2
Управление выходом стохастической дифференциальной системы по квадратичному критерию. II. Численное решение уравнений динамического программирования	
А. В. Босов, А. И. Стефанович	9
Об одном классе задач фильтрации на многообразиях	
К. А. Рыбаков	16
О числе максимальных независимых элементов частичных порядков (случай цепей)	
Е. В. Дюкова, Г. О. Масляков, П. А. Прокофьев	25
Анализ уязвимости многополюсных сетей при структурных повреждениях	
Ю. Е. Малашенко, И. А. Назарова, Н. М. Новикова	33
Классификация физической активности человека с помощью локальных аппроксимирующих моделей	
Д. А. Аникеев, Г. О. Пенкин, В. В. Стрижов	40
Обращение однородных операторов с помощью стабилизированной жесткой пороговой обработки при неизвестной дисперсии шума	
О. В. Шестаков	49
Об унимодальности функции дохода системы массового обслуживания типа $G M s$ с управляемой очередью	
Я. М. Агаларов, В. Г. Ушаков	55
Априорные Фреше и масштабированное обратное хи-распределение в байесовских моделях баланса	
А. А. Кудрявцев, С. И. Палионная, В. С. Шоргин	62
Полиномиальные алгоритмы вычисления локальных аффинностей квадратичных булевых функций	
О. А. Логачев, А. А. Сукаев, С. Н. Федоров	67
Оптимизация гиперпараметров нейронных сетей с использованием высокопроизводительных вычислений для предсказания осадков	
А. К. Горшенин, В. Ю. Кузьмин	75
Синтез геоданных в пространственных инфраструктурах на основе связанных данных	
С. К. Дулин, Н. Г. Дулина, О. С. Кожунова	82
Целенаправленное развитие систем лингвистических знаний: выявление и заполнение лакун	
И. М. Зацман	91
Ресурсные системы массового обслуживания с произвольным обслуживанием	
А. В. Горбунова, В. А. Наумов, Ю. В. Гайдамака, К. Е. Самуйлов	99
Сравнительный анализ показателей эффективности модели беспроводной сети межмашинного взаимодействия, работающей в рамках двух политик разделения радиоресурсов	
Е. В. Маркова, А. А. Гольская, И. Л. Дзантиев, И. А. Гудкова, С. Я. Шоргин	108
Об авторах	117
Правила подготовки рукописей	120
Requirements for manuscripts	123

ИНТЕРПОЛЯЦИОННОЕ АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ В СЛОЖНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ

И. Н. Синицын¹

Аннотация: Разработаны методы интерполяционного аналитического моделирования (ИАМ) в сложных непрерывных и дискретных скалярных и приводимых к ним векторных стохастических системах (СтС). Выделены типовые классы сложных СтС. Методы основаны на интерполяционном решении эволюционного уравнения для одномерной характеристической функции (х.ф.). Получены уравнения чувствительности для оценки х.ф. к параметрам СтС. В основу базовых алгоритмов ИАМ положена теорема отсчетов Котельникова. Рассмотрены практические вопросы выбора интерполяционных формул Котельникова и оценки количества интерполяционных отсчетов. Для СтС с известной аналитической природой в качестве дальнейшего развития следует рассмотреть применение сплайнт-вейвлет методов для офлайн алгоритмов, а для онлайн алгоритмов — фильтрационных подходов, при этом особое внимание следует уделить многомерным распределениям.

Ключевые слова: одномерная плотность вероятности (п.в.); одномерная характеристическая функция (х.ф.); стохастическая система (СтС); стохастический процесс (СтП)

DOI: 10.14357/19922264190101

1 Введение

Среди известных методов аналитического моделирования стохастических процессов (СтП) в дифференциальных СтС важное место занимают методы, основанные на прямом численном решении эволюционных уравнений для n -мерной х.ф. [1]. Мальчиковым в [2, 3] для гауссовских СтС на основе теоремы отсчетов Котельникова [4] разработан метод ИАМ для одномерной х.ф.

В [5] дано обобщение [2, 3] для негауссовских СтС на основе замены негауссовской СтС эквивалентной гауссовской (нормальной) СтС. Дадим обобщение [5] на случай непрерывных и дискретных СтС, не приводимых к гауссовским. Разделы 2 и 3 посвящены методу ИАМ соответственно для непрерывных и дискретных СтС. В разд. 4 описаны практические способы выбора интерполяционных формул и чисел отсчетов. Заключение содержит выводы и некоторые возможные обобщения.

2 Метод интерполяционного аналитического моделирования для непрерывных стохастических систем

2.1. Следуя [5], введем следующие определения и допущения.

1. Пусть скалярный СтП $Y(t) = Y_t$ определяется скалярным стохастическим дифференциальным уравнением Ито следующего вида [1]:

$$dY_t = a(Y_t, \Theta, t) dt + b(Y_t, \Theta, t) dW_0(\Theta, t) + \int_{R_0^1} c(Y_t, \Theta, t, v) P^0(\Theta, dt, dv), \quad Y(t_0) = Y_0. \quad (1)$$

Здесь $W_0 = W_0(\Theta, t)$ — скалярный винеровский СтП интенсивности $\nu_0 = \nu_0(\Theta, t)$; Θ — n^Θ -мерный вектор параметров; $P^0(\Theta, \mathcal{J}, \mathcal{A})$ представляет собой для любого множества \mathcal{A} простой пуассоновский СтП, причем $P^0(\Theta, \mathcal{J}, \mathcal{A}) = P(\Theta, \mathcal{J}, \mathcal{A}) - \mu_P(\Theta, \mathcal{J}, \mathcal{A})$, где $\mu_P(\Theta, \mathcal{J}, \mathcal{A})$ — его математическое ожидание, равное

$$\begin{aligned} \mu_P &= \mu_P(\Theta, \mathcal{J}, \mathcal{A}) = M P(\Theta, \mathcal{J}, \mathcal{A}) = \\ &= \int_{\mathcal{J}} \nu_P(\Theta, \tau, \mathcal{A}) d\tau, \end{aligned}$$

а $\nu_P = \nu_P(\Theta, \mathcal{J}, \mathcal{A})$ — интенсивность соответствующего пуассоновского потока событий; $\mathcal{J} = (t_1, t_2]$; интегрирование по v распространяется на пространство R^1 с выколотым началом координат; $a = a(Y_t, \Theta, t)$, $b = b(Y_t, \Theta, t)$ и $c = c(Y_t, \Theta, t, v)$ — скалярные нелинейные функции отмеченных аргументов.

Предположим, что при фиксированном Θ выполнены условия существования и единствен-

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

ности СтП, определяемого (1) при соответствующем начальном условии [1, 6]. Кроме того, функции a , b и c и характеристики СтП W_0 и P^0 будем считать дифференцируемыми по Θ .

В дальнейшем для белого шума $V = \dot{W}$ (понимаемого в строгом смысле и являющегося производной по времени от произвольного СтП с независимыми приращениями $W(t)$) будем использовать уравнение:

$$\dot{Y}_t = a(Y_t, \Theta, t) + b(Y_t, \Theta, t)V, \quad V = \dot{W}. \quad (2)$$

2. Существует одномерная плотность $f = f(y; \Theta, t)$, а уравнение для одномерной х.ф. $g = g(\lambda; \Theta, t)$ при фиксированном Θ имеет вид [1]:

$$\frac{\partial g(\lambda; \Theta, t)}{\partial t} = \int_{-\infty}^{\infty} [i\lambda a(y, \Theta, t) + \chi(\lambda; y, \Theta, t)] e^{i\lambda y} f(y; \Theta, t) dy \quad (3)$$

при начальном условии

$$g(\lambda; \Theta, t_0) = g_0(\lambda; \Theta). \quad (4)$$

Здесь $f = f(y; \Theta, t)$ — одномерная плотность вероятности (п.в.), связанная с х.ф. $g(\lambda; \Theta, t)$ следующим Фурье преобразованием:

$$f(y; \Theta, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda y} g(\lambda; \Theta, t) d\lambda; \quad (5)$$

$$\begin{aligned} \chi(\lambda; y, \Theta, t) = & -\frac{\lambda^2 \nu_0(\Theta, t)}{2} b^2(y, \Theta, t) + \\ & + \int_{R'_0} \exp\{[i\lambda c(y, \Theta, t, v)] - 1 - \\ & - i\lambda c(y, \Theta, t, v)\} \nu_P(\Theta, t, dv). \end{aligned}$$

Для стационарных дифференциальных СтС (1) уравнения для стационарных х.ф. $g^* = g^*(\lambda; \Theta)$ и п.в. $f^*(y; \Theta)$ имеют вид:

$$\int_{-\infty}^{\infty} [i\lambda a^*(y, \Theta) + \chi(\lambda; y, \Theta)] e^{i\lambda y} f^*(y; \Theta) dy = 0; \quad (6)$$

$$f^*(y; \Theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda y} g^*(\lambda; \Theta) dy.$$

Здесь звездочкой отмечены стационарные значения соответствующих функций.

Для СтС (2) уравнения (3) и (6) имеют следующий вид:

$$\begin{aligned} \frac{\partial y(\lambda; \Theta, t)}{\partial t} = & \int_{-\infty}^{\infty} [i\lambda a(y, \Theta, t) + \\ & + \chi(b(y, \Theta, t)\lambda; t)] \exp(i\lambda y) f(y, \Theta, t) dy; \quad (7) \\ & \int_{-\infty}^{\infty} [i\lambda a^*(y, \Theta) + \chi(b^*(y, \Theta)\lambda; t)] \times \\ & \times \exp(i\lambda y) f^*(y, \Theta) dy = 0. \end{aligned}$$

3. Пусть х.ф. $g = g(\lambda; \Theta, t)$ для СтС (1) выражается через значение в фиксированных точках отсчета, применив для этого какую-нибудь интерполяционную формулу [6]:

$$g(\lambda; \Theta, t) = \sum_{l=-N}^N g_l(\lambda_l; \Theta, t) \alpha_l(\lambda) + R_N. \quad (8)$$

Здесь R_N — остаточный член; λ_l — некоторые фиксированные значения аргумента λ в точках отсчета х.ф.; $\alpha_l(\lambda)$ — известные функции, удовлетворяющие условию

$$\alpha_l(\lambda_r) = \delta_{lr} \quad (l, r = 0, \pm 1, \dots, \pm N),$$

где δ_{lr} — символ Кронекера.

Таким образом, задача сводится к вычислению значений х.ф. в точках отсчета х.ф. После определения х.ф. $g(\lambda; \Theta, t)$ п.в. определяется преобразованием Фурье (5).

2.2. Рассмотрим метод ИАМ для СтС (1). Подставляя в (5) выражение (8), найдем:

$$\begin{aligned} f(y; \Theta, t) \approx & \sum_{l=-N}^N g_l(\lambda_l; \Theta, t) \beta_l(y), \\ \beta_l = & \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda y} \alpha_l(\lambda) d\lambda, \quad (9) \end{aligned}$$

где функции $\beta_l(y)$ являются преобразованием Фурье функций $\alpha_l(\lambda)$ и, следовательно, известны. А значит, искомая п.в. выражается через значения х.ф. в точках отсчета. Полагая в (3) $\lambda = \lambda_r$ и подставляя в него (9), получаем:

$$\frac{\partial g_r}{\partial t} = \dot{g}_r = \sum_{l=-N}^N \varepsilon_{lr} g_l. \quad (10)$$

Здесь $g_l = g_l(\lambda_r; \Theta, t)$, а ε_{lr} равны

$$\varepsilon_{lr} = \varepsilon_{lr}^0 + \varepsilon'_{lr} + \varepsilon''_{lr}, \quad (11)$$

где

$$\varepsilon_{lr}^0 = \varepsilon_{lr}^0(\lambda_r; \Theta, t) = i\lambda_r \int_{-\infty}^{\infty} a(y, \Theta, t) \beta_l(y) e^{i\lambda_r y} dy;$$

$$\begin{aligned} \varepsilon'_{lr} &= \varepsilon'_{lr}(\lambda_r; \Theta, t) = \\ &= -\frac{\lambda_r^2}{2} \nu_0(\Theta, t) \int_{-\infty}^{\infty} b^2(y, \Theta, t) \beta_l(y) \exp(i\lambda_r y) dy; \\ \varepsilon''_{lr} &= \varepsilon''_{lr}(\lambda_r; \Theta, t) = \int_{-\infty}^{\infty} \int_{R_0^q} \{ \exp[i\lambda_r c(y, \Theta, t, v)] - \\ &- 1 - i\lambda_r c(y, \Theta, t, v) \} \beta_l(y) \exp[i\lambda_r y] dv dy. \end{aligned} \quad (12)$$

Замечание 2.1. Уравнение (10) линейно, причем интегралы (12) могут быть вычислены заранее, так как подынтегральные функции известны.

Следуя [5], выполним с (10) следующие преобразования. Перейдем от комплексных величин к действительным, положив

$$g_r = g_{r1} + i g_{r2}; \quad (13)$$

$$\left. \begin{aligned} \beta_l &= \beta_{l1} + i \beta_{l2}; \\ \beta_{l1} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \alpha_l(\lambda) \cos \lambda y d\lambda; \\ \beta_{l2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \alpha_l(\lambda) \sin \lambda y d\lambda; \end{aligned} \right\} \quad (14)$$

$$\varepsilon_{lr} = \varepsilon_{lr1} + i \varepsilon_{lr2},$$

где

$$\begin{aligned} \varepsilon_{lr1} &= -\lambda_r \int_{-\infty}^{\infty} [\beta_{l1}(y) \sin \lambda_r y + \\ &+ \beta_{l2}(y) \cos \lambda_r y] a(y, t) dy; \\ \varepsilon_{lr2} &= -\lambda_r \int_{-\infty}^{\infty} [\beta_{l1}(y) \cos \lambda_r y + \\ &+ \beta_{l2}(y) \sin \lambda_r y] a(y, t) dy. \end{aligned} \quad (15)$$

После подстановки (13)–(15) в (11) и приравняв действительные и мнимые части, получим систему линейных обыкновенных дифференциальных уравнений для значений х.ф. в точках отсчета в действительной форме:

$$\left. \begin{aligned} \dot{g}_{r1} &= G_{r1} = \sum_{l=-N}^N (\varepsilon_{lr1} g_{l1} - \varepsilon_{lr2} g_{l2}); \\ \dot{g}_{r2} &= G_{r2} = \sum_{l=-N}^N (\varepsilon_{lr2} g_{l1} + \varepsilon_{lr1} g_{l2}). \end{aligned} \right\} \quad (16)$$

Сократим число уравнений в (16), учтя известные свойства х.ф. [1]:

$$g(0, \Theta, t) = 1; \quad g(-\lambda; \Theta, t) = \overline{g(\lambda; \Theta t)}, \quad (17)$$

где черта — символ комплексного сопряжения. На основании (17) находим:

$$\left. \begin{aligned} g_{01} &= 1; & g_{02} &= 0; \\ g_{-r,1} &= g_{r1}; & g_{-r,2} &= -g_{r2}. \end{aligned} \right\} \quad (18)$$

Отсюда следует, что (16) надо решать лишь для g_{r1} и g_{r2} при $r > 0$.

Наконец, выделим слагаемое, соответствующее $l = 0$, далее разобьем полученную сумму на две суммы и сменим порядок суммирования на $\overline{1, N}$. Тогда придем к следующим окончательным уравнениям:

$$\begin{aligned} \dot{g}_{r1} &= G_{r1} = \\ &= \sum_{h=1}^N [(\varepsilon_{hr1} + \varepsilon_{-hr1}) g_{h1} - (\varepsilon_{hr2} - \varepsilon_{-hr2}) g_{h2}]; \\ \dot{g}_{r2} &= G_{r2} = \sum_{h=1}^N [(\varepsilon_{hr2} - \varepsilon_{-hr2}) g_{h1} + \\ &+ (\varepsilon_{hr1} + \varepsilon_{-hr1}) g_{h2}] \quad (r = \overline{1, N}). \end{aligned} \quad (19)$$

В случае стационарного распределения уравнения (19) принимают вид:

$$G_{r1}^* = 0; \quad G_{r2}^* = 0. \quad (20)$$

Выражение (5) для п.в. после аналогичных преобразований с х.ф. принимает вид:

$$\begin{aligned} f(y, \Theta, t) &= \beta_{01}(y) + \sum_{h=1}^N [\beta_{h1}(y) + \beta_{-h1}(y)] g_{h1} - \\ &- \sum_{h=1}^N [\beta_{h2}(y) - \beta_{-h2}(y)] g_{h2}. \end{aligned} \quad (21)$$

В качестве начальных условий для (19) следует принять следующие количества:

$$g_{rj}(\lambda_r; t_0) = g_{rj0} \quad (j = 1, 2). \quad (22)$$

Так как х.ф. является преобразованием Фурье для п.в., то теорема отсчетов Котельникова [4] полностью определяется последовательностью ее значений в дискретном ряде точек. Поэтому, если распределение СтП Y_t сосредоточено на интервале $2\Delta = 2\Delta_y$, то согласно [4] $\alpha_r = \alpha_r(\lambda)$ и $\beta_r = \beta_r(y)$ можно выбрать в виде:

$$\alpha_r = \frac{\sin \Delta(\lambda - \lambda_r)}{\Delta(\lambda - \lambda_r)}, \quad \lambda_r = \frac{\pi r}{\Delta}; \quad (23)$$

$$\beta_r = \begin{cases} \frac{1}{2\Delta} e^{-i\lambda_r y} & \text{при } |y| \leq \Delta; \\ 0 & \text{при } |y| > \Delta, \end{cases} \quad (24)$$

при этом п.в. будет определяться формулой:

$$f(y, \Theta, t) = \begin{cases} \frac{1}{\Delta} \sum_{r=1}^N (g_{r1} \cos \lambda_r y + g_{r2} \sin \lambda_r y) & \text{при } |y| \leq \Delta; \\ 0 & \text{при } |y| > \Delta. \end{cases}$$

Чувствительность метода ИАМ к параметрам Θ оценивается согласно [5, 7] путем вычисления следующих первых векторных функций чувствительности:

$$\nabla^\Theta g_{rj} = \nabla^\Theta g_{rj}(\lambda_r; \Theta, t) \quad (j = 1, 2), \quad (25)$$

где ∇^Θ — символ вектора частных производных по векторному параметру Θ . В силу (25) соответствующие уравнения для функций чувствительности имеют вид:

$$\dot{\nabla}^\Theta g_{rj} = \nabla^\Theta G_{rj} \quad (j = 1, 2) \quad (26)$$

при нулевых начальных условиях и

$$\nabla^\Theta G_{rj}^* = 0 \quad (j = 1, 2) \quad (27)$$

в стационарном случае.

Таким образом, приходим к следующему обобщению [5].

Теорема 1. Пусть СтС (1) допускает одномерное нестационарное распределение, х.ф. которого удовлетворяет уравнению (3) при начальном условии (4). Тогда в основе метода ИАМ х.ф. по Котельникову лежат согласно (23) и (24) линейные обыкновенные дифференциальные уравнения (19) при начальных условиях (22). Для моделирования п.в. применяется формула (21). В случае, когда имеет место стационарное одномерное распределение, используются уравнения (20). При этом чувствительность ИАМ оценивается согласно (25)–(27).

2.3. Уравнения ИАМ для СтС (2) имеют вид (19)–(21), если принять

$$\varepsilon_{lr} = \varepsilon_{lr}^0 + \varepsilon_{lr}^1, \quad (28)$$

где

$$\begin{aligned} \varepsilon_{lr}^0 &= i\lambda_r \int_{-\infty}^{\infty} a(y, \Theta, t) \beta_l(y) e^{i\lambda_r y} dy; \\ \varepsilon_{lr}^1 &= \int_{-\infty}^{\infty} \chi(b(y, \Theta, t); \lambda; \Theta, t) \beta_l(y) e^{i\lambda_r y} dy. \end{aligned} \quad (29)$$

В результате имеем следующий алгоритм метода ИАМ.

Теорема 2. Пусть СтС (2) допускает одномерное нестационарное распределение, х.ф. которого удовлетворяет уравнению (7) при начальном условии (4).

Тогда в основе метода ИАМ х.ф. по Котельникову согласно (23) и (24) лежат линейные обыкновенные дифференциальные уравнения (19)–(21) при условиях (28)–(29). Для моделирования п.в. применяются формулы (18). В стационарном случае используются уравнения (20). При этом чувствительность метода ИАМ оценивается согласно (25)–(27).

2.4. Теперь рассмотрим метод ИАМ для векторных СтС. Следуя [2, 3], воспользуемся принципом эквивалентности векторной СтС скалярной СтС. В задачах практики обычно не требуется определения х.ф. и п.в. всех составляющих вектора Y_t , а достаточно определить одномерное распределение некоторых из них. Поэтому будем искать распределения отдельных составляющих.

Для определения одномерных х.ф. и п.в. некоторой h -й составляющей вектора Y_t заменим исходную многомерную СтС эквивалентной одномерной СтС, описываемой стохастическим дифференциальным уравнением Ито (или уравнением с θ -дифференциалом [1]), выходная переменная \tilde{Y}_{ht} которого имеет те же х.ф. и п.в., что и h -я составляющая Y_{ht} .

В результате приходим к следующему утверждению.

Теорема 3. Пусть векторные СтС вида (1) и (2) приводятся к эквивалентной системе скалярных СтС в форме Ито. Тогда в основе метода ИАМ лежат уравнения теорем 1 и 2 для каждой скалярной СтС.

Замечание 2.2. Выходные переменные Y_{ht} и \tilde{Y}_{ht} в общем случае (2) — различные СтП, поскольку \tilde{Y}_{ht} является марковским СтП, а Y_{ht} — составляющая марковского процесса и, следовательно, сама по себе может и не быть марковским СтП. Однако одномерные распределения у них одинаковы.

Замечание 2.3. Для квазилинейных СтС с аддитивным шумом замена многомерной системы одномерной допустима в случае асимптотической устойчивости линейной детерминированной части. Для комплексно-сопряженных корней эквивалентная система будет уже двумерной. Случаи кратных корней требуют специального рассмотрения.

3 Метод интерполяционного аналитического моделирования для дискретных стохастических систем

Следуя [1], рассмотрим сначала нелинейные дискретные СтС, описываемые разностными стохастическими уравнениями вида:

$$Y_{k+1} = \omega_k(Y_k, \Theta, V_k^d) \quad (k = 1, 2, \dots); \quad (30)$$

$$Y_{k+1} = a_k(Y_k, \Theta) + b_k(Y_k, \Theta) V_k^d \quad (k = 1, 2, \dots), \quad (31)$$

где ω_k , a_k и b_k — скалярные функции. Характеристические функции $h_k = h_k(\rho, \Theta)$ векторных случайных величин V_k^d и соответствующие им п.в. $\eta_k = \eta_k(v, \Theta)$ будем считать известными.

Для одномерных п.в. и х.ф., определяемых известными формулами

$$f_k(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda y} g_k(\lambda) d\lambda;$$

$$g_k(\lambda) = M \exp\{i\lambda Y_k\}$$

применительно к уравнению (30) имеют место следующие рекуррентные формулы [1]:

$$g_{k+1}(\lambda) = M \exp\{i\lambda \omega_k(Y_k, \Theta, V_k)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\lambda \omega_k(y, v, \Theta)} f_k(y) h_k(v, \Theta) dy dv. \quad (32)$$

Общая рекуррентная формула для одномерного распределения в (31) немедленно получается из соответствующей формулы (32) и приводит к следующему результату:

$$g_{k+1} = M \exp\{i\lambda a_k(Y_k, \Theta) + i\lambda b_k(Y_k, \Theta) V_k\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\lambda a_k(y, \Theta) + i\lambda b_k(y, \Theta)v} f_k(y) h_k(v, \Theta) dy dv = M [\exp\{i\lambda a_k(Y_k, \Theta)\} + h_k(b_k(Y_k, \Theta) \lambda)]. \quad (33)$$

Для СтС (31) метод ИАМ приводит к следующим рекуррентным формулам:

$$g_{r1, k+1} = G_{r1, k}^d = \sum_{h=1}^N [(\varepsilon_{hr1, k} + \varepsilon_{-hr1, k}) g_{h1, k} - (\varepsilon_{hr2, k} - \varepsilon_{-hr2, k}) g_{h2, k}]; \quad (34)$$

$$g_{r2, k+1} = G_{r2, k}^d = \sum_{h=1}^N [(\varepsilon_{hr2, k} - \varepsilon_{-hr2, k}) g_{h1, k} + (\varepsilon_{hr1, k} + \varepsilon_{-hr1, k}) g_{h2, k}]; \quad (35)$$

$$g_{rj}(\lambda_r; k = 1) = g_{rj, 1} \quad (j = 1, 2); \quad (36)$$

$$\varepsilon_{lr, k} = \varepsilon_{lr, k}^0 + \varepsilon_{lr, k}^1,$$

где

$$\varepsilon_{lr, k}^0 = i\lambda_r \int_{-\infty}^{\infty} a_k(y_k, \Theta) \beta_l(y_k) \exp(i\lambda_r y_k) dy_k;$$

$$\varepsilon_{lr, k}^1 = \int_{-\infty}^{\infty} \exp\{\exp(i\lambda a_k)\} + h_k(b_k \lambda) dy_k;$$

$$f_k(y; \Theta, k) = \begin{cases} \frac{1}{\Delta} \sum_{r=1}^N g_{r1, k} \cos \lambda_r y + \\ + g_{r1, k} \sin \lambda_r y \quad \text{при } |y| \leq \Delta; \\ 0 \quad \text{при } |y| > \Delta; \end{cases} \quad (37)$$

$$\Delta^\Theta g_{rj, k+1} = \nabla^\Theta G_{rj, k}^d. \quad (38)$$

Здесь аргументы для краткости опущены.

Таким образом, приходим к следующему утверждению.

Теорема 4. Пусть дискретная СтС (31) допускает одномерное нестационарное распределение, х.ф. которого удовлетворяет разностному уравнению (33) при начальном условии (36). Тогда в основе метода ИАМ х.ф. по Котельникову согласно (23) и (24) лежат рекуррентные формулы (34) и (35) при начальных условиях (36). Для моделирования п.в. применяется формула (37). В стационарном случае используются (34) и (35), где $g_{rj, k+1} = g_{rj, k}$ ($j = 1, 2$). При этом чувствительность МИАМ оценивается согласно (38).

Замечание 3.1. Общий случай СтС (30) поддается разработке только при дополнительных ограничениях на функции ω_k .

Замечание 3.2. До сих пор СтС (30) и (31) предполагались скалярными. В векторном случае можно непосредственно воспользоваться результатами п. 2.4.

4 Практические способы выбора интерполяционных формул и чисел отсчетов

4.1. В задачах аналитического моделирования СтП в непрерывных и дискретных СтС выбор формул интерполяции х.ф. в (8) определяется интервалом значений Y_t , в котором сосредоточена его п.в.

Наряду с теоремой Котельникова [4] для х.ф. различают следующие интерполяционные функции [8]:

- (1) интерполяционная формула Лагранжа;
- (2) разделенные разности и интерполяционная формула Ньютона;
- (3) итерационно-интерполяционная формула Эйткена.

Выбор функций $\alpha_r(\lambda)$ и $\beta_l(y)$ посредством этих интерполяционных формул встречает следующие трудности: λ_{\max} (значение аргумента, определяющее область, вне которой х.ф. практически можно считать равной нулю) заранее определить трудно. Кроме того, формулы для $\beta_r = \beta_r(y)$ при этом оказываются аналитически сложными.

Для сложных СтС в случае равностоящих значений аргумента можно рекомендовать интерполяционные формулы для х.ф. с центральными разностями различного порядка, восходящие к Стирлингу, Бесселю, Эверетту и Стефенсону [8].

4.2. Необходимое приемлемое число N членов ряда, обеспечивающее достаточную точность расчетов R_N в (8) существенно зависит от удачного определения интервала 2Δ . Если выбранный интервал значительно больше, чем область фактического распределения СтП Y_t , то может потребоваться значительное число членов ряда.

В [2, 3] показано, что для гауссовского шума V достаточно точные результаты получаются уже при $N = 3$, а для большинства негауссовских п.в. — при выборе $\Delta = (3 \dots 5)\sqrt{\tilde{D}}$, где \tilde{D} — приближенное значение дисперсии СтП Y_t .

В том случае, когда заранее определить интервал Δ нет возможности, то чтобы обеспечить в каждый момент времени разложение п.в. Y_t на интервале $(3 \dots 5)\sqrt{\tilde{D}}$ можно в исходных уравнениях (1) и (2) сделать линейную замену переменных:

$$Y_t = Z_t\sqrt{c_2} + c_1.$$

Здесь c_1 и c_2 — приближенные значения математического ожидания $c_1 = M Y_t$ и дисперсии $c_2 = D Y_t$, определяемые приближенно, например методами параметризации распределений (нормальной аппроксимации, моментов, семиинвариантов, ортогонального разложение и др. [1]). При этом переменная Z_t имеет близкое к нулю математическое ожидание $c_1 = m_z \approx 0$ и близкую к единице дисперсию $c_2 = D_z \approx 1$. После такой замены уравнение (1) примет соответственно вид:

$$dZ_t = a^z(Z_t, c_y, \Theta, t) dt + b^z(Z_t, c_y, \Theta, t) dW_0(\Theta, t) + \int_{R_0^q} c^z(Z_t, c_y, \Theta, t, v) P^0(\Theta, dt, dv), \quad (39)$$

где $c_y = \{c_j\}$ — вектор параметров ортогонального разложения п.в. СтП Y_t ;

$$\left. \begin{aligned} a^z &= a^z(Z_t, c_y, \Theta, t) = \\ &= \frac{1}{\sqrt{c_2}} a(Z_t\sqrt{c_2} + c_1, \Theta, t) - Z_t \frac{\dot{c}_2}{2c_2} - \frac{\dot{c}_1}{\sqrt{c_2}}; \\ b^z &= b^z(Z_t\sqrt{c_2} + c_1, \Theta, t) = \\ &= \frac{1}{\sqrt{c_2}} b(Z_t\sqrt{c_2} + c_1, \Theta, t); \\ c^z &= c^z(Z_t\sqrt{c_2} + c_1, \Theta, t, v). \end{aligned} \right\} \quad (40)$$

При применении приближенного метода нормальной аппроксимации (МНА) [1], когда $c^y = \{c_1, c_2\}$ в силу (39) и (40) используются следующие уравнения для c_1 и c_2 :

$$\dot{c}_1 = A^{c_1}(c_1, c_2, \Theta, t); \quad \dot{c}_2 = A^{c_2}(c_1, c_2, \Theta, t), \quad (41)$$

где A^{c_1} и A^{c_2} известные функции отмеченных функций.

4.3. При вычислении п.в. переменной $Z = Z_t$ согласно теоремам 1 и 2 интервал разложения можно принять равным $3 \dots 5$. Пользуясь известной формулой вычисления линейной функции случайного аргумента [1], получим искомую п.в.:

$$f(y, \Theta) = \frac{1}{\sqrt{c_2}} f_z\left(\frac{y - c_1}{\sqrt{c_2}}\right),$$

где f_z — п.в. переменной Z_t . При этом c_1 и c_2 определяются из (41).

Замечание 4.1. В случае СтС (1), когда применяется нелинейная замена переменных $z = \varphi(Y_t, t)$, в качестве параметров c^y выбираются коэффициенты ортогонального разложения одномерной п.в.; уравнения для Z_t составляются путем применения обобщенной формулы Ито [1] для СтС (1).

Замечание 4.2. Для дифференциальных СтС описанные практические способы выбора интерполяционных формул и чисел отсчетов сохраняют силу.

5 Заключение

- 5.1. Для сложных непрерывных и дискретных скалярных и приводимых к ним векторным СтС разработаны методы ИАМ. Рассмотрены типовые классы сложных СтС. Разработанные методы основаны на интерполяционном решении эволюционного уравнения для одномерной х.ф. Получены уравнения для оценки чувствительности х.ф. к параметрам СтС.
- 5.2. В основу базовых алгоритмов ИАМ (теоремы 1–4) положены теорема отсчетов Котельникова. Рассмотрены практические вопросы выбора интерполяционных формул Котельникова и оценки количества интерполяционных отсчетов.
- 5.3. Для СтС с известной аналитической природой в качестве дальнейшего развития следует рассмотреть применение сплайн-вейвлет методов для офлайн алгоритмов, а для онлайн алгоритмов фильтрационных подходов [1]. При этом особое внимание следует уделить многомерным распределениям.

Литература

1. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
2. Мальчиков С. В. Приближенный метод определения законов распределения фазовых координат нелинейных

- автоматических систем // Автоматика и телемеханика, 1970. № 5. С. 43–50.
3. Казаков И. Е., Мальчиков С. В. Анализ стохастических систем в пространстве состояний. — М.: Наука, 1983. 348 с.
 4. Котельников В. А. Теория потенциальной помехоустойчивости. — М.: Госэнергоиздат, 1956. 412 с.
 5. Синицын И. Н. Метод интерполяционного аналитического моделирования одномерных распределений в стохастических системах // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 56–62.
 6. Справочник по теории вероятностей и математической статистике / Под ред. В. С. Королюка, Н. И. Портенко, А. В. Скорохода, А. Ф. Турбина. — М.: Наука, 1985. 640 с.
 7. Евланов Л. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 585 с.
 8. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров) / Пер с англ. — М.: Наука, 1974. 832 с. (Korn G., Korn T. *Mathematical handbook for scientists and engineers*. — New York – San Francisco – Toronto – London – Sydney: McGraw Hill Book Co., 1968. 1147 p.)

Поступила в редакцию 16.08.18

INTERPOLATIONAL ANALYTICAL MODELING IN COMPLEX STOCHASTIC SYSTEMS

I. N. Sinitsyn

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: Methods and algorithms for interpolational analytical modeling (IAM) of stochastic processes in complex continuous and discrete stochastic systems (StS) which are scalar (or vector StS reducible to scalar) are developed. Several types of continuous and discrete StS are considered. The IAM methods are based on evolutionary equations numerical interpolation solution for one-dimensional characteristic function (c.f.). Special attention is paid to c.f. sensitivity analysis. In basic IAM algorithms, the Kotel’nikov theorem was implemented. Some practical questions concerning interpolational formulae and number of intervals for interpolation are discussed. In the future, IAM for StS with known analytical structure for offline modeling will be based on spline-wavelet methods and for online modeling — on filtration approaches. Special attention should be paid to multidimensional distributions.

Keywords: one-dimensional characteristic functions (c.f.); one-dimensional probability density (p.d.); stochastic processes (StP); stochastic system (StS)

DOI: 10.14357/19922264190101

References

1. Pugachev, V. S., and I. N. Sinitsyn. 2000, 2004. *Teoriya stokhasticheskikh sistem* [Stochastic systems: Theory and applications]. Moscow: Logos. 1000 p.
2. Mal’chikov, S. V. 1970. *Priblizhennyi metod opredeleniya zakonov raspredeleniya fazovykh koordinat nelineynykh avtomaticheskikh sistem* [Approximate method for phase coordinates analysis in nonlinear control systems]. *Automat. Rem. Contr.* 5:43–50.
3. Kazakov, I. E., and S. V. Mal’chikov. 1983. *Analiz stokhasticheskikh sistem v prostranstve sostoyaniy* [Stochastic systems analysis in state space]. Moscow: Nauka. 348 p.
4. Kotel’nikov, V. A. 1956. *Teoriya potentsial’noy pomekho-ustoychivosti* [Theory of potential noiseproof]. Moscow: Gosenergoizdat. 412 p.
5. Sinitsyn, I. N. 2018. Metod interpolyatsionnogo analiticheskogo modelirovaniya odnomernykh raspredeleniy v stokhasticheskikh sistemakh [Method of interpolational analytical modeling of processes in stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(1):56–62.
6. Korolyuk, V. S., N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, eds. 1985. *Spravochnik po teorii veroyatnostey v matematicheskoy statistike* [Handbook on probability theory and mathematical statistics]. Moscow: Nauka. 640 p.
7. Evlanov, L. G., and V. M. Konstantinov. 1976. *Sistemy so sluchaynymi parametrami* [Systems with random parameters]. Moscow: Nauka. 585 p.
8. Korn, G., and T. Korn. 1968. *Mathematical handbook for scientists and engineers*. New York – San Francisco – Toronto – London – Sydney: McGraw Hill Book Co. 1147 p.

Received August 16, 2018

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, 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; sinitsin@dol.ru

УПРАВЛЕНИЕ ВЫХОДОМ СТОХАСТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ СИСТЕМЫ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ. II. ЧИСЛЕННОЕ РЕШЕНИЕ УРАВНЕНИЙ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ*

А. В. Босов¹, А. И. Стефанович²

Аннотация: Представлена вторая часть исследования задачи оптимального управления для диффузионного процесса Ито и линейного управляемого выхода. Оптимальное управление выходом $dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t$ стохастической дифференциальной системы с состоянием $dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t$ и квадратичным критерием качества, определяемое функцией Беллмана вида $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$, рассчитывается путем приближенного решения сеточными методами дифференциальных уравнений для коэффициентов α_t , $\beta_t(y)$ и $\gamma_t(y)$. Подробно рассмотрен модельный пример, опирающийся на простую дифференциальную модель для показателя RTT (Round-Trip Time) сетевого протокола TCP (Transmission Control Protocol). Приводятся результаты численного эксперимента, позволяющие оценить трудности практической реализации оптимального решения и обозначить задачи дальнейшего исследования.

Ключевые слова: стохастическое дифференциальное уравнение; оптимальное управление; динамическое программирование; функция Беллмана; уравнение Риккати; линейные уравнения параболического типа

DOI: 10.14357/19922264190102

1 Введение

В работе [1] получены аналитические соотношения, описывающие оптимальное решение в задаче управления линейным выходом стохастической дифференциальной системы по квадратичному критерию качества. Оптимизируемая динамическая система описывается двумя уравнениями: нелинейным стохастическим дифференциальным уравнением Ито для состояния и линейным уравнением для управляемого выхода. Квадратичный целевой функционал обеспечил возможность решения задачи методом динамического программирования: получены аналитические выражения для функции Беллмана и, как следствие, для оптимального управления и для значения функционала качества. Соответствующие соотношения содержат, как и всегда в подобных задачах, решения определенных дифференциальных уравнений (обыкновенных и в частных производных). Формальная постановка задачи и основной теоретический результат кратко сформулированы в следующем разделе статьи. В данной работе исследование этой же задачи продолжено, но внимание сосредоточено на аспектах практической реализации результата.

С фундаментальной точки зрения в сравнении с исходной постановкой задачи оптимизации в стохастической системе поиск решений дифференциальных уравнений — это удовлетворительный итог исследования, достаточный для перехода в практическую область, т.е. проведения расчетов, модельных как минимум, а то и с реальными данными. На самом деле нюансы численной реализации полученного в [1] оптимального управления в совокупности сами по себе представляют исследовательский вызов. В полном объеме ответить на него пока не удастся, но начальные результаты, приближенные алгоритмы, модельные расчеты получены и составляют предмет настоящей работы.

Итоги подводятся в заключении, где в том числе обозначены перспективы дальнейших исследований.

2 Оптимальное решение задачи управления выходом

Следуя [1], будем предполагать рассматриваемые далее случайные функции скалярными. Ска-

* Работа выполнена при частичной поддержке РФФИ (проект 16-07-00677).

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

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

лярная модель позволяет обойти трудности численной реализации, неизбежно возникающие при росте размерности. Это принципиально важно для данного этапа исследования, поскольку дальнейшие усилия предполагаются в направлении совершенствования предложенных здесь методов приближенного численного решения, в том числе с учетом возможности роста размерности.

Итак, рассматривается состояние y_t стохастической дифференциальной системы, описываемое нелинейным стохастическим дифференциальным уравнением Ито:

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

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 [2].

С состоянием этой системы связан выход, который описывается функцией z_t , линейно связанной с y_t :

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z, \quad (1)$$

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 [2], что обеспечивает существование решения уравнения (1) для любого допустимого управления.

Используется целевой функционал следующего вида:

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

$$U_0^T = \{u_t, 0 \leq t \leq T\},$$

где S_t , G_t и H_t — неотрицательные функции.

Решение задачи поиска u_t^* — допустимого управления, доставляющего минимум квадратичному функционалу $J(U_0^T)$, составляют следующие соотношения.

Функция Беллмана $V_t(y, z)$ может быть представлена в виде:

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

а оптимальное управление

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

Коэффициенты α_t , $\beta_t(y)$ и $\gamma_t(y)$ задаются следующими дифференциальными уравнениями:

$$\begin{aligned} \frac{\partial \alpha_t}{\partial t} + 2\alpha_t \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) + \\ + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) g_t^2 + \\ + G_t - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t^2 = 0, \\ \alpha_T = S_T g_T^2 + G_T; \quad (3) \end{aligned}$$

$$\begin{aligned} \frac{\partial \beta_t(y)}{\partial y} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t(y)}{\partial y^2} + A_t(y) \frac{\partial \beta_t(y)}{\partial y} + \\ + 2\alpha_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ + \beta_t(y) \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) - \\ - 2 \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t y - \\ - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t \beta_t(y) = 0, \\ \beta_T(y) = -2S_T s_T g_T y; \quad (4) \end{aligned}$$

$$\begin{aligned} \frac{\partial \gamma_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \gamma_t(y)}{\partial y^2} + \sigma_t^2 \alpha_t + A_t(y) \frac{\partial \gamma_t(y)}{\partial y} + \\ + \beta_t(y) \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 - \\ - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2(y) = 0, \\ \gamma_T(y) = S_T s_T^2 y^1. \quad (5) \end{aligned}$$

Уравнение (3) является уравнением Риккати и имеет единственное неотрицательное решение для всех $0 \leq t \leq T$, так как предполагается $S_t h_t^2 + H_t > 0$. Уравнения (4) и (5) представляют собой линейные уравнения в частных производных второго порядка, относятся к параболическому типу, поскольку $\Sigma_t^2(y) > 0$. Предполагается, что данные уравнения имеют на рассматриваемом интервале $0 \leq t \leq T$ хотя бы одно ограниченное решение, которое и определяет оптимальное решение рассматриваемой задачи оптимизации.

3 Модельный эксперимент и приближенное численное решение

Практическая реализация оптимального управления (2), конечно, возможна только приближенно, рассматривать варианты, когда уравнения (3) и (4) могут быть решены аналитически, бесперспективно. Приближенное решение очевидным образом получается в результате численного интегрирования уравнений (3) и (4), а при необходимости и (5). Для реализации этого требуется модельный пример, который позволит, с одной стороны, поэкспериментировать с численными процедурами, сформировать представление об их точности и ресурсоемкости. С другой стороны, при выборе модельного примера следует ориентироваться на потенциальные прикладные области применения представленного результата. В качестве такой возможной области видится область телекоммуникаций — популярный современный источник приложений для различных теоретических задач. Из моделей, исследуемых в этой области, была выбрана простая модель для показателя RTT сетевого протокола TCP, предложенная в [3]. Опуская детали выбора параметров в используемой диффузионной модели RTT, укажем только итоговое уравнение рассматриваемого модельного примера:

$$dy_t = (1 - 0,1y_t) dt + 0,5\sqrt{y_t} dv_t, \quad y_0 = Y \sim N(15, 9). \quad (6)$$

Здесь $N(M, D)$ — нормальное распределение со средним M и дисперсией D .

Надо отметить, что такое уравнение (с точностью до варьируемых параметров) изначально известно как модель Кокса—Ингерсона—Росса (Cox—Ingersoll—Ross model) [4] и описывает эволюцию процентных ставок. Конечно, для описания реальных данных RTT модель типа (6) представляет не более чем начальное приближение, но оно удачно уже потому, что дает основу для дальнейших обобщений, уточнений модели (см., например, [5]), поэтому удобна и для целей настоящей работы. Свойства процесса y_t из (6) хорошо изучены, в частности имеется свойство неотрицательности выборочных значений, наличие эргодичности, известно предельное распределение и переходная вероятность. Отметим, что начальные условия в (6) выбраны так, что M и D отличаются от моментных характеристик предельного распределения, т. е. управление ведется в рамках переходного процесса, а не в стационарном режиме.

Выход для (6) задается уравнением

$$dz_t = 0,1y_t dt - z_t dt + u_t dt + dw_t, \quad z_0 = Z \sim N(9, 9), \quad (7)$$

целевой функционал

$$J(U_0^T) = E \left\{ \int_0^T \left((y_t - z_t)^2 + z_t^2 + u_t^2 \right) dt + (y_T - z_T)^2 + z_T^2 \right\}.$$

Моделировалось $N = 1000$ траекторий y_t , z_t^* и u_t^* для $T = 5$ и 50 (через z_t^* обозначен выход (7), рассчитанный для $u_t = u_t^*$). Кроме того, моделировались траектории z_t^{prog} и u_t^{prog} для наилучшего программного управления $u_t^{\text{prog}} = E\{u_t^*\}$ и z_t^0 и u_t^0 для неуправляемого выхода (7), т. е. для $u_t = 0$. В каждом случае путем осреднения по пучку траекторий оценивались величины $J(U_0^t)$. Соответственно, в результате можно увидеть не только конечные значения целевой функции для разных вариантов управления $J((U^*)_0^T)$, $J((U^{\text{prog}})_0^T)$ и $J((U^0)_0^T)$, но и их формирование в динамике.

Для определения оптимального управления численно решались уравнения для α_t и $\beta_t(y)$. Сначала было получено численное решение уравнения (3). Это уравнение Риккати, оно имеет единственное решение, и вычисление этого решения не представляет труда, например неявным методом Эйлера.

Далее решалось уравнение (4). Как уже отмечалось, нахождение решений уравнений (4) и при необходимости (5) (это уравнение можно использовать для определения качества оптимального управления) представляет определенные трудности. Для целей данной работы достаточно использовать традиционный подход к численному решению — метод конечных разностей, тем более что варианты этого метода для уравнения параболического типа давно и хорошо изучены [6].

Первое, что требуется сделать, — это ограничить область значения аргумента y для формирования численной схемы расчета. С этой целью использовались N смоделированных методом Эйлера траекторий (6): выборочные значения использовались для оценивания функций математического ожидания и дисперсионной, далее задавалась 3σ -трубка, а в расчетах граница значений y задавалась путем добавления 10% к границе 3σ -трубки. Рисунок 1 иллюстрирует характерные траектории (6), среднее значение процесса и вычисленные границы (закрашенная серым область).

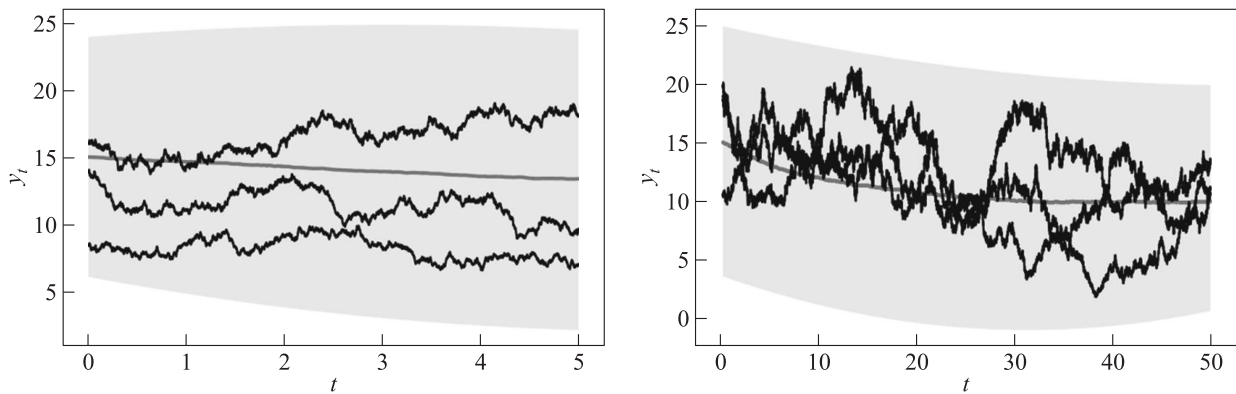


Рис. 1 Выборочные траектории и характеристика границы

Определение границы создает следующую проблему, обычную при применении любого сеточного метода, а именно: отсутствие граничных условий для рассматриваемых уравнений. Этих условий нет, и нет физических оснований для их обоснованного выбора. Из-за этого приходится выбирать граничные условия довольно волюнтаристски и надеяться на устойчивость решения в отношении этих граничных условий.

Выбор граничного условия и анализ его устойчивости выполнялся следующим образом. Были взяты два традиционных варианта граничных условий — в задаче Дирихле: $\beta_t(y) = 0$, т.е. условие поглощения, и в задаче Неймана: $\partial\beta_t(y)/\partial y = 0$, т.е. условия отражения, для всех граничных точек. Далее уравнение решалось методом конечных разностей с использованием явной и неявной численных схем сначала для одного граничного условия, затем для второго, и полученные решения сравнивались. Устойчивость в таком случае означает совпадение (близость) решений в заданной области за исключением границы.

Для численного интегрирования в области $[0, T] \cup [y, \bar{y}]$ (y и \bar{y} — определенные в результате моделирования границы) формируется сетка $\{t_k, y_i\}$ с шагами δ_t (для интегрирования по переменной t) и δ_y (для интегрирования по переменной y). С использованием обозначения β_i^k для значения $\beta_t(y)$ в узле (t_k, y_i) ее производные аппроксимировались следующим образом:

$$\frac{\partial\beta_t(y)}{\partial t} \approx \frac{\beta_i^{k+1} - \beta_i^k}{\delta_t}; \quad \frac{\partial\beta_t(y)}{\partial y} \approx \frac{\beta_{i+1}^K - \beta_{i-1}^K}{2\delta_y};$$

$$\frac{\partial^2\beta_t(y)}{\partial y^2} \approx \frac{\beta_{i+1}^K - 2\beta_i^K + \beta_{i-1}^K}{\delta_y^2},$$

где $K = k$ для явной схемы и $K = k + 1$ — для неявной.

Шаги интегрирования δ_t и δ_y при анализе устойчивости выбирались разными, например явная схе-

ма обчисчивалась для $\delta_t = 0,0002$ и $\delta_y = 0,1$, в том числе с учетом того, что для нее самой вопрос устойчивости ограничивает вариации шагов. Главное, что в итоге подтвердилась устойчивость результатов численного интегрирования $\beta_t(y)$ по отношению к граничным условиям по крайней мере для выбранной модели.

Итоговый расчет выполнен для неявной схемы с шагами $\delta_t = 0,001$ (в том числе для приближенного вычисления α_t) и $\delta_y = 0,01$. Результаты расчетов представлены на рис. 2–4, в частности приведены примеры траекторий для выхода и управлений $z_t^*, u_t^*, z_t^{\text{prog}}, u_t^{\text{prog}}$ и z_t^0, u_t^0 .

Рисунок 4 демонстрирует ожидаемый проигрыш неуправляемой системы z_t^0, u_t^0 . Кроме того, обращает на себя внимание высокое качество программного управления и выхода $z_t^{\text{prog}}, u_t^{\text{prog}}$. Объективных оснований последнее обстоятельство, конечно, не имеет и объясняется исключительно выбранными параметрами рассматриваемого модельного примера, прежде всего интенсивностями возмущений — коэффициентов при v_t и w_t , что в свою очередь сделано для того, чтобы на рис. 4 изменения целевого функционала для разных стратегий управления можно было представить в одном масштабе.

Наконец, отметим, что существенных технических трудностей при выполнении расчетов не возникло. Без дополнительных усилий сетка размещалась в оперативной памяти: для интегрирования с выбранными шагами в области $[0, 50] \cup [-10, 40]$ было использовано всего лишь 2 Гб памяти, что позволило провести расчеты на обычном персональном компьютере. При этом размещение сетки в оперативной памяти обеспечило довольно быстрое проведение расчетов. Гораздо больше временных ресурсов требуется для моделирования пучков траекторий, осреднения — сравнения свойств управлений. Конечно, с ростом размерности ситу-

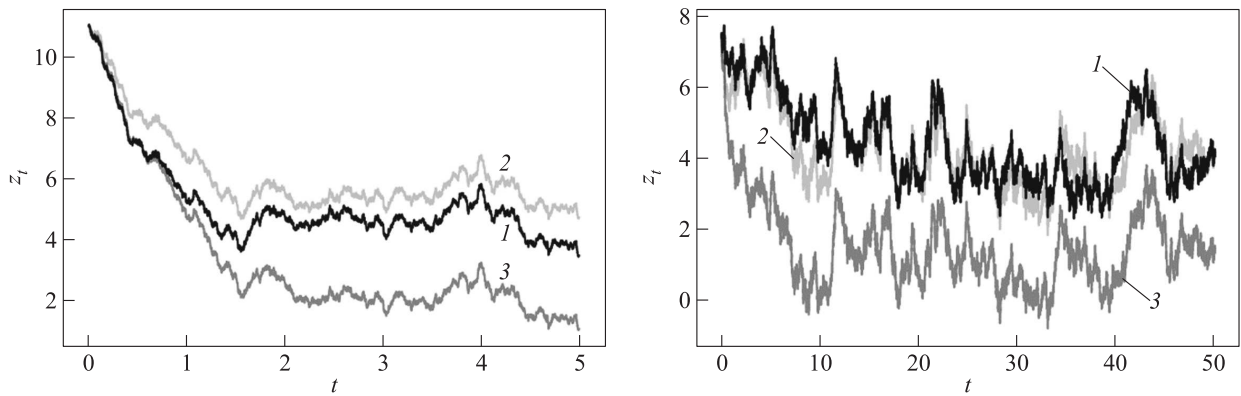


Рис. 2 Выборочные траектории z_t^* (1), z_t^{prog} (2) и z_t^0 (3)

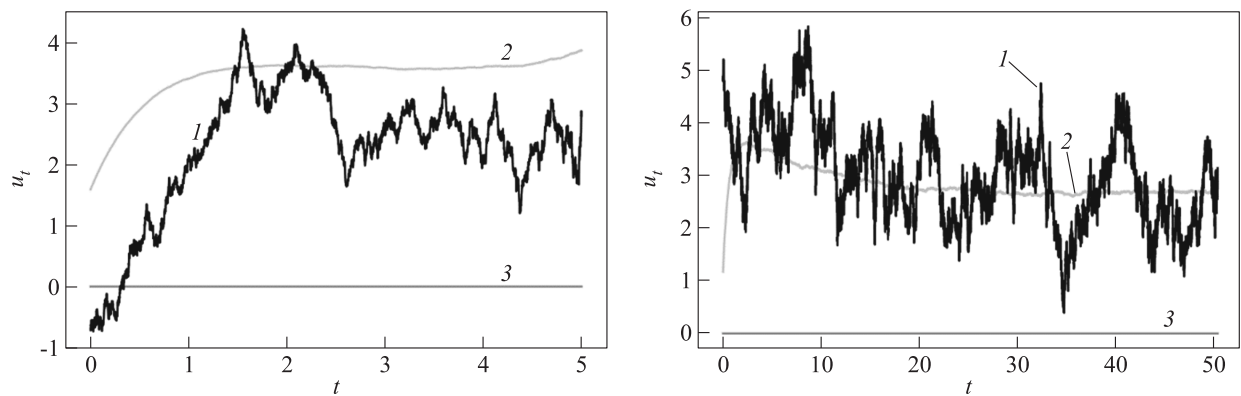


Рис. 3 Выборочные траектории u_t^* (1), u_t^{prog} (2) и u_t^0 (3)

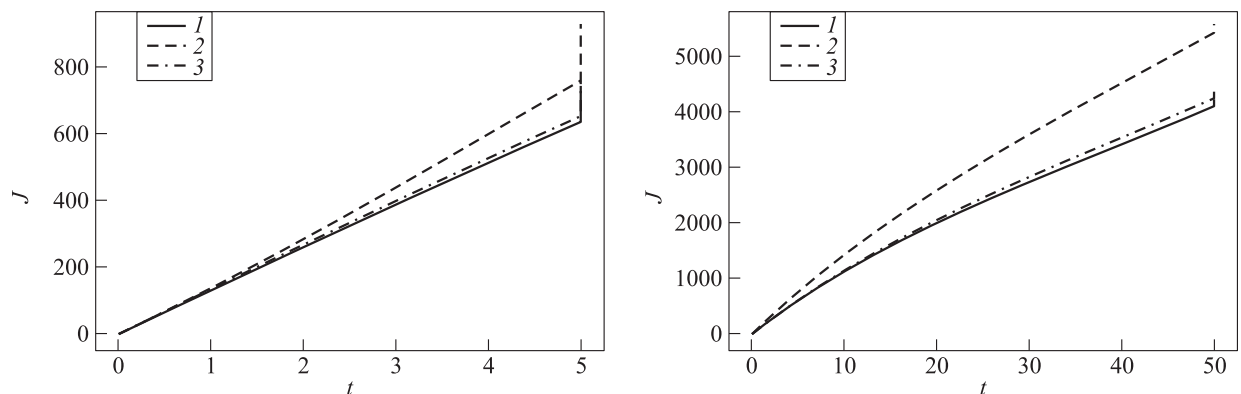


Рис. 4 Динамика целевого функционала: 1 — $J((U^*)_0^t)$; 2 — $J((U^0)_0^t)$; 3 — $J((U^{\text{prog}})_0^t)$

ация принципиально изменится и затраты на численное интегрирование дифференциальных уравнений вырастут на порядки.

4 Заключение

Настоящая статья содержит вторую, практическую часть исследования задачи оптимизации линейного выхода нелинейной дифференциальной системы по квадратичному критерию. В отношении

представленного результата следует отметить, что полученное решение представляет определенные вычислительные трудности при его приближенной реализации путем численного решения дифференциальных уравнений для соответствующих коэффициентов.

Отсутствие физически обоснованных граничных условий потребовало значительных усилий для анализа численного решения на границе. Кроме того, очевидное обобщение постановки на много-

мерный случай еще более усугубит вычислительные проблемы. Таким образом, затраты вычислительных ресурсов даже на модельные расчеты в скалярном случае оказались слишком значительными, чтобы не учитываться в теоретической части решения. По этой причине, прежде чем двигаться с рассмотренной задачей дальше, обобщать постановку на многомерный случай, нужно предложить более действенные инструменты в части численной реализации полученных точных решений. Возможности для этого имеются, а соответствующие результаты должны составить ближайшие перспективы дальнейших исследований.

Представленная в данной работе задача оптимизации пока не исследовалась детально на предмет практического применения. Выполненные модельные расчеты — это, скорее, первый шаг в направлении поиска реальных приложений. При этом основным представляется поиск моделей, обладающих практической ценностью и достаточно исследованных в рамках задач анализа. Источником таких моделей видится область инфотелекоммуникаций, активно развиваемая в последние годы, в частности модели сетевых транспортных протоколов на основе скачкообразных марковских процессов [7–10]. Применение рассмотренной оптимизационной задачи к таким моделям также может стать содержательным продолжением представленного в данной работе исследования.

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. doi: 10.14357/19922264180314.
2. Флеминг У., Рушел П. 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.)
3. Bohacek S., Rozovskii B. A diffusion model of roundtrip time // Comput. Stat. Data An., 2004. Vol. 45. Iss. 1. P. 25–50. doi: 10.1016/S0167-9473(03)00114-2.
4. Cox J. C., Ingersoll J. E., Ross S. A. A theory of the term structure of interest rates // Econometrica, 1985. Vol. 53. Iss. 2. P. 385–407. doi: 10.2307/1911242.
5. Bohacek S. A stochastic model of TCP and fair video transmission // Proc. IEEE INFOCOM, 2003. P. 1134–1144. doi: 10.1109/INFOCOM.2003.1208950.
6. Саульев В. К. Интегрирование уравнений параболического типа методом сеток. — М.: Физматлит, 1960. 324 с.
7. Борисов А. В., Миллер Г. Б. Анализ и фильтрация специальных марковских процессов в дискретном времени. II. Оптимальная фильтрация // Автоматика и телемеханика, 2005. № 7. С. 112–125. doi: 10.1007/s10513-005-0153-7.
8. Борисов А. В., Миллер Б. М., Семенхин К. В. Фильтрация марковского скачкообразного процесса по наблюдениям мультивариантного точечного процесса // Автоматика и телемеханика, 2015. № 2. С. 34–60. doi: 10.1134/S0005117915020034.
9. Borisov A., Bosov A., Miller G. Modeling and monitoring of RTP link on the receiver side // Internet of things, smart spaces, and next generation networks and systems / Eds. S. I. Balandin, S. D. Andreev, Y. Koucheryavy. — Lecture notes in computer science ser. — Springer, 2015. Vol. 9247. P. 229–241. doi: 10.1007/978-3-319-23126-6_21.
10. Борисов А. В. Применение методов оптимальной фильтрации для оперативного оценивания состояний сетей массового обслуживания // Автоматика и телемеханика, 2016. № 2. С. 115–141. doi: 10.1134/S0005117916020053.

Поступила в редакцию 07.06.18

STOCHASTIC DIFFERENTIAL SYSTEM OUTPUT CONTROL BY THE QUADRATIC CRITERION. II. DYNAMIC PROGRAMMING EQUATIONS NUMERICAL SOLUTION

A. V. Bosov and A. I. Stefanovich

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 second part of the optimal control problem investigation for the Ito diffusion process and the controlled linear output is presented. Optimal control for output $dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t$ of the stochastic differential system $dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t$ and quadratic quality criterion defined by Bellman function having form $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$ is determined numerically by an approximate solution to

the grid methods of differential equations for the coefficients α_t , $\beta_t(y)$, and $\gamma_t(y)$. A model experiment based on a simple differential presentation for the RTT (Round-Trip Time) parameter of the TCP (Transmission Control Protocol) network protocol is considered in detail. The results of numerical simulation are given and allow one to assess the difficulties in the practical implementation of the optimal solution and define the tasks of further research.

Keywords: stochastic differential equation; optimal control; dynamic programming; Bellman function; Riccati equation; linear differential equations of parabolic type

DOI: 10.14357/19922264190102

Acknowledgments

This work was partially supported by the Russian Science Foundation (grant 16-07-00677).

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. doi: 10.14357/19922264180314.
2. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p.
3. Bohacek, S., and B. Rozovskii. 2004. A diffusion model of roundtrip time. *Comput. Stat. Data An.* 45(1):25–50. doi: 10.1016/S0167-9473(03)00114-2.
4. Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53:385–407. doi: 10.2307/1911242.
5. Bohacek, S. 2003. A stochastic model of TCP and fair video transmission. *Proc. IEEE INFOCOM*. 1134–1144. doi: 10.1109/INFCOM.2003.1208950.
6. Sauliev, V. K. 1960. *Integrirovaniye uravneniy parabolicheskogo tipa metodom setok* [Integration of parabolic equations by the grid method]. Moscow: Fizmatlit. 324 p.
7. Borisov, A. V., and G. B. Miller. 2005. Analysis and filtration of special discrete-time Markov processes. II. Optimal filtration. *Autom. Rem. Contr.* 66(7):1125–1136.
8. Borisov, A. V., G. B. Miller, and K. V. Semenikhin. 2015. Filtering of the Markov jump process given the observations of multivariate point process. *Autom. Rem. Contr.* 76(2):219–240.
9. Borisov, A., A. Bosov, and G. Miller. 2015. Modeling and monitoring of RTP link on the receiver side. *Internet of things, smart spaces, and next generation networks and systems*. Eds. S. I. Balandin, S. D. Andreev, and Y. Koucheryavy. Lecture notes in computer science ser. Springer, 2015. 9247:229–241. doi: 10.1007/978-3-319-23126-6_21.
10. Borisov, A. V. 2016. Application of optimal filtering methods for on-line of queueing network states. *Autom. Rem. Contr.* 77(2):277–296.

Received June 7, 2018

Contributors

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

Stefanovich Alexey I. (b. 1983) — principal specialist, 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; AStefanovich@frccsc.ru

ОБ ОДНОМ КЛАССЕ ЗАДАЧ ФИЛЬТРАЦИИ НА МНОГООБРАЗИЯХ*

К. А. Рыбаков¹

Аннотация: Цель статьи состоит в описании стохастических дифференциальных систем, траектории которых находятся на гладком многообразии, в приложении к задаче оптимальной фильтрации. Дополнительным условием является принадлежность этому же многообразию не только траекторий системы, но и результата оценивания этих траекторий на основе косвенных измерений, а именно: решения задачи оптимальной фильтрации по критерию минимума среднеквадратической ошибки оценивания. Рассматриваются системы как диффузионного типа, так и диффузионно-скачкообразного типа, т. е. при наличии как винеровских, так и пуассоновских возмущений. Результатом являются условия на коэффициенты уравнения для случайного процесса, траектории которого требуется оценить. В основе полученных условий лежит понятие первого интеграла стохастических дифференциальных уравнений, а также некоторые его свойства.

Ключевые слова: инвариант; многообразие; оптимальная фильтрация; оценивание; случайный процесс; стохастическая дифференциальная система

DOI: 10.14357/19922264190103

1 Введение

В теории стохастических динамических систем задача фильтрации имеет важное значение, она состоит в нахождении оценки ненаблюдаемого вектора состояния системы по результатам его косвенных измерений. Критерии оптимальности оценки можно задавать различным образом, во многих приложениях ограничиваются критерием минимума среднеквадратической ошибки оценивания. Такие задачи решаются для непрерывных, дискретных и непрерывно-дискретных систем, в том числе и в случае, когда вектор состояния системы принадлежит заданному многообразию. Решению задач фильтрации на многообразиях посвящен целый ряд работ, опубликованных в последнее время [1–5].

Принадлежность вектора состояния стохастической динамической системы многообразию означает, что в системе выполняется некоторый закон сохранения. В более общем случае рассматривается принадлежность пары «время + состояние» динамическому многообразию. Методы описания и построения подобных непрерывных стохастических систем подробно изложены в [6–9]. Но этот закон сохранения для оценки вектора состояния при решении задачи фильтрации выполняться не будет, кроме специального класса систем.

Перейдем к иллюстрирующему примеру. Пусть векторный случайный процесс $X(t) = [X_1(t) X_2(t)]^T$ со значениями в \mathbb{R}^2 удовлетворяет линейному стохастическому дифференциальному уравнению Ито

$$dX(t) = \underbrace{\begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}}_F X(t) dt + \underbrace{\begin{bmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & 0 \end{bmatrix}}_S X(t) dW(t),$$

в котором $W(t)$ — скалярный винеровский процесс. Соответствующее линейное стохастическое дифференциальное уравнение Стратоновича имеет вид:

$$d_{1/2}X(t) = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_A X(t) dt + \underbrace{\begin{bmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & 0 \end{bmatrix}}_S X(t) d_{1/2}W(t).$$

Аналитическое решение этих уравнений можно записать, используя матричные экспоненты [10]: $X(t) = \exp At \exp SW(t) X(t_0)$, $t \geq t_0$. Несложно проверить, что собственные числа матриц A и S — комплексно-сопряженные с нулевой действительной частью, определители матриц $\exp At$ и $\exp SW(t)$ равны единице и эти матрицы задают ортогональные линейные преобразования на плоскости, а именно: повороты вокруг начала координат. Такие преобразования сохраняют расстояние между точками в \mathbb{R}^2 : $|X(t)| = |X(t_0)|$, поэтому траектории случайного процесса $X(t)$ принадлежат круговому цилиндру в $\mathbb{T} \times \mathbb{R}^2$, а фазовые траектории — это окружности с центром в начале координат (точка покоя — центр), радиус которых определяется начальными данными $X(t_0)$.

* Работа выполнена при поддержке РФФИ (проект 17-08-00530-а).

¹ Московский авиационный институт (национальный исследовательский университет), rkoffice@mail.ru

Математическое ожидание для случайного процесса $X(t)$ также можно выразить с помощью матричной экспоненты: $\mathbb{E}X(t) = \exp Ft \mathbb{E}X(t_0)$ (здесь и далее \mathbb{E} — знак математического ожидания), собственные числа матрицы F — комплексно-сопряженные с отрицательной действительной частью и для математического ожидания верно соотношение:

$$e^t |\mathbb{E}X(t)| = e^{t_0} |\mathbb{E}X(t_0)|.$$

Таким образом, траектории случайного процесса и его математическое ожидание принадлежат разным многообразиям $T \times \mathbb{R}^2$. Кривая математического ожидания принадлежит конусу в $T \times \mathbb{R}^2$ — динамическому многообразию, а проекция этой кривой на фазовую плоскость — это логарифмическая спираль (точка покоя — устойчивый фокус).

Если же рассмотреть задачу оптимальной фильтрации, т. е. оценивать траектории случайного процесса по результатам косвенных измерений, то результаты оценивания по критерию минимума среднеквадратической ошибки не будут удовлетворять ни одному из указанных инвариантных соотношений и, следовательно, закон сохранения, который выполняется для траекторий случайного процесса $X(t)$, не будет выполняться для результатов оценивания этих траекторий. Результаты оценивания будут находиться в области, ограниченной круговым цилиндром и конусом.

Цель статьи состоит в описании класса стохастических дифференциальных систем, для которых одному и тому же многообразию принадлежат не только траектории системы, но и результат решения задачи оптимальной фильтрации по критерию минимума среднеквадратической ошибки оценивания.

Статья помимо введения и заключения содержит 5 разделов. В разд. 2 сформулирована задача оптимальной фильтрации на гладком многообразии. В разд. 3 приведены необходимые и достаточные условия принадлежности траекторий стохастической системы заданному многообразию. Раздел 4 содержит основной результат статьи — необходимые и достаточные условия принадлежности решения задачи оптимальной фильтрации по критерию минимума среднеквадратической ошибки оценивания многообразию, которому принадлежат оцениваемые траектории. В разд. 5 результаты, полученные в разд. 4, обобщаются на стохастические системы при наличии пуассоновских возмущений. В разд. 6 приводится модельный пример стохастической системы, для которой одному и тому же многообразию принадлежат не только траектории этой системы, но и результат решения задачи оптимальной фильтрации.

2 Постановка задачи оптимальной фильтрации

В работе рассматривается стохастическая система наблюдения, задаваемая стохастическими дифференциальными уравнениями Ито:

$$dX(t) = f(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad X(t_0) = X_0; \quad (1)$$

$$dY(t) = c(t, X(t)) dt + \zeta(t) dV(t), \quad Y(t_0) = Y_0, \quad (2)$$

где $X \in \mathbb{R}^n$ — ненаблюдаемый вектор состояния; $Y \in \mathbb{R}^m$ — вектор измерений; $t \in T$, $T = [t_0, T]$ — заданный отрезок времени; $W(t)$ и $V(t)$ — s - и d -мерные независимые винеровские процессы; $f(t, x)$, $c(t, x)$, $\sigma(t, x)$ и $\zeta(t)$ — заданные вектор-функции и матричные функции соответствующих размеров; X_0 и Y_0 — начальный вектор состояния и начальный вектор измерений. Распределение X_0 известно, а Y_0 , как правило, — нулевой вектор. Функции $f(t, x)$, $c(t, x)$, $\sigma(t, x)$ и $\zeta(t)$ удовлетворяют условиям существования и единственности решения стохастических дифференциальных уравнений [11]. Кроме того, $\eta(t) = \zeta(t) \zeta^T(t)$ — невырожденная матрица, для которой существует обратная матрица $q(t) = \eta^{-1}(t)$.

Предполагается, что траектории случайного процесса $X(t)$ принадлежат гладкому многообразию $\mathcal{M} \subset T \times \mathbb{R}^n$, которое определяется соотношением $\mathcal{M} = \{(t, x) \in T \times \mathbb{R}^n : M(t, x) = C = const\}$. Здесь $M(t, x)$ — скалярная функция, не равная постоянной, непрерывно дифференцируемая по переменной t и дважды непрерывно дифференцируемая по координатам вектора x . Стохастическую систему, которая задается уравнением (1), будем называть инвариантной. Для инвариантной системы почти наверное $(t, X(t)) \in \mathcal{M}$, если $(t_0, X_0) \in \mathcal{M}$, т. е. $M(t, X(t)) = M(t_0, X_0)$. Подобных ограничений на случайный процесс $Y(t)$ не накладывается.

Задача оптимальной фильтрации состоит в нахождении оценки $\hat{X}(t)$ по результатам измерений $Y_0^t = \{Y(\tau), \tau \in [t_0, t]\}$, т. е. $\hat{X}(t) = \psi(t, Y_0^t)$, где $\psi(t, Y_0^t)$ — функция, обеспечивающая в каждый момент времени $t \in T$ выполнение условия:

$$\mathbb{E} \left[\left(X(t) - \hat{X}(t) \right)^T \left(X(t) - \hat{X}(t) \right) \right] \rightarrow \min_{\psi(t, \cdot)},$$

т. е. решается задача оптимальной фильтрации по критерию минимума среднеквадратической ошибки оценивания.

Решение этой задачи записывается в виде апостериорного математического ожидания [12]:

$$\hat{X}(t) = \mathbb{E} [X(t) | Y_0^t] = \int_{\mathbb{R}^n} xp(t, x | Y_0^t) dx,$$

или [13]:

$$\hat{X}(t) = \frac{\mathbb{E}[\omega(t)X(t)]}{\mathbb{E}\omega(t)}, \quad (3)$$

где $p(t, x|Y_0^t)$ — апостериорная плотность вероятности вектора состояния X ; $\omega(t)$ — весовая функция:

$$\omega(t) = \exp \left\{ \int_{t_0}^t c^T(\tau, X(\tau))q(\tau) dY(\tau) - \frac{1}{2} \int_{t_0}^t c^T(\tau, X(\tau))q(\tau)c(\tau, X(\tau)) d\tau \right\}.$$

Отметим, что на последнем соотношении основан непрерывный фильтр частиц и его различные варианты [13, 14].

3 Условия инвариантности

Запишем стохастическое дифференциальное уравнение в форме Стратоновича для случайного процесса $X(t)$:

$$d_{1/2}X(t) = a(t, X(t)) dt + \sigma(t, X(t)) d_{1/2}W(t), \quad X(t_0) = X_0. \quad (4)$$

В этом уравнении $a(t, x)$ — вектор-функция той же размерности, что и функция $f(t, x)$:

$$a(t, x) = f(t, x) - \frac{1}{2} \sum_{l=1}^s \frac{\partial \sigma_{*l}(t, x)}{\partial x} \sigma_{*l}(t, x), \quad (5)$$

где $\sigma_{*l}(t, x)$ — столбец матричной функции $\sigma(t, x)$ с номером $l, l = 1, 2, \dots, s$.

Для функций $M(t, x)$, $a(t, x)$ и $\sigma(t, x)$, определяющих гладкое многообразие \mathcal{M} и уравнение (4), на траекториях случайного процесса $X(t)$ должны выполняться следующие условия:

$$\frac{\partial M(t, x)}{\partial t} + \sum_{i=1}^n a_i(t, x) \frac{\partial M(t, x)}{\partial x_i} = 0; \quad (6)$$

$$\sum_{i=1}^n \sigma_{il}(t, x) \frac{\partial M(t, x)}{\partial x_i} = 0, \quad l = 1, 2, \dots, s. \quad (7)$$

Эти условия эквивалентны равенству нулю дифференциала Стратоновича для случайного процесса $M(t, X(t))$ [7]. Условия для функции $f(t, x)$ можно получить из равенства нулю дифференциала Ито для случайного процесса $M(t, X(t))$ или подставить (5) в (6). Такие условия приведены в [6–9], в этих же работах подробно изложена теория первых интегралов стохастических дифференциальных уравнений (функция $M(t, x)$ — первый интеграл уравнений (1) и (4)), приложение этой теории к за-

дачам программного управления стохастическими динамическими системами, там же приведены многочисленные примеры.

Отметим, что с геометрической точки зрения (6) — условие ортогональности блочного вектора $[1 \ a^T(t, x)]^T$ и обобщенного градиента $\nabla_{t,x}M(t, x)$, а (7) — условие ортогональности каждого столбца матрицы $\sigma(t, x)$ и градиента $\nabla M(t, x)$ в $\mathbb{R}^n \forall t \in T$. В случае $M(t, x) = M(x)$ условие (6) — это условие ортогональности вектора $a(t, x)$ и градиента $\nabla M(t, x) = \nabla M(x)$ в $\mathbb{R}^n \forall t \in T$.

4 Условия инвариантности в среднем

Опишем стохастическую систему, задаваемую стохастическим дифференциальным уравнением Ито (1) и соответствующим стохастическим дифференциальным уравнением Стратоновича (4), с дополнительным условием:

$$(t, \mathbb{E}X(t)), (t, \mathbb{E}[X(t)|Y_0^t]) \in \mathcal{M},$$

т. е. заданному многообразию \mathcal{M} принадлежат не только траектории случайного процесса $X(t)$, но и априорное, а также апостериорное математическое ожидание случайного процесса $X(t)$. Для такого класса стохастических систем решение задачи оптимальной фильтрации принадлежит тому же многообразию, что и оцениваемые траектории, т. е. $(t, \hat{X}(t)) \in \mathcal{M}$.

Для этого определим линейную по вектору $x \in \mathbb{R}^n$ функцию

$$M(t, x) = (\vartheta(t), x) = \vartheta_1(t)x_1 + \dots + \vartheta_n(t)x_n, \quad (8)$$

где $\vartheta(t)$ — дифференцируемая вектор-функция, координаты которой одновременно не обращаются в нуль: $|\vartheta(t)| > 0$, $(\vartheta(t), x)$ — скалярное произведение в \mathbb{R}^n . Функция $\vartheta(t) = \nabla M(t, x)$ задает вектор нормали к гиперплоскости $M(t, x) = C$ в $\mathbb{R}^n \forall t \in T$. В $T \times \mathbb{R}^n$ многообразие $M(t, x) = C$ гиперплоскостью в общем случае не является. Тогда

$$\begin{aligned} \mathbb{E}M(t, X(t)) &= \mathbb{E}[\vartheta_1(t)X_1(t) + \dots + \vartheta_n(t)X_n(t)] = \\ &= \vartheta_1(t)\mathbb{E}X_1(t) + \dots + \vartheta_n(t)\mathbb{E}X_n(t) = M(t, \mathbb{E}X(t)). \end{aligned}$$

Аналогично

$$\begin{aligned} \mathbb{E}[M(t, X(t))|Y_0^t] &= \\ &= \mathbb{E}[\vartheta_1(t)X_1(t) + \dots + \vartheta_n(t)X_n(t) | Y_0^t] = \\ &= \vartheta_1(t)\mathbb{E}[X_1(t)|Y_0^t] + \dots + \vartheta_n(t)\mathbb{E}[X_n(t)|Y_0^t] = \\ &= M(t, \mathbb{E}[X(t)|Y_0^t]); \end{aligned}$$

следовательно, если $M(t, X(t)) = C$, то $M(t, \mathbb{E}X(t)) = C$ и $M(t, \mathbb{E}[X(t)|Y_0^t]) = C$.

Несмотря на линейность по вектору $x \in \mathbb{R}^n$ функции $M(t, x)$, стохастическая система, траектории которой принадлежат многообразию \mathcal{M} , может быть как линейной, так и нелинейной. Чтобы конструктивно описать такую систему, определим $n - 1$ линейно независимых векторов N_1, \dots, N_{n-1} , ортогональных градиенту $\nabla M(t, x)$. Эти векторы являются функциями переменной $t \in \mathbb{T}$ со значениями в \mathbb{R}^n , зависимость от t для краткости опущена.

Векторы N_1, \dots, N_{n-1} образуют базис линейного подпространства $M(t, x) = 0$ в $\mathbb{R}^n \forall t \in \mathbb{T}$. Их можно выбрать, например, следующим образом:

$$N_i = \left[E_i^T - \frac{\vartheta_i(t)}{\vartheta_n(t)} \right]^T, \quad i = 1, \dots, n - 1,$$

где E_i — единичные векторы в \mathbb{R}^{n-1} (столбцы единичной матрицы E порядка $n - 1$). Чтобы такое определение было корректным, дополнительно потребуем, чтобы $\vartheta_n(t) \neq 0$ на \mathbb{T} , тем самым обеспечив и выполнение условия $|\vartheta(t)| > 0$.

Кроме того, определим n линейно независимых векторов $\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_{n-1}$, ортогональных обобщенному градиенту $\nabla_{t,x} M(t, x)$. Вектор \tilde{N}_0 является функцией точки $(t, x) \in \mathbb{T} \times \mathbb{R}^n$, а остальные векторы — функции переменной $t \in \mathbb{T}$ со значениями в \mathbb{R}^{n+1} :

$$\tilde{N}_0 = \left[1 \ 0 \ \dots \ 0 - \frac{\vartheta_0(t, x)}{\vartheta_n(t)} \right]^T;$$

$$\tilde{N}_i = \left[0 \ E_i^T - \frac{\vartheta_i(t)}{\vartheta_n(t)} \right]^T, \quad i = 1, \dots, n - 1,$$

где

$$\vartheta_0(t, x) = \frac{\partial M(t, x)}{\partial t} = \left(\frac{\partial \vartheta(t)}{\partial t}, x \right) =$$

$$= \frac{\partial \vartheta_1(t)}{\partial t} x_1 + \dots + \frac{\partial \vartheta_n(t)}{\partial t} x_n.$$

Далее обозначим через \mathcal{N} линейную оболочку векторов N_1, \dots, N_{n-1} : $\mathcal{N} = \text{Lin}\{N_1, \dots, N_{n-1}\}$, а через \mathcal{N}_0 — линейное многообразие $N_0 + \mathcal{N}$: $\mathcal{N}_0 = \{ \mathcal{V} : \mathcal{V} = N_0 + N, N \in \mathcal{N} \}$, где вектор N_0 является функцией точки $(t, x) \in \mathbb{T} \times \mathbb{R}^n$ со значениями в \mathbb{R}^n :

$$N_0 = \left[0 \ \dots \ 0 - \frac{\vartheta_0(t, x)}{\vartheta_n(t)} \right]^T.$$

По построению произвольная линейная комбинация векторов N_1, \dots, N_{n-1} ортогональна градиенту $\nabla M(t, x)$. Следовательно, равенство (7) можно переписать в виде:

$$\sigma_{*l}(t, x) \in \mathcal{N}, \quad l = 1, 2, \dots, s, \quad (9)$$

или $\sigma_{*l}(t, x) = q_1^l(t, x)N_1 + \dots + q_{n-1}^l(t, x)N_{n-1}$, где скалярные функции $q_1^l(t, x), \dots, q_{n-1}^l(t, x)$ могут быть выбраны произвольно при дополнительных условиях существования решения уравнения (4), они представляют собой коэффициенты разложения столбца $\sigma_{*l}(t, x)$ по линейно независимой системе векторов N_1, \dots, N_{n-1} — базису линейного подпространства \mathcal{N} .

Равенство (6) с учетом введенных обозначений можно переписать следующим образом:

$$a(t, x) \in \mathcal{N}_0, \quad (10)$$

или $a(t, x) = N_0 + q_1^a(t, x)N_1 + \dots + q_{n-1}^a(t, x)N_{n-1}$, где скалярные функции $q_1^a(t, x), \dots, q_{n-1}^a(t, x)$, как и ранее введенные функции $q_r^l(t, x)$, могут быть выбраны произвольно при дополнительных условиях существования решения уравнения (4). Если $M(t, x) = M(x)$, то вектор N_0 является нулевым и, следовательно, \mathcal{N} и \mathcal{N}_0 совпадают.

Аналог условия (10) для блочного вектора $[1 \ a^T(t, x)]^T$ записывается в виде:

$$[1 \ a^T(t, x)]^T \in \tilde{N}_0 + \tilde{\mathcal{N}}, \quad \tilde{\mathcal{N}} = \text{Lin}\{\tilde{N}_1, \dots, \tilde{N}_{n-1}\}.$$

При выполнении условия (9) разность между вектор-функциями $a(t, x)$ и $f(t, x)$ согласно (5) представляет собой вектор, состоящий из суммы компонент вида:

$$\frac{1}{2} \frac{\partial (q_k^l(t, x)N_k)}{\partial x} q_r^l(t, x)N_r =$$

$$= \frac{1}{2} q_r^l(t, x)N_k [\nabla q_k^l(t, x)]^T N_r,$$

$$k, r = 1, \dots, n - 1, \quad l = 1, 2, \dots, s,$$

где произведение $[\nabla q_k^l(t, x)]^T N_r$ — это скалярная функция. Таким образом, если $a(t, x) \in \mathcal{N}_0$ и $\sigma_{*l}(t, x) \in \mathcal{N}$, то $f(t, x) \in \mathcal{N}_0$, или $f(t, x) = N_0 + q_1^f(t, x)N_1 + \dots + q_{n-1}^f(t, x)N_{n-1}$, где скалярные функции $q_1^f(t, x), \dots, q_{n-1}^f(t, x)$ могут быть выбраны произвольно при дополнительных условиях существования решения уравнения (1). Для функции $M(t, x)$, отличной от (8), условие $f(t, x) \in \mathcal{N}_0$, вообще говоря, не выполняется; соответствующий пример приведен во введении.

Теорема 1. Для того чтобы траектории стохастической дифференциальной системы, заданной уравнением Ито (1), принадлежали многообразию \mathcal{M} , которое определяется функцией (8): $(t, X(t)) \in \mathcal{M}$, и при этом $(t, \mathbb{E}X(t)), (t, \mathbb{E}[X(t)|Y_0^t]) \in \mathcal{M}$, необходимо и достаточно, чтобы коэффициенты этого уравнения удовлетворяли условиям:

$$f(t, x) \in \mathcal{N}_0; \quad \sigma_{*l}(t, x) \in \mathcal{N}, \quad l = 1, 2, \dots, s,$$

на траекториях случайного процесса $X(t)$.

Доказательство теоремы следует из предыдущих рассуждений. Если стохастическая дифференциальная система задается уравнением Стратоновича, то используется условие $a(t, x) \in \mathcal{N}_0$.

Условия теоремы, а также то, что векторы вида

$$\begin{aligned} & \frac{\partial(q_k^l(t, x)N_k)}{\partial x} q_r^j(t, x)N_r = \\ & = q_r^l(t, x)N_k [\nabla q_r^j(t, x)]^T N_r, \\ & \quad k, r = 1, \dots, n-1, \quad l, j = 1, 2, \dots, s, \end{aligned}$$

коллинеарны вектору $N_k \forall t \in T$, обеспечивают отсутствие погрешности, связанной с отклонением численного решения от многообразия \mathcal{M} , при численном интегрировании стохастического дифференциального уравнения (1) или (4) и дополнительном требовании $\vartheta(t) = const$, т. е. $M(t, x) = M(x)$. Это связано с видом соответствующих разностных схем, для которых $X_{k+1} = X_k + \Delta X_k$. Здесь ΔX_k — случайный вектор, зависящий от шага численного интегрирования h , пары (t_k, X_k) и принадлежащий $\mathcal{N} = \mathcal{N}_0$. Например, для метода Эйлера — Маруямы

$$\Delta X_k = hf(t_k, X_k) + \sqrt{h}\sigma(t_k, X_k)\Delta W_k,$$

где ΔW_k — s -мерный случайный вектор, координаты которого независимы и имеют стандартное нормальное распределение. Дискретные моменты времени t_k определяются при разбиении отрезка T с шагом h .

Погрешности такого типа анализировались в работе [15] на примере методов Эйлера — Маруямы, Мильштейна и Платена, методов типа Рунге — Кутты и типа Розенброка.

5 Условия инвариантности при пуассоновских возмущениях

Рассмотрим систему наблюдения в более общей постановке, которая принята в [1–5], а именно:

$$\begin{aligned} dX(t) &= f(t, X(t), Y(t), \rho) dt + \\ & \quad + \sigma(t, X(t), Y(t), \rho) dW(t) + \\ & \quad + \int_{\Theta} \gamma(t, X(t-), Y(t-), \rho, \theta) \nu(dt \times d\theta), \\ X(t_0) &= X_0; \end{aligned} \quad (11)$$

$$\begin{aligned} dY(t) &= c(t, X(t), Y(t), \rho) dt + \\ & \quad + \zeta(t, X(t), Y(t), \rho) dV(t) + \\ & \quad + \int_{\Theta} \delta(t, X(t-), Y(t-), \rho, \theta) \mu(dt \times d\theta), \\ Y(t_0) &= Y_0, \end{aligned} \quad (12)$$

где $X \in \mathbb{R}^n$ — ненаблюдаемый вектор состояния; $Y \in \mathbb{R}^m$ — вектор измерений; $\rho \in P \subset \mathbb{R}^l$ — вектор параметров; $t \in T$, $T = [t_0, T]$ — заданный отрезок времени; $W(t)$ и $V(t)$ — s - и d -мерные независимые винеровские процессы; $f(t, x, y, \rho)$, $\sigma(t, x, y, \rho)$, $\gamma(t, x, y, \rho, \theta)$, $c(t, x, y, \rho)$, $\zeta(t, x, y, \rho)$ и $\delta(t, x, y, \rho, \theta)$ — заданные вектор-функции и матричные функции соответствующих размеров; ν и μ — пуассоновские меры на $T \times \Theta$, $\Theta \subseteq \mathbb{R}^k$, с заданными характеристическими мерами, которые определяют интенсивность соответствующих пуассоновских потоков и законы распределения вектора θ для уравнений (11) и (12).

Наличие пуассоновской компоненты в уравнении, описывающем ненаблюдаемый вектор состояния, не влияет на условия (9) и (10), но требует дополнительного условия для функции $\gamma(t, x, y, \rho, \theta)$, которое должно выполняться на траекториях случайных процессов $X(t)$ и $Y(t)$:

$$M(t, x) = M(t, x + \gamma(t, x, y, \rho, \theta)).$$

В силу линейности по вектору $x \in \mathbb{R}^n$ функции $M(t, x)$ имеем $M(t, \gamma(t, x, y, \rho, \theta)) = 0$, что эквивалентно ортогональности вектор-функции $\gamma(t, x, y, \rho, \theta)$ и градиента $\nabla M(t, x)$ в $\mathbb{R}^n \forall t \in T$, т. е. вектор $\gamma(t, x, y, \rho, \theta)$ можно разложить по линейно независимой системе векторов N_1, \dots, N_{n-1} — базису линейного подпространства \mathcal{N} . Переменные y , ρ и θ входят в это условие как параметры, они могут быть любыми с учетом их области определения:

$$\begin{aligned} \gamma(t, x, y, \rho, \theta) &= \\ &= q_1^\gamma(t, x, y, \rho, \theta)N_1 + \dots + q_{n-1}^\gamma(t, x, y, \rho, \theta)N_{n-1}. \end{aligned}$$

Переменные y и ρ войдут как параметры и в условия (9) и (10):

$$\begin{aligned} f(t, x, y, \rho) &= \\ &= N_0 + q_1^f(t, x, y, \rho)N_1 + \dots + q_{n-1}^f(t, x, y, \rho)N_{n-1}; \\ \sigma_{*l}(t, x, y, \rho) &= \\ &= q_1^l(t, x, y, \rho)N_1 + \dots + q_{n-1}^l(t, x, y, \rho)N_{n-1}, \\ & \quad l = 1, 2, \dots, s. \end{aligned}$$

Если уравнения для случайных процессов $X(t)$ и $Y(t)$ записать в форме Стратоновича, то для функции $a(t, x, y, \rho)$ — соответствующего коэффициента

в уравнении для случайного процесса $X(t)$ — будем иметь:

$$a(t, x, y, \rho) = N_0 + q_1^a(t, x, y, \rho)N_1 + \dots + q_{n-1}^a(t, x, y, \rho)N_{n-1}.$$

Здесь все коэффициенты при векторах N_1, \dots, N_{n-1} могут быть выбраны произвольно при дополнительных условиях существования решения стохастических дифференциальных уравнений.

Теорема 2. *Для того чтобы траектории стохастической дифференциальной системы, заданной уравнением Ито (11), принадлежали многообразию \mathcal{M} , которое определяется функцией (8): $(t, X(t)) \in \mathcal{M}$, и при этом $(t, \mathbb{E}X(t)), (t, \mathbb{E}[X(t)|Y_0^t]) \in \mathcal{M}$, необходимо и достаточно, чтобы коэффициенты этого уравнения удовлетворяли условиям:*

$$\begin{aligned} f(t, x, y, \rho) &\in \mathcal{N}_0; \\ \sigma_{*l}(t, x, y, \rho) &\in \mathcal{N}, \quad l = 1, 2, \dots, s; \\ \gamma(t, x, y, \rho, \theta) &\in \mathcal{N} \end{aligned}$$

на траекториях случайных процессов $X(t)$ и $Y(t)$ и $\forall \rho \in P, \theta \in \Theta$.

Отметим, что поскольку ограничений на случайный процесс $Y(t)$ в этой работе не накладывается, то зависимость коэффициентов уравнения (11) от вектора измерений, зависимость матричной функции при дифференциале $dV(t)$ от ненаблюдаемого вектора состояния и вектора измерений, а также наличие пуассоновской компоненты в уравнении измерителя не усложняют условий инвариантности, которые сформулированы выше. Однако все перечисленные факторы усложняют алгоритмы нахождения оптимальной оценки $\hat{X}(t)$, в частности формула (3) здесь не применима.

6 Модельный пример

В качестве примера рассмотрим двумерную стохастическую систему вида (1), траектории которой принадлежат гладкому многообразию $\mathcal{M} \subset T \times \mathbb{R}^2$. Это многообразие задается уравнением $M(t, x) = -2x_1 + e^{-t}x_2 = C = const$, т.е. функция $M(t, x)$ имеет вид (8) при $n = 2$, $x = [x_1 \ x_2]^T$ и $\vartheta(t) = [-2 \ e^{-t}]^T$. Таким образом,

$$\begin{aligned} \vartheta_0(t, x) &= \frac{\partial M(t, x)}{\partial t} = -e^{-t}x_2; \\ \vartheta_1(t) &= -2; \quad \vartheta_2(t) = e^{-t}. \end{aligned}$$

Согласно методике, изложенной выше, определим два вектора:

$$\begin{aligned} N_0 &= \left[0 \quad -\frac{\vartheta_0(t, x)}{\vartheta_2(t)} \right]^T = [0 \ x_2]^T; \\ N_1 &= \left[1 \quad -\frac{\vartheta_1(t)}{\vartheta_2(t)} \right]^T = [1 \ 2e^t]^T. \end{aligned}$$

Опишем стохастическую дифференциальную систему, траектории которой принадлежат заданному гладкому многообразию \mathcal{M} . Положим размерность винеровского процесса $s = 1$. Тогда согласно теореме 1 имеем $f(t, x) \in \mathcal{N}_0$, $\sigma(t, x) \in \mathcal{N}$, где \mathcal{N}_0 — линейное многообразие, а \mathcal{N} — линейное подпространство, построенные на векторах N_0 и N_1 , а именно:

$$f(t, x) = N_0 + q^f(t, x)N_1; \quad \sigma(t, x) = q(t, x)N_1,$$

где скалярные функции $q^f(t, x)$ и $q(t, x)$ могут быть выбраны произвольно при дополнительных условиях существования решения стохастического дифференциального уравнения с коэффициентами $f(t, x)$ и $\sigma(t, x)$. Например, зададим эти функции так, чтобы уравнение (1) было линейным:

$$q^f(t, x) = e^{-t}(x_1 - x_2); \quad q(t, x) = e^{-t}.$$

Тогда

$$\begin{aligned} f(t, x) &= [e^{-t}(x_1 - x_2) \quad 2x_1 - x_2]^T; \\ \sigma(t, x) &= [e^{-t} \ 2]^T. \end{aligned}$$

Кроме функций $f(t, x)$ и $\sigma(t, x)$, определяющих уравнение ненаблюдаемого вектора состояния системы, зададим функции $c(t, x)$ и $\zeta(t)$, которые входят в уравнение измерителя (2):

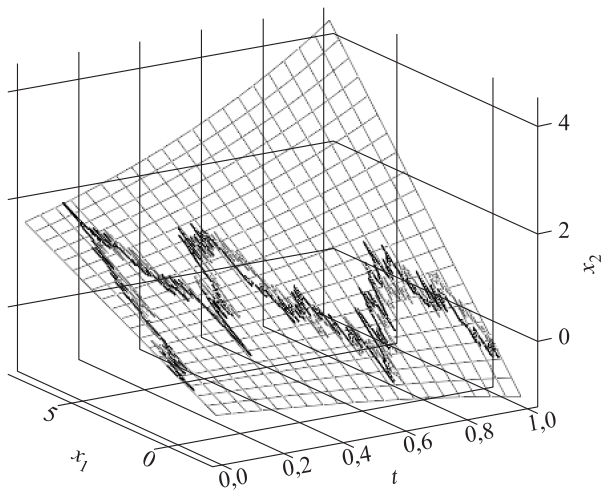
$$c(t, x) = x_1 + x_2; \quad \zeta(t) = 0,05.$$

Следовательно, уравнения (1) и (2) для рассматриваемого примера имеют вид:

$$\begin{aligned} dX(t) &= \begin{bmatrix} e^{-t} & -e^{-t} \\ 2 & -1 \end{bmatrix} X(t) dt + \begin{bmatrix} e^{-t} \\ 2 \end{bmatrix} dW(t), \\ X(t) &= [X_1(t) \ X_2(t)]^T; \\ dY(t) &= (X_1(t) + X_2(t)) dt + 0,05 dV(t). \end{aligned}$$

С помощью выбора функций $q^f(t, x)$ и $q(t, x)$ можно сформировать нелинейные уравнения стохастической системы, однако ограничимся линейным случаем и воспользуемся фильтром Калмана–Бьюси для нахождения оптимальной оценки вектора состояния по критерию минимума среднеквадратической ошибки.

Зададим отрезок времени $T = [0, 1]$ и нулевые начальные данные: $X_1(0) = X_2(0) = Y(0) = 0$. Результаты моделирования траектории случайного



Выборочная траектория случайного процесса $X(t)$ и ее оценка

процесса $X(t)$ методом Эйлера–Маруямы с шагом численного интегрирования $h = 10^{-3}$ показаны на рисунке, на нем же показан результат оценивания с помощью фильтра Калмана–Бьюси (оцениваемая траектория показана черным, а ее оценка — серым, ось времени направлена вправо, ось x_1 — влево, ось x_2 — вверх), а также изображена поверхность $M(t, x) = -2x_1 + e^{-t}x_2 = 0$ (нуль в правой части является следствием нулевых начальных данных для оцениваемой траектории).

Кроме построения одной траектории и нахождения ее оценки была проведена серия вычислительных экспериментов: моделировались ансамбли из 1000 траекторий случайных процессов $X(t)$ и $Y(t)$ методом Эйлера–Маруямы с шагами численного интегрирования h , равными 10^{-2} , 10^{-3} и 10^{-4} . Для каждой траектории $X(t)$ была найдена оптимальная оценка $\hat{X}(t)$ с помощью фильтра Калмана–Бьюси по соответствующей траектории $Y(t)$. На основе этих результатов моделирования были вычислены оценки среднего значения для величин $|M(1, X(1)) - M(0, X(0))|$ и $|M(1, \hat{X}(1)) - M(0, X(0))|$, которые показывают отклонения траектории и ее оценки, полученных численно, от заданного многообразия в момент времени $T = 1$. Результаты представлены в виде таблицы.

Отклонения от заданного многообразия

h	$\mathbb{E} M(1, X(1)) - M(0, X(0)) $	$\mathbb{E} M(1, \hat{X}(1)) - M(0, X(0)) $
10^{-2}	0,006488	0,006821
10^{-3}	0,000668	0,000685
10^{-4}	0,000064	0,000066

Из полученных данных видно, что при уменьшении шага численного интегрирования h в 10 раз средние отклонения уменьшаются почти в 10 раз. Это соответствует первому порядку слабой сходимости метода Эйлера–Маруямы, т. е. отклонение траекторий и их оценок от заданного многообразия вызваны погрешностью при численном решении стохастических дифференциальных уравнений.

7 Заключение

В статье описан класс стохастических дифференциальных систем, для которых одному и тому же многообразию принадлежат не только траектории системы, но и результат решения задачи оптимальной фильтрации по критерию минимума среднеквадратической ошибки оценивания, а именно: приведены необходимые и достаточные условия принадлежности траекторий стохастической системы и их оценок заданному гладкому многообразию. Рассмотрены системы как диффузионного типа, так и диффузионно–скачкообразного типа, т. е. при наличии как винеровских, так и пуассоновских возмущений. Приведен модельный пример линейной стохастической дифференциальной системы, для которой траектории и оценки этих траекторий с помощью фильтра Калмана–Бьюси принадлежат динамическому многообразию. В дальнейшем планируется рассмотреть многообразия меньшей размерности (здесь рассмотрены многообразия, размерность которых на единицу меньше размера оцениваемого вектора состояния), а также обобщить полученные результаты для стохастических дифференциальных систем с изменениями структуры, это даст возможность перейти к задачам фильтрации на кусочно–гладких многообразиях.

Литература

1. Сеницын И. Н. Ортогональные субоптимальные фильтры для нелинейных стохастических систем на многообразиях // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 34–44.
2. Сеницын И. Н. Нормальные и ортогональные субоптимальные фильтры для нелинейных стохастических систем // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 45–50.

- ческих систем на многообразиях // Системы и средства информатики, 2016. Т. 26. № 1. С. 199–226.
3. Синицын И. Н., Синицын В. И., Корепанов Э. Р. Эллипсоидальные субоптимальные фильтры для нелинейных стохастических систем на многообразиях // Системы и средства информатики, 2016. Т. 26. № 2. С. 79–97.
 4. Синицын И. Н., Синицын В. И., Корепанов Э. Р. Модифицированные эллипсоидальные условно-оптимальные фильтры для нелинейных стохастических систем на многообразиях // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 101–111.
 5. Синицын И. Н., Синицын В. И., Сергеев И. В., Корепанов Э. Р. Методы эллипсоидальной фильтрации процессов в нелинейных стохастических системах на многообразиях // Автоматика и телемеханика, 2018. № 1. С. 147–161.
 6. Дубко В. А. Вопросы теории и применения стохастических дифференциальных уравнений. — Владивосток: ДВО АН СССР, 1989. 185 с.
 7. Дубко В. А. Стохастические дифференциальные уравнения. Избранные разделы. — Киев: Логос, 2012. 68 с.
 8. Карачанская Е. В. Случайные процессы с инвариантами. — Хабаровск: ТОГУ, 2014. 148 с.
 9. Карачанская Е. В. Интегральные инварианты стохастических систем и программное управление с вероятностью 1. — Хабаровск: ТОГУ, 2015. 148 с.
 10. Аверина Т. А. Аналитические и численные решения трех систем стохастических дифференциальных уравнений с инвариантами // Аналитические и численные методы моделирования естественно-научных и социальных проблем: Тр. XII Междунар. научн.-технич. конф. — Пенза: ПГУ, 2017. С. 3–8.
 11. Øksendal B., Sulem A. Applied stochastic control of jump diffusions. — Berlin: Springer, 2005. 214 p.
 12. Синицын И. Н. Фильтры Калмана и Пугачева. — М.: Логос, 2007. 776 с.
 13. Bain A., Crisan D. Fundamentals of stochastic filtering. — New York, NY, USA: Springer, 2009. 394 p.
 14. Рыбаков К. А. Статистические методы анализа и фильтрации в непрерывных стохастических системах. — М.: МАИ, 2017. 176 с.
 15. Аверина Т. А., Карачанская Е. В., Рыбаков К. А. Statistical analysis of diffusion systems with invariants // Russ. J. Numer. Anal. M., 2018. Vol. 33. Iss. 1. P. 1–13.

Поступила в редакцию 19.04.18

ON A CLASS OF FILTERING PROBLEMS ON MANIFOLDS

K. A. Rybakov

Moscow Aviation Institute (National Research University), 4 Volokolamskoye Shosse, Moscow 125993, Russian Federation

Abstract: The goal of the paper is to describe stochastic differential systems whose trajectories belong to a smooth manifold as an application to the optimal filtering problem. An additional condition is that not only system trajectories belong to the given manifold, but also the estimation results for these trajectories (solution of the optimal filtering problem with the minimum mean-squared error) belong to this manifold. Diffusion and jump-diffusion systems are considered. These systems can be driven by the Wiener process and the Poisson process. The main result is the conditions on coefficients of the equation for the estimated random process. These conditions are obtained on the basis of the first integral concept for the stochastic differential equation and some of its properties.

Keywords: invariant; estimation; manifold; optimal filtering; random process; stochastic differential system

DOI: 10.14357/19922264190103

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 17-08-00530-a).

References

1. Sinitsyn, I. N. 2016. Ortogonal'nye suboptimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraziyakh [Orthogonal suboptimal filters for nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya* — *Inform. Appl.* 10(1):34–44.
2. Sinitsyn, I. N. 2016. Normal'nye i ortogonal'nye suboptimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraziyakh [Normal and orthogonal conditionally optimal filters for nonlinear stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki* — *Systems and Means of Informatics* 26(1):199–226.
3. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2016. Ellipsoidal'nye suboptimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraziyakh [Ellipsoidal conditionally optimal filters for nonlinear stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki* — *Systems and Means of Informatics* 26(1):199–226.

- tems on manifolds]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(2):79–97.
4. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2017. Modifitsirovannye ellipsoidal'nye uslovno-optimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraznykh [Modified ellipsoidal conditionally optimal filters for nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):101–111.
 5. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, and E. R. Korepanov. 2018. Methods of ellipsoidal filtration in nonlinear stochastic systems on manifolds. *Autom. Rem. Contr.* 79(1):117–127.
 6. Dubko, V. A. 1989. *Voprosy teorii i primeneniya stokhasticheskikh differentsial'nykh uravneniy* [Problems of theory and application of stochastic differential equations]. Vladivostok: Akad. Nauk SSSR. 185 p.
 7. Dubko, V. A. 2012. *Stokhasticheskie differentsial'nye uravneniya. Izbrannye razdely* [Stochastic differential equations. Selected topics]. Kiev: Logos. 68 p.
 8. Karachanskaya, E. V. 2014. *Sluchaynye protsessy s invariantami* [Random process with invariants]. Khabarovsk: Pacific National University. 148 p.
 9. Karachanskaya, E. V. 2015. *Integralnye invarianty stokhasticheskikh sistem i programmnoe upravlenie s veroyatnost'yu 1* [Integral invariants of stochastic systems and program control with probability 1]. Khabarovsk: Pacific National University. 148 p.
 10. Averina, T. A. 2017. Analiticheskie i chislennye resheniya trekh sistem stokhasticheskikh differentsial'nykh uravneniy s invariantami [Analytic and numerical solutions of three systems of stochastic differential equations with invariants]. *12th Scientific and Technical Conference (International) “Analytical and Numerical Methods for Modeling Natural-Scientific and Social Problems” Proceedings*. Penza. 3–8.
 11. Øksendal, B., and A. Sulem. 2005. *Applied stochastic control of jump diffusions*. Berlin: Springer. 214 p.
 12. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. Moscow: Logos. 776 p.
 13. Bain, A., and D. Crisan. 2009. *Fundamentals of stochastic filtering*. New York, NY: Springer. 394 p.
 14. Rybakov, K. A. 2017. *Statisticheskie metody analiza i fil'tratsii v nepreryvnykh stokhasticheskikh sistemakh* [Statistical methods of analysis and filtering for continuous stochastic systems]. Moscow: MAI. 176 p.
 15. Averina, T. A., E. V. Karachanskaya, and K. A. Rybakov. 2018. Statistical analysis of diffusion systems with invariants. *Russ. J. Numer. Anal. M.* 33(1):1–13.

Received April 19, 2018

Contributor

Rybakov Konstantin A. (b. 1979) — Candidate of Sciences (PhD) in physics and mathematics, associate professor, Moscow Aviation Institute (National Research University), 4 Volokolamskoye Shosse, Moscow 125993, Russian Federation; rkoffice@mail.ru

О ЧИСЛЕ МАКСИМАЛЬНЫХ НЕЗАВИСИМЫХ ЭЛЕМЕНТОВ ЧАСТИЧНЫХ ПОРЯДКОВ (СЛУЧАЙ ЦЕПЕЙ)*

Е. В. Дюкова¹, Г. О. Масляков², П. А. Прокофьев³

Аннотация: Рассматривается одна из центральных труднорешаемых задач логического анализа данных — дуализация над произведением частичных порядков. Исследуется важный частный случай, когда каждый порядок является цепью. Если мощность каждой цепи равна двум, то рассматриваемая задача — это построение сокращенной дизъюнктивной нормальной формы монотонной булевой функции, заданной конъюнктивной нормальной формой (КНФ), что эквивалентно перечислению неприводимых покрытий булевой матрицы. При условии, что число строк булевой матрицы по порядку меньше числа столбцов, известна асимптотика типичного числа неприводимых покрытий. В настоящей работе аналогичный результат получен для дуализации над произведением цепей, когда мощность каждой цепи больше двух. Получение подобных асимптотических оценок является технически сложной задачей и необходимо, в частности, для обоснования существования асимптотически оптимальных алгоритмов для задачи монотонной дуализации и различных обобщений этой задачи.

Ключевые слова: задача дуализации; произведение частичных порядков; цепь; покрытие булевой матрицы; упорядоченное покрытие целочисленной матрицы; асимптотически оптимальный алгоритм

DOI: 10.14357/19922264190104

1 Введение

Логический анализ данных основан на решении сложных в вычислительном плане задач, что обусловлено применением дискретного аппарата. Как правило, возникают задачи, которые в теории алгоритмической сложности называют труднорешаемыми. Особой сложностью отличаются перечислительные задачи, в которых требуется найти (перечислить) все решения, при этом число решений растет экспоненциально с ростом размера задачи (размера входа). Одной из центральных перечислительных задач считается дуализация над произведением частичных порядков. Ниже приведена ее формулировка.

Пусть $P = P_1 \times \dots \times P_n$, где P_1, \dots, P_n — конечные частично упорядоченные множества. Считается, что элемент $y = (y_1, \dots, y_n) \in P$ следует за элементом $x = (x_1, \dots, x_n) \in P$, если y_i следует за x_i при $i = 1, 2, \dots, n$. Для обозначения того, что $y \in P$ следует за $x \in P$ и $y \neq x$, далее используется запись $x \prec y$. Пусть $R \subseteq P$, $R^+ = R \cup \{x \in P \mid \exists a \in R, a \prec x\}$. Задача построения двойственного к R множества $I(R)$, состоящего из элементов $a \in P \setminus R^+$, таких что для любого $x \in P \setminus R^+$, $x \neq a$, отношение $a \prec x$ не выполняется, называется дуа-

лизацией над произведением частичных порядков. Элементы множества $I(R)$ называются *максимальными независимыми от R элементами P* .

Важность дуализации обусловлена большим числом приложений, среди которых прежде всего следует выделить логический анализ данных в распознавании (машинное обучение по прецедентам) и поиск ассоциативных правил в базах данных (data mining).

Одним из наиболее востребованных является случай, когда каждое P_i представляет собой цепь, т. е. любые два элемента в P_i сравнимы. Если $P_i = \{0, 1\}$ при $i \in \{1, 2, \dots, n\}$ и $0 \prec 1$, то рассматриваемая задача сводится к построению сокращенной дизъюнктивной нормальной формы монотонной булевой функции от n переменных, заданной КНФ из $|R|$ элементарных дизъюнкций (дуализация монотонной КНФ). Эквивалентной задачей является поиск неприводимых покрытий булевой матрицы из $|R|$ строк и n столбцов (дуализация булевой матрицы).

Теоретические оценки эффективности алгоритмов дуализации базируются на оценке сложности одного шага [1]. Наиболее эффективным считается алгоритм, который имеет полиномиальный от размера входа шаг. Однако полиномиальные алгорит-

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 19-01-00430-а).

¹ Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; Московский государственный университет им. М. В. Ломоносова, edjukova@mail.ru

² Московский государственный университет им. М. В. Ломоносова, gleb-mas@mail.ru

³ Институт машиноведения им. А. А. Благонравова Российской академии наук, p-prok@mail.ru

мы удалось построить лишь для некоторых частных случаев дуализации монотонной КНФ, поэтому требования к алгоритму были ослаблены. Сформировались два основных направления исследований.

Первое направление нацелено на построение так называемых инкрементальных алгоритмов, когда алгоритму разрешено просматривать решения, найденные на предыдущих шагах. При этом оценка сложности шага алгоритма дается для худшего случая (для самого сложного варианта задачи). В [2] построен инкрементальный алгоритм дуализации монотонной КНФ с квазиполиномиальным шагом, определяемым фактически не только размером входа задачи, но и размером ее выхода. В [3, 4] для случая, когда каждое P_i является цепью и $|P_i| \geq 2$, на базе алгоритма, предложенного в [2], построен квазиполиномиальный инкрементальный алгоритм. Подход интересен в основном для теории, поскольку в худшем случае число решений дуализации (размер выхода задачи) растет экспоненциально с ростом размера ее входа.

Второе направление основано на построении асимптотически оптимальных алгоритмов дуализации булевой матрицы (впервые предложено в [5]). В этом случае алгоритму разрешено делать лишние полиномиальные шаги при условии, что их число почти всегда должно быть достаточно мало по сравнению с числом всех решений задачи (числом неприводимых покрытий булевой матрицы). В результате удалось построить алгоритмы дуализации булевой матрицы, эффективные в типичном случае (эффективные для почти всех вариантов задачи). Эти алгоритмы лидируют по скорости счета [6].

Теоретическое обоснование асимптотически оптимальных алгоритмов дуализации булевой матрицы базируется на получении асимптотик типичных значений числа всех неприводимых покрытий и числа неприводимых покрытий типичной длины. Технические основы получения подобных оценок заложены в работах [5, 7].

В [8] рассмотрен случай, когда $P_i = \{0, 1, \dots, k - 1\}$, $k \geq 2$, $i = 1, 2, \dots, n$, и элементы в P_i упорядочены в порядке возрастания. Показано, что задача перечисления $I(R)$ эквивалентна построению некоторого специального подмножества множества неприводимых покрытий булевой матрицы из $|R|$ строк и kn столбцов. Для поиска элементов множества $I(R)$ построен алгоритм RUNC-M+, который представляет собой модификацию асимптотически оптимального алгоритма поиска неприводимых покрытий булевой матрицы RUNC-M из [6].

В настоящей работе обоснована асимптотическая оптимальность алгоритма RUNC-M+ при условии, что $|R|$ на порядок меньше n при $n \rightarrow \infty$. С целью получения требуемых асимптотических

оценок введено понятие упорядоченного тупикового покрытия целочисленной матрицы, являющееся обобщением понятия неприводимого покрытия булевой матрицы. Установлено взаимно однозначное соответствие между множеством упорядоченных тупиковых покрытий матрицы, строками которой являются наборы из R , и множеством $I(R)$. Получена асимптотика типичных значений числа упорядоченных тупиковых покрытий и числа упорядоченных тупиковых покрытий типичной длины целочисленной матрицы в случае большого числа столбцов. Из полученных асимптотических оценок, в частности, следует, что величина $|I(R)|$ (здесь и далее $|A|$ — мощность A) почти всегда при $n \rightarrow \infty$ асимптотически равна числу шагов алгоритма RUNC-M+.

2 Основные понятия

Пусть $P_i = \{0, 1, \dots, k - 1\}$, $k \geq 2$, $i = 1, 2, \dots, n$, и элементы в P_i упорядочены в порядке возрастания. Введем обозначения: L — матрица, в которой n столбцов и элементы принадлежат множеству $\{0, 1, \dots, k - 1\}$, $k \geq 2$; E_k^r , $r \leq n$, — множество всех наборов вида $(\sigma_1, \dots, \sigma_r)$, в которых $\sigma_i \in \{0, 1, \dots, k - 1\}$, $k \geq 2$, при $i = 1, 2, \dots, r$.

Рассмотрим $\sigma \in E_k^r$, $\sigma = (\sigma_1, \dots, \sigma_r)$, $\sigma_i < k - 1$, $i = 1, 2, \dots, r$. Через $Q_i(\sigma)$, $i \in \{1, 2, \dots, r\}$, обозначим множество наборов $(\beta_1, \dots, \beta_r)$ в E_k^r , таких что $\beta_i = \sigma_i + 1$ и $\beta_j \leq \sigma_j$ при $j \in \{1, 2, \dots, r\} \setminus \{i\}$.

Пусть H — набор из r различных столбцов матрицы L . Множество различных строк подматрицы матрицы L , образованной столбцами набора H , можно рассматривать как некоторое подмножество E^H наборов из E_k^r . Набор столбцов H называется *упорядоченным тупиковым σ -покрытием* матрицы L , если выполнены два следующих условия:

- (1) E^H не содержит набор $(\beta_1, \dots, \beta_r) \in E_k^r$, в котором $\beta_j \leq \sigma_j$ при $j \in \{1, 2, \dots, r\}$;
- (2) если $i \in \{1, 2, \dots, r\}$, то E^H содержит хотя бы один набор из $Q_i(\sigma)$.

Если выполнено условие 1, то набор столбцов H называется *упорядоченным σ -покрытием* матрицы L . Если выполнено условие 2, то набор столбцов H называется *упорядоченным σ -совместимым набором столбцов* матрицы L . Упорядоченное (тупиковое) $(0, 0, \dots, 0)$ -покрытие булевой матрицы называется (*неприводимым*) *покрытием*.

Квадратную подматрицу порядка r матрицы L назовем *упорядоченной σ -подматрицей*, если для множества ее различных строк E , рассматриваемого как некоторое подмножество наборов из E_k^r , выполнено $E \cap Q_i(\sigma) \neq \emptyset$ при $i \in \{1, 2, \dots, r\}$.

Обозначим через L_R матрицу, строками которой являются элементы множества R .

Пусть $\sigma = (\sigma_1, \dots, \sigma_n)$ — набор из E_k^n , в котором элемент с номером t , $t \in \{j_1, \dots, j_r\}$, не является максимальным в P_t , а элемент с номером t , $t \notin \{j_1, \dots, j_r\}$, является максимальным в P_t . Очевидным является

Утверждение 1. Набор σ является максимальным независимым от R элементом множества P тогда и только тогда, когда набор столбцов матрицы L_R с номерами j_1, \dots, j_r является упорядоченным тупиковым $(\sigma_{j_1}, \dots, \sigma_{j_r})$ -покрытием.

Таким образом, задача нахождения максимальных независимых от R элементов множества P (перечисления множества $I(R)$) сводится к задаче нахождения упорядоченных тупиковых покрытий матрицы L_R .

Пусть M_{mn}^k — совокупность всех матриц размера $m \times n$ с элементами из $\{0, 1, \dots, k-1\}$, $k \geq 2$. Представляют интерес типичные значения числа тупиковых упорядоченных покрытий и длины тупикового упорядоченного покрытия для матрицы из M_{mn}^k . Выявление типичной ситуации связано с высказыванием типа «для почти всех матриц L из M_{mn}^k при $n \rightarrow \infty$ выполнено свойство Р», причем свойство Р может также иметь предельный характер. Например, если на матрицах из M_{mn}^k заданы две функции $F(L)$ и $G(L)$ с положительными значениями, то можно говорить, что для почти всех матриц L из M_{mn}^k выполнено $F(L) \sim G(L)$ ($F(L)$ асимптотически равно $G(L)$), если существуют две положительные бесконечно малые при $n \rightarrow \infty$ функции $\alpha(n)$ и $\beta(n)$, такие что для всех достаточно больших n имеет место

$$1 - \frac{|M|}{|M_{mn}^k|} \leq \alpha(n),$$

где M — множество матриц L из M_{mn}^k , для которых

$$1 - \beta(n) \leq \frac{F(L)}{G(L)} \leq 1 + \beta(n).$$

3 Асимптотика типичного числа упорядоченных тупиковых покрытий целочисленной матрицы в случае большого числа столбцов

Обозначим: ϕ_d , $d > 0$, — интервал $((1/2) \log_d mn - (1/2) \log_d \log_d mn - \log_d \log_d \log_d n, (1/2) \log_d mn - (1/2) \log_d \log_d mn + \log_d \log_d \log_d n)$; E_{k-1}^r — множество наборов $(\sigma_1, \dots, \sigma_r)$ в E_k^r , таких

что $\sigma_i < k-1$, $i = 1, 2, \dots, r$; $\Pi_r(\sigma) = (\sigma_1 + 1)^{r-1} \dots (\sigma_r + 1)^{r-1}$, $\sigma \in E_{k-1}^r$.

Пусть $L \in M_{mn}^k$, $\sigma \in E_{k-1}^r$. Положим $B(L, \sigma)$ — множество всех упорядоченных тупиковых σ -покрытий матрицы L ; $S(L, \sigma)$ — множество всех упорядоченных σ -подматриц матрицы L ;

$$\Sigma_1(L) = \sum_{r=1}^n \sum_{\sigma \in E_{k-1}^r} |B(L, \sigma)|;$$

$$\Sigma_2(L) = \sum_{r=1}^n \sum_{\sigma \in E_{k-1}^r} |S(L, \sigma)|.$$

Теорема 1. Если $m^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, то для почти всех матриц L из M_{mn}^k при $n \rightarrow \infty$ справедливо

$$\Sigma_1(L) \sim \Sigma_2(L) \sim \sum_{r \notin \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}$$

и длины почти всех упорядоченных тупиковых покрытий матрицы L принадлежат интервалу ϕ_d .

Доказательство теоремы 1 опирается на ряд приводимых ниже лемм 1–9.

Введем обозначения: W_r^n , $r \leq n$, — множество всех наборов вида $\{j_1, \dots, j_r\}$, где $j_q \in \{1, 2, \dots, n\}$ при $q = 1, 2, \dots, r$ и $j_1 < \dots < j_r$; V_r^m , $r \leq m$, — множество всех упорядоченных наборов вида $\{i_1, \dots, i_r\}$, где $i_u \in \{1, 2, \dots, m\}$ при $t = 1, 2, \dots, r$ и $i_{u_1} \neq i_{u_2}$ при $u_1, u_2 = 1, 2, \dots, r$.

Пусть $\sigma \in E_{k-1}^r$, $v \in V_r^m$, $v = \{i_1, \dots, i_r\}$, $w \in W_r^n$, $M_{(v,w,\sigma)}$ — совокупность всех матриц L в M_{mn}^k , таких что в подматрице, образованной столбцами с номерами из w , строка с номером i_t принадлежит $Q_t(\sigma)$ при $t = 1, 2, \dots, r$.

Лемма 1. Если $\sigma \in E_{k-1}^r$, $v \in V_r^m$, $w \in W_r^n$, то

$$|M_{(v,w,\sigma)}| = \Pi_r(\sigma) k^{mn-r^2}.$$

Доказательство. Действительно, строки с номерами из v можно выбрать $\Pi_r(\sigma) k^{r(n-r)}$ способами, остальные строки — $k^{n(m-r)}$ способами. Лемма доказана.

Пусть $M_{(v,w,\sigma)}^*$ — совокупность всех таких матриц L в $M_{(v,w,\sigma)}$, для которых набор столбцов с номерами из w является упорядоченным σ -покрытием и $L \neq M_{(v',w,\sigma)}$ при $v' \neq v$.

Лемма 2. Если $\sigma \in E_{k-1}^r$, $v \in V_r^m$, $w \in W_r^n$, то

$$M_{(v,w,\sigma)}^* \geq \Pi_r(\sigma) k^{mn-r^2} \left(1 - \frac{(r+1)(k-1)^r}{k^r}\right)^{m-r}.$$

Доказательство. Действительно, строки с номерами из v можно выбрать $\Pi_r(\sigma) k^{r(n-r)}$ способами, а остальные строки можно выбрать $(k^n - (r(k-1)^{r-1} + (k-1)^r) k^{n-r})^{m-r}$ способами.

Лемма 3. Если $\sigma' \in E_{k-1}^r$, $\sigma'' \in E_{k-1}^l$, $v_1 \in V_r^m$, $v_2 \in V_l^m$, $w_1 \in W_r^n$, $w_2 \in W_l^n$ и наборы v_1 и v_2 пересекаются по a ($a > 0$) элементам, а наборы w_1 и w_2 пересекаются по b ($b > 0$) элементам, то

$$|M_{(v_1, w_1, \sigma')} \cup M_{(v_2, w_2, \sigma'')}| \leq \Pr_r(\sigma') \Pi_l(\sigma'') (k-1)^b d^{ab} k^{mn-r^2-l^2}.$$

Доказательство. Оценим, сколькими способами можно построить матрицу из $M = M_{(v_1, w_1, \sigma')} \cap M_{(v_2, w_2, \sigma'')}$.

Сначала выберем те элементы, которые расположены на пересечении строк с номерами из v_1 и столбцов с номерами из w_1 (не более чем $\Pr_r(\sigma')$ способов). Затем выберем элементы, которые расположены на пересечении строк с номерами из v_2 и столбцов с номерами из w_2 , учитывая, что ab из них расположены одновременно на пересечении строк с номерами из v_1 и столбцов с номерами из w_1 (не более чем $\Pi_l(\sigma'')(k-1)^{b(1-a)}$ способов). Произвольным способом доопределим остальные элементы матрицы ($k^{mn-r^2-l^2+ab}$ способов). Из сказанного следует требуемая оценка для M .

Лемма 4. Если $m^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, то имеет место

$$\begin{aligned} \sum_{r=1}^n \sum_{\sigma \in E_{k-1}^r} \Pr_r(\sigma) C_n^r C_m^r r! k^{-r^2} &\sim \\ &\sim \sum_{r \in \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pr_r(\sigma) C_n^r C_m^r r! k^{-r^2}, \quad n \rightarrow \infty. \end{aligned}$$

Доказательство. Положим

$$\begin{aligned} p &= \frac{1}{2} \log_d mn - \frac{1}{2} \log_d \log_d mn; \\ q &= \log_d \log_d \log_d n; \\ a_r &= \sum_{\sigma \in E_{k-1}^r} \Pr_r(\sigma) C_n^r C_m^r r! k^{-r^2}; \\ \Pr_r &= \sum_{\sigma \in E_{k-1}^r} \Pr_r(\sigma). \end{aligned}$$

1. Пусть $r \geq p+q-1$. Рассмотрим

$$\begin{aligned} (\sigma_1 + 1)^r \cdots (\sigma_{r+1} + 1)^r &= (\sigma_1 + 1)^{r-1} \cdots \\ \cdots (\sigma_r + 1)^{r-1} (\sigma_1 + 1) \cdots (\sigma_r + 1) &(\sigma_{r+1} + 1)^r. \end{aligned}$$

Отсюда и из того, что число членов в \Pr_{r+1} в $k-1$ раз больше числа членов в сумме \Pr_r , получаем:

$$\frac{\Pr_{r+1}}{\Pr_r} \leq (k-1)^{2r+1}.$$

Следовательно,

$$\begin{aligned} \frac{a_{r+1}}{a_r} &= \frac{(n-r)(m-r)}{r+1} d^{-2r-1} \leq \\ &\leq \frac{mn}{p} d^{-2p-2q+1} \leq_n 4d^{-2q+1}. \end{aligned}$$

2. Пусть $r \leq p-q+1$ и пусть $r_0(k)$ — наименьшее целое r , $r \geq 2$, при котором

$$\frac{\exp(1)r(k-2)^{r-2}}{(k-1)^{r-2}} \leq 1.$$

Покажем, что при $r \geq r_0(k)$ имеет место

$$\Pr_{r-1}(\sigma) \leq \exp(1)(k-1)^{(r-1)(r-2)}.$$

Нетрудно убедиться в справедливости оценки при $k \leq 3$ и любом $r \geq 2$.

Пусть $k > 3$ и для меньших значений k при указанном ограничении на r доказываемая оценка справедлива. Представляя сумму $\Pr_{r-1}(\sigma)$ в виде полинома Ньютона, получаем согласно предположению индукции

$$\begin{aligned} \Pr_{r-1}(\sigma) &= (1 + 2^{r-2} + \cdots + (k-1)^{r-2})^{r-1} \leq \\ &\leq (k-1)^{(r-1)(r-2)} \left(1 + \frac{\exp(1)(k-2)^{r-2}}{(k-1)^{r-2}}\right)^{r-1} \leq \\ &\leq \exp(1)(k-1)^{(r-1)(r-2)}. \end{aligned}$$

Очевидно, что при любом $k \geq 2$ и любом $r \geq 1$

$$(k-1)^{r(r-1)} \leq \Pr_r(\sigma).$$

Таким образом, при $r \geq r_0(k)$ имеем:

$$\begin{aligned} \frac{a_{r-1}}{a_r} &= \frac{\exp(1)(k-1)^{(r-1)(r-2)} r k^{2r-1}}{(k-1)^{r(r-1)} (n-r+1)(m-r+1)} \leq_n \\ &\leq_n \frac{\exp(1)(k-1) p d^{2p-2q+1}}{(n-p)(m-p)} \leq_n \\ &\leq_n \frac{2}{1-1/\alpha} \exp(1)(k-1) d^{-2q+1}. \end{aligned}$$

При $2 \leq r < r_0(k)$, пользуясь тем, что $\Pr_{r-1}(\sigma)/\Pr_r(\sigma) \leq 1$, имеем:

$$\frac{a_{r-1}}{a_r} = \frac{r k^{2r-1}}{(n-p)(m-p)} \leq_n \frac{c}{mn}, \quad c = const, \quad c > 0.$$

Следовательно,

$$\sum_{r \in [p+q, n]} a_r = o \left(\sum_{r \in \phi_d} a_r \right); \quad \sum_{r \in [1, p-q]} a_r = o \left(\sum_{r \in \phi_d} a_r \right).$$

Лемма 5. Если $r, l \leq c \log_d n$, $c < 1$, то имеет место

$$\sum_{b=0}^{\min(r,l)} (k-1)^b d^{lb} C_n^r C_r^b C_{n-r}^{l-b} \leq C_n^r C_n^l (1 + \delta(n)),$$

где $\delta(n) \rightarrow 0$ при $n \rightarrow \infty$.

Доказательство. Обозначим:

$$\lambda_b = \frac{(k-1)^b d^{lb} C_n^r C_r^b C_{n-r}^{l-b}}{C_n^r C_{n-r}^l}.$$

Так как

$$\frac{C_r^b C_{n-r}^{l-b}}{C_{n-r}^l} \leq \left(\frac{rl}{n-r-l} \right)^b$$

и по условию $r, l \leq c \log_d n$, $c < 1$, то

$$\lambda_b \leq \left(\frac{(k-1) \log_d^2 n}{n^{1-c}(1-2 \log_d n/n)} \right)^b$$

и оцениваемая сумма не превосходит $C_n^r C_{n-r}^l (1 + \delta(n))$, где $\delta(n) \rightarrow 0$, $n \rightarrow \infty$. Отсюда, пользуясь неравенством $C_{n-r}^l \leq C_n^l$, получаем утверждение леммы.

Будем считать $M_{mn}^k = \{L\}$ пространством элементарных событий, в котором каждое событие L происходит с вероятностью $1/|M_{mn}^k| = 1/k^{mn}$. Через $M X(L)$ будем обозначать математическое ожидание случайной величины $X(L)$, через $D X(L)$ — дисперсию случайной величины $X(L)$.

Лемма 6 [6]. Пусть для случайных величин $X_1(L)$ и $X_2(L)$, определенных на M_{mn}^k , выполнено $X_1(L) \geq X_2(L) \geq 0$ и при $n \rightarrow \infty$ верно

$$M X_1(L) \sim M X_2(L); \quad \frac{D X_2(L)}{(M X_2(L))^2} \rightarrow 0.$$

Тогда для почти всех матриц L из M_{mn}^k имеет место:

$$X_2(L) \sim X_1(L) \sim M X_2(L), \quad n \rightarrow \infty.$$

Пусть $\sigma \in E_{k-1}^r$, $v \in V_r^m$ и $w \in W_r^n$. На множестве $M_{mn}^k = \{L\}$ рассмотрим случайную величину $\eta_{(v,w)}^\sigma(L)$, равную 1, если L принадлежит $M_{(v,w,\sigma)}^\sigma$, и равную 0 иначе. Оценим вероятность события $\eta_{(v,w)}^\sigma(L) = 1$, обозначаемую далее через $P(\eta_{(v,w)}^\sigma(L) = 1)$. Очевидно, в силу леммы 1

$$P(\eta_{(v,w)}^\sigma(L) = 1) = \frac{|M_{(v,w,\sigma)}^\sigma|}{|M_{mn}^k|} = \Pi_r(\sigma) k^{-r^2}. \quad (1)$$

Положим

$$\eta_1(L) = \sum_{r=1}^n \sum_{v \in V_r^m} \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} \eta_{(v,w)}^\sigma(L);$$

$$\eta_2(L) = \sum_{r \in \phi_d} \sum_{v \in V_r^m} \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} \eta_{(v,w)}^\sigma(L).$$

Нетрудно видеть, что $\eta_1(L) = \Sigma_2(L)$ и $\eta_2(L) \leq \eta_1(L)$. В силу (1)

$$M \eta_1(L) = \sum_{r=1}^n \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2};$$

$$M \eta_2(L) = \sum_{r \in \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}.$$

Из полученных оценок для $M \eta_1(L)$, $M \eta_2(L)$ и леммы 4 сразу следует

Лемма 7. Если $m^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, то имеет место

$$M \eta_1(L) \sim M \eta_2(L) \sim \sum_{r \in \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}, \quad n \rightarrow \infty.$$

Пусть $\sigma \in E_{k-1}^r$, $w \in W_r^n$. На множестве $M_{mn}^k = \{L\}$ рассмотрим случайную величину $\xi_w^\sigma(L)$, равную 1, если L принадлежит $B(L, \sigma)$, и равную 0 иначе. Вероятность события $\xi_w^\sigma(L) = 1$ обозначим через $P(\xi_w^\sigma(L) = 1)$.

Оценим $P(\xi_w^\sigma(L) = 1)$ сверху, пользуясь леммой 1. Нетрудно видеть, что

$$P(\xi_w^\sigma(L) = 1) \leq \sum_{v \in V_r^m} \frac{|M_{(v,w,\sigma)}^\sigma|}{|M_{mn}^k|} = \Pi_r(\sigma) C_m^r r! k^{-r^2}. \quad (2)$$

С другой стороны, в силу леммы 2

$$P(\xi_w^\sigma(L) = 1) \geq \sum_{v \in V_r^m} \frac{|M_{(v,w,\sigma)}^*|}{|M_{mn}^k|} = \Pi_r(\sigma) C_m^r r! k^{-r^2} \times \left(1 - \frac{r(k-1)^{r-1} + (k-1)^r}{k^r} \right)^{m-r}. \quad (3)$$

Положим

$$\xi_1(L) = \sum_{r=1}^n \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} \xi_w^\sigma(L);$$

$$\xi_2(L) = \sum_{r \in \phi_d} \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} \xi_w^\sigma(L).$$

Нетрудно видеть, что $\xi_1(L) = \Sigma_1(L)$ и $\xi_2(L) \leq \xi_1(L)$.

Имеем

$$M \xi_1(L) = \sum_{r=1}^n \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} P(\xi_w^\sigma(L) = 1);$$

$$M \xi_2(L) = \sum_{r \in \phi_d} \sum_{w \in W_r^n} \sum_{\sigma \in E_{k-1}^r} P(\xi_w^\sigma(L) = 1).$$

Следовательно, в силу (2)

$$M \xi_2(L) \leq M \xi_1(L) \leq \sum_{r=1}^n \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}. \quad (4)$$

Пользуясь (3) и тем, что $mr(k-1)^r/k^r \leq \log_d^2 n/n^c$, $c > 0$, при $r \in \phi_d$, получаем:

$$M \xi_1(L) \geq M \xi_2(L) \geq F(n) \sum_{r \in \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}, \quad (5)$$

где $F(n) \rightarrow 1$ при $r \in \phi_d$, $n \rightarrow \infty$.

Из (4), (5) и леммы 4 сразу следует

Лемма 8. Если $m^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, то имеет место

$$M \xi_1(L) \sim M \xi_2(L) \sim \sum_{r \in \phi_d} \sum_{\sigma \in E_{k-1}^r} \Pi_r(\sigma) C_n^r C_m^r r! k^{-r^2}, \quad n \rightarrow \infty.$$

Лемма 9. Если $m^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, то имеет место

$$\frac{D \eta_2(L)}{M \eta_2(L)^2} \rightarrow 0; \quad \frac{D \xi_2(L)}{(M \xi_2(L))^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Доказательство. Нетрудно видеть, что

$$M(\eta_2(L))^2 = \sum_{r,l \in \phi_d} \sum_{\substack{v_1 \in V_r^m, v_2 \in V_l^m, \\ w_1 \in W_r^n, w_2 \in W_l^n}} \sum_{\substack{\sigma' \in E_{k-1}^r, \\ \sigma'' \in E_{k-1}^l}} \frac{|M|}{k^{mn}},$$

где $M = M_{(v_1, w_1, \sigma')} \cap M_{(v_2, w_2, \sigma')}$. Отсюда, пользуясь леммами 3 и 5, получаем

$$M(\eta_2(L))^2 \leq \sum_{r,l \in \phi_d} \sum_{\substack{\sigma' \in E_{k-1}^r, \\ \sigma'' \in E_{k-1}^l}} \Pi_r(\sigma') \Pi_l(\sigma'') \times$$

$$\times \sum_{b=0}^{\min(r,l)} (k-1)^b d^{lb} C_n^r C_r^b C_{n-r}^{l-b} C_m^r r! C_m^l l! k^{-r^2-l^2} \leq$$

$$\leq \sum_{r,l \in \phi_d} \sum_{\substack{\sigma' \in E_{k-1}^r, \\ \sigma'' \in E_{k-1}^l}} \Pi_r(\sigma') \Pi_l(\sigma'') \times$$

$$\times C_n^r C_n^l C_m^r r! C_m^l l! k^{-r^2-l^2} (1 + \delta(n)), \quad (6)$$

где $\delta(n) \rightarrow 0$ при $n \rightarrow \infty$.

С другой стороны, в силу леммы 7 при $n \rightarrow \infty$ имеем:

$$(M \eta_2(L))^2 \sim \sum_{r,l \in \phi_d} \sum_{\substack{\sigma' \in E_{k-1}^r, \\ \sigma'' \in E_{k-1}^l}} \Pi_r(\sigma') \Pi_l(\sigma'') \times$$

$$\times C_n^r C_n^l C_m^r r! C_m^l l! k^{-r^2-l^2}. \quad (7)$$

Из (6), (7) и равенства $D \eta_2(L) = M(\eta_2(L))^2 - (M \eta_2(L))^2$ следует утверждение доказываемой леммы.

Утверждение теоремы 1 следует непосредственно из лемм 6–9.

Следствие 1. Если $m^\alpha \leq n \leq 2^m$, $\alpha > 1$, то для почти всех матриц L из M_{mn}^2 при $n \rightarrow \infty$ справедливо

$$\Sigma_1(L) \sim \Sigma_2(L) \sim \sum_{r \in \phi_2} C_n^r C_m^r r! 2^{-r^2}$$

и длины почти всех неприводимых покрытий матрицы L принадлежат интервалу ϕ_2 .

Приведенные в следствии 1 оценки типичных значений количественных характеристик множества неприводимых покрытий булевой матрицы первоначально получены в [9] с использованием понятий теории нормальных форм булевых функций. Идеино близки результаты более ранней работы [5], в которой предложен асимптотически оптимальный тестовый алгоритм распознавания и при обосновании этого алгоритма получены асимптотики типичных значений числа неприводимых покрытий и числа неприводимых покрытий типичной длины для матрицы из некоторого специального подмножества множества M_{mn}^2 .

Замечание 1. Пусть N_{mn}^k — подмножество в M_{mn}^k , содержащее все матрицы из M_{mn}^k с попарно различными строками. В утверждении теоремы 1 можно заменить M_{mn}^k на N_{mn}^k , так как нетрудно показать, что при $m^2 = o(k^n)$, $n \rightarrow \infty$, почти все матрицы из M_{mn}^k представляют собой матрицы с попарно различными строками.

Замечание 2. Из теоремы 1 и замечания 1 следует, что асимптотически оптимальным является алгоритм дуализации над произведением цепей RUNC-M+, описанный в [8]. Этот алгоритм перечисляет с полиномиальной задержкой некоторое множество упорядоченных совместимых наборов столбцов матрицы L_R , содержащее множество всех упорядоченных тупиковых покрытий этой матрицы. Число шагов алгоритма RUNC-M+ не превосходит $\Sigma_2(L_R)$.

Литература

1. Johnson D. S., Yannakakis M., Papadimitriou C. H. On general all maximal independent sets // *Inform. Process. Lett.*, 1988. Vol. 27. Iss. 3. P. 119–123.
2. Fredman M., Khachiyan L. On the complexity of dualization of monotone disjunctive normal forms // *J. Algorithm.*, 1996. Vol. 21. P. 618–628.
3. Boros E., Elbassioni K., Gurvich V., Khachiyan L., Makino K. Dual-bounded generating problems: All minimal integer solutions for a monotone system of linear inequalities // *SIAM J. Comput.*, 2002. Vol. 31. Iss. 5. P. 1624–1643.
4. Elbassioni K. Algorithms for dualization over products of partially ordered sets // *SIAM J. Discrete Math.*, 2009. Vol. 23. Iss. 1. P. 487–510.
5. Дюкова Е. В. Об асимптотически оптимальном алгоритме построения тупиковых тестов // *Докл. Акад. наук СССР*, 1977. Т. 233. № 4. С. 527–530.
6. Дюкова Е. В., Прокофьев П. А. Об асимптотически оптимальных алгоритмах дуализации // *Ж. вычисл. матем. матем. физ.*, 2015. Т. 55. № 5. С. 895–910.
7. Носков В. Н., Слепян В. А. О числе тупиковых тестов для одного класса таблиц // *Кибернетика*, 1972. № 1. С. 60–65.
8. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О дуализации над произведением частичных порядков // *Машинное обучение и анализ данных*, 2017. Т. 3. № 4. С. 239–249.
9. Дюкова Е. В. О сложности реализации некоторых процедур распознавания // *Ж. вычисл. матем. матем. физ.*, 1987. Т. 27. № 1. С. 114–127.

Поступила в редакцию 15.11.18

ON THE NUMBER OF MAXIMAL INDEPENDENT ELEMENTS OF PARTIALLY ORDERED SETS (THE CASE OF CHAINS)

E. V. Djukova^{1,2}, G. O. Maslyakov², and P. A. Prokofyev³

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 42 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

³Mechanical Engineering Research Institute of the Russian Academy of Sciences, 4 Bardina Str., Moscow 119334, Russian Federation

Abstract: One of the central intractable problems of logical data analysis is considered — dualization over the product of partially ordered sets. The authors investigate an important special case where each order is a chain. If the power of each chain is two, then the problem under consideration is to construct the reduced disjunctive normal form of a monotone Boolean function given by a conjunctive normal form. This is equivalent to enumerating irreducible covers of a Boolean matrix. Provided the growth of the row’s number of the Boolean matrix to be less than the growth of the column’s number, the asymptotic for the typical number of irreducible covers is known. In the present work, a similar result is obtained for the dualization over the product of chains when the power of each chain is more than two. Obtaining such asymptotic estimates is a technically complex task and is necessary, in particular, to justify the existence of asymptotically optimal algorithms for the problem of monotonic dualization and various generalizations of this problem.

Keywords: problem of dualization; product of partially ordered sets; chain; covering of a Boolean matrix; ordered covering of an integer matrix; asymptotically optimal algorithm

DOI: 10.14357/19922264190104

Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (project 19-01-00430-a).

References

1. Johnson, D., M. Yannakakis, and C. Papadimitriou. 1988. On generating all maximal independent sets. *Inform. Process. Lett.* 27(3):119–123.
2. Fredman, M., and L. Khachiyan. 1996. On the complexity of dualization of monotone disjunctive normal forms. *J. Algorithm.* 21(3):618–628.
3. Boros, E., K. Elbassioni, V. Gurvich, L. Khachiyan, and K. Makino. 2002. Dual-bounded generating prob-

- lems: All minimal integer solutions for a monotone system of linear inequalities. *SIAM J. Comput.* 31(5):1624–1643.
4. Elbassioni, K. 2009. Algorithms for dualization over products of partially ordered sets. *SIAM J. Discrete Math.* 23(1):487–510.
 5. Djukova, E. 1977. On an asymptotically optimal algorithm for constructing irredundant tests. *Sov. Math. Dokl.* 18(2):423–426.
 6. Djukova, E., and P. Prokofyev. 2015. Asymptotically optimal dualization algorithms. *Comp. Math. Math. Phys.* 55(5):891–905.
 7. Noskov, V., and V. Slepian. 1972. O chisle tupikovykh testov dlya odnogo klassa tablits [On the number of irredundant tests for a class of tables]. *Kibernetika [Cybernetics]* 1:60–65.
 8. Djukova, E., G. Masliakov, and P. Prokofyev. 2017. O dualizatsii nad proizvedeniem chastichnykh poryadkov [About product over partially ordered sets]. *Mashinnoe obuchenie i analiz dannykh [Machine Learning Data Anal.]* 3(4):239–249.
 9. Dyukova, E. 1987. On the complexity of implementation of some recognition procedures. *Comp. Math. Math. Phys.* 27(1):74–83.

Received November 15, 2018

Contributors

Djukova Elena V. (b. 1945) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 42 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; edjukova@mail.ru

Maslyakov Gleb O. (b. 1996) — student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; gleb-mas@mail.ru

Prokofyev Petr A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, scientist, Mechanical Engineering Research Institute of the Russian Academy of Sciences, 4 Bardina Str., Moscow 119334, Russian Federation; p_prok@mail.ru

АНАЛИЗ УЯЗВИМОСТИ МНОГОПОЛЮСНЫХ СЕТЕЙ ПРИ СТРУКТУРНЫХ ПОВРЕЖДЕНИЯХ

Ю. Е. Малашенко¹, И. А. Назарова², Н. М. Новикова³

Аннотация: Предложен метод получения информативных оценок изменений функциональных возможностей многостокковой сетевой системы после потенциальной аварии. В рамках формализма модели передачи однопродуктового потока изучается множество достижимых векторов-решений, удовлетворяющих стандартным условиям сохранения и ограничениям на потоки по дугам. Для анализа изначального состояния системы для каждой стоковой вершины, отдельно и независимо от остальных, вычисляется максимальный поток. Соответствующий минимальный разрез отделяет эту стоковую вершину от источника. Все дуги найденного минимального разреза модельно удаляются и в поврежденной таким образом сети оцениваются возможности передачи потоков в другие стоковые вершины — рассчитываются предельно-допустимые для вершины потоки, которые сравниваются с их изначальными значениями. Оценки ущерба проводятся для различных минимальных разрезов. Определяется влияние таких структурных повреждений на величины потоков для всех стоковых вершин и строятся агрегированные характеристики подверженности вершины влиянию структурных повреждений.

Ключевые слова: структурная уязвимость сети; подверженность влиянию критических повреждений; многополюсная потоковая модель

DOI: 10.14357/19922264190105

1 Введение

В [1] на примере многопользовательской сетевой модели изучались возможности передачи всем и каждому из пользователей запрашиваемых величин потоков при повреждениях элементов исходной сетевой системы. Последовательно определялись гарантированные оценки величин потоков, передаваемых в сети после аварии. В [2] был предложен метод определения векторов ущерба для пользователей сети при выходе из строя некоторых дуг. Предлагалось одновременно оценивать как уязвимость, так и живучесть сети. Вычислялись величины потоков, которые можно гарантированно доставить пользователям после аварии.

Данное исследование продолжает изучение уязвимости сети, начатое на основе общей методологии [3] в работе [4]. Рассматриваются предельные функциональные возможности сети и их изменения при повреждениях. В начальный момент для каждой стоковой вершины вычисляется предельно допустимый поток и определяется минимальный разрез [5], отделяющий соответствующую вершину от сети. После удаления из сети всех дуг найденного минимального разреза вновь вычисляются максимальные потоки для всех стоковых вершин и сравниваются с их исходными значениями. Для каждого

из полученных разрезов (отделяющих одну из стоковых вершин) подсчитываются величины ущерба для всех остальных потребителей от подобного повреждения сети и проводится оценка его влияния на данного потребителя.

По итогам сформированного множества оценок для каждого потребителя строятся характеристики его подверженности влиянию структурных повреждений рассматриваемого класса (далее — критических).

В работе заложена возможность учета различных критических структурных повреждений и их влияния на функционирование сети. Полученные характеристики позволяют анализировать уязвимость стоковых вершин по отношению к удалению дуг минимальных разрезов содержательно разных типов. При этом в настоящей работе (в отличие от [6]) предлагается совместно рассматривать минимальные разрезы сразу двух типов: по числу входящих в разрез дуг и по сумме значений их пропускной способности.

Работа построена следующим образом. В разд. 2 даются основные определения и вводятся обозначения. В разд. 3 строится набор учитываемых критических структурных повреждений сети. В разд. 4 производится расчет характеристик подверженности стоковых вершин сети их влия-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, malash09@ccas.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, n.novikova@umail.ru

нию (на случай возникновения). В заключении работы обсуждается место проведенных исследований в группе публикаций по близкой тематике.

2 Многополюсная потоковая модель

Сеть передачи единственного вида продукта в многополюсной системе будем описывать ориентированным графом $\overline{\mathcal{G}} = \langle \overline{\mathcal{V}}, \overline{\mathcal{L}} \rangle$ без петель, который определяется множествами вершин (узлов) $\overline{\mathcal{V}} = \{v_1, v_2, \dots, v_N\}$, $|\overline{\mathcal{V}}| = N$, и направленных дуг $\overline{\mathcal{L}} = \{l_{ij} \mid i \in \mathcal{N}, j \in \mathcal{N}, i \neq j\}$, соединяющих вершины, где \mathcal{N} — множество индексов вершин; $l_{ij} = (v_i, v_j)$ — дуга, ведущая из вершины v_i в вершину v_j ; $|\overline{\mathcal{L}}| = L$. Здесь и далее модуль в применении к множеству показывает его мощность (число элементов множества).

Обозначим через $\mathcal{V}_{\mathcal{S}}$ и $\mathcal{V}_{\mathcal{R}}$ множества вершин графа $\overline{\mathcal{G}}$, являющихся соответственно источниками и стоками для потока, который передается по многополюсной сети; $\mathcal{N}_{\mathcal{S}}$ и $\mathcal{N}_{\mathcal{R}}$ — множества индексов вершин-источников и вершин-стоков:

$$\mathcal{V}_{\mathcal{S}} = \{v_i \mid i \in \mathcal{N}_{\mathcal{S}}\}, \quad |\mathcal{V}_{\mathcal{S}}| = S, \quad \mathcal{V}_{\mathcal{S}} \subset \overline{\mathcal{V}}, \\ \mathcal{N}_{\mathcal{S}} \subset \mathcal{N}, \quad S \geq 1;$$

$$\mathcal{V}_{\mathcal{R}} = \{v_i \mid i \in \mathcal{N}_{\mathcal{R}}\}, \quad |\mathcal{V}_{\mathcal{R}}| = K, \quad \mathcal{V}_{\mathcal{R}} \subset \overline{\mathcal{V}}, \\ \mathcal{N}_{\mathcal{R}} \subset \mathcal{N}, \quad K \geq 1;$$

$$\mathcal{V}_{\mathcal{S}} \cap \mathcal{V}_{\mathcal{R}} = \emptyset.$$

Считается, что на дугах графа $\overline{\mathcal{G}}$ заданы веса — значения d_{ij} пропускной способности дуг l_{ij} , определяющие максимально допустимую величину потока по дуге, $d_{ij} \geq 0 \forall l_{ij} \in \overline{\mathcal{L}}$. К графу $\overline{\mathcal{G}}$ добавим:

v_0 — фиктивный источник потока бесконечной мощности — и фиктивные дуги (v_0, v_j) , $j \in \mathcal{N}_{\mathcal{S}}$, соединяющие v_0 с исходными узлами-источниками. Для каждой дуги l_{0j} формально определим верхнее ограничение d_{0j} , которое соответствует величине максимального потока из узла-источника v_j в систему. Будем называть дуги (v_0, v_j) , $j \in \mathcal{N}_{\mathcal{S}}$, дугами-источниками и обозначим их множество как $\hat{\mathcal{L}} = \{l_{0j} \mid j \in \mathcal{N}_{\mathcal{S}}\}$, $d_{0j} \geq 0 \forall l_{0j} \in \hat{\mathcal{L}}$;

v_{N+1} — единственный фиктивный узел-сток бесконечного объема — и фиктивные дуги (v_i, v_{N+1}) , $i \in \mathcal{N}_{\mathcal{R}}$, соединяющие каждый узел-сток с v_{N+1} . Назовем направленные дуги (v_i, v_{N+1}) , $i \in \mathcal{N}_{\mathcal{R}}$, дугами-стоками, или стоковыми дугами, и обозначим их множество через $\tilde{\mathcal{L}} = \{l_{i(N+1)} \mid i \in \mathcal{N}_{\mathcal{R}}\}$, $|\tilde{\mathcal{L}}| = K$. Перенумеруем

стоковые дуги и стоковые вершины натуральными числами от 1 до K и установим взаимно однозначное соответствие $l_k = l_{k(N+1)}$, $k = \overline{1, K}$, $k \in \mathcal{N}_{\mathcal{R}}$. Пусть y_k — величина потока по стоковой дуге l_k из стоковой вершины v_k в вершину v_{N+1} .

Ориентированный граф, который определяется множествами вершин $\mathcal{V} = \overline{\mathcal{V}} \cup \{v_0, v_N\}$ и дуг $\mathcal{L} = \hat{\mathcal{L}} \cup \overline{\mathcal{L}}$, обозначим $\mathcal{G} = \langle \mathcal{V}, \mathcal{L} \rangle$. Для графа \mathcal{G} введем: x_{ij} — поток по дуге l_{ij} , $l_{ij} \in \mathcal{L}$, протекающий в соответствии с ее направлением; \mathcal{N}_j^- — множество индексов узлов, из которых исходят дуги, ведущие в j -й узел, $\mathcal{N}_j^- \subset \mathcal{N} \cup \{0\}$; \mathcal{N}_j^+ — множество индексов узлов, в которые ведут дуги, исходящие из j -го узла, $\mathcal{N}_j^+ \subset \mathcal{N}$. Вектор потоков $\mathbf{x} = \langle x_{0j}, \dots, x_{ij}, \dots, x_{iN} \rangle$ по дугам $l_{ij} \in \mathcal{L}$, где $i \in \mathcal{N} \cup \{0\}$, $j \in \mathcal{N}$, $i \neq j$, должен удовлетворять условию сохранения потока в транзитных узлах

$$\sum_{i \in \mathcal{N}_j^-} x_{ij} = \sum_{i \in \mathcal{N}_j^+} x_{ji}, \quad j \in \mathcal{N}, \quad (1)$$

условию сохранения потока в стоковых узлах

$$\sum_{i \in \mathcal{N}_k^-} x_{ik} = \sum_{i \in \mathcal{N}_k^+} x_{ki} + y_k, \quad y_k \geq 0, \quad k \in \mathcal{N}_{\mathcal{R}}, \quad (2)$$

и ограничению на пропускную способность соответствующих дуг

$$0 \leq x_{ij} \leq d_{ij}, \quad l_{ij} \in \mathcal{L}. \quad (3)$$

Вектор $d = \{d_{ij} \mid d_{ij} \geq 0, l_{ij} \in \mathcal{L}\}$ задает максимально допустимые величины потоков по дугам. Пусть $\mathcal{X}(d) = \{\mathbf{x} \mid \text{выполняются (1)–(3)}\}$ — множество допустимых потоков в сети по всем дугам, кроме стоковых. Вектор $\mathbf{y} = \langle y_1, \dots, y_k, \dots, y_K \rangle$, $k = \overline{1, K}$, покомпонентно определяет из (2) величину потока, который передается по каждой стоковой дуге сети в соответствии с некоторым допустимым потоком $\mathbf{x} \in \mathcal{X}(d)$. Обозначим через $\mathcal{Y}(d)$ множество всех достижимых векторов-потоков по стоковым дугам \mathbf{y} : $\mathcal{Y}(d) = \{\mathbf{y} \mid \exists \mathbf{x} \in \mathcal{X}(d) : \langle \mathbf{x}, \mathbf{y} \rangle \text{ удовлетворяют (1)–(3)}\}$.

3 Критические структурные повреждения

Структурное повреждение w определяется подмножеством дуг сети, пропускная способность которых полагается равной нулю, $w = \{l_{ij} \mid d_{ij} = 0\}$. Пусть $I(w)$ — список индексов этих дуг (номеров пар вершин начала и конца дуги). Структурное повреждение считается *критическим*, если

при удалении соответствующих дуг максимально возможный поток, хотя бы по одной стоковой дуге, оказывается равным нулю (далее — КС-повреждение). Рассмотрим специальные типы КС-повреждений, совпадающие с минимальными разрезами, отделяющими хотя бы одну стоковую вершину от источника.

Сначала в исходном графе сети положим пропускные способности всех дуг сети (кроме стоковых) равными единице и вычислим для некоторой стоковой вершины v_a максимальный поток по исходящей стоковой дуге l_a .

Задача 1. Для выделенной стоковой вершины $v_a \in \mathcal{V}_{\mathcal{R}}$ найти

$$\bar{y}_a^0 = \max_{y \in \mathcal{Y}(d)} y_a$$

при условиях $y_i = 0$ для всех $i \neq a, i \in \mathcal{N}_{\mathcal{R}}$,
 $d_{ij} = 1$ для всех $l_{ij} \in \mathcal{L}$.

Оптимальное решение \bar{y}_a^0 — максимальный поток по дуге l_a — численно равен пропускной способности минимального разреза [5], отделяющего вершину v_a . В данном случае \bar{y}_a^0 равно числу дуг в минимальном разрезе. Обозначим соответствующее подмножество дуг через Q_a , тогда $|Q_a| = \bar{y}_a^0$.

Введем $Q(a) = \{Q_a^1, Q_a^2, \dots, Q_a^{r(a)}\}$ — множество всех минимальных разрезов для вершины v_a . Для каждой стоковой вершины $v_m, m \in \mathcal{N}_{\mathcal{R}}$, решим задачу 1 и найдем все соответствующие множества $Q(m)$. При удалении всех дуг некоторого разреза Q_m^i вершина v_m оказывается отделена от источника, а максимальный поток по стоковой дуге l_m становится равным нулю. Все повреждения $Q(m)$ объединим в одно множество

$$\mathcal{Q} = \bigcup_{m \in \mathcal{N}_{\mathcal{R}}} Q(m).$$

Минимальные разрезы из \mathcal{Q} назовем КС-повреждениями первого типа.

Определение 1. Монопольным режимом передачи потока из фиктивного узла-источника v_0 в фиктивный сток v_{N+1} по стоковой дуге l_a будем называть такие допустимые векторы потоков $x \in \mathcal{X}(d)$, при которых потоки по всем остальным стоковым дугам полагаются равными нулю.

В качестве КС-повреждений второго типа рассмотрим минимальные разрезы из решения следующей задачи.

Задача 2. Для выделенной стоковой вершины $v_a \in \mathcal{V}_{\mathcal{R}}$ найти

$$y_a^0 = \max_{y \in \mathcal{Y}(d)} y_a$$

при дополнительных условиях $y_i = 0$

для всех $i \neq a, i \in \mathcal{N}_{\mathcal{R}}$.

Оптимальное решение y_a^0 — максимальный поток, достижимый при монопольном режиме передачи из v_0 по стоковой дуге l_a в фиктивный узел-сток v_{N+1} (при условии, что потоки по всем остальным стоковым дугам полагаются равными нулю), y_a^0 будем называть МРМ-поток. МРМ-поток соответствует по крайней мере один минимальный разрез (МРМ-разрез).

Обозначим МРМ-разрез — подмножество дуг l_{ij} , суммарная пропускная способность которых равна максимальному потоку y_a^0 и при удалении которых вершина v_a и сток v_0 оказываются в различных связных компонентах графа сети, через \tilde{H}_a , тогда

$$\sum_{(i,j) \in I(\tilde{H}_a)} d_{ij} = y_a^0,$$

где $I(\tilde{H}_a)$ — список индексов дуг, образующих разрез \tilde{H}_a . Одновременное удаление дуг из МРМ-разреза \tilde{H}_a отделяет вершину v_a от сети, а МРМ-поток по стоковой дуге l_a оказывается нулевым. Введем $H(a) = \{H_a^1, H_a^2, \dots, H_a^{n(a)}\}$ — множество всех МРМ-разрезов, соответствующих МРМ-поток y_a^0 .

Последовательно для каждой стоковой вершины $v_m \in \mathcal{V}_{\mathcal{R}}, m \in \mathcal{N}_{\mathcal{R}}$, решим задачу 2 и сформируем вектор МРМ-потоков $y^0 = \langle y_1^0, y_2^0, \dots, y_m^0, \dots, y_K^0 \rangle$, а также все множества $H(m) = \{H_m^1, H_m^2, \dots, H_m^{n(m)}\}$. Все повреждения $H(m)$ объединим в множество \mathcal{H} КС-повреждений второго типа,

$$\mathcal{H} = \bigcup_{m \in \mathcal{N}_{\mathcal{R}}} H(m).$$

Рассмотрим множество $\mathcal{Q} \cup \mathcal{H}$, в нем могут содержаться разрезы с одинаковым набором дуг. Исключим из $\mathcal{Q} \cup \mathcal{H}$ повторяющиеся разрезы и сформируем множество $\mathcal{W} = \{w^1, w^2, \dots, w^T\}$, в котором каждый элемент w^i — минимальный разрез из $\mathcal{Q} \cup \mathcal{H}$, однако $I(w^i)$ отличается от любого $I(w^j), w^j \in \mathcal{W}, i \neq j$. Список дуг (и их индексов), образующих w^i , является уникальным — не совпадает ни с одним другим для $w^j \in \mathcal{W}$.

4 Анализ уязвимости и оценки повреждений

Проанализируем изменения МРМ-потоков при удалении всех дуг, принадлежащих некоторому КС-повреждению из множества \mathcal{W} .

Задача 3. Для выделенной стоковой вершины $v_a \in \mathcal{V}_{\mathfrak{R}}$ найти

$$y_a^0(w') = \max_{y \in \mathcal{Y}(d)} y_a$$

при условиях $y_i(w') = 0$ для всех $i \neq a$, $i \in \mathcal{N}_{\mathfrak{R}}$,
 $d_{ij} = 0 \quad \forall (i, j) \in I(w')$.

Оптимальное решение задачи 3 — значение МРМ-потока по стоковой дуге l_a при удалении из сети всех дуг из выбранного КС-повреждения w' .

Последовательно решим задачу 3 для выбранного КС-повреждения w' и каждой стоковой дуги l_k , $k \in \mathcal{N}_{\mathfrak{R}}$. Для поврежденной сети сформируем вектор МРМ-потоков $\mathbf{y}^0(w') = \langle y_1^0(w'), y_2^0(w'), \dots, y_k^0(w'), \dots, y_K^0(w') \rangle$ и вычислим значения индикатора стоковых дуг, МРМ-потоки по которым равны нулю,

$$\forall k \in \mathcal{N}_{\mathfrak{R}} : \rho_k(w') = \begin{cases} 1, & \text{если } y_k^0(w') = 0; \\ 0, & \text{если } y_k^0(w') > 0. \end{cases}$$

Для КС-повреждения w' и найденных значений подсчитаем число

$$\nu^0(w') = \sum_{k=1}^K \rho_k(w')$$

и долю

$$\rho(w') = \frac{\nu^0(w')}{K}$$

стоковых дуг, МРМ-потоки по которым равны нулю.

Рассмотрим множество стоковых дуг $\mathcal{L}^{\pm}(w')$, МРМ-поток по которым после КС-повреждения w' больше нуля, $\mathcal{L}^{\pm}(w') = \{l_k \mid k \in \mathcal{N}_{\mathfrak{R}}, y_k^0(w') > 0\}$. В разд. 3 при решении задачи 2 были получены значения y_k^0 для МРМ-потоков в исходной неповрежденной сети. Определим ущерб для k -й стоковой дуги из $\mathcal{L}^{\pm}(w')$ от КС-повреждения w' (не важно первого или второго типа) как падение МРМ-потока по ней и дадим оценку относительного ущерба за счет снижения МРМ-потока:

$$\psi_k(w') = \frac{y_k^0 - y_k^0(w')}{y_k^0}, \quad y_k^0 > 0, \quad k \in \mathcal{N}_{\mathfrak{R}}.$$

Для найденных значений $\psi_k(w')$ определим значение медианы $\psi(w')$. Медиана $\psi(w')$ делит множество $\mathcal{L}^{\pm}(w')$ на две части — подмножество $\mathcal{L}^+(w')$, состоящее из стоковых дуг, для которых ущерб $\psi_k(w')$ больше медианы, $\mathcal{L}^+(w') = \{l_k \mid \psi_k(w') > \psi(w'), l_k \in \mathcal{L}^{\pm}(w')\}$, и подмножество $\mathcal{L}^-(w')$, состоящее из стоковых дуг, для которых

ущерб не больше медианы, $\mathcal{L}^-(w') = \{l_k \mid \psi_k(w') \leq \psi(w'), l_k \in \mathcal{L}^{\pm}(w')\}$ (при нечетном числе элементов в $\mathcal{L}^{\pm}(w')$ в качестве медианы выбирается среднее значение, а при четном — полусумма двух средних по порядку величин). Введем индикаторную функцию $\varphi_k(w')$ и вычислим ее значения для дуг из $\mathcal{L}^{\pm}(w')$:

$$\varphi_k(w') = \begin{cases} 0, & \text{если } l_k \in \mathcal{L}^-(w'); \\ 1, & \text{если } l_k \in \mathcal{L}^+(w'). \end{cases}$$

Доля дуг, которые принадлежат подмножеству $\mathcal{L}^+(w')$, составляет

$$\delta(w') = \frac{\sum_{k=1}^K \varphi_k(w')}{K} \leq \frac{1}{2} (1 - \rho(w')).$$

Анализ результатов решения задач 2 и 3, полученных для КС-повреждений, позволяет получить подробную информацию о последствиях любого повреждения сети из рассматриваемого класса структурных повреждений. Действительно, пара чисел — значений параметров $\langle \rho(w'), \psi(w') \rangle$ — в агрегированном виде описывают изменения функциональных возможностей сетевой системы после КС-повреждения w' : $\rho(w')$ — доля нулевых МРМ-потоков; $\psi(w')$ — медианная величина ущерба для потребителей. У некоторых стоковых дуг ущерб оказывается больше значения $\psi(w')$, и их доля составляет не более $(1 - \rho(w'))/2$ от их общего числа K . При четном числе $|\mathcal{L}^{\pm}(w')|$ и несовпадении средних по порядку величин ущерба ($\psi(w') \neq \psi_k(w')$) КС-повреждение w' разбивает все множество стоковых дуг, а также значения соответствующих МРМ-потоков в пропорциях (долях) $\langle \rho(w'), (1 - \rho(w'))/2, (1 - \rho(w'))/2 \rangle$. При нечетном числе разбиение аналогично, но надо использовать величины $\delta(w')$.

МРМ-потоки по стоковым дугам характеризуют возможности сетевой системы до и после повреждения. В целях получения сравнительных оценок уязвимости как подверженности потребителей сети влиянию повреждений проанализируем изменения всех стоковых потоков для каждого КС-повреждения из множества \mathcal{W} . Для произвольной стоковой вершины определим, в каком подмножестве она оказывается после каждого КС-повреждения. Для всех стоковых вершин и всех повреждений из \mathcal{W} вычислим все значения индикаторных функций $\rho_k(w^i)$ и $\varphi_k(w^i)$, $k \in \mathcal{N}_{\mathfrak{R}}$, $i = \overline{1, T}$. Для каждой v_k , $k \in \mathcal{N}_{\mathfrak{R}}$, подсчитаем долю тех КС-повреждений $w^i \in \mathcal{W}$, для которых МРМ-поток по стоковой дуге l_k будет нулевым:

$$\rho_k(\mathcal{W}) = \sum_{w^i \in \mathcal{W}} \frac{\rho_k(w^i)}{T}.$$

Подсчитаем также долю КС-повреждений из \mathcal{W} , при которых ущерб потребителя потока в вершине v_k (относительные потери МРМ-потока по дуге l_k), $k \in \mathcal{N}_{\mathbb{R}}$, больше значения соответствующей медианы:

$$\varphi_k(\mathcal{W}) = \sum_{w^i \in \mathcal{W}} \frac{\varphi_k(w^i)}{T}.$$

Используем полученные результаты для построения стандартной диаграммы, представляющей найденные оценки уязвимости по отношению ко всем учитываемым в модели повреждениям наглядно в виде точек на плоскости, соответствующих паре введенных агрегированных характеристик для каждой стоковой вершины. Это позволит ранжировать потребителей потока сети по подверженности влиянию КС-повреждений заданного класса (\mathcal{W}). Ранжирование проводится по двум критериям ($\rho_k(\mathcal{W}), \varphi_k(\mathcal{W})$), поэтому на диаграмме (см. рисунок) используются именно такие координаты.

Точки диаграммы с рисунка, находящиеся на выпуклой оболочке или «в относительной близости» от нее, соответствуют стоковым вершинам, наиболее уязвимым (в смысле указанных двух критериев) по отношению к КС-повреждениям первого и второго типа. При удалении дуг какого-либо минимального разреза из \mathcal{W} потока в такие стоковые вершины в большем числе случаев оказываются равными нулю и/или потери МРМ-потока оказываются больше медианных значений. Таким образом, найденные стоковые вершины (и стоящие за ними потребители потоков сети) сильнее других подвержены влиянию КС-повреждений первого и второго типа.

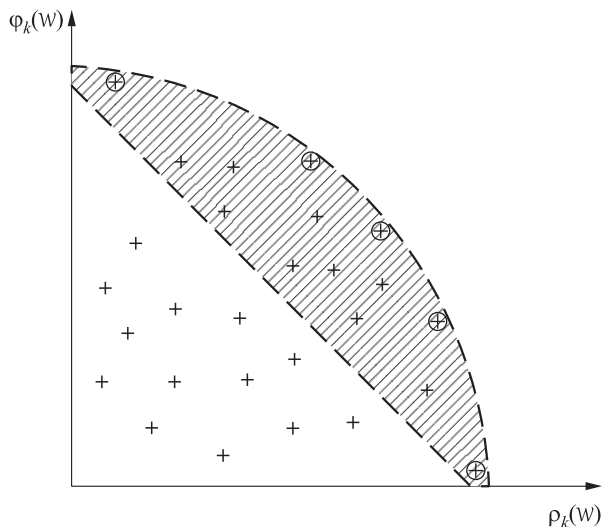


Диаграмма $(\rho_k(\mathcal{W}), \varphi_k(\mathcal{W}))$

5 Заключение

Поиск потребителей сети, наиболее подверженных влиянию ее повреждений, осуществляется в данной работе с помощью эффективных потоковых алгоритмов. Для этого авторы ограничили рассмотрение только такими КС-повреждениями, которые представляются базовыми при решении задачи за потенциального противника, разрушающего сеть с целью лишить возможности получения потока хотя бы одного ее пользователя. Учет всех комбинаций базовых разрезов и/или не только базовых (т. е. минимальных в том или ином смысле) разрезов существенно усложняет поиск. В таком случае рекомендуется подход имитационного моделирования, когда наборы дуг сети, образующие разрез, задаются сценарно исследователем (пример подобной программно реализованной системы для конкретного класса многополюсных сетей приведен в [7]). Также при наличии дополнительной экспертной информации можно ограничиться лишь некоторыми комбинациями различных разрезов. При этом базовые разрезы все равно должны исследоваться в полном объеме.

В данной работе авторы ориентируются на анализ влияния повреждений на конечных потребителей передаваемого потока. Поэтому после того как разрез отобран, в характеристику уязвимости не включаются ни число дуг, образующих разрез, ни их пропускная способность, т. е. все разрезы 1-го и 2-го типа для проводимого анализа равноправны (в отличие от характеристик уязвимости, введенных в [2, 4]). Отметим также, что не принимается во внимание вес разреза в следующем смысле. Для сетевых систем (в частности, с графом-деревом) характерно наличие на одном разрезе сети целой «грозди» потребителей, для каждого из которых этот разрез — минимальный в задаче на максимум потока именно ему. В предложенной статье влияние таких разрезов на каждого потребителя рассматривается лишь один раз, хотя в принципе указанный разрез может возникать чаще (при желании «противника» повредить любому из соседей по сети данного потребителя). Другой подход см. в статье [6].

Поскольку рассматриваемые модели являются однопродуктовыми, используются потоковые методы анализа уязвимости, разработанные авторами ранее. Тем не менее нельзя не упомянуть теоретико-графовые способы оценки структурной уязвимости, которые в контексте настоящей статьи дают некоторые критерии (см., например, в [8]), позволяющие выделить комбинации базовых разрезов для дополнительного включения в множество учитываемых повреждений. Однако при этом необходимо принимать во внимание и мощность

повреждения, чтобы не получить тривиальных результатов.

Предлагаемые методы достаточно общие, чтобы применять их и в сетях более сложной структуры: многопродуктовых, динамических, стохастических. В частности, их можно рекомендовать и для анализа уязвимости потоковых сетей, рассмотренных в [9].

Литература

1. Малащенко Ю. Е., Назарова И. А., Новикова Н. М. Метод анализа функциональной уязвимости потоковых сетевых систем // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 47–54.
2. Малащенко Ю. Е., Назарова И. А., Новикова Н. М. Диаграммы уязвимости потоковых сетевых систем // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 11–18.
3. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971. 384 с.
4. Малащенко Ю. Е., Назарова И. А., Новикова Н. М. Анализ разрезных повреждений в многополюсных сетях // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 35–41.
5. Йенсен П., Барнес Д. Потокное программирование / Пер. с англ. — М.: Радио и связь, 1984. 392 с. (Jensen P. A., Barnes J. W. Network flow programming. — New York, NY, USA: Wiley, 1980. 408 p.)
6. Назарова И. А., Малащенко Ю. Е., Новикова Н. М. Управление топливно-энергетической системой при крупномасштабных повреждениях. III. Критические и стационарные режимы // Изв. РАН. Теория и системы управления, 2018. № 4. С. 107–121.
7. Козлов М. В., Малащенко Ю. Е., Назарова И. А. и др. Управление топливно-энергетической системой при крупномасштабных повреждениях. I. Сетевая модель и программная реализация // Изв. РАН. Теория и системы управления, 2017. № 6. С. 50–73.
8. Кочкаров А. А., Салтагаров М. Б., Эльканова Л. М. Дискретная модель структурного разрушения сложных систем // Проблемы управления, 2007. Вып. 5. С. 21–26.
9. Кузнецов О. П., Жиякова Л. Ю. Двусторонние ресурсные сети — новая потоковая модель // Докл. Академии наук, 2010. Т. 433. Вып. 5. С. 609–612.

Поступила в редакцию 08.11.18

VULNERABILITY ANALYSIS OF MULTIPOLAR NETWORKS AFTER STRUCTURAL DAMAGES

Yu. E. Malashenko, I. A. Nazarova, and N. M. Novikova

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

Abstract: A method to obtain informative estimates of functionality changes in a multistock network system after possible damages has been proposed. Within the framework of the transmission model of single-commodity flow, the authors study the set of achievable flow vectors satisfying the standard conditions of conservation and constraints on the flows along arcs. To analyze the initial state of the system, for each stock arc, the maximum flow has been calculated separately and independently of the flow value across the remaining stock arcs. The corresponding minimal cut separates this stock vertex from the source. All arcs of the found minimal cut are removed model-wisely and in the network which is damaged in this way, the possibilities of transmission flows to other stock vertices have been estimated. The maximum flows for the vertex have been calculated and compared with their initial values. Estimates of losses have been compared for various minimal cuts. Influences of these structural damages have been constructed for all stock vertices. The vertices’ aggregated characteristics of the structural damage effect have been calculated.

Keywords: structural network vulnerability; sensitivity to critical damage; multipolar flow model

DOI: 10.14357/19922264190105

References

1. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2017. Metod analiza funktsional’noy uyazvimosti potokovykh setevykh sistem [Method of the analysis of the functional vulnerability of flow network systems]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):47–54.
2. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2018. Diagrammy uyazvimosti potokovykh setevykh sistem [Diagram of the functional vulnerability of flow network systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(1):11–18.
3. Germeier, Yu. B. 1971. *Vvedenie v teoriyu issledovaniya operatsiy* [An introduction to operations research theory]. Moscow: Nauka. 384 p.

4. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2018. Analiz razreznykh povrezhdeniy v mnogopolyusnykh setyakh [Analysis of cutting damages to multipolar networks]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):35–41.
5. Jensen, P. A., and J. W. Barnes. 1980. *Network flow programming*. New York, NY: Wiley. 408 p.
6. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2018. Fuel and energy system control at large-scale damages. III. Emergency and stationary modes. *J. Comput. Sys. Sc. Int.* 57(4):581–593.
7. Kozlov, M. V., Yu. E. Malashenko, I. A. Nazarova, *et al.* 2017. Fuel and energy system control at large-scale damages. I. Network model and software implementation. *J. Comput. Sys. Sc. Int.* 56(6):945–968.
8. Kochkarov, A. A., M. B. Salpagarov, and L. M. Elkanova. 2007. Diskretnaya model' strukturnogo razrusheniya slozhnykh sistem [A discrete model of compound systems destruction]. *Problemy upravleniya [Control Sciences]* 5:21–26.
9. Kuznetsov, O. P., and L. Y. Zhilyakova. 2010. Bidirectional resource networks: A new flow model. *Dokl. Math.* 82(1):643–646.

Received November 8, 2018

Contributors

Malashenko Yuri E. (b. 1946) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

Nazarova Irina A. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

Novikova Natalya M. (b. 1953) — Doctor of Science in physics and mathematics, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; n_novikova@umail.ru

КЛАССИФИКАЦИЯ ФИЗИЧЕСКОЙ АКТИВНОСТИ ЧЕЛОВЕКА С ПОМОЩЬЮ ЛОКАЛЬНЫХ АППРОКСИМИРУЮЩИХ МОДЕЛЕЙ*

Д. А. Анিকেев¹, Г. О. Пенкин², В. В. Стрижов³

Аннотация: Исследуется проблема классификации временных рядов акселерометра мобильного телефона. Классу физической активности соответствует сегмент временного ряда. Сегменту сопоставляется его признаковое описание. Оно порождается аппроксимирующим сплайном. Элементами вектора признаков являются коэффициенты при базисных функциях сплайнов. Вычислительный эксперимент находит оптимальные параметры аппроксимации и параметры модели классификации согласно максимуму правдоподобия логистической модели классификации.

Ключевые слова: временные ряды; классификация; сплайн; локальная аппроксимация; признаковое пространство

DOI: 10.14357/19922264190106

1 Введение

Цель работы заключается в решении задачи классификации временных рядов сложной структуры [1, 2], для которых признаковое описание не задано явно. Такие временные ряды встречаются в задачах классификации звуковых сигналов [3], данных с акселерометра [4, 5], построения прогнозных финансовых моделей [6]. В [7, 8] дается обзор методов анализа временных рядов за последнее десятилетие. Задача построения признакового пространства требует выбора адекватной гипотезы порождения временного ряда. Многие из исследуемых временных рядов описываются моделью авторегрессии [9, 10] или моделью анализа сингулярного спектра.

В данной работе изучается классификация сегментов временных рядов по их локальному описанию. Под локальным описанием будем понимать параметры сплайна, аппроксимирующего сегмент ряда [11–13].

Предполагается, что класс физической активности описывается уникальной аппроксимирующей кривой, моделью, которая задана суперпозицией параметрических функций. В [2, 5] показано, что использование параметров аппроксимирующих моделей в качестве признакового

описания сегмента дает ожидаемое качество классификации физической активности. При этом остается открытой проблема использования значения параметров локальных аппроксимирующих моделей. Точность классификации зависит от этих параметров. Требуется найти оптимальные параметры локальных моделей.

Построение моделей, использующих в качестве признакового описания параметры локальных моделей, рассмотрено в [14].

В данной работе для решения задачи классификации временных рядов поставлена задача поиска оптимальной локальной аппроксимирующей модели. Сегмент временного ряда аппроксимируется сплайном (алгебраическим, сглаживающим, адаптивным регрессионным или сплайном с динамическими узлами). При помощи логистической регрессии строится отображение из вектора признаков, состоящего из коэффициентов сплайна, в метку класса.

В качестве прикладной задачи рассматривается классификация физической активности человека по данным с акселерометра. В вычислительном эксперименте сравнивается точность классификации в пространствах признаков, построенных различными локальными моделями, найдены оптимальные параметры таких моделей.

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 17-07-1574 и 19-07-1155) и Правительства Российской Федерации (соглашение № 05.Y09.21.0018).

¹ Московский физико-технический институт, dmitriy.anikeev@phystech.edu

² Московский физико-технический институт, penkin.gr@gmail.com

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

2 Постановка задачи классификации сегментов временного ряда

Поставим задачу построения признакового пространства, описывающего сегменты временных рядов с целью их классификации. Задачу сформулируем в терминах построения суперпозиции функций. Задан временной ряд, полученный в результате измерений акселерометра. Для построения выборки \mathcal{D} ряд разбивается на сегменты. Сегмент временного ряда — конечная последовательность значений $\mathbf{x}(t) = [x_1, \dots, x_p]^T$ в моменты времени $\mathbf{t} = [t_1, \dots, t_p]^T$. Задана выборка $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, где $y \in \mathbb{Y}$ — метка класса из конечного множества. Требуется найти оптимальный классификатор $f(\mathbf{x}_i)$ из условия

$$\hat{f} = \arg \min_f CV(f, \mathcal{D}),$$

где

$$CV(f, \mathcal{D}) = \frac{1}{R} \sum_{r=1}^R [f(\mathbf{x}_i) \neq y_i]$$

при $\mathbf{x}_i \in \mathcal{D} \setminus \mathcal{D}_r$.

Выборка \mathcal{D} случайно разбивается R раз на контрольную \mathcal{D}_r и тестовую $\mathcal{D} \setminus \mathcal{D}_r$. Функцией ошибки выступает число неверно классифицированных объектов \mathbf{x}_i на тестовой выборке. Предлагается представить f в виде суперпозиции функций $f = g(a(t))$. Параметризованное отображение

$$a : (\mathbf{t}, \mathbf{b}) \mapsto \mathbf{x}$$

отображает сегмент временного ряда $\mathbf{x} \in \mathbf{X}$ в пространство параметров аппроксимирующего сплайна, а

$$g : (\mathbf{b}, \mathbf{w}) \mapsto y$$

является логистической регрессией и отображает признаковое описание в метку класса. Вектор \mathbf{b} назовем вектором параметров аппроксимирующей модели, а \mathbf{w} — вектором параметров классификатора g . Оптимальные параметры (\mathbf{b}, \mathbf{w}) находятся минимизацией ошибок:

$$\begin{aligned} \hat{\mathbf{b}} &= \arg \min_{\mathbf{b}} \|a(\mathbf{t}, \mathbf{b}) - \mathbf{x}\|_2^2; \\ \hat{\mathbf{w}} &= \arg \min_{\mathbf{w}} CV(f(a, \hat{\mathbf{b}}, \mathbf{w}), \mathcal{D}), \end{aligned} \quad (1)$$

где a — сплайн, аппроксимирующий сегмент временного ряда, а $\|\cdot\|_2^2$ — квадратичная ошибка аппроксимации.

3 Выбор локальных аппроксимирующих моделей

Рассмотрим различные методы построения признакового пространства. Зафиксируем сегмент $\mathbf{x} = [x_1, \dots, x_p]^T$ временного ряда, порожденный измерениями, получаемыми с носимого устройства в моменты времени t_1, \dots, t_p .

3.1 Алгебраический сплайн

Разобьем сегмент $[t_1, t_p]$ на K равных отрезков. Зададим $\{h_1(t), \dots, h_N(t)\}$ — базисные функции сплайна. Гладким сплайном называется функция $a(t)$. Она непрерывна и имеет непрерывные производные вплоть до порядка, называемого гладкостью сплайна, причем на каждом из K отрезков

$$a_k(t) = \sum_{n=1}^N \theta_{k,n} h_n \left(t - \frac{t_k - t_{k-1}}{2} \right), \quad k \in \{1, \dots, K\}, \quad (2)$$

где коэффициенты $\theta_{k,n}$ при базисных функциях $\theta_{k,n} \in \mathbb{R}$.

Оптимальный с точки зрения аппроксимации сплайн ищется из условия (1). В разновидностях сплайнов, аппроксимирующих функцию, накладываются дополнительные условия на функцию. Матрица $\Theta_{K \times N}$ векторизуется, образуя элемент признакового пространства. Для алгебраического сплайна порядка N базисными функциями являются мономы $h_n = t^n$, $n \in \{1, \dots, N\}$.

В качестве параметров модели выступают

$$\mathbf{b} = [\Theta_{K \times N}, N, K]. \quad (3)$$

По каждому сегменту строится $K \times (N + 1)$ параметров. В данной работе рассматриваются только квадратичный и кубический сплайны.

3.2 Сглаживающий сплайн

Сглаживающий сплайн минимизирует на отрезке $[t_{\min}, t_{\max}]$ одновременно квадратичную невязку и производную $a(t)$ порядка m . В данной работе используется вторая производная и модель вида (2). Таким образом, минимизируется линейная комбинация

$$(1-\lambda) \sum_{k=1}^K (x(t_k) - a(t_k))^2 + \lambda \int_{t_{\min}}^{t_{\max}} (D^2 f(t))^2 dt \rightarrow \min_{\theta}$$

где λ — параметр гладкости; N — число базисных функций; K — число узлов. В качестве базисных

функций выступают мономы не выше заданной степени. Параметрами аппроксимирующей модели являются

$$\mathbf{b} = [\Theta_{K \times N}, \lambda, K, N]. \quad (4)$$

3.3 Адаптивные регрессионные сплайны

В одномерном случае адаптивные регрессионные сплайны (MARS) выражаются через кусочно-линейные базисные функции:

$$(\tau - t)_+ = \begin{cases} \tau - t, & \text{если } \tau > t; \\ 0 & \text{иначе;} \end{cases}$$

$$(t - \tau)_+ = \begin{cases} t - \tau, & \text{если } \tau < t; \\ 0 & \text{иначе.} \end{cases}$$

Используемые для аппроксимации базисные функции имеют вид:

$$h_j(t) = \prod_{k=1}^{K_j} \left((-1)^{s(j)} (t - \tau_{k_j}) \right)_+,$$

где K_j — общее число усеченных линейных функций в m -й базисной функции. На каждом из M шагов алгоритма в множество базисных функций (изначально содержащем одну) добавляется функция вида

$$h_m(t) = \hat{C} h_j(t) (\tau_k - t)_+ + \hat{C}' h_j(t) (t - \tau_k)_+$$

такая, что ошибка (1) минимальна. Таким образом, модель имеет вид:

$$a(t) = c_0 + \sum_{j=1}^M c_j h_j(t) + \varepsilon, \quad (5)$$

где ε — невязка.

В результате добавления модель приобретает избыточное число базисных функций. Необходимо удалить из множества базисных функций те, удаление которых вносит наименьший вклад в увеличение ошибки (1). Параметрами данной модели выступают

$$\mathbf{b} = [\mathbf{H}, M]. \quad (6)$$

Оптимальное число базовых функций

$$\hat{M} = \arg \min_{M \in \{1, \dots, M_{\max}\}} \text{GCV}(M)$$

определяется с помощью критерия обобщенного скользящего контроля

$$\text{GCV}(M) = \frac{1}{R} \sum_{i=1}^R \frac{(x(\tau_i) - a_M(\tau_i))^2}{(1 - C(M)/R)^2},$$

где R — размер выборки; $C(M)$ — оценка штрафов модели:

$$C(M) = \text{trace} \left(\mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \right) + 1,$$

в которой \mathbf{H} — матрица с элементами $h_j(\tau_i)$.

3.4 Сплайны с динамическими узлами

Пусть временной сегмент $[t_1, t_p]$ разбит на несколько отрезков $[t_1, t_{n_1}]$, $[t_{n_1}, t_{n_2}]$, \dots , $[t_{n_{K-1}}, t_p]$. Рассмотрим аппроксимацию сегмента временного ряда на каждом отрезке $[t_{n_{k-1}}, t_{n_k}]$ с помощью полинома степени N . Тогда $K \times (N + 1)$ коэффициентов матрицы $\Theta_{K \times N}$ образуют элемент признакового пространства. Предлагается определять концы отрезков последовательно, используя критерий Фишера. Метод заключается в последовательном определении t_{n_k} ($t_{n_0} = t_1$). Положим текущую длину отрезка $l = 1$. Построим полином $y'_k(t)$ соответствующей степени, минимизирующий RSS (regression sum of squares) на отрезке $[t_{n_k}, t_{n_k+l}]$. Полученные значения $x(t)$ и $y'_k(t)$ в целых точках на этом отрезке образуют две выборки X и Y . Отношение дисперсий

$$F = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2}$$

определяет схожесть дисперсий данных выборок. Если дисперсии выборок близки, увеличиваем число l на 1 (кривая хорошо аппроксимирует временной ряд). Формально определяем критерий остановки с помощью p -значения. Если $p(F)$ ниже уровня значимости α — определяем $n_{k+1} = n_k + l$. Продолжим получать последовательность узлов t_{n_k} , пока сегмент не закончится. Число $\alpha \in [0; 0,5]$ будем интерпретировать как параметр метамоделли.

Поиск аппроксимирующего сплайна сведем к задаче условной минимизации среднеквадратичного отклонения (1) при соответствующих условиях гладкости функции

$$\frac{\partial^j y_{k+1}(t_{n_k})}{\partial t^j} = \frac{\partial^j y_k(t_{n_k})}{\partial t^j} \text{ для всех } k < K, \quad n < N.$$

Также рассматривается аппроксимация при условиях равенства только значений сплайна на концах смежных отрезков (кусочно-гладкая). Результаты аппроксимации при параметре $\alpha = 0,07$ и различных степенях полинома показаны на рис. 1.

Критерий Фишера устанавливает разное число узлов на разных сегментах. Способ построения признакового пространства в таком случае заключается

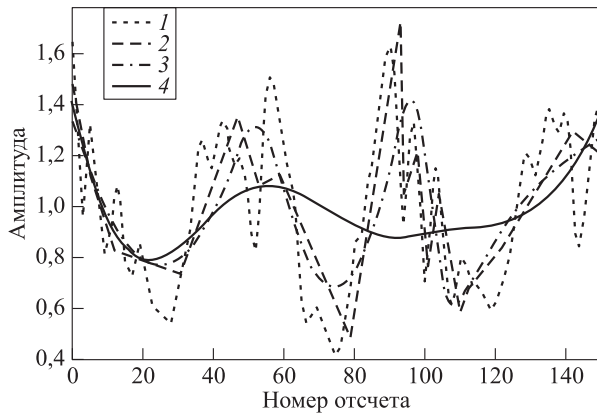


Рис. 1 Аппроксимация сплайном с динамическими узлами: 1 — без аппроксимации; 2–4 — степени 1–3

в дополнении всех векторов нулевыми компонентами до вектора одинаковой размерности. В качестве альтернативного метода рассмотрим следующую перестановку компонент получившегося вектора.

Назовем $\|y - g\|_2^2$ относительным изменением импульса на сегменте, где $g = (0; 0; 9,8)$ м/с² — ускорение свободного падения. Отсортируем отрезки по убыванию относительных изменений импульса, запишем в соответствующем порядке параметры сплайнов и дополним все векторы нулевыми компонентами до векторов одинаковой размерности. Поскольку производится классификация данных с акселерометра, такая сортировка позволяет соотносить между собой отрезки времени с наибольшими изменениями импульса датчика для различных сегментов.

Параметрами модели в этом случае являются

$$b = [\Theta_{K \times N}, \alpha, K, N]. \quad (7)$$

3.5 Альтернативные локальные модели

Рассмотрим некоторые альтернативные методы построения признакового пространства.

Дискретное преобразование Фурье. В качестве признакового описания временного ряда берутся коэффициенты прямого преобразования Фурье:

$$\theta(x) = [\theta_1, \dots, \theta_{2p}];$$

$$\theta_{2k-1} + i\theta_{2k} = \sum_{j=1}^p x^j \exp\left(-\frac{2\pi i}{t}kj\right).$$

Тогда обратное преобразование задает аппроксимацию:

$$a(x^k) = \frac{1}{t} \sum_{j=1}^t (\theta_{2j-1} + i\theta_{2j}) \exp\left(-\frac{2\pi i}{t}kj\right).$$

Коэффициенты преобразования, образующие признаковое пространство, отыскиваются с помощью линейной регрессии временного ряда на столбцы матрицы Фурье. Применение этого метода для классификации рассматривается в статье [2].

Анализ сингулярного спектра. Элементы признакового пространства — сингулярные числа траекторной матрицы X_N , отвечающие за величины различных частот спектра сегмента

$$X^T X = V \Lambda V^T, \quad \Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}. \quad (8)$$

Тогда собственные числа матрицы $X^T X$ образуют вектор признаков Λ . Значение N назначается равным математическому ожиданию длины сегмента по критерию, описанному в разделе сегментации. Результаты классификации в таком пространстве описаны в [5].

4 Сегментация временных рядов

Исходные данные представляют собой размеченные несегментированные отрезки произвольной длины. Для построения выборки необходимо их сегментировать: разбить временной ряд x_1, \dots, x_L на сегменты вида $[x_1, \dots, x_p]^T$ длины p . Рассматриваются два варианта разбиения квазипериодических рядов на синфазные сегменты. Алгоритмы выделения периодов рассмотрены в [15, 16].

Выделение главных компонент заключается в разложении временного ряда $x = \hat{x} + \tilde{x} + \epsilon$, где \hat{x} — тренд; \tilde{x} — периодическая часть; ϵ — вектор невязок. Построим траекторную матрицу из элементов временного ряда x_1, \dots, x_L :

$$X = \begin{bmatrix} x_1 & \dots & x_p \\ \vdots & \ddots & \vdots \\ x_{L-p+1} & \dots & x_L \end{bmatrix}.$$

По первым собственным значениям матрицы (8)

$$X^T X = V \Lambda V^T, \quad \Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\},$$

восстановим периодическую зависимость $\tilde{x} = x_1 + \dots + x_d$, где $x_i = \sqrt{\lambda_i} v_i (X v_i)^T$. Искомый период ряда $\hat{p} = \arg \min_p \|x - \tilde{x} - \hat{x}\|_2^2$.

Метод поиска локального максимума временного ряда. Для заданных чисел s и p в каждом отрезке ищется индекс начала сегмента согласно условию $\hat{s} = \arg \max_{j < s} x_j$ и в качестве сегмента берется $[x_{\hat{s}}, \dots, x_{\hat{s}+p}]^T$. Для получения синфазных отрезков квазипериодических рядов предлагается использовать параметр s , больший длины периода. Данный метод неустойчив к выбросам, но требует малой вычислительной мощности. Сложность алгоритма — линейная по длине исходного сегмента.

5 Вычислительный эксперимент

В качестве вычислительного эксперимента выбрана задача классификации типов физической активности человека по данным с акселерометра.

Данные WISDM. Данные представляют собой размеченные несегментированные трехмерные временные ряды, полученные с датчиков акселерометра. Частота измерений составляла 20 Гц. В выборке представлены классы sitting (225), standing (275), walking (2890), jogging (1631), upstairs (801) и downstairs (657). Дробление производилось на временные сегменты длины 50. Для выравнивания сегментов использовался поиск максимального

значения за следующие несколько отсчетов, который задает точку начала сегмента.

Сегментация временного ряда проводилась методом поиска максимума с параметрами $m = 20, k = 100$. В качестве классифицирующего алгоритма использована логистическая регрессия из библиотеки sklearn. Результаты классификации показаны на рис. 2.

Точность классификации для квадратичной аппроксимации составляет 0,25, что ниже результатов остальных исследованных методов аппроксимации.

Зависимость точности модели от параметра p для сглаживающего сплайна можно наблюдать на

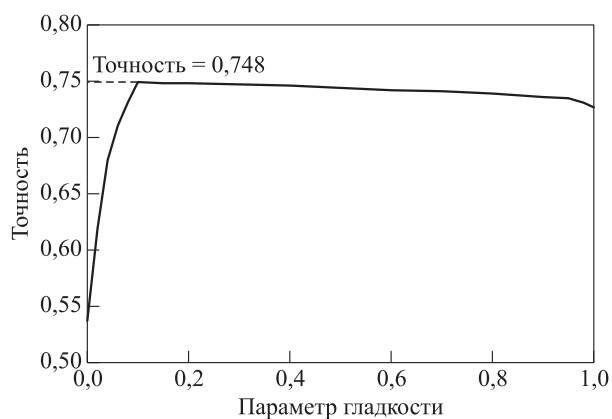
Downstairs	0,362	0,078	0,011	0,073	0,181	0,295
Jogging	0,058	0,725	0,013	0,071	0,053	0,081
Sitting	0,0	0,018	0,88	0,084	0,0	0,018
Standing	0,007	0,0	0,007	0,971	0,007	0,007
Upstairs	0,146	0,081	0,005	0,111	0,356	0,301
Walking	0,077	0,02	0,0	0,04	0,054	0,808
	Downstairs	Jogging	Sitting	Standing	Upstairs	Walking

(а)

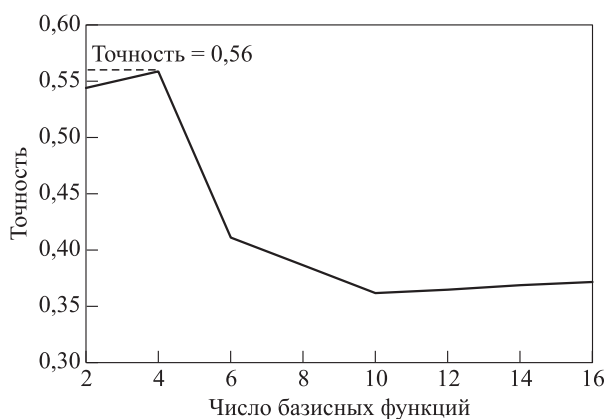
Downstairs	0,012	0,079	0,274	0,21	0,004	0,421
Jogging	0,008	0,06	0,224	0,17	0,009	0,529
Sitting	0,01	0,079	0,264	0,184	0,033	0,431
Standing	0,023	0,049	0,256	0,194	0,017	0,461
Upstairs	0,018	0,08	0,269	0,201	0,003	0,428
Walking	0,011	0,058	0,217	0,169	0,006	0,538
	Downstairs	Jogging	Sitting	Standing	Upstairs	Walking

(б)

Рис. 2 Матрица ковариации для кубического (а) и квадратичного (б) сплайна



(а)



(б)

Рис. 3 Зависимость точности классификации от параметра сглаживания (а) и от числа базисных функций (б)

Сравнение моделей локальной аппроксимации WISDM и USC-HAD

Аппроксимирующая модель	Оптимальный параметр	Точность
Выборка WISDM		
Quadratic (2)	(3)	0,2540
Cubic (2)	(3)	0,7305
Smoothing spline (2)	$\lambda = 0,25$ (4)	0,748
MARS (5)	$M = 4$ (6)	0,56
Выборка USC-HAD		
Quadratic (2)	(3)	0,587
Cubic (2)	(3)	0,926
Smoothing spline (2)	$\lambda = 0,4$ (4)	0,960
MARS (5)	$M = 4$ (6)	0,835
Динамические узлы Sort (2)	$K = 2, \alpha = 0,015$ (7)	0,935
Динамические узлы Unsorted (2)	$K = 2, \alpha = 0,01$ (7)	0,926

рис. 3, а. Как видно из данного графика, максимальная точность достигается при параметре сглаживания $p = 0,25$.

Для построения адаптивного регрессионного сплайна была выбрана библиотека ARESLAB. Для того чтобы размерность пространства параметров была одинакова для всех сегментов, аппроксимация происходила без фазы назад. Параметр такой модели — число базисных функций. График зависимости точности классификации от параметра метамодели показан на рис. 3, б. Рассмотрены различные отображения h временных рядов в признаковое пространство. Данные отображения можно сравнить с тождественным, т. е. таким, при котором в качестве матрицы плана выступает матрица значений временного ряда в фиксированные моменты времени. Точность и оптимальные параметры моделей представлены для сравнения в таблице.

Данные USC-HAD. Данные представляют собой размеченные временные ряды различной длины с датчиков акселерометра. Даны проекции ускорения на три оси с частотой 100 Гц. В выборке представлены классы walking forward, jumping up, walking left, sitting, walking right, standing, walking upstairs, sleeping, walking downstairs, elevator up, running forward, elevator down.

Рассмотрим теперь аппроксимацию гладкими сплайнами с динамическими узлами и кусочно-гладкими, о которых шла речь в подразд. 3.4. После аппроксимации ими сегментов имеется возможность менять порядок получаемых признаков. Можно задать его пропорциональным длине кривой между двумя соседними узлами или же пропорционально относительному изменению импульса. Также можно оставить его без изменения.

Все вышеизложенные варианты дают всего 6 различных вариантов (с точностью до степени аппроксимирующего многочлена) задания призна-

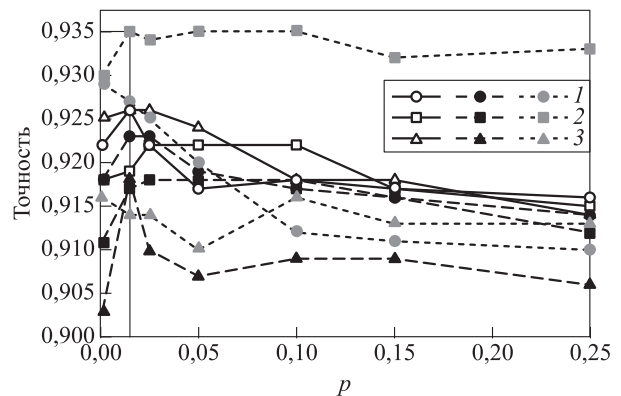


Рис. 4 Точность классификации для сплайнов с динамическими узлами: пустые значки — неотсортированные; черные — неотсортированные; серые значки — отсортированные по энергии; 1 — линейный метод; 2 — квадратичный; 3 — кубический метод

кового пространства. На рис. 4 изображено качество классификации в зависимости от значения параметра p .

Чтобы результаты классификации для этой выборки можно было сравнить с предыдущей, выделим также 6 классов, по своей природе совпадающих или близких к WISDM. Возьмем классы 1, 6, 7, 8, 9 и 10. Будем их для краткости называть walking, running, jumping, sitting, standing и sleeping. Сегментация временного ряда также производилась методом поиска максимума с параметрами $m = 50, k = 200$. Результаты сравнения моделей представлены в таблице.

На рис. 5 показана зависимость точности классификации от параметра сглаживания (рис. 5, а) и от числа базисных функций (рис. 5, б). Зависимость погрешности аппроксимации SMAPE от числа базисных функций для регрессионного сплайна показана на рис. 6.

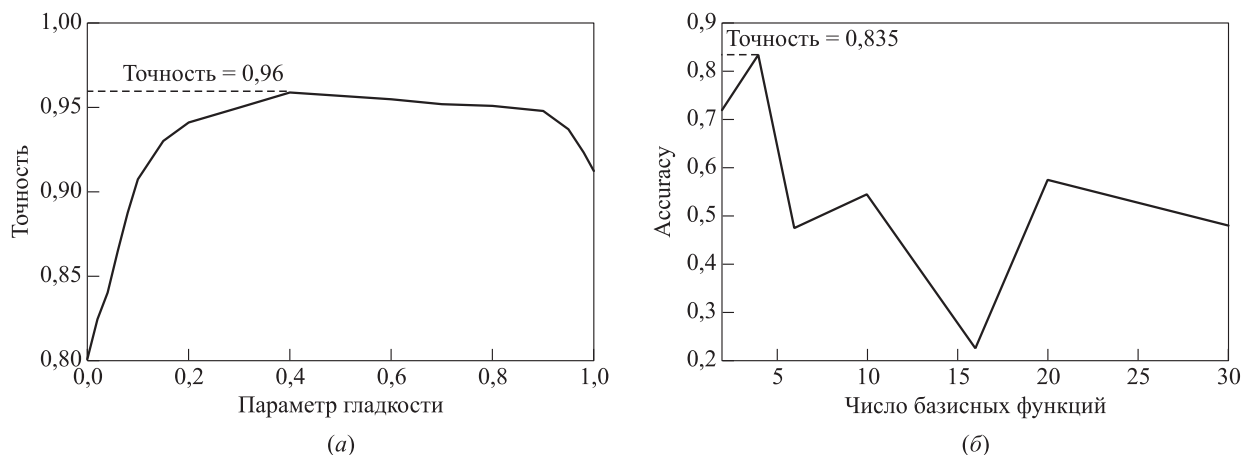


Рис. 5 Зависимость точности классификации от параметра сглаживания (а) и от числа базисных функций (б)

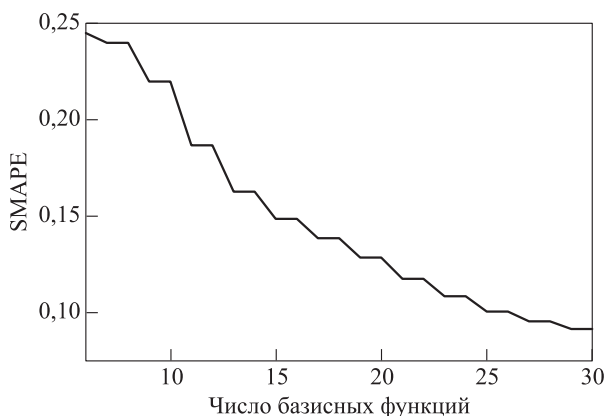


Рис. 6 Зависимость погрешности SMAPE от числа базисных функций

6 Заключение

В работе были описаны различные параметрические модели аппроксимации временных рядов. Проанализирована их эффективность при построении классификатора. Предложен метод классификации временных рядов при помощи построения пространства коэффициентов локальных аппроксимирующих моделей. Сравнялся ряд таких моделей. Квадратичный сплайн оказался непригоден для классификации, остальные отображения показали результат, сравнимый с тождественным отображением. Оптимальные параметры классификатора повышают точность на 3%–5%. Предложенный метод позволяет получить классификатор без затрат времени на ручную генерацию признаков.

Литература

1. Farmer J. D., Sidorowich J. J. Predicting chaotic time series // Phys. Rev. Lett., 1987. Vol. 59. No. 8. P. 845–848.

2. Карасиков М. Е., Стрижов В. В. Классификация временных рядов в пространстве параметров порождающих моделей // Информатика и её применения, 2008. Т. 10. Вып. 4. С. 121–131.

3. Мясников Л. Л., Мясникова Е. Н. Автоматическое распознавание звуковых образов. — Л.: Энергия, 1970. 183 с.

4. Kwapisz J. R., Weiss G. M., Moore S. A. Activity recognition using cell phone accelerometers // ACM SigKDD Explorations Newsletter, 2011. Vol. 12. No. 2. P. 72–82.

5. Кузнецов М. П., Ивкин Н. П. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1471–1483.

6. Tsay R. S. Analysis of financial time series. — Hoboken, NJ, USA: Wiley, 2006. 712 p.

7. Fu C. A review on time series data mining // Eng. Appl. Artif. Intel., 2011. Vol. 24. P. 164–181.

8. Esling P., Agon C. Time series data mining // ACM Comput. Surv., 2012. Vol. 45. No. 1. P. 1–34.

9. Coull B. A., Staundenmayer J. Self modeling regression for multivariate curve data // Stat. Sinica, 2004. Vol. 14. P. 695–711.

10. Weber G.-W., Batmaz I., Köksal G., et al. CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimisation // Inverse Probl. Sci. En., 2012. Vol. 20. No. 3. P. 371–400. doi: 10.1080/17415977.2011.624770.

11. Истомин И. А., Котляров О. Л., Лоскутов А. Ю. К проблеме обработки временных рядов: расширение возможностей метода локальной аппроксимации посредством сингулярного спектрального анализа // Теоретическая и математическая физика, 2005. Т. 142. № 1. С. 148–160.

12. Dette H., Melas V. B., Pepelyshev A. Optimal design for smoothing splines // Ann. I. Stat. Math., 2007. Vol. 63. No. 5. P. 981–1003.

13. Целых В. Р. Многомерные адаптивные регрессионные сплайны // Машинное обучение и анализ данных, 2012. Т. 1. № 3. С. 272–278.

14. Gholizadel S., Salajegheh E. Optimal design of structures subjected to time history loading by swarm intelligence and an advanced metamodels // *Comput. Method. Appl. M.*, 2009. Vol. 198. P. 2936–2949.
15. Vasko K., Toivonen H. Estimating the number of segments in time series data using permutation tests // *Conference (International) on Data Mining Proceedings.* — IEEE, 2006. P. 466–473.
16. Motrenko A. P., Strijov V. V. Extracting fundamental periods to segment biomedical signals // *IEEE J. Biomed. Health*, 2015. Vol. 20. No. 6. P. 1466–1476.

Поступила в редакцию 23.05.18

LOCAL APPROXIMATION MODELS FOR HUMAN PHYSICAL ACTIVITY CLASSIFICATION

D. A. Anikeyev¹, G. O. Penkin¹, 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 Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The research is devoted to the time series classification. The time series is measured by an accelerometer of a wearable device. A class of physical activity is defined by its feature description of a time segment. To construct this description, the authors propose to use parameters of various approximation splines (algebraic, smoothing, adaptive regression, or spline with dynamic nodes). The logistic regression is used as a classifier. It delivers desired quality of the activity recognition. The authors analyze the space of the local approximation parameters. Classification accuracy depends on the method of this space construction. The computational experiment finds the optimal approximation parameters and parameters of the classifier.

Keywords: local approximation model; time series; classification; splines; feature space

DOI: 10.14357/19922264190106

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (projects 17-07-1574 and 19-07-1155).

References

1. Farmer, J. D., and J. J. Sidorowich. 1987. Predicting chaotic time series. *Phys. Rev. Lett.* 59(8):845–848.
2. Karasikov, M. E., and V. V. Strijov. 2008. Klassifikatsiya vremennykh ryadov v prostranstve parametrov porozhdayushchikh modeley [Feature-based time-series classification]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):128–138.
3. Myasnikov, L. L., and E. N. Myasnikova. 1970. *Avtomaticheskoe raspoznavanie zvukovykh obrazov* [Automat-ic speech recognition]. Moscow: Energiya. 183 p.
4. Kwapisz, J. R., G. M. Weiss, and S. A. Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12(2):72–82.
5. Kuznetsov, M. P., and N. P. Ivkin. 2015. Algoritm klas-sifikatsii vremennykh ryadov akselerometra po kombinirovannomu priznakovomu opisaniyu [Time series clas-sification algorithm using combined feature description]. *Mashinnoe obuchenie i analiz dannykh* [Machine Learning Data Anal.] 1(11):1471–1483.
6. Tsay, R. S. 2010. *Analysis of financial time series*. Hoboken, NJ: Wiley. 712 p.
7. Fu, C. 2011. A review on time series data mining. *Eng. Appl. Artif. Intel.* 24:164–181.
8. Esling, P., and C. Agon. 2012. Time series data mining. *ACM Comput. Surv.* 45(1):1–34.
9. Coull, B. A., and J. Staundenmayer. 2004. Self model-ing regression for multivariate curve data. *Stat. Sinica* 14:695–711.
10. Weber, G.-W., I. Batmaz, G. Köksal, et al. 2012. CMARS: A new contribution to nonparametric regression with mul-tivariate adaptive regression splines supported by contin-uous optimisation. *Inverse Probl. Sci. En.* 20(3):371–400. doi: 10.1080/17415977.2011.624770.
11. Istomin, I. A., O. L. Kotlyarov, and A. Yu. Loskutov. 2005. The problem of processing time series: Extending possi-bilities of the local approximation method using singular spectrum analysis. *Theor. Math. Phys.* 142(1):128–137.

12. Dette, H., V. B. Melas, and A. Pepelyshev. 2011. Optimal design for smoothing splines *Ann. I. Stat. Math.* 63(5):981–1003.
13. Tselykh, V. R. 2012. Mnogomernye adaptivnye regressionnye splayny [Multivariate adaptive regression splines]. *Mashinnoe obuchenie i analiz dannykh* [Machine Learning Data Anal.] 1(3):272–278.
14. Gholizadel, S., and E. Salajegheh. 2009. Optimal design of structures subjected to time history loading by swarm intelligence and an advanced metamodels. *Comput. Method. Appl. M.* 198:2936–2949.
15. Vasko, K., and H. Toivonen. 2002. Estimating the number of segments in time series data using permutation tests. *Conference (International) on Data Mining Proceedings* 466–473.
16. Motrenko, A. P., and V. V. Strijov. 2016. Extracting fundamental periods to segment biomedical signals. *IEEE J. Biomed. Health* 20(6):1466–1476.

Received May 23, 2018

Contributors

Anikeyev Dmitriy A. (b. 1995) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; dmitriy.anikeyev@phystech.edu

Penkin Grigoriy O. (b. 1993) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; penkin.gr@gmail.com

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 Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; strijov@ccas.ru

ОБРАЩЕНИЕ ОДНОРОДНЫХ ОПЕРАТОРОВ С ПОМОЩЬЮ СТАБИЛИЗИРОВАННОЙ ЖЕСТКОЙ ПОРОГОВОЙ ОБРАБОТКИ ПРИ НЕИЗВЕСТНОЙ ДИСПЕРСИИ ШУМА*

О. В. Шестаков¹

Аннотация: При обращении линейных однородных операторов обычно необходимо использовать методы регуляризации, поскольку наблюдаемые данные, как правило, зашумлены. Для подавления шума часто используется пороговая обработка вейвлет-коэффициентов функции наблюдаемого сигнала. Пороговая обработка стала популярным инструментом подавления шума благодаря своей простоте, вычислительной эффективности и возможности адаптации к функциям, имеющим на разных участках разную степень регулярности. Рассматривается предложенный недавно стабилизированный метод жесткой пороговой обработки, в котором устранены основные недостатки мягкой и жесткой пороговой обработки, и исследуются статистические свойства этого метода. В модели данных с аддитивным гауссовским шумом с неизвестной дисперсией проведен анализ несмещенной оценки среднеквадратичного риска и показано, что при определенных условиях данная оценка является асимптотически нормальной, при этом дисперсия предельного распределения зависит от способа оценивания дисперсии шума.

Ключевые слова: вейвлеты; пороговая обработка; несмещенная оценка риска; асимптотическая нормальность; сильная состоятельность

DOI: 10.14357/19922264190107

1 Введение

В медицинских, физических, астрономических и других научных проблемах часто возникает задача получить представление об объекте, который описывается некоторой функцией f , имея возможность наблюдать только функцию Kf , где K — некоторый линейный оператор. При этом часто нельзя просто применить к наблюдаемым данным обратный оператор K^{-1} , поскольку эти данные, как правило, содержат шум и задача обращения оператора K некорректно поставлена. К тому же обычно дисперсия шума неизвестна и ее необходимо оценивать по наблюдаемым данным.

Одним из популярных инструментов при регуляризации процедуры обращения служит вейвлет-разложение с последующей пороговой обработкой вейвлет-коэффициентов. Наиболее распространенные виды пороговой обработки — жесткая и мягкая. В работе [1] был предложен метод стабилизированной жесткой пороговой обработки, который объединяет в себе преимущества этих двух видов. В ситуации, когда дисперсия шума предполагается известной, в работе [2] доказана асимптотическая нормальность оценки среднеквадратичного риска пороговой обработки.

В данной работе исследуется влияние способов оценивания дисперсии шума на характеристики предельного распределения оценки среднеквадратичного риска. Для метода мягкой пороговой обработки подобные исследования проводились в работах [3, 4].

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

В данной работе рассматривается метод обращения линейных однородных операторов, основанный на вейвлет-вейвлет-разложении [5]. Линейный оператор K называется однородным, если

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

для любого x_0 и любого $a > 0$. Параметр α называется показателем однородности. Примерами линейных однородных операторов служат оператор интегрирования, преобразование Гильберта и преобразование Абеля.

*Работа выполнена при частичной финансовой поддержке РФФИ (проект 19-07-00352).

¹Московский 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

Относительно наблюдаемой функции Kf будем предполагать, что она определена на конечном отрезке и равномерно регулярна по Липшицу с некоторым показателем $\gamma > 0$. Вейвлет-разложение Kf представляет собой ряд по ортонормированному базису

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

где $\psi(t)$ — некоторая материнская вейвлет-функция, а $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$. Индекс j в (1) называется масштабом, а индекс k — сдвигом. Если вейвлет-функция обладает определенными свойствами регулярности [6], то для коэффициентов разложения в (1) справедливо

$$|\langle Kf, \psi_{j,k} \rangle| \leq \frac{C_f}{2^{j(\gamma+1/2)}}, \quad (2)$$

где C_f — некоторая положительная константа.

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

$$f = \sum_{j,k \in \mathbb{Z}} \beta_{j,k} \langle Kf, \psi_{j,k} \rangle u_{j,k}, \quad (3)$$

где $u_{j,k} = K^{-1} \psi_{j,k} / \beta_{j,k}$, $\beta_{j,k} = 2^{\alpha j} \beta_{00}$, $\beta_{00} = \|K^{-1} \psi\|$ (функции $u_{j,k}$, как и $\psi_{j,k}$, представляют собой сдвиги и растяжения одной материнской функции u и называются вейглетами). При соответствующем выборе $\psi(t)$ последовательность $\{u_{j,k}\}$ образует устойчивый базис [7]. Формула (3) и есть основа метода вейглет-вейвлет-разложения.

Пороговая обработка эмпирических коэффициентов

При фактических измерениях значения функции сигнала регистрируются в дискретных отчетах, при этом такие значения, как правило, зашумлены. Рассмотрим следующую модель данных с шумом:

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

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

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

где $\epsilon_{j,k}^W$ независимы и распределены так же, как и ϵ_i , а $\mu_{j,k} = 2^{J/2} \langle Kf, \psi_{j,k} \rangle$ [6].

Для подавления шума и построения оценки функции сигнала к коэффициентам $Y_{j,k}$ обычно применяется функция жесткой пороговой обработки $\rho_H(y, T) = x \mathbf{I}(|y| > T)$ или мягкой пороговой обработки $\rho_S(y, T) = \mathbf{sgn}(x) (|y| - T)_+$ с порогом T . При таком подходе обнуляются коэффициенты, абсолютная величина которых ниже порога, так как в силу (2) основная часть полезного сигнала содержится в относительно небольшом числе больших по модулю коэффициентов.

Каждому из этих видов пороговой обработки присущи свои недостатки. Жесткая пороговая функция разрывна, и это приводит к отсутствию устойчивости при выборе порога [8] и невозможности построения несмещенной оценки среднеквадратичного риска [9]. При мягкой пороговой обработке в оценке функции появляется дополнительное смещение. Чтобы частично избежать этих недостатков, в работе [1] был предложен новый вид пороговой обработки, представляющий собой сглаженный (стабилизированный) аналог жесткой пороговой обработки. В этом методе оценки $\mu_{j,k}$ вычисляются по формулам:

$$\hat{\mu}_{j,k} = \mathbb{E} [\rho_H(Y_{j,k} + \lambda \xi_{j,k}, T_j) | Y_{j,k}],$$

где случайные величины $\xi_{j,k}$ имеют стандартное нормальное распределение и не зависят от $Y_{j,k}$, а $\lambda > 0$ — параметр стабилизации, отвечающий за степень сглаживания. Вычисляя математическое ожидание, получаем:

$$\hat{\mu}_{j,k} = Y_{j,k} \left[\Phi \left(-\frac{T_j + Y_{j,k}}{\lambda} \right) + 1 - \Phi \left(\frac{T_j - Y_{j,k}}{\lambda} \right) \right] + \lambda \left[\phi \left(\frac{T_j - Y_{j,k}}{\lambda} \right) - \phi \left(\frac{T_j + Y_{j,k}}{\lambda} \right) \right].$$

Достоинством такого метода является бесконечная дифференцируемость $\hat{\mu}_{j,k}$ по $Y_{j,k}$, что приводит к более робастным оценкам [1]. Заметим также, что при $\lambda \rightarrow 0$ получается обычный метод жесткой пороговой обработки. В данной работе параметр λ предполагается фиксированным, а в качестве T_j для каждого масштаба j выбирается порог $T_j = \sigma \sqrt{2 \ln 2^j}$. Такой порог получил название «универсальный», так как он не зависит от наблюдаемых данных. И при жесткой, и при мягкой пороговой обработке этот порог обеспечивает близость среднеквадратичного риска к минимальному [6].

3 Несмещенная оценка среднеквадратичного риска

Среднеквадратичный риск метода пороговой обработки определяется по формуле:

$$R_J(\sigma) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbb{E}(\hat{\mu}_{j,k}(\sigma) - \mu_{j,k})^2. \quad (4)$$

В [1] показано, что при стабилизированной жесткой пороговой обработке

$$\begin{aligned} & \mathbb{E}(\hat{\mu}_{j,k}(\sigma) - \mu_{j,k})^2 = \\ & = \mathbb{E} \left[(Y_{j,k} - \hat{\mu}_{j,k}(\sigma))^2 + 2\sigma^2 \frac{\partial}{\partial Y_{j,k}} \hat{\mu}_{j,k}(\sigma) \right] - \sigma^2, \end{aligned}$$

где

$$\begin{aligned} & \frac{\partial}{\partial Y_{j,k}} \hat{\mu}_{j,k}(\sigma) = \\ & = \Phi \left(-\frac{T_j + Y_{j,k}}{\lambda} \right) + 1 - \Phi \left(\frac{T_j - Y_{j,k}}{\lambda} \right) + \\ & + \frac{T_j}{\lambda} \left[\phi \left(\frac{T_j - Y_{j,k}}{\lambda} \right) + \phi \left(\frac{T_j + Y_{j,k}}{\lambda} \right) \right]. \end{aligned}$$

Таким образом, величина

$$\begin{aligned} \hat{R}_J(\sigma) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 & \left[(Y_{j,k} - \hat{\mu}_{j,k}(\sigma))^2 + \right. \\ & \left. + 2\sigma^2 \frac{\partial}{\partial Y_{j,k}} \hat{\mu}_{j,k}(\sigma) - \sigma^2 \right] \quad (5) \end{aligned}$$

является несмещенной оценкой R_J , не зависящей от ненаблюдаемых значений $\mu_{j,k}$.

В работе [2] доказано следующее утверждение, устанавливающее асимптотическую нормальность оценки (5) и позволяющее строить асимптотические доверительные интервалы для риска (4).

Теорема 1. Пусть K — линейный однородный оператор с показателем однородности $\alpha > 0$, а Kf задана на конечном отрезке и равномерно регулярна по Липшицу с показателем $\gamma > 0$. Тогда

$$\mathbb{P} \left(\frac{\hat{R}_J(\sigma) - R_J(\sigma)}{D_J} < x \right) \Rightarrow \Phi(x),$$

где

$$D_J^2 = \frac{2\sigma^4 \beta_{0,0}^4}{2^{4\alpha+1} - 1} 2^{(4\alpha+1)J}.$$

4 Виды оценок дисперсии шума

Как правило, дисперсия σ^2 неизвестна и вместо ее точного значения необходимо использовать некоторую оценку $\hat{\sigma}^2$, которая обычно строится по половине всех вейвлет-коэффициентов для $j = J - 1$, так как в силу (2) эти коэффициенты фактически содержат только шум. При этом порог вычисляется по формуле $\hat{T}_j = \hat{\sigma} \sqrt{2 \ln 2^j}$.

В качестве оценки σ^2 (или σ) в данной работе рассматривается выборочная дисперсия

$$\hat{\sigma}_S^2 = \frac{1}{2^{J-1}} \sum_{k=0}^{2^{J-1}-1} Y_{J-1,k}^2 - \bar{Y}^2, \quad (6)$$

где

$$\bar{Y} = \frac{1}{2^{J-1}} \sum_{k=0}^{2^{J-1}-1} Y_{J-1,k},$$

а также соответствующим образом нормированный выборочный интерквартильный размах $\hat{\sigma}_R$ и выборочное абсолютное медианное отклонение $\hat{\sigma}_M$, которые определяются следующим образом:

$$\hat{\sigma}_R = \frac{Y_{(J-1,3/4)} - Y_{(J-1,1/4)}}{2\xi_{3/4}}; \quad (7)$$

$$\hat{\sigma}_M = \frac{\text{med}_{0 \leq k \leq 2^{J-1}-1} |Y_{J-1,k} - \text{med}_{0 \leq l \leq 2^{J-1}-1} Y_{J-1,l}|}{\xi_{3/4}}. \quad (8)$$

Здесь $Y_{(J-1,1/4)}$ и $Y_{(J-1,3/4)}$ — выборочные квантили порядка 1/4 и 3/4, построенные по выборке из половины всех вейвлет-коэффициентов при $j = J - 1$; $\xi_{3/4}$ — теоретическая квантиль порядка 3/4 стандартного нормального распределения ($\xi_{3/4} \approx 0,6745$); med обозначает выборочную медиану.

Выборочная дисперсия служит самой популярной оценкой величины σ^2 , и в случае отсутствия выбросов она наиболее предпочтительна. Однако в случае, когда оценка дисперсии строится по выборке сигнала, естественно ожидать, что выборка не будет однородной. Преимущество использования последних двух оценок заключается в их робастности, т. е. нечувствительности к выбросам.

5 Предельная дисперсия оценки среднеквадратичного риска

Способ оценивания дисперсии шума влияет на вид предельной дисперсии оценки среднеквадратичного риска. Подобный эффект наблюдается и при мягкой пороговой обработке [4].

Теорема 2. Пусть Kf задана на конечном отрезке и равномерно регулярна по Липшицу с показателем

$\gamma > 1/4$, а оценка дисперсии шума задана соотношением (6). Тогда

$$P\left(\frac{\widehat{R}_J(\widehat{\sigma}_S) - R_J(\sigma)}{D_J} < x\right) \Rightarrow \Phi_{\Upsilon_1}(x),$$

где $\Phi_{\Upsilon_1}(x)$ — функция распределения нормального закона с нулевым средним и дисперсией

$$\Upsilon_1^2 = \frac{1}{2^{4\alpha+1}} + \frac{2^{4\alpha+1} - 1}{2^{4\alpha+1} (2^{2\alpha+1} - 1)^2}.$$

Доказательство. Обозначим

$$\widehat{U}_J(\sigma) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \left[(Y_{j,k} - \widehat{\mu}_{j,k}(\sigma))^2 + 2\sigma^2 \frac{\partial}{\partial Y_{j,k}} \widehat{\mu}_{j,k}(\sigma) \right]$$

и запишем $\widehat{R}_J(\widehat{\sigma}_S) - R_J(\sigma)$ в виде

$$\begin{aligned} \widehat{R}_J(\widehat{\sigma}_S) - R_J(\sigma) &= \\ &= \left[\widehat{U}_J(\widehat{\sigma}_S) - \widehat{U}_J(\sigma) \right] + \left[\widehat{R}_J(\sigma) - R_J(\sigma) \right] + \\ &+ \frac{2^{(2\alpha+1)J} - 1}{2^{2\alpha+1} - 1} (\sigma^2 - \widehat{\sigma}_S^2) \equiv S_1 + S_2 + S_3. \end{aligned}$$

Повторяя рассуждения из работ [3, 4] и учитывая, что если $\gamma > 1/4$, то выполнено $2^{J/2} \overline{Y}^2 \xrightarrow{P} 0$ при $J \rightarrow \infty$ [4], можно показать, что

$$P\left(\frac{S_2 + S_3}{D_J} < x\right) \Rightarrow \Phi_{\Upsilon_1}(x).$$

Докажем, что $D_J^{-1} S_1 \xrightarrow{P} 0$ при $J \rightarrow \infty$. Пусть $C_\delta > 0$ — некоторая константа, а $\delta_J = C_\delta J^{1/2} 2^{-J/2}$. Запишем

$$S_1 = \mathbf{1}(|\sigma^2 - \widehat{\sigma}_S^2| > \delta_J) S_1 + \mathbf{1}(|\sigma^2 - \widehat{\sigma}_S^2| \leq \delta_J) S_1 \equiv S'_1 + S''_1.$$

Для произвольного $\varepsilon > 0$

$$P(S'_1 > \varepsilon) \leq P(|\sigma^2 - \widehat{\sigma}_S^2| > \delta_J).$$

При выполнении условий теоремы, если константа C_δ достаточно велика, то найдется константа $\tilde{C}_\delta > 0$ такая, что [4]

$$P(|\sigma^2 - \widehat{\sigma}_S^2| > \delta_J) \leq \tilde{C}_\delta 2^{-J/2}.$$

Следовательно, $S'_1 \xrightarrow{P} 0$ при $J \rightarrow \infty$.

Обозначим слагаемые в сумме S''_1 через $F_{j,k}(\widehat{\sigma}_S)$. Пусть $A_j = \sqrt{A \ln 2^j}$, где $0 < A < 2(\sigma^2 - \delta_J)$. Имеем:

$$\begin{aligned} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} F_{j,k}(\widehat{\sigma}_S) &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbf{1}(|Y_{j,k}| \leq A_j) F_{j,k}(\widehat{\sigma}_S) + \\ &+ \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbf{1}(|Y_{j,k}| > A_j) F_{j,k}(\widehat{\sigma}_S) \equiv W_1 + W_2. \end{aligned}$$

Рассмотрим W_1 . Учитывая определения $\widehat{\mu}_{j,k}(\sigma)$, $(\partial/\partial Y_{j,k}) \widehat{\mu}_{j,k}(\sigma)$ и A_j , можно убедиться, что найдутся константы $C_1 > 0$ и $\theta > 0$ такие, что

$$|\mathbf{1}(|Y_{j,k}| \leq A_j) F_{j,k}(\widehat{\sigma}_S)| \leq C_1 J^{5/2} 2^{(2\alpha-\theta)j-J/2} \text{ п.в.}$$

Следовательно, $D_J^{-1} W_1 \rightarrow 0$ п.в. при $J \rightarrow \infty$.

Далее для слагаемых W_2 имеем:

$$\begin{aligned} |\mathbf{1}(|Y_{j,k}| > A_j) F_{j,k}(\widehat{\sigma}_S)| &\leq \\ &\leq C_2 J^{3/2} 2^{2\alpha j - J/2} \mathbf{1}(|Y_{j,k}| > A_j) |Y_{j,k}|^2 \text{ п.в.,} \end{aligned}$$

где $C_2 > 0$ — некоторая константа. Учитывая распределение $Y_{j,k}$, нетрудно убедиться, что

$$\mathbb{E} \frac{1}{D_J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} J^{3/2} 2^{2\alpha j - J/2} \mathbf{1}(|Y_{j,k}| > A_j) |Y_{j,k}|^2 \rightarrow 0$$

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

$$D_J^{-1} W_2 \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

Таким образом, $D_J^{-1} S_1 \xrightarrow{P} 0$ при $J \rightarrow \infty$.

Теорема доказана.

Рассмотрим теперь ситуацию, когда в качестве оценки σ используется величина $\widehat{\sigma}_R$ или $\widehat{\sigma}_M$. В этом случае повышаются требования к гладкости функции сигнала.

Теорема 3. Пусть Kf задана на конечном отрезке и равномерно регулярна по Липшицу с показателем $\gamma > 1/2$, а оценка дисперсии шума $\widehat{\sigma}$ задана соотношением (7) или соотношением (8). Тогда

$$P\left(\frac{\widehat{R}_J(\widehat{\sigma}) - R_J(\sigma)}{D_J} < x\right) \Rightarrow \Phi_{\Upsilon_2}(x),$$

где $\Phi_{\Upsilon_2}(x)$ — функция распределения нормального закона с нулевым средним и дисперсией

$$\begin{aligned} \Upsilon_2^2 &= 1 + \frac{2^{4\alpha+1} - 1}{4(2^{2\alpha+1} - 1)^2 \xi_{3/4}^2 (\phi(\xi_{3/4}))^2} - \\ &- \frac{2^{4\alpha+1} - 1}{2^{2\alpha-1} (2^{2\alpha+1} - 1)}. \end{aligned}$$

Доказательство. Как и в предыдущей теореме, запишем $\widehat{R}_J(\widehat{\sigma}) - R_J(\sigma) = S_1 + S_2 + S_3$. Учитывая,

что $\gamma > 1/2$, и поступая, как в работах [2, 4, 10], с использованием разложения Бахадура для выборочных квантилей [11] и выборочного абсолютного медианного отклонения [12], можно показать, что

$$P\left(\frac{S_2 + S_3}{D_J} < x\right) \Rightarrow \Phi_{\Gamma_2}(x).$$

Используя экспоненциальные неравенства для выборочных квантилей [11] и выборочного абсолютного медианного отклонения [12], получаем, что при выполнении условий теоремы найдется такая константа $C_\delta > 0$, что при $\delta_J = C_\delta J^{1/2} 2^{-J/2}$ для некоторой константы $\tilde{C}_\delta > 0$ выполнено:

$$P(|\hat{\sigma}_R - \sigma| > \delta_J) \leq \tilde{C}_\delta 2^{-J/2};$$

$$P(|\hat{\sigma}_M - \sigma| > \delta_J) \leq \tilde{C}_\delta 2^{-J/2}.$$

Далее, повторяя рассуждения предыдущей теоремы, заключаем, что $D_J^{-1} S_1 \xrightarrow{P} 0$ при $J \rightarrow \infty$.

Теорема доказана.

Литература

1. Huang H.-C., Lee T.C.M. Stabilized thresholding with generalized sure for image denoising // IEEE 17th Conference (International) on Image Processing Proceedings. — IEEE, 2010. P. 1881–1884.
2. Shestakov O.V. Nonlinear regularization of inverse problems for linear homogeneous transforms by the stabilized hard thresholding // J. Math. Sci., 2018. Vol. 234. No. 6. P. 780–785.
3. Кудрявцев А. А., Шестаков О. В. Асимптотика оценки риска при вейвлет-вейвлет разложении наблюдаемого сигнала // Т-Comm — телекоммуникации и транспорт, 2011. № 2. С. 54–57.
4. Кудрявцев А. А., Шестаков О. В. Асимптотическое распределение оценки риска пороговой обработки вейвлет-коэффициентов сигнала при неизвестном уровне шума // Т-Comm — телекоммуникации и транспорт, 2011. № 5. С. 24–30.
5. Abramovich F., Silverman B.W. Wavelet decomposition approaches to statistical inverse problems // Biometrika, 1998. Vol. 85. No. 1. P. 115–129.
6. Mallat S. A Wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.
7. Lee N. Wavelet-vaguelette decompositions and homogeneous equations. — West Lafayette, IN, USA: Purdue University, 1997. PhD Thesis. 103 p.
8. Breiman L. Heuristics of instability and stabilization in model selection // Ann. Stat., 1996. Vol. 24. No. 6. P. 2350–2383.
9. Jansen M. Noise reduction by wavelet thresholding. — Lecture notes in statistics ser. — New York, NY, USA: Springer Verlag, 2001. Vol. 161. 196 p.
10. Шестаков О. В. О скорости сходимости оценки риска пороговой обработки вейвлет-коэффициентов к нормальному закону при использовании робастных оценок дисперсии // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 122–128.
11. Serfling R. Approximation theorems of mathematical statistics. — New York, NY, USA: John Wiley & Sons, 1980. 371 p.
12. Serfling R., Mazumder S. Exponential probability inequality and convergence results for the median absolute deviation and its modifications // Stat. Probabil. Lett., 2009. Vol. 79. No. 16. P. 1767–1773.

Поступила в редакцию 14.12.18

INVERSION OF HOMOGENEOUS OPERATORS USING STABILIZED HARD THRESHOLDING WITH UNKNOWN NOISE VARIANCE

O. V. Shestakov

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

Abstract: When inverting linear homogeneous operators, it is necessary to use regularization methods, since observed data are usually noisy. For noise suppression, threshold processing of wavelet coefficients of the observed signal function is often used. Threshold processing has become a popular noise suppression tool due to its simplicity, computational efficiency, and ability to adapt to functions that have different degrees of regularity at different domains. The paper discusses the recently proposed stabilized hard thresholding method that eliminates the main drawbacks of soft and hard thresholding methods and studies statistical properties of this method. In the data model

with an additive Gaussian noise with unknown variance, an unbiased estimate of the mean square risk is analyzed and it is shown that under certain conditions, this estimate is asymptotically normal and the variance of the limit distribution depends on the type of estimate of noise variance.

Keywords: wavelets; threshold processing; unbiased risk estimate; asymptotic normality; strong consistency

DOI: 10.14357/19922264190107

Acknowledgments

This research was partly supported by the Russian Foundation for Basic Research (project No. 19-07-00352).

References

- Huang, H.-C., and T. C. M. Lee. 2010. Stabilized thresholding with generalized sure for image denoising. *IEEE 17th Conference (International) on Image Processing*. IEEE. 1881–1884.
- Shestakov, O. V. 2018. Nonlinear regularization of inverse problems for linear homogeneous transforms by the stabilized hard thresholding. *J. Math. Sci.* 234(6):780–785.
- Kudryavtsev, A. A., and O. V. Shestakov. 2011. Asimptotika otsenki riska pri veyglet-veyvlet razlozhenii nablyudamogo signala [The average risk assessment of the wavelet decomposition of the signal]. *T-Comm — Telecommunications and Their Application in Transport Industry* 2:54–57.
- Kudryavtsev, A. A., and O. V. Shestakov. 2011. Asimptoticheskoe raspredelenie otsenki riska porogovoy obrabotki veyglet-koeffitsientov signala pri neizvestnom urovne shuma [Asymptotic distribution of the risk estimate of the signal vaguelette coefficients thresholding at the unknown noise level]. *T-Comm — Telecommunications and Their Application in Transport Industry* 5:24–30.
- Abramovich, F., and B. W. Silverman. 1998. Wavelet decomposition approaches to statistical inverse problems. *Biometrika* 85(1):115–129.
- Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
- Lee, N. 1997. Wavelet-vaguelette decompositions and homogenous equations. West Lafayette, IN: Purdue University. PhD Thesis. 103 p.
- Breiman, L. 1996. Heuristics of instability and stabilization in model selection. *Ann. Stat.* 24(6):2350–2383.
- Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. New York, NY: Springer Verlag. Vol. 161. 196 p.
- Shestakov, O. V. 2012. O skorosti skhodimosti otsenki riska porogovoy obrabotki veyvlet-koeffitsientov k normal'nomu zakonu pri ispol'zovanii robastnykh otsenok dispersii [On the rate of convergence to the normal law of risk estimate for wavelet coefficients thresholding when using robust variance estimates]. *Informatika i ee Primeneniya — Inform. Appl.* 6(2):122–128.
- Serfling, R. 1980. *Approximation theorems of mathematical statistics*. New York, NY: John Wiley & Sons. 371 p.
- Serfling, R., and S. Mazumder. 2009. Exponential probability inequality and convergence results for the median absolute deviation and its modifications. *Stat. Probabil. Lett.* 79(16):1767–1773.

Received December 14, 2018

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

ОБ УНИМОДАЛЬНОСТИ ФУНКЦИИ ДОХОДА СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ ТИПА $G|M|s$ С УПРАВЛЯЕМОЙ ОЧЕРЕДЬЮ

Я. М. Агаларов¹, В. Г. Ушаков²

Аннотация: Рассматривается задача максимизации среднего дохода в системе массового обслуживания (СМО) типа $G|M|s$ в единицу времени на множестве чистых стационарных пороговых стратегий с одной точкой переключения режима ограничения доступа. Функция дохода зависит от следующих параметров, измеряемых в стоимостных единицах: плата, получаемая за обслуживание заявок; затраты на техническое обслуживание прибора; вычет из дохода за задержку заявок в очереди; штраф за необслуженные заявки. Доказано, что функция дохода является унимодальной на множестве рассматриваемых пороговых стратегий. Предложен алгоритм расчета оптимального порогового значения и соответствующего значения максимального дохода. Приведены результаты вычислительного эксперимента, иллюстрирующего работу предложенного алгоритма.

Ключевые слова: многоканальная система массового обслуживания; пороговая стратегия; максимизация дохода

DOI: 10.14357/19922264190108

1 Введение

Для повышения эффективности работы современных вычислительных сетей используют алгоритмы управления потоками (ограничения нагрузки), из которых наиболее часто применяются различные модификации порогового управления (пороговых стратегий) [1]. Исследованием и разработкой пороговых стратегий управления потоками занимались практически с начала зарождения вычислительных сетей с целью защиты связанных и вычислительных ресурсов от перегрузок. Одним из основных методов исследования эффективности пороговых стратегий является математическое моделирование с использованием аппарата теории очередей, которая изучает СМО различного типа. Большинство работ, в которых рассмотрены СМО с пороговой стратегией управления потоками, посвящено расчету характеристик системы (средней длины очереди, среднего времени пребывания, вероятности отклонения заявки, загруженности приборов и т. п.) при заданной пороговой стратегии.

В ряде работ, посвященных данной тематике, ставится задача оптимизации пороговой стратегии в смысле максимизации дохода системы, представленного в виде стоимостного функционала (см.,

например, [2–7]). Хотя практический интерес к такой постановке задачи в смысле повышения эффективности систем, как представляется, не ниже, чем к задачам расчета характеристик систем при фиксированной пороговой стратегии, вопрос существования эффективных методов и алгоритмов поиска оптимальных пороговых стратегий для СМО остается открытым, за исключением простых СМО ($M|M|1$, $M|M|n$) и простых целевых функций (допустимая средняя задержка заявок в системе, допустимая интенсивность потери). Для более сложных СМО (например, $G|M|1$, $M|G|1$) с более сложными целевыми функциями результаты исследований ограничиваются математическими постановками задач и эвристическими алгоритмами их решения.

Из недавно опубликованных в отечественной литературе работ, посвященных пороговой стратегии управления очередью с одной точкой переключения, отметим [3], где для системы $G|M|n$ сформулирована задача максимизации дохода системы на множестве пороговых стратегий с одним переключением, фиксированной платой за своевременное обслуживание и штрафом за невыполнение этого условия для допущенной в систему заявки. В работе предложен эвристический алго-

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

²Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vgushakov@mail.ru

ритм поиска оптимальной стратегии и выдвинута гипотеза о том, что для систем $G|G|n$ существует единственное решение указанной задачи в виде простой пороговой стратегии.

Аналогичная задача рассмотрена в работе [4] для системы $M|D|1$, где также приведена математическая постановка задачи и получена нижняя оценка для оптимального порогового значения.

Постановки задач, наиболее близких к рассматриваемой в настоящей статье, приведены в работах [5–7]. В [5] решение задачи максимизации среднего дохода, получаемого системой $M|G|1$ в единицу времени, ищется на множестве смешанных пороговых стратегий при входной нагрузке меньше единицы. В работе доказано, что если решение задачи существует, то оно принадлежит множеству чистых стратегий. В работах [6, 7] аналогичная задача решена для СМО типа $M|G|1$ и $G|M|1$ соответственно.

В данной статье рассматривается задача максимизации среднего дохода СМО типа $G|M|s$ на множестве чистых пороговых стратегий с одной точкой переключения режима доступа заявок в систему. Ниже приведены результаты исследования, касающегося вопросов существования решения задачи и метода поиска оптимальной стратегии.

2 Постановка задачи

Рассматривается СМО типа $G|M|s$ с накопителем бесконечной емкости и s приборами обслуживания, в которую поступает рекуррентный поток заявок с функцией распределения $H(t)$. Время обслуживания заявки распределено по экспоненциальному закону с параметром $\mu > 0$. Поступившая заявка допускается в накопитель системы (занимает любое свободное место в накопителе), если в момент ее поступления число заявок в системе меньше k , $k \geq s$ — некоторое заданное целое число (ниже тривиальный случай $k < s$ не рассматривается). Такую процедуру доступа заявок в систему называют в литературе пороговой стратегией управления доступом, и в дальнейшем для краткости назовем ее просто стратегией. Обозначим эту стратегию соответствующим пороговым значением k . Если заявка допущена в накопитель, она занимает любое свободное место в накопителе и обслуживается на приборе в порядке поступления. Заявка покидает систему только при завершении обслуживания, освободив одновременно прибор и накопитель, а на освободившийся прибор поступает очередная заявка из накопителя (если таковая есть). Система получает доход, который формируется из следующих составляющих:

C_0v — плата, получаемая системой, если поступившая заявка обслужена системой (принята в накопитель), $v \geq 0$ — время занятия заявкой прибора обслуживания;

$C_1 \geq 0$ — величина штрафа, который платит система, если поступившая заявка отклонена;

$C_2 \geq 0$ — вычет из дохода системы за единицу времени ожидания заявки в системе;

$C_3 \geq 0$ — затраты системы в единицу времени на техническое обслуживание системы, $C_3 < C_0$.

Всюду ниже под доходом системы будем понимать суммарный доход с учетом всех указанных выше составляющих. Отметим, что процесс обслуживания заявок в данной системе описывается цепью Маркова, где переходы цепи определяются моментами поступления заявок и состояние системы есть число заявок, находящихся в системе в момент поступления (см., например, [8, 9]).

Введем обозначения:

$\{\pi_i^k, 0 \leq i \leq k\}$ — стационарное распределение вероятностей цепи при стратегии k (π_i^k — стационарная вероятность того, что цепь находится в состоянии i);

\tilde{g}^k — предельное среднее значение дохода системы в единицу времени;

g^k — предельное среднее значение суммарного дохода системы, усредненного по числу поступивших заявок;

q_i^k — средний доход, получаемый системой в состоянии i при стратегии k , $i \geq 0$;

$\bar{v} = \int_0^\infty t dH(t)$ — среднее время между соседними моментами поступления заявок.

Из определения вложенной цепи Маркова следует:

$$\tilde{g}^k = \frac{1}{\bar{v}} g^k, \quad (1)$$

где

$$g^k = \sum_{i=0}^k \pi_i^k q_i^k.$$

Ставится задача нахождения оптимальной стратегии k^0 , такой что

$$\max_{k \geq s} g^k = g^{k^0}.$$

3 Метод решения

Стационарные вероятности состояний $\pi_j^k, 0 \leq j \leq k$, вычисляются по формулам (см. [8, 9]):

$$\pi_j^k = R_j^k \pi_k^k, \quad j = 0, \dots, k-1,$$

$$\pi_k^k = \left(1 + \sum_{i=0}^{k-1} R_i^k\right)^{-1}, \quad (2)$$

где

$$R_k^k = 1; \quad R_{k-1}^k = \frac{1-r_0}{r_0};$$

$$R_{j-1}^k = \frac{R_j^k(1-r_1) - \sum_{i=j+1}^{k-1} R_i^k r_{i+1-j} - r_{k-j}}{r_0}, \quad s \leq j \leq k-1;$$

$$R_{s-2}^k = \frac{R_{s-1}^k(1-r_0) - \sum_{i=s}^{k-1} R_i^k b_{i,s-1} - b_{k,s-1}}{a_{s-2,s-1}};$$

$$r_i = \int_0^\infty \frac{(s\mu t)^i}{i!} e^{-\mu t s} dH(t); \quad (3)$$

$$R_{j-1}^k = \frac{R_j^k(1-a_{jj}) - \sum_{i=j+1}^{s-1} R_i^k a_{ij} - \sum_{i=s}^{k-1} R_i^k b_{ij} - b_{kj}}{a_{j-1,j}}, \quad 1 \leq j \leq s-2;$$

$$a_{ij} = \int_0^\infty \binom{i+1}{j} (1-e^{-\mu t})^{i+1-j} e^{-\mu t j} dH(t), \quad j \leq i+1 \leq s-1;$$

$$b_{ij} = \int_0^\infty \binom{s}{j} e^{-\mu t j} \left(\int_0^t \frac{(s\mu y)^{i-s}}{(i-s)!} (e^{-\mu y} - e^{-\mu t})^{s-j} s\mu dy \right) dH(t), \quad j < s \leq i+1 \leq k.$$

Положим

$$Q_i^k = \sum_{j=0}^i \pi_j^k; \quad A_{k+1} = \frac{1-Q_{s-1}^{k+1}}{1-Q_{s-2}^k};$$

$$F(k) = \frac{1}{1-A_{k+1}} \left(\sum_{i=s}^k \pi_i^{k+1} \left(\bar{v} - \frac{1}{\mu s} \sum_{m=i+2-s}^\infty (m-(i+1-s))r_m \right) \pi_{k+1}^{k+1} + \pi_{k+1}^{k+1} \left(\bar{v} - \frac{1}{\mu s} \sum_{m=k+2-s}^\infty (m-(k+1-s))r_m \right) \right);$$

$$G(k) = (C_0 - C_3) \bar{v} - C_2 F(k), \quad k \geq s. \quad (4)$$

Справедливы следующие утверждения.

Лемма 1. Среднее значение дохода, получаемого системой при стратегии k в состоянии i , равно

$$q_i^k = \left\{ \begin{array}{l} \frac{C_2}{\mu s} \left(\frac{1}{2} \sum_{m=1}^{i+2-s} m(m+2s-2i-3)r_m - \right. \\ \left. - \frac{1}{2}(i+1-s)(i+2-s) \sum_{m=i+3-s}^\infty r_m \right) + \\ \left. + (C_0 - C_3) \bar{v}, \quad s-1 \leq i \leq k-1; \right. \\ \left. (C_0 - C_3) \bar{v}, \quad 0 \leq i \leq s-2; \right. \\ \left. q_k^k = q_{k-1}^k - C_0 \bar{v} - C_1. \right. \end{array} \right. \quad (5)$$

Лемма 2. Справедливы равенства:

$$q_{i+1}^k = q_i^k - \frac{C_2}{\mu s} \sum_{m=1}^{i+2-s} m r_m - \frac{C_2(i+2-s)}{\mu s} \sum_{m=i+3-s}^\infty r_m, \quad s-1 \leq i \leq k-2. \quad (6)$$

Лемма 3. Справедливы соотношения:

$$\left. \begin{array}{l} \pi_{j+1}^{k+1} = A_{k+1} \pi_j^k, \quad s-1 \leq j \leq k; \\ g^k - g^{k+1} = (1 - A_{k+1})(g^k - G(k)), \quad k \geq s. \end{array} \right\} \quad (7)$$

Лемма 4. Функция $G(k)$ не возрастает по $k \geq s$.

Теорема 1. Справедливы следующие утверждения:

- (1) если $\inf_{k \geq s} G(k) < \sup_{k \geq s} g^k$, то при любых значениях параметров $C_i \geq 0, i = 0, 1, 3$, и $C_2 > 0$ существует оптимальная стратегия k_0 такая, что $s \leq k_0 < \infty$. В противном случае $k_0 = \infty$;
- (2) если $g^s \geq G(s)$, то $k^0 = s$;
- (3) если $C_2 = 0$ и $g^s < G(s)$, то $k^0 = \infty$;
- (4) условие $g^{k^0-1} < g^{k^0} \leq g^{k^0+1}$ является необходимым и достаточным для того, чтобы $s \leq k^0 < \infty$.

При выполнении условий п. 4 теоремы можно предложить следующий алгоритм нахождения оптимальной стратегии.

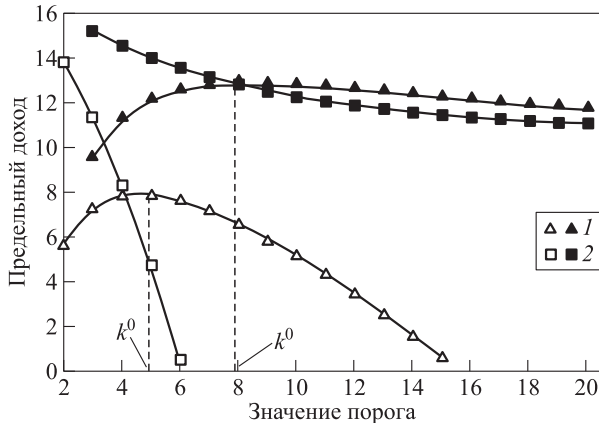
1. Положить $k = s$.
2. Вычислить g^s и $G(s)$.
3. Если $C_2 = 0$ и $g^s < G(s)$, то положить $k^0 = \infty$ и перейти к п. 8.
4. Если $g^s \geq G(s)$, то положить $k^0 = s$ и перейти к п. 8.
5. Вычислить g^{k+1} .
6. Если выполняется неравенство $g^k < g^{k+1}$, то перейти к п. 7, иначе положить $k^0 = k$ и перейти к п. 8.
7. Увеличить k на единицу и перейти к п. 5.
8. Конец алгоритма.

Доказательства лемм 1–4 и теоремы 1 представлены в приложении.

4 Пример

В качестве примера применения указанного выше алгоритма нахождения оптимальной стратегии рассмотрим СМО $H_2|M|s$ с функцией распределения входящего потока $H(t) = \sum_{i=1}^2 f_i (1 - e^{-\lambda_i t})$, $f_1 = 0,4, \lambda_1 = 0,8, \lambda_2 = 2$. Стоимостные параметры равны: $C_0 = 20, C_1 = 5, C_2 = 1,5, C_3 = 0,01$.

На рисунке приведены зависимости функций g^k и $G(k)$ от порогового значения.



Зависимость предельного дохода g^k (1) и функции $G(k)$ (2) от порогового значения ($\mu = 0,5$): пустые значки — $s = 2$; заштрихованные значки — $s = 3$

Приложение

Доказательство леммы 1. Фиксируем состояние $s \leq i \leq k - 1$, и пусть время нахождения системы в состоянии i равно v . Найдем выражение для суммарного среднего времени ожидания всех заявок в очереди в состоянии i , т.е. в интервале времени $(0, v]$.

Рассмотрим случайные величины вида $W_l = \sum_{j=1}^l \tau_j$, $l \geq 1$, где τ_j — независимые экспоненциально распределенные случайные величины (СВ) с параметром $s > 0$. Пусть B_m — событие вида $(W_m \leq v, W_{m+1} > v)$, $m > 0$, B_0 — событие $(W_1 > v)$. Заметим, что $B_m, m \geq 0$, — несовместные события и в совокупности составляют полную группу событий. Известно (см., например, [8]), что совместное распределение величин $W_l, l \leq m$, при условии выполнения события B_m совпадает с распределением порядковых статистик из выборки объема m из равномерного на $(0, v]$ распределения, а маргинальным условным распределением СВ W_l/v является бета-распределение с плотностью

$$f(x|B_m) = \begin{cases} \frac{m!}{(l-1)!(m-l)!} x^{l-1} (1-x)^{m-l}, & 0 < x < l; \\ 0 & \text{в противном случае.} \end{cases}$$

Отсюда следует, что $M(W_l|B_m) = (l/m + 1)v$.

Обозначим: $\bar{W}_{обсл/m}$ — среднее суммарное время ожидания заявок, завершивших обслуживание или приступивших к обслуживанию в состоянии i при условии B_m ; $\bar{W}_{необсл/m}$ — среднее суммарное время ожидания в очереди заявок, не поступивших на обслуживание в состоянии i при условии B_m . Заметим, что при $i < s$ для всех $m > 0$ имеют место равенства $\bar{W}_{обсл/m} = \bar{W}_{необсл/m} = 0$. При $s \leq i \leq k - 1, m \leq i - s + 1$ получим:

$$\bar{W}_{обсл/m} = \frac{mv}{2}; \quad \bar{W}_{необсл/m} = (i + 1 - s - m)v. \quad (8)$$

Далее, при $m > i + 1 - s, s \leq i \leq k - 1$,

$$\bar{W}_{обсл/m} = \frac{(i + 1 - s)(i - s + 2)v}{2(m + 1)}. \quad (9)$$

При $i = k$ и $i = k - 1$ значения $\bar{W}_{обсл/m}$ и $\bar{W}_{необсл/m}$ совпадают.

Обозначим $q_i^k(v)$ величину суммарного дохода системы в состоянии i при условии, что время пребывания в нем равно v . Из (8) и (9) при $s \leq i \leq k - 1$ имеем:

$$\begin{aligned} q_i^k(v) &= C_0 v - C_3 v - \\ &- \sum_{m=0}^{i-s+1} e^{-\mu sv} \frac{(\mu sv)^m}{m!} (C_2 (\bar{W}_{обсл/m} + \bar{W}_{необсл/m})) - \\ &- \sum_{m=i+2-s}^{\infty} e^{-\mu sv} \frac{(\mu sv)^m}{m!} C_2 \bar{W}_{обсл/m} = C_0 v - C_3 v + \\ &+ C_2 \sum_{m=0}^{i-s+1} \left(\frac{m}{2} - i - 1 + s \right) e^{-\mu sv} \frac{(\mu s)^m v^{m+1}}{m!} - \\ &- C_2 \frac{(i + 1 - s)(i + 2 - s)}{2} \sum_{m=i+2-s}^{\infty} e^{-\mu sv} \frac{(\mu s)^m v^{m+1}}{(m + 1)!}. \end{aligned}$$

Следовательно,

$$\begin{aligned} q_i^k &= \int_0^{\infty} q_i^k(v) dH(v) = C_0 \bar{v} - C_3 \bar{v} + \\ &+ \frac{C_2}{2\mu s} \sum_{m=0}^{i-s+1} m(m + 1) \int_0^{\infty} \frac{(\mu sv)^{m+1}}{(m + 1)!} e^{-\mu sv} dH(v) - \\ &- \frac{C_2(i + 1 - s)}{\mu s} \sum_{m=0}^{i-s+1} (m + 1) \int_0^{\infty} \frac{(\mu sv)^{m+1}}{(m + 1)!} e^{-\mu sv} dH(v) - \\ &- \frac{C_2(i + 1 - s)(i + 2 - s)}{2\mu s} \times \\ &\times \sum_{m=i+2-s}^{\infty} \int_0^{\infty} \frac{(\mu sv)^{m+1}}{(m + 1)!} e^{-\mu sv} dH(v). \end{aligned}$$

Отсюда и из (3) следует (5) при $s \leq i \leq k - 1$. Справедливость (5) при $i = k$ очевидна.

Доказательство леммы 2. При $s-1 \leq i \leq k-2$ из (9) имеем:

$$\begin{aligned} q_{i+1}^k - q_i^k &= \frac{C_2}{2\mu s} \sum_{m=1}^{i+2-s} m(m-1)r_m + \\ &+ \frac{C_2}{2\mu s} (i+2-s)(i+3-s)r_{i+3-s} - \frac{C_2(i+2-s)}{\mu s} \sum_{m=1}^{i+2-s} mr_m - \\ &- \frac{C_2}{\mu s} (i+2-s)(i+3-s)r_{i+3-s} - \\ &- \frac{C_2(i+2-s)(i+3-s)}{2\mu s} \sum_{m=i+4-s}^{\infty} r_m - \\ &- \frac{C_2}{2\mu s} \sum_{m=1}^{i+2-s} m(m-1)r_m + \\ &+ \frac{C_2(i+1-s)}{\mu s} \sum_{m=1}^{i+2-s} mr_m + \\ &+ \frac{C_2(i+1-s)(i+2-s)}{2\mu s} \sum_{m=i+4-s}^{\infty} r_m + \\ &+ \frac{C_2}{2\mu s} (i+1-s)(i+2-s)r_{i+3-s} = \\ &= -\frac{C_2}{\mu s} \sum_{m=1}^{i+2-s} mr_m - \frac{C_2(i+2-s)}{\mu s} \sum_{m=i+3-s}^{\infty} r_m. \end{aligned}$$

Доказательство леммы 3.

Первое соотношение в лемме 3 следует из (2). Теперь из (1) имеем:

$$\begin{aligned} g^k - g^{k+1} &= \sum_{i=0}^k \pi_i^k q_i^k - \sum_{i=0}^{k+1} \pi_i^{k+1} q_i^{k+1} = \\ &= (C_0 - C_3) \bar{v} \sum_{i=0}^{s-2} \pi_i^k + \sum_{i=s-1}^k \pi_i^k q_i^k - \sum_{i=s}^{k+1} \pi_i^{k+1} q_i^{k+1} - \\ &- (C_0 - C_3) \bar{v} \sum_{i=0}^{s-1} \pi_i^{k+1} = (C_0 - C_3) \bar{v} (Q_{s-2}^k - Q_{s-1}^{k+1}) + \\ &+ \sum_{i=s-1}^k \pi_i^k q_i^k - A_{k+1} \sum_{i=s}^{k+1} \pi_{i-1}^k q_i^{k+1}. \end{aligned}$$

Обозначим $\Delta_{i+1}^{k+1} = q_{i+1}^{k+1} - q_i^{k+1}$, $s-1 \leq i \leq k$. Заметим, что $q_i^{k+1} = q_i^k$, $0 \leq i \leq k-1$, $q_{k+1}^{k+1} = q_k^{k+1} - C_1 - C_0 = q_k^k + \Delta_k^{k+1}$. Имеем:

$$\begin{aligned} g^k - g^{k+1} &= (C_0 - C_3) \bar{v} (Q_{s-2}^k - Q_{s-1}^{k+1}) + \sum_{i=s-1}^k \pi_i^k q_i^k - \\ &- A_{k+1} \left(\sum_{i=s-1}^{k-1} \pi_i^k (q_i^k + \Delta_{i+1}^{k+1}) + \pi_k^k (q_k^k + \Delta_k^{k+1}) \right) = \\ &= \frac{Q_{s-1}^{k+1} - Q_{s-2}^k}{1 - Q_{s-2}^k} \left(g^k - (C_0 - C_3) \bar{v} - \right. \\ &\left. - \frac{1 - Q_{s-1}^{k+1}}{Q_{s-1}^{k+1} - Q_{s-2}^k} \left(\sum_{i=s-1}^{k-1} \pi_i^k \Delta_{i+1}^{k+1} + \pi_k^k \Delta_k^{k+1} \right) \right). \end{aligned}$$

Используя (6), для Δ_{i+1}^{k+1} получим

$$\begin{aligned} \Delta_{i+1}^{k+1} &= -\frac{C_2}{\mu s} \sum_{m=1}^{i+2-s} mr_m - \frac{C_2(i+2-s)}{\mu s} \sum_{m=i+3-s}^{\infty} r_m = \\ &= \left(\frac{1}{\mu s} \sum_{m=i+3-s}^{\infty} (m-i-2+s)r_m - \bar{v} \right) C_2, \end{aligned}$$

$$s-1 \leq i \leq k-1;$$

$$\Delta_k^k = -C_1 - C_0;$$

$$\Delta_k^{k+1} = -C_2 \left(\bar{v} - \frac{1}{\mu s} \sum_{m=i+2-s}^{\infty} (m-k-1+s)r_m \right).$$

Отсюда

$$\begin{aligned} g^k - g^{k+1} &= \frac{Q_{s-1}^{k+1} - Q_{s-2}^k}{1 - Q_{s-2}^k} \left(g^k - (C_0 - C_3) \bar{v} + \right. \\ &+ C_2 \frac{1 - Q_{s-2}^{k+1}}{Q_{s-1}^{k+1} - Q_{s-2}^k} \left(\pi_{k+1}^{k+1} \times \right. \\ &\times \left. \left(\bar{v} - \frac{1}{\mu s} \sum_{m=k+2-s}^{\infty} (m-k-1+s)r_m \right) + \right. \\ &\left. \left. + \sum_{i=s}^k \pi_i^{k+1} \left(\bar{v} - \frac{1}{\mu s} \sum_{m=k+2-s}^{\infty} (m-i-1+s)r_m \right) \right) \right). \end{aligned}$$

Подставляя в последнее равенство (4), получаем (7).

Доказательство леммы 4.

Положим $f_k = (1 - A_{k+1}) F(k)$. Тогда

$$\begin{aligned} f_k &= \sum_{i=s}^{k+1} \pi_i^{k+1} \bar{v} - \frac{1}{\mu s} \sum_{i=s}^k \pi_i^{k+1} \sum_{m=i+2-s}^{\infty} (m-i-1+s)r_m - \\ &- \frac{1}{\mu s} \pi_{k+1}^{k+1} \sum_{m=i+2-s}^{\infty} (m-k-1+s)r_m = \\ &= \frac{1}{\mu s} \sum_{i=s}^k \pi_i^{k+1} \sum_{m=1}^{i+1-s} mr_m + \frac{1}{\mu s} \pi_{k+1}^{k+1} \sum_{m=1}^{k+1-s} mr_m + \\ &+ \sum_{i=s}^k \pi_i^{k+1} \left(1 - \sum_{m=0}^{i+1-s} r_m \right) \frac{i+1-s}{\mu s} + \\ &+ \pi_{k+1}^{k+1} \left(1 - \sum_{m=0}^{k+1-s} r_m \right) \frac{k+1-s}{\mu s}. \end{aligned}$$

Заметим, что правая часть последнего равенства выражает среднее значение длительности времени, в течение которого в произвольно взятом состоянии при стратегии $k+1$ все приборы заняты.

Отсюда следует, что f_k возрастает по k . Докажем, что A_{k+1} возрастает по $k \geq s$. Далее из-за громоздкости доказательства, полученного для рекуррентного входящего потока, приводим аналогичное, но гораздо более короткое доказательство для пуассоновского входящего потока. Пусть λ — его интенсивность. Известно, что

$$\pi_j^k = \begin{cases} \pi_0^k \frac{\alpha^j}{j!}, & 1 \leq j \leq s; \\ \pi_0^k \frac{\alpha^s}{s!} \left(\frac{\alpha}{s} \right)^{s-j}, & s \leq j \leq k; \end{cases}$$

$$\pi_0^k = \left(1 + \sum_{j=1}^s \frac{\alpha^j}{j!} + \frac{\alpha^s}{s!} \sum_{l=1}^{k-s} \left(\frac{\alpha}{s} \right)^l \right)^{-1}, \quad \alpha = \frac{\lambda}{\mu}.$$

Отсюда

$$\begin{aligned} & \frac{1 - Q_{s-1}^{k+2}}{1 - Q_{s-2}^{k+1}} - \frac{1 - Q_{s-1}^{k+1}}{1 - Q_{s-2}^k} = \\ & = \frac{\pi_0^{k+2} (\alpha^s / s!) \sum_{i=0}^{k+2-s} (\alpha/s)^i}{\pi_0^{k+1} \left(\alpha^{s-1} / ((s-1)!) + (\alpha^s / s!) \sum_{i=0}^{k+1-s} (\alpha/s)^i \right)} - \\ & - \frac{\pi_0^{k+1} (\alpha^s / s!) \sum_{i=0}^{k+1-s} (\alpha/s)^i}{\pi_0^k \left(\alpha^{s-1} / ((s-1)!) + (\alpha^s / s!) \sum_{i=0}^{k-s} (\alpha/s)^i \right)}. \end{aligned} \quad (10)$$

Далее

$$\begin{aligned} & \frac{\alpha^s}{s!} \sum_{i=0}^{k+2-s} \left(\frac{\alpha}{s} \right)^i \left(\frac{\alpha^{s-1}}{(s-1)!} + \frac{\alpha^s}{s!} \sum_{i=0}^{k-s} \left(\frac{\alpha}{s} \right)^i \right) = \\ & = \frac{\alpha^s}{s!} \sum_{i=0}^{k+1-s} \left(\frac{\alpha}{s} \right)^i \left(\frac{\alpha^{s-1}}{(s-1)!} + \frac{\alpha^s}{s!} \sum_{i=0}^{k+1-s} \left(\frac{\alpha}{s} \right)^i \right). \end{aligned} \quad (11)$$

Действительно, разность левой и правой частей (11) равна

$$\begin{aligned} & \frac{\alpha^s}{s!} \frac{\alpha^{s-1}}{(s-1)!} \left(\sum_{i=0}^{k+2-s} \left(\frac{\alpha}{s} \right)^i \left(1 + \sum_{i=0}^{k-s} \left(\frac{\alpha}{s} \right)^{i+1} \right) - \right. \\ & \left. - \sum_{i=0}^{k+1-s} \left(\frac{\alpha}{s} \right)^i \left(1 + \sum_{i=0}^{k+1-s} \left(\frac{\alpha}{s} \right)^{i+1} \right) \right) = \\ & = \frac{\alpha^s}{s!} \frac{\alpha^{s-1}}{(s-1)!} \left(\left(\frac{\alpha}{s} \right)^{k+2-s} + \right. \\ & \left. + \frac{\alpha}{s} \left(\left(\frac{\alpha}{s} \right)^{k+1-s} \left(\frac{\alpha}{s} - 1 \right) \sum_{i=0}^{k+1-s} \left(\frac{\alpha}{s} \right)^i - \right. \right. \\ & \left. \left. - \left(\frac{\alpha}{s} \right)^{2k+3-2s} \right) \right) = 0 \end{aligned}$$

Следовательно, знак правой части в (10) совпадает со знаком $\pi_0^{k+2} \pi_0^k - (\pi_0^{k+1})^2$. Знак последнего выражения противоположен знаку разности:

$$\begin{aligned} & \left(1 + \sum_{j=1}^s \frac{\alpha^j}{j!} + \frac{\alpha^s}{s!} \sum_{l=1}^{k+2-s} \left(\frac{\alpha}{s} \right)^l \right) \times \\ & \times \left(1 + \sum_{j=1}^s \frac{\alpha^j}{j!} + \frac{\alpha^s}{s!} \sum_{l=1}^{k-s} \left(\frac{\alpha}{s} \right)^l \right) - \\ & - \left(1 + \sum_{j=1}^s \frac{\alpha^j}{j!} + \frac{\alpha^s}{s!} \sum_{l=1}^{k+1-s} \left(\frac{\alpha}{s} \right)^l \right)^2. \end{aligned}$$

Эта разность равна:

$$\begin{aligned} & \frac{\alpha^s}{s!} \left(\frac{\alpha}{s} \right)^{k+1-s} \left(\frac{\alpha}{s} - 1 \right) \left(1 + \sum_{j=1}^s \frac{\alpha^j}{j!} + \right. \\ & \left. + \frac{\alpha^s}{s!} \sum_{l=1}^{k+1-s} \left(\frac{\alpha}{s} \right)^l \right) - \left(\frac{\alpha^s}{s!} \right)^2 \left(\frac{\alpha}{s} \right)^{2k+3-2s} = \\ & = \frac{\alpha^s}{s!} \left(\frac{\alpha}{s} \right)^{k+1-s} \left(\sum_{j=1}^s \frac{\alpha^j}{s(j-1)!} - \sum_{j=1}^s \frac{\alpha^j}{j!} \right). \end{aligned}$$

Но

$$\sum_{j=1}^s \frac{\alpha^j}{s(j-1)!} - \sum_{j=1}^s \frac{\alpha^j}{j!} < 0.$$

Итак, лемма 4 полностью доказана.

Доказательство теоремы 1 аналогично доказательству теоремы 1 в [7].

Литература

1. *Welzl M.* Network congestion control: Managing internet traffic. — New York, NY, USA: Wiley, 2005. 282 p.
2. *Жерновский Ю. В.* Решение задач оптимального синтеза для некоторых марковских моделей обслуживания // Информационные процессы, 2010. Т. 10. № 3. С. 257–274.
3. *Коновалов М. Г.* Об одной задаче оптимального управления нагрузкой на сервер // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 34–43.
4. *Агаларов Я. М.* Пороговая стратегия ограничения доступа к ресурсам в системе массового обслуживания $M|D|1$ с функцией штрафов за несвоевременное обслуживание заявок // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 55–64.
5. *Гришунина Ю. Б.* Оптимальное управление очередью в системе $M|G|1|\infty$ с возможностью ограничения приема заявок // Автоматика и телемеханика, 2015. № 3. С. 79–93.
6. *Агаларов Я. М.* Максимизация среднего стационарного дохода системы массового обслуживания типа $M|G|1$ // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 25–32.
7. *Агаларов Я. М., Шоргин В. С.* Об одной задаче максимизации дохода системы массового обслуживания типа $G|M|1$ с пороговым управлением очередью // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 55–64.
8. *Karlin S.* A first course in stochastic processes. — New York – London: Academic Press, 1968. 502 p.
9. *Бочаров П. П., Печинкин А. В.* Теория массового обслуживания. — М.: РУДН, 1995. 529 с.

Поступила в редакцию 05.11.18

ON THE UNIMODALITY OF THE INCOME FUNCTION OF A TYPE $G|M|s$ QUEUEING SYSTEM WITH CONTROLLED QUEUE

Ya. M. Agalarov¹ and V. G. Ushakov^{1,2}

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

²Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninsky Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: The problem of maximizing the average income in a queueing system of type $G|M|s$ on a set of pure stationary threshold strategies with single point switching access restriction mode is considered. The income function depends on the following parameters, measured in value units: the fee received for servicing requests, the cost of maintenance of the device, the deduction of income for the delay applications in the queue, the penalty for unserved applications. It is proved that the income function is unimodal on the set of considered threshold strategies. An algorithm for calculating the optimal threshold value and the corresponding maximum value income is proposed. The results of the computational experiment that illustrate the work of the proposed algorithm are given.

Keywords: multichannel queueing system; threshold management; maximizing income

DOI: 10.14357/19922264190108

References

1. Welzl, M. 2005. *Network congestion control*. New York, NY: Wiley. 282 p.
2. Zhernovij, Ju. V. 2010. Reshenie zadach optimal'nogo sinteza dlya nekotorykh markovskikh modeley obsluzhivaniya [Solution of optimum synthesis problem for some Markov models of service]. *Informatsionnye processy* [Information Processes] 10(3):257–274.
3. Konovalov, M. G. 2013. Ob odnoy zadache optimal'nogo upravleniya nagruzkoy na server [About one task of overload control]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):34–43.
4. Agalarov, Ya. M. 2015. Porogovaya strategiya ogranicheniya dostupa k resursam v sisteme massovogo obsluzhivaniya $M/D/1$ s funktsiey shtrafov za nesvoevremennoe obsluzhivanie zayavok [The threshold strategy for restricting access in the $M/D/1$ queueing system with penalty function for late service]. *Informatika i ee Primeneniya — Inform. Appl.* 9(3):55–64.
5. Grishunina, Y. B. 2015. Optimal control of queue in the $M|G|1|\infty$ system with possibility of customer admission restriction. *Automat. Rem. Contr.* 76(3):433–445.
6. Agalarov, Ya. M. 2017. Maksimizatsiya srednego statsionarnogo dokhoda sistemy massovogo obsluzhivaniya tipa $M/G/1$ [Maximization of average stationary profit in $M/G/1$ queueing system]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):25–32.
7. Agalarov, Ya. M., and V. S. Shorgin. 2017. Ob odnoy zadache maksimizatsii dokhoda sistemy massovogo obsluzhivaniya tipa $G/M/1$ s porogovym upravleniem ochered'yu [About the problem of profit maximization in $G/M/1$ queueing systems with threshold control of the queue]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):55–64.
8. Karlin, S. 1968. *A first course in stochastic processes*. New York – London: Academic Press. 502 p.
9. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.

Received November 5, 2018

Contributors

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

Ushakov Vladimir G. (b. 1952) — 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 Leninsky 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; vgushakov@mail.ru

АПРИОРНЫЕ ФРЕШЕ И МАСШТАБИРОВАННОЕ ОБРАТНОЕ ХИ-РАСПРЕДЕЛЕНИЕ В БАЙЕСОВСКИХ МОДЕЛЯХ БАЛАНСА*

А. А. Кудрявцев¹, С. И. Палионная², В. С. Шоргин³

Аннотация: Статья продолжает ряд работ авторов в области моделирования систем массового обслуживания с применением байесовского подхода. Постановка задачи распространяется на более широкий класс прикладных исследований — изучение индекса баланса факторов, влияющих на функционирование системы. Предполагается, что параметры модели разделены на два класса, к одному из которых относятся те, что оказывают позитивное влияние на функционирование сложного агрегата, а к другому — препятствующие функционированию. Эффективность работы исследуемой системы, естественно, зависит от соотношения позитивного и негативного факторов, называемого индексом баланса. В рамках байесовского подхода предполагается, что факторы суть случайные величины с известными априорными распределениями. Во многих прикладных задачах свою адекватность демонстрируют распределения из гамма-класса. В статье рассматриваются смеси частных случаев обобщенного гамма-распределения — распределение Фреше и масштабированное обратное хи-распределение.

Ключевые слова: байесовский подход; масштабированное обратное хи-распределение; распределение Фреше; гамма-экспоненциальная функция; модели баланса; смешанные распределения

DOI: 10.14357/19922264190109

1 Введение

В современном мире процессы, координирующие различные сферы человеческой деятельности, настолько усложнились, что определение критериев эффективности систем путем детерминированного анализа практически не представляется возможным. В связи с этим вводятся различные показатели, рейтинги и индексы, позволяющие оценить состояние системы, сэкономив при этом временные, материальные и финансовые ресурсы.

Все факторы, так или иначе влияющие на функционирование системы, можно разделить на позитивные (способствующие корректному функционированию системы) и негативные (препятствующие работе системы). Далее для удобства записи будем называть их *p*- и *n*-факторами соответственно. Однако на эффективность работы системы влияют не столько абсолютные значения этих факторов, сколько их соотношение. Поэтому для исследования функционирования сложных агрегатов, включая сложные модифицируемые информационно-телекоммуникационные системы, естествен-

но перейти к рассмотрению отношения *n*- к *p*-фактору модели. Такая величина $\rho = \lambda/\mu$, где λ и μ соответственно обозначают *n*- и *p*-факторы, называется *индексом баланса* [1].

С течением времени меняется состояние среды, окружающей систему. По этой причине *n*- и *p*-факторы рассматриваемой модели также изменяются, причем зачастую непредсказуемым образом. Это дает предпосылки для рассмотрения факторов, а следовательно, и индексов баланса как случайных величин. Стоит также отметить, что глобальные изменения окружающей среды происходят крайне редко, поэтому можно считать, что законы, оказывающие влияние на факторы и индексы конкретной модели, остаются неизменными. Это позволяет перейти к применению байесовского метода с заданными априорными распределениями исходных параметров.

Ниже приводятся результаты для вероятностных характеристик индекса баланса в случае, когда в качестве априорных распределений факторов рассматриваются распределение Фреше и масштабированное обратное хи-распределение.

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00577).

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, nubigena@mail.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, sofipalioonnaya@gmail.com

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, VShorgin@ipiran.ru

2 Основные результаты

Рассмотрим гамма-экспоненциальную функцию [2]:

$$\text{Ge}_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta), \quad x \in \mathbb{R},$$

$$\alpha \geq 0, \quad \beta > 0.$$

Для дальнейших вычислений понадобится следующее утверждение.

Лемма 1. Пусть $\alpha, \theta, r, s > 0$. Тогда

$$\int_0^{\infty} y^{-r-1} e^{-(\alpha/y)^s - (\theta/y)^2} dy =$$

$$= \begin{cases} \frac{\theta^{-r}}{2} \text{Ge}_{s/2, r/2} \left(- \left(\frac{\alpha}{\theta} \right)^s \right), & s < 2; \\ \frac{\alpha^{-r}}{s} \text{Ge}_{2/s, r/s} \left(- \left(\frac{\theta}{\alpha} \right)^2 \right), & s > 2; \\ \frac{(\alpha^2 + \theta^2)^{-r/2}}{2} \Gamma \left(\frac{r}{2} \right), & s = 2. \end{cases}$$

Доказательство. Рассмотрим случай $s < 2$. Используя теорему Лебега о предельном переходе, получаем:

$$\int_0^{\infty} y^{-r-1} e^{-(\alpha/y)^s - (\theta/y)^2} dy =$$

$$= \frac{\alpha^{-r}}{s} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^{\infty} t^{r/s+k-1} e^{-(\theta/\alpha t^s)^2} dt =$$

$$= \frac{\theta^{-r}}{2} \sum_{k=0}^{\infty} \frac{(-(\alpha/\theta)^s)^k}{k!} \int_0^{\infty} z^{(r+sk)/2-1} e^{-z} dz =$$

$$= \frac{\theta^{-r}}{2} \sum_{k=0}^{\infty} \frac{(-(\alpha/\theta)^s)^k}{k!} \Gamma \left(\frac{r+sk}{2} \right).$$

Случай $s > 2$ рассматривается аналогично. Случай $s = 2$ напрямую следует из определения гамма-функции. Лемма доказана.

Рассмотрим случайную величину ξ , имеющую масштабированное обратное хи-распределение $I\chi(q, \theta)$ с плотностью

$$f_{\xi}(x) = \frac{2\theta^q e^{-\theta^2/x^2}}{\Gamma(q/2)x^{q+1}}, \quad \theta > 0, \quad q > 0, \quad x > 0,$$

и случайную величину η , имеющую распределение Фреше $\text{Fr}(u, \alpha)$ [3] с нулевым минимумом, плотность которого имеет вид:

$$f_{\eta}(x) = \frac{u\alpha^u e^{-(\alpha/x)^u}}{x^{u+1}}, \quad \alpha > 0, \quad u > 0, \quad x > 0.$$

Легко видеть, что для случайных величин ξ и η справедливо следующее утверждение.

Лемма 2. Для случайных величин ξ и η , имеющих соответственно масштабированное обратное хи-распределение $I\chi(q, \theta)$ и распределение Фреше $\text{Fr}(u, \alpha)$, для $z \in \mathbb{R}$ выполняются соотношения:

$$E\xi^z = \frac{\theta^z \Gamma(q/2 - z/2)}{\Gamma(q/2)}, \quad z < q;$$

$$E\eta^z = \alpha^z \Gamma(1 - z/u), \quad z < u.$$

Леммы 1 и 2 дают возможность вычислить вероятностные характеристики индекса баланса ρ в байесовской модели.

Теорема 1. Пусть n -фактор λ имеет масштабированное обратное хи-распределение $I\chi(q, \theta)$, а p -фактор μ имеет распределение Фреше $\text{Fr}(u, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты индекса баланса $\rho = \lambda/\mu$ имеют вид:

$$f_{\rho}(x) =$$

$$= \begin{cases} \frac{2\theta^q}{\alpha^q \Gamma(q/2) x^{q+1}} \text{Ge}_{2/u, q/u+1} \left(- \left(\frac{\theta}{\alpha x} \right)^2 \right), & u > 2; \\ \frac{u\alpha^u x^{u-1}}{\theta^u \Gamma(q/2)} \text{Ge}_{u/2, (q+u)/2} \left(- \left(\frac{\alpha x}{\theta} \right)^u \right), & u < 2; \end{cases}$$

$$F_{\rho}(x) =$$

$$= \begin{cases} \frac{q}{2\Gamma(q/2)} \int_{(\theta/(\alpha x))^q}^{\infty} \text{Ge}_{2/u, q/u+1}(-z^{2/q}) dz, & u > 2; \\ 1 - \frac{\text{Ge}_{u/2, q/2}(-(\alpha x/\theta)^u)}{\Gamma(q/2)}, & u < 2; \end{cases}$$

$$E\rho^z = \frac{(\theta/\alpha)^z \Gamma(q/2 - z/2) \Gamma(z/u + 1)}{\Gamma(q/2)}, \quad z < q.$$

При $u = 2$ распределение индекса баланса ρ совпадает с масштабированным распределением Бурра [4] с параметрами $(2, q/2, \theta/\alpha)$.

Доказательство. Поскольку

$$f_{\rho}(x) = \int_0^{\infty} \frac{2u\alpha^u \theta^q e^{-(\theta/(xy))^2 - (\alpha/y)^u}}{\Gamma(q/2)x^{q+1}y^{q+u+1}} dy,$$

вид плотности $f_{\rho}(x)$ вытекает из леммы 1 для всех $u > 0$.

Для функции распределения ρ при $u > 2$ справедливо для $x > 0$

$$\begin{aligned}
 F_\rho(x) &= \\
 &= \frac{2(\theta/\alpha)^q}{\Gamma(q/2)} \int_0^x y^{-q-1} \text{Ge}_{2/u, q/u+1} \left(- \left(\frac{\theta}{\alpha y} \right)^2 \right) dy = \\
 &= \frac{2(\theta/\alpha)^q}{\Gamma(q/2)} \int_0^x y^{-q-1} \times \\
 &\quad \times \sum_{l=0}^{\infty} \frac{(-1)^l y^{-2l}}{(\theta/\alpha)^{-2l} l!} \Gamma \left(\frac{2l}{u} + \frac{q}{u} + 1 \right) dy = \\
 &= \frac{2}{q\Gamma(q/2)} \int_{(\theta/(\alpha x))^q}^{\infty} \sum_{l=0}^{\infty} \frac{(-z^{2/q})^l}{l!} \Gamma \left(\frac{2l}{u} + \frac{q}{u} + 1 \right) dz.
 \end{aligned}$$

В случае $u < 2$ имеем:

$$\begin{aligned}
 F_\rho(x) &= \frac{u(\alpha/\theta)^u}{\Gamma(q/2)} \times \\
 &\times \int_0^x y^{u-1} \sum_{l=0}^{\infty} \frac{(-(\alpha y/\theta)^u)^l}{l!} \Gamma \left(\frac{lu + u + q}{2} \right) dy = \\
 &= \frac{u(\alpha/\theta)^u}{\Gamma(q/2)} \sum_{l=0}^{\infty} \frac{(-1)^l}{(\theta/\alpha)^{ul} l!} \Gamma \left(\frac{lu + u + q}{2} \right) \times \\
 &\quad \times \int_0^x y^{ul+u-1} dy = \\
 &= \frac{x^u (\alpha/\theta)^u}{\Gamma(q/2)} \sum_{l=0}^{\infty} \frac{(-1)^l x^{ul}}{(\theta/\alpha)^{ul} l!} \frac{\Gamma((lu + u + q)/2)}{l + 1} = \\
 &= 1 - \frac{1}{\Gamma(q/2)} \sum_{m=0}^{\infty} \frac{(-1)^m x^{um}}{(\theta/\alpha)^{um} m!} \Gamma \left(\frac{mu + q}{2} \right).
 \end{aligned}$$

Для нахождения моментов ρ достаточно воспользоваться независимостью случайных величин λ и μ и леммой 2. Теорема доказана.

Рассмотрим симметричный случай априорных распределений.

Теорема 2. Пусть n -фактор λ имеет распределение Фреше $\text{Fr}(v, \theta)$, а p -фактор μ имеет масштабированное обратное хи-распределение $I\chi(p, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты индекса баланса $\rho = \lambda/\mu$ имеют вид:

$$\begin{aligned}
 f_\rho(x) &= \\
 &= \begin{cases} \frac{v\theta^v}{\alpha^v \Gamma(p/2) x^{v+1}} \text{Ge}_{v/2, (v+p)/2} \left(- \left(\frac{\theta}{\alpha x} \right)^v \right), & v < 2; \\ \frac{2\alpha^p x^{p-1}}{\theta^p \Gamma(p/2)} \text{Ge}_{2/v, p/v+1} \left(- \left(\frac{\alpha x}{\theta} \right)^2 \right), & v > 2; \end{cases} \\
 F_\rho(x) &= \\
 &= \begin{cases} \frac{1}{\Gamma(p/2)} \int_{(\theta/(\alpha x))^v}^{\infty} \text{Ge}_{v/2, (v+p)/2}(-z) dz, & v < 2; \\ \frac{2\alpha^p x^p}{v\theta^p \Gamma(p/2)} \text{Ge}_{2/v, p/v} \left(- \left(\frac{\alpha x}{\theta} \right)^2 \right), & v > 2; \end{cases} \\
 E\rho^z &= \frac{(\theta/\alpha)^z \Gamma(p/2 + z/2) \Gamma(1 - z/v)}{\Gamma(p/2)}, \quad z < v.
 \end{aligned}$$

При $v = 2$ распределение индекса баланса ρ совпадает с распределением Дагума [5] с параметрами $(2, \theta/\alpha, p/2)$.

Доказательство. Аналогично предыдущей теореме для получения выражения для плотности ρ при всех $v > 0$ достаточно воспользоваться леммой 1.

Для функции распределения ρ при $v < 2$ справедливо для $x > 0$:

$$\begin{aligned}
 F_\rho(x) &= \frac{v(\theta/\alpha)^v}{\Gamma(p/2)} \int_0^x y^{-v-1} \times \\
 &\quad \times \sum_{k=0}^{\infty} \frac{(-\theta^v)^k}{(\alpha y)^{vk} k!} \Gamma \left(\frac{kv + v + p}{2} \right) dy = \\
 &= \frac{1}{\Gamma(p/2)} \int_{(\theta/\alpha)^v x^{-v}}^{\infty} \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \Gamma \left(\frac{kv + v + p}{2} \right) dz.
 \end{aligned}$$

В случае $v > 2$ имеем:

$$\begin{aligned}
 F_\rho(x) &= \frac{2(\alpha/\theta)^p}{\Gamma(p/2)} \int_0^x y^{p-1} \times \\
 &\quad \times \sum_{k=0}^{\infty} \frac{(-1)^k (\alpha y)^{2k}}{\theta^{2k} k!} \Gamma \left(\frac{2k}{v} + \frac{p}{v} + 1 \right) dy = \\
 &= \frac{2(\alpha/\theta)^p}{\Gamma(p/2)} \sum_{k=0}^{\infty} \frac{(-1)^k \alpha^{2k}}{\theta^{2k} k!} \Gamma \left(\frac{2k}{v} + \frac{p}{v} + 1 \right) \frac{x^{2k+p}}{2k+p} = \\
 &= \frac{2(\alpha/\theta)^p x^p}{v\Gamma(p/2)} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(\theta/\alpha)^{2k} k!} \Gamma \left(\frac{2k+p}{v} \right).
 \end{aligned}$$

Соотношение для моментов следует из леммы 2 и независимости случайных величин λ и μ . Теорема доказана.

Замечание. Очевидно, что теоремы 1 и 2 несложно переформулировать для случая смеси априорного распределения Фреше с масштабированным обратным χ^2 -распределением, имеющим плотность

$$f(x) = \frac{\theta^{q/2} e^{-\theta/x}}{\Gamma(q/2)x^{q/2+1}}, \quad \theta > 0, \quad q > 0, \quad x > 0,$$

и являющимся частным случаем обратного гамма-распределения, рассмотренного в [6].

3 Заключение

Методы, использованные при получении результатов данной работы, дают возможность исследовать отношения случайных величин, имеющих обратные распределения из гамма-класса, в терминах гамма-экспоненциальной функции $Ge_{\alpha, \beta}(x)$ в случаях, когда распределение смеси отлично от обобщенного бета-распределения второго рода.

Литература

1. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
2. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
3. Fréchet M. Sur la loi de probabilité de l'écart maximum // Ann. Soc. Polonaise Math., 1927. Vol. 6. P. 93–116.
4. Burr I. W. Cumulative frequency functions // Ann. Math. Stat., 1942. Vol. 13. No. 2. P. 215–232.
5. Dagum C. A new model of personal income-distribution-specification and estimation // Econ. Appl., 1977. Vol. 30. No. 3. P. 413–437.
6. Кудрявцев А. А., Палионная С. И., Шоргин В. С. Априорное обратное гамма-распределение в байесовских моделях массового обслуживания // Системы и средства информатики, 2018. Т. 28. № 4. С. 52–58.

Поступила в редакцию 28.12.18

A PRIORI FRECHET AND SCALED INVERSE CHI DISTRIBUTION IN BAYESIAN BALANCE MODELS

A. A. Kudryavtsev¹, S. I. Palionnaia¹, and V. S. Shorgin²

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: This article continues a series of authors’ works in the field of modeling queuing systems using the Bayesian approach. The problem’s statement is extended to a wider class of applied research — the study of the balance index of factors affecting the functioning of the system. It is assumed that the model parameters are divided into two classes, one of which includes factors that have a positive impact on the functioning of a complex aggregate and the other includes those that interferes with the functioning. The effectiveness of the system under study, of course, depends on the ratio of positive and negative factors, called the balance index. In the framework of the Bayesian approach, it is assumed that the factors are random variables with known *a priori* distributions. In a wide range of applied problems, it is reasonable to use gamma-type distributions. In this paper, the mixtures of particular generalized gamma distribution cases — the Fréchet distribution and the scaled inverse chi distribution — are considered.

Keywords: Bayesian approach; scaled inverse chi distribution; Fréchet distribution; gamma-exponential function; balance models; mixed distributions

DOI: 10.14357/19922264190109

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

References

1. Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
2. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
3. Fréchet, M. 1927. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polonaise Math.* 6:93–116.
4. Burr, I. W. 1942. Cumulative frequency functions. *Ann. Math. Stat.* 13(2):215–232.
5. Dagum, C. 1977. A new model of personal income-distribution-specification and estimation. *Econ. Appl.* 30(3):413–437.
6. Kudryavtsev, A. A., S. I. Palionnaia, and V. S. Shorgin. 2018. Apriornoe obratnoe gamma-raspredelenie v bayesovskikh modelyakh massovogo obsluzhivaniya [A priori inverse gamma distribution in Bayesian queuing models]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):52–58.

Received December 28, 2018

Contributors

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

Palionnaia Sofia I. (b. 1995) — student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; sofiapalionnaya@gmail.com

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, 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; VShorgin@ipiran.ru

ПОЛИНОМИАЛЬНЫЕ АЛГОРИТМЫ ВЫЧИСЛЕНИЯ ЛОКАЛЬНЫХ АФФИННОСТЕЙ КВАДРАТИЧНЫХ БУЛЕВЫХ ФУНКЦИЙ*

О. А. Логачев¹, А. А. Сукаев², С. Н. Федоров³

Аннотация: Аффинная нормальная форма позволяет рассматривать произвольную булеву функцию на определенных плоскостях (так называемых локальных аффинностях) как аффинную. Данное представление — по сути, аффинная аппроксимация — булевых функций может помочь в решении систем нелинейных уравнений над полем из двух элементов. Задача решения таких систем (специального вида), среди прочего, используется в ряде методов синтеза и анализа средств обеспечения информационной безопасности. В статье описывается способ нахождения локальных аффинностей для квадратичных булевых функций, основанный на теореме Диксона. Тем самым решается задача построения аффинных нормальных форм для таких функций. Кроме того, обсуждаются вопросы эффективности подобных алгоритмов. Основная цель данной статьи — подготовить базу для готовящейся к публикации работы, предлагающей метод решения систем квадратичных булевых уравнений с помощью «аппроксимирования» соответствующих функций их аффинными нормальными формами.

Ключевые слова: булева функция; система квадратичных булевых уравнений; разбиение векторного пространства; плоскость; локальная аффинность; теорема Диксона; аффинная нормальная форма; алгебраический криптоанализ

DOI: 10.14357/19922264190110

1 Введение

Центральная идея алгебраического криптоанализа состоит в том, чтобы описать используемые в анализируемой криптосхеме преобразования системы алгебраических уравнений (с некоторой секретной информацией в качестве неизвестных) над конечным полем и затем решить эту систему. В данной статье рассматриваются только булевы системы уравнений, хотя часть представленных здесь результатов может иметь место и для систем алгебраических уравнений над произвольными конечными полями.

Из теории сложности вычислений известно, что вычислительная задача определения совместности систем нелинейных булевых уравнений является NP-полной [1, 2], а вычислительная задача решения систем нелинейных булевых уравнений является NP-трудной [1, 2]. Однако в специальных случаях эти задачи могут решаться эффективно (см., например, [2, 3]).

Кроме того, существуют полиномиальные алгоритмы построения по произвольной системе уравнений системы с фиксированной алгебраической степенью [4, § 11.4.2], что позволяет, в частности, ограничиться рассмотрением только квадратичных систем уравнений.

Можно выделить несколько основных классов методов, используемых в криптоанализе для решения (или оценки трудоемкости решения) систем полиномиальных булевых уравнений: использование базисов Грёбнера [4, section 12.2], применение программных систем поиска выполняющего набора булевой формулы (SAT-solvers) [5], вероятностные и теоретико-кодовые методы [6], а также методы линеаризации [4, section 12.3]. Основная идея методов линеаризации состоит в применении «линейных» методов к нелинейным системам, т. е. в построении систем линейных уравнений, решение которых дает возможность найти решение исходной нелинейной системы.

Важным параметром метода линеаризации служит число переменных в синтезируемых линейных системах уравнений. Как правило, речь идет об увеличении (не более чем полиномиальном) количества переменных. Метод, основанный на рассмотренных в данной работе идеях, по своей сути, осуществляет линеаризацию, но при этом он оставляет число переменных неизменным.

Этот метод решения квадратичных систем булевых уравнений использует локальные аффинности уравнений системы и состоит из двух этапов. Первый этап (предварительный) содержательно пред-

* Работа выполнена при частичной поддержке РФФИ (проект 18-29-03124 мк).

¹Московский государственный университет им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, logol@iisi.msu.ru

²Московский государственный университет им. М. В. Ломоносова, asukaev@gmail.com

³Московский государственный университет им. М. В. Ломоносова, s.n.feodorov@yandex.ru

ставляет собой описание семейств локальных аффинностей уравнений. Второй этап метода заключается собственно в решении исходной системы посредством анализа систем линейных уравнений, полученных с помощью этих локальных аффинностей.

Настоящая работа (в силу ограниченности объема публикации) посвящена исследованию первого этапа и, в частности, вопросам его эффективности. Результаты исследований с оценкой эффективности и описанием параметров второго этапа предлагаемого метода предполагается опубликовать в одном из следующих выпусков журнала.

2 Необходимые понятия и обозначения

В данной работе булев куб $\{0, 1\}^n$ отождествляется с n -мерным векторным пространством V_n над полем из двух элементов \mathbb{F}_2 . Векторы из V_n будет удобнее записывать *строками* длины n . Значок \top используется для операции транспонирования матриц. Всюду далее x обозначает вектор (x_1, x_2, \dots, x_n) .

Знак \oplus будет использоваться для записи суммы по модулю 2 булевых переменных и операций сложения в \mathbb{F}_2 и покомпонентного сложения в V_n .

Множество всех невырожденных аффинных преобразований (отображений в себя) пространства V_n обозначается через $GA(V_n)$. В матричном представлении действие элемента $\alpha \in GA(V_n)$ на векторах пространства имеет вид $\alpha(x) = xA \oplus b$, где x пробегает V_n ; A — невырожденная $(n \times n)$ -матрица над \mathbb{F}_2 ; $b \in V_n$.

Множество всех булевых функций от n переменных обозначим через

$$\mathcal{F}_n = \{f: V_n \rightarrow \mathbb{F}_2\}.$$

Как известно, произвольную булеву функцию f от переменных x_1, \dots, x_n можно представить (единственным образом) в виде полинома Жегалкина:

$$f(x) = \bigoplus_{\varepsilon \in \{0,1\}^n} a_\varepsilon x^\varepsilon,$$

где $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, $a_\varepsilon \in \mathbb{F}_2$ и $x^\varepsilon = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ (считаем, $x_i^0 = 1$, $x_i^1 = x_i$). Далее под булевой функцией будет, как правило, подразумеваться ее запись в виде полинома.

Если φ — некоторое преобразование пространства V_n , то его действие на функцию $f \in \mathcal{F}_n$ будем определять и обозначать так: $f^\varphi(x) = f(\varphi(x))$.

¹В алгебре такие функции называют аффинно-квадратичными. Квадратичными при этом называют функции, представляемые однородными полиномами второй степени.

В частности, в случае аффинных преобразований пространства будет рассматриваться множество $\text{Orb}_f(GA(V_n)) = \{f^\varphi \mid \varphi \in GA(V_n)\}$ — орбита функции f относительно действия группы $GA(V_n)$. Имея в виду, что произведение $\alpha_1\alpha_2$ элементов из $GA(V_n)$ есть композиция $\alpha_1 \circ \alpha_2(x) = \alpha_1(\alpha_2(x))$, заметим, что действие $\alpha_1\alpha_2$ на произвольную функцию $f \in \mathcal{F}_n$ корректно определять следующим образом:

$$f^{\alpha_1\alpha_2}(x) = (f^{\alpha_1})^{\alpha_2}(x) = f^{\alpha_1}(\alpha_2(x)) = f(\alpha_1\alpha_2(x)),$$

поскольку α_i действуют на булеву функцию преобразованием ее аргумента.

Алгебраической степенью булевой функции f от n переменных называют величину

$$\deg f = \max \left\{ \sum_{i=1}^n \varepsilon_i \mid a_\varepsilon = 1 \right\}$$

(суммирование — в \mathbb{Z}), т.е. максимальное число различных переменных в мономах данного представления.

В множестве \mathcal{F}_n всех булевых функций от n переменных выделим подмножество

$$\mathcal{A}_n = \{f \in \mathcal{F}_n \mid \deg f \leq 1\}.$$

Составляющие это подмножество функции называются линейными (в математической логике и кибернетике) или аффинно-линейными (в алгебре), однако по сложившейся в криптологии традиции в данной работе они называются *аффинными*, т.е. понимаются как частный случай аффинного отображения n -мерного пространства в одномерное.

Булеву функцию f с $\deg f \leq 2$ будем называть *квадратичной*¹. По определению квадратичная функция $f \in \mathcal{F}_n$ (ее полином Жегалкина) имеет вид:

$$f(x) = \bigoplus_{1 \leq i < j \leq n} q_{ij} x_i x_j \oplus \bigoplus_{1 \leq k \leq n} l_k x_k \oplus c,$$

где $q_{ij}, l_k, c \in \mathbb{F}_2$.

В настоящей работе рассматриваются системы уравнений

$$\left. \begin{aligned} f_1(x_1, \dots, x_n) &= 0; \\ f_2(x_1, \dots, x_n) &= 0; \\ &\vdots \\ f_m(x_1, \dots, x_n) &= 0 \end{aligned} \right\} \quad (1)$$

с квадратичными булевыми функциями f_i , $1 \leq i \leq m$, и $m > n$.

В матричном виде квадратичная функция записывается следующим образом:

$$f(x) = xQ_f x^T \oplus l_f x^T \oplus c,$$

где Q_f — верхнетреугольная $(n \times n)$ -матрица с нулевой главной диагональю; $l_f \in \mathbb{F}_2^n$; $c \in \mathbb{F}_2$. Рассматривают также симметричную матрицу

$$\tilde{Q}_f = Q_f \oplus Q_f^T.$$

Она определяет билинейную форму

$$q_f(u, v) = u\tilde{Q}_f v^T = f(u \oplus v) \oplus f(u) \oplus f(v) \oplus c,$$

называемую ассоциированной с квадратичной функцией f .

Булева билинейная форма $q(u, v)$, $u, v \in V_n$, удовлетворяющая условиям

$$q(u, u) = 0; \quad q(u, v) = q(v, u),$$

называется *симплектической*. Такие билинейные формы находятся во взаимно однозначном соответствии с булевыми симметричными матрицами с нулевой главной диагональю, называемыми *симплектическими матрицами*.

Таким образом, для произвольной квадратичной булевой функции f матрица \tilde{Q}_f — симплектическая. Также очевидно, что билинейная форма $q_f(u, v)$, ассоциированная с f , является симплектической.

Предложение 1 [6, лемма 3.3.1; 7, § 15.2, лемма 3]. *Ранг симплектической матрицы четен.*

Плоскость π в V_n — это множество вида $v + L$, где v и L — соответственно вектор и подпространство пространства V_n . Другими словами, плоскость — аффинное подпространство в V_n . Размерность плоскости совпадает с размерностью соответствующего подпространства: $\dim \pi = \dim L$. Как известно, любая плоскость является решением некоторой системы линейных уравнений, и наоборот: решение произвольной системы линейных уравнений — плоскость в соответствующем пространстве.

Сужение функции $f \in \mathcal{F}_n$ на плоскость π будем обозначать через $f|_\pi$. Таким образом, $f|_\pi: \pi \rightarrow \mathbb{F}_2$ и $f|_\pi(u) = f(u)$ для всех $u \in \pi$.

3 Локальная аффинность и аффинная нормальная форма булевой функции

В этом разделе вводятся понятия, связанные с представлением произвольной булевой функции

совокупностью аффинных функций, заданных для определенных плоскостей в векторном пространстве. Более общее изложение этой теории можно найти в работе [8].

Локальной аффинностью функции $f \in \mathcal{F}_n$ будем называть такую плоскость π , что $f|_\pi$ можно продолжить до аффинной функции, т. е. существует $l \in \mathcal{A}_n$ со свойством $f|_\pi = l|_\pi$. Очевидно, для любой булевой функции существует разбиение пространства V_n на ее локальные аффинности.

Возьмем произвольное разбиение $\Pi = \{\pi_1, \dots, \pi_\lambda\}$ пространства V_n на плоскости, являющиеся локальными аффинностями булевой функции f от n переменных. Будем называть *аффинной нормальной формой* функции f выражение вида

$$f(x) = \bigoplus_{j=1}^{\lambda} \chi_{\pi_j}(x) l_j(x), \quad (2)$$

где для каждого j , $1 \leq j \leq \lambda$, функция l_j аффинна и $f|_{\pi_j}(x) = l_j|_{\pi_j}(x)$, а χ_{π_j} — характеристическая функция (индикатор) множества π_j . Функции l_j из этого выражения для краткости назовем *аффинными аппроксимациями* функции f . *Длиной аффинной нормальной формы* называется число плоскостей в разбиении Π , далее она будет обозначаться через $\lambda(\Pi)$.

Замечание 1. Характеристические функции плоскостей известны также под именем «мультиаффинных функций» [2], играющих важную роль при описании классов эффективно решаемых систем булевых уравнений.

Характеристическая функция плоскости в пространстве V_n имеет вполне определенный вид. Любая плоскость π , как уже отмечалось, может быть задана как множество решений системы d линейных уравнений (для некоторого d):

$$\left. \begin{aligned} h_1(x_1, \dots, x_n) &= 0; \\ h_2(x_1, \dots, x_n) &= 0; \\ &\vdots \\ h_d(x_1, \dots, x_n) &= 0, \end{aligned} \right\} \quad (3)$$

где все $h_i(x) \in \mathcal{A}_n$. Поскольку вектор x принадлежит плоскости π тогда и только тогда, когда все h_i , $1 \leq i \leq d$, обращаются в нуль на нем, характеристическая функция π выражается следующим образом:

$$\chi_\pi(x) = \prod_{i=1}^d (h_i(x) \oplus 1).$$

Если система линейных уравнений задана в матричной форме: $xH \oplus (b_1, \dots, b_d) = 0$, $b_i = h_i(0)$, то выражение будет иметь вид:

$$\chi_\pi(x) = \prod_{i=1}^d (xH_i \oplus b_i \oplus 1),$$

где H_i — столбцы матрицы H .

Как видно из определения, аффинная нормальная форма представляет собой в некотором смысле кусочно-аффинную аппроксимацию булевой функции. На каждой локальной аффинности π_j из разбиения Π все, кроме одного, слагаемые в выражении (2) обращаются в нуль, и функция принимает вид $f(x) = \chi_{\pi_j}(x)l_j(x) = l_j(x)$ для всех $x \in \pi_j$.

Возможность заменить на плоскости π_j квадратичное уравнение $f(x) = 0$ линейным уравнением $l_j(x) = 0$ вместе с дописанной к нему системой (3) будет использоваться при решении систем полиномиальных уравнений в следующей статье. Как сказано в замечании 1, функции $\chi_\pi(x)$, а также и слагаемые в аффинной нормальной форме (2) являются мультиаффинными функциями. Теоретико-сложностные вопросы, связанные, в частности, с решением систем мультиаффинных уравнений, а также оценки числа таких функций рассматриваются в работе [9].

При аффинном преобразовании пространства аффинные нормальные формы сохраняются в том смысле, что выражение, полученное после применения преобразования к этой форме, тоже будет аффинной нормальной формой для некоторой функции.

Предложение 2. Пусть $\varphi \in \text{GA}(V_n)$ и $f(x) = \bigoplus_{j=1}^{\lambda(\Pi)} \chi_{\pi_j}(x)l_j(x)$ — некоторая аффинная нормальная форма функции f . Тогда

$$f^\varphi(x) = f(\varphi(x)) = \bigoplus_{j=1}^{\lambda(\Pi)} \chi_{\pi_j}(\varphi(x))l_j(\varphi(x))$$

есть аффинная нормальная форма функции f^φ .

Доказательство. Множество $\Pi' = \{\pi'_j = \varphi^{-1}(\pi_j) \mid \pi_j \in \Pi\}$ является разбиением пространства V_n на $\lambda(\Pi)$ плоскостей, поскольку φ — невырожденное аффинное преобразование. Заметим, что $\varphi(x) \in \pi_j$ тогда и только тогда, когда $x \in \varphi^{-1}(\pi_j) = \pi'_j$. Поэтому $\chi_{\pi'_j}^\varphi$ — характеристическая функция плоскости π'_j . Выражение для f^φ в новых обозначениях выглядит следующим образом:

$$f^\varphi(x) = \bigoplus_{j=1}^{\lambda(\Pi')} \chi_{\pi'_j}^\varphi(x)l_j^\varphi(x).$$

Так как, очевидно, функции $l_j^\varphi(x) = l_j(\varphi(x))$ аффинны, полученное выражение представляет собой аффинную нормальную форму.

4 Теорема Диксона и приведение квадратичных функций к каноническому виду

Благодаря теореме Диксона можно для любой квадратичной булевой функции f найти ее каноническое представление, в котором она выглядит наиболее просто. Как будет видно ниже, это представление — элемент из орбиты данной функции $\text{Orb}_f(\text{GA}(V_n))$. Канонический вид квадратичной функции, в свою очередь, подсказывает простой способ построения ее аффинной нормальной формы.

Теорема 1 [10, § 199]. Для любой квадратичной функции $f \in \mathcal{F}_n$ с ненулевой матрицей \tilde{Q}_f существует аффинное преобразование $\alpha \in \text{GA}(V_n)$, которое приводит f к одному из канонических представлений:

$$f^\alpha(x) = x_1x_2 \oplus x_3x_4 \oplus \dots \oplus x_{2r-1}x_{2r} \oplus c$$

или

$$f^\alpha(x) = x_1x_2 \oplus x_3x_4 \oplus \dots \oplus x_{2r-1}x_{2r} \oplus x_{2r+1},$$

где $2r = \text{rank } \tilde{Q}_f$ и $c \in \mathbb{F}_2$.

Доказательство этого утверждения помимо авторского варианта можно найти также в [6, § 3.3; 7, § 15.2].

На практике приведение полинома Жегалкина квадратичной булевой функции к каноническому виду можно осуществить следующим способом.

Предположим, не ограничивая общности, что в полиноме Жегалкина функции f присутствует моном x_1x_2 (иначе с помощью аффинного преобразования координат «перенумеруем» переменные). Представим функцию в виде:

$$f(x) = x_1x_2 \oplus x_1l_1(x_3, \dots, x_n) \oplus x_2l_2(x_3, \dots, x_n) \oplus q_1(x_3, \dots, x_n),$$

где $l_1, l_2 \in \mathcal{A}_{n-2}$, а q_1 — некоторая квадратичная функция. Возьмем отображение φ_2 пространства V_n , задаваемое равенством:

$$\varphi_2(x) = (x_1 \oplus l_2(x_3, \dots, x_n), x_2 \oplus l_1(x_3, \dots, x_n), x_3, \dots, x_n),$$

и рассмотрим следующую функцию:

$$f^{(2)}(x) = x_1x_2 \oplus q_2(x_3, \dots, x_n),$$

где $q_2 = q_1 \oplus l_1l_2$ — квадратичная функция. Заметим, что $(f^{(2)})^{\varphi_2} = f$.

для канонического представления функции f_i можно аффинным преобразованием, обратным к φ_i , получить аффинную нормальную форму для исходной функции:

$$f_i(x) = g_i^{\varphi_i^{-1}}(x) = \bigoplus_{\delta \in \mathbb{F}_2^{r_i}} \chi_{\pi_{i,\delta}}(x) l_{i,\delta}(x),$$

где $\pi_{i,\delta} = \varphi_i(\pi'_{i,\delta})$ и $l_{i,\delta}(x) = l'_{i,\delta}(\varphi_i^{-1}(x))$.

Разумеется, если алгебраическая степень какой-либо функции f_i оказалась равной 1, то искать ничего не нужно: ее полином Жегалкина является ее аффинной нормальной формой для тривиального разбиения $\Pi_i = \{V_n\}$.

Замечание 2. Подобный способ построения аффинной нормальной формы можно использовать и непосредственно для квадратичной¹ функции f в ее исходном виде. Нужно просто фиксировать значения переменных так, чтобы в каждом мономе оставалось не более одной свободной переменной. Для этого удобнее рассмотреть матрицу Q_f , выбрать в ней столбец или строку с максимальным числом единиц среди всех столбцов и строк (пусть это будет k -я строка) и зафиксировать x_k . Затем то же проделать, исключив из рассмотрения k -ю строку и k -й столбец матрицы, и так далее, пока единицы в матрице не кончатся. Однако, несмотря на то что здесь имеет место экономия на приведении функции к каноническому виду, такой способ представляется менее эффективным в следующем смысле. Канонический вид квадратичной функции содержит минимальное число мономов степени 2, поэтому для исходной (неканонической) функции придется фиксировать, как правило, большее число переменных. Но с каждой дополнительно зафиксированной переменной размерность локальных аффинностей функции f уменьшается на 1, а их число, соответственно, увеличивается вдвое.

6 «Локальные» системы линейных уравнений

Идея метода решения систем квадратичных булевых уравнений состоит в следующем. Пусть для всех f_i , $1 \leq i \leq m$, определены некоторые аффинные нормальные формы

$$f_i(x) = \bigoplus_{j=1}^{\lambda(i)} \chi_{\pi_{ij}}(x) l_{ij}(x).$$

Исходя из этих аффинных нормальных форм, можно для каждой пары i, j записать эквивалентную уравнению $f_i = 0$ на π_{ij} систему линейных уравнений:

$$\begin{aligned} l_{ij}(x) &= 0; \\ h_{ij}^1(x) &= 0; \\ &\vdots \\ h_{ij}^{d(i,j)}(x) &= 0, \end{aligned}$$

в которой первое уравнение выражает равенство $f_i = 0$ через аффинную аппроксимацию $l_{ij}(x)$ функции $f_i(x)$ на плоскости π_{ij} , а остальные $d(i, j)$ уравнений задают эту плоскость.

Замечание 3. Если аффинная нормальная форма получена описанным выше способом — через канонический вид квадратичной функции, — то характеристическая функция будет иметь вид:

$$\chi_{\pi_{i,\delta}}(x) = \prod_{k=1}^{r_i} (\varphi_i^{-1}(x) e_{2k-1}^T \oplus \delta_k \oplus 1),$$

где e_{2k-1} — $(2k - 1)$ -й базисный вектор, т.е. $\varphi_i^{-1}(x) e_{2k-1}^T$ — $(2k - 1)$ -я компонента вектора $\varphi_i^{-1}(x)$. Значит, соответствующую плоскость задают уравнения $\{\varphi_i^{-1}(x) e_{2k-1}^T \oplus \delta_k = 0 \mid 1 \leq k \leq \leq r_i\}$.

Таким образом, имеется набор «локальных» линейных систем для каждого уравнения исходной системы и для каждой его локальной аффинности. Метод состоит в том, чтобы подобрать комбинацию «локальных» систем разных квадратичных уравнений, в совокупности дающую решение исходной системы. Если решение квадратичной системы единственно (а это естественное предположение для криптоанализа), ровно одна такая комбинация будет иметь решение, и от того, как быстро удастся ее обнаружить, зависит эффективность метода.

7 О трудоемкости построения аффинной нормальной формы

Напомним, что для функций из системы (1) $r_i = = (1/2)\text{rank } \tilde{Q}_{f_i} \leq n/2$, $1 \leq i \leq m$, — параметр, введенный в разд. 5. Алгоритм приведения m функций к каноническому виду (см. разд. 4) имеет трудоемкость, оцениваемую выражением $O(\sum_{i=1}^m n^2 r_i)$, а учитывая неравенство $r_i \leq n/2$, имеем $O(mn^3)$.

При построении аффинных нормальных форм для функции f_i в разд. 5 потребуется порядка $r_i 2^{r_i} + + n^3$ операций. Значит, для всех m функций имеем оценку $O(mn^3 + \sum_{i=1}^m r_i 2^{r_i})$.

Таким образом, в худшем случае, когда все $r_i = = n/2$ или даже когда хотя бы $r_i = O(n)$ для некоторого i , предложенный алгоритм экспоненциален. Однако можно рассчитывать, что во встречающихся на практике системах квадратичных уравнений

¹Для функций более высоких степеней такой подход тоже работает, но описать его строго сложнее и полученные таким образом локальные аффинности, скорее всего, будут слишком маленькой размерности.

параметр r_i растёт (с увеличением n) медленнее, и тогда можно говорить о полиномиальности алгоритма построения аффинных нормальных форм.

В случае, когда система вида (1) переопределённая, т.е. $n \ll m$ (переопределённые системы достаточно часто рассматриваются в задачах информатики, теории кодирования и криптографии), можно рассчитывать на существование подсистемы (из l уравнений с номерами i_1, \dots, i_l), для которой трудоёмкость построения аффинных нормальных форм меньше, чем экспоненциальная. Например, когда $r_{i_j} = O(\sqrt{n})$, $1 \leq j \leq l$, оценка соответствующей трудоёмкости для системы (1) имеет субэкспоненциальный характер.

Рассмотрим в качестве еще одного примера класс \mathcal{K}_m систем m квадратичных булевых уравнений от n неизвестных вида (1), где $m = m(n)$ — некоторый полином от n и где $r_i = O(\log_2 n)$ для всех i , $1 \leq i \leq m$.

Предложение 3. Для систем (1) квадратичных булевых уравнений из класса \mathcal{K}_m существует полиномиальный (по n) алгоритм построения аффинных нормальных форм для функций f_i .

Для доказательства этого утверждения достаточно рассмотреть предложенный в статье алгоритм построения аффинных нормальных форм для квадратичных булевых функций. В полученной выше оценке $O(mn^3 + \sum_{i=1}^m r_i 2^{r_i})$ данные в условии ограничения на m и на r_i дают полиномиальную оценку трудоёмкости алгоритма.

Литература

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи / Пер. с англ. — М.: Мир, 1982.
2. Горшков С. П., Тарасов А. В. Сложность решения систем булевых уравнений. — М.: Курс, 2017. 192 с.
3. Смирнов В. Г. Некоторые классы эффективно решаемых систем булевых уравнений // Труды по дискретной математике, 2000. Т. 3. С. 269–282.
4. Bard G. V. Algebraic cryptanalysis. — Springer, 2009. 389 p.
5. Bard G., Courtois N., Jefferson C. Efficient methods for conversion and solution of sparse systems of low-degree multivariate polynomials over GF(2) via SAT-solvers // Cryptology ePrint Archive. Report 2007/024. <http://eprint.iacr.org/2007/024.pdf>.
6. Логачев О. А., Сальников А. А., Смышляев С. В., Яценко В. В. Булевы функции в теории кодирования и криптологии. — М.: ЛЕНАНД, 2015. 576 с.
7. Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А. Теория кодов, исправляющих ошибки / Пер. с англ. — М.: Связь, 1979. 743 с. (MacWilliams F. J., Sloane N. J. A. The theory of error-correcting codes. — North-Holland mathematical library ser. — North-Holland Publishing Co., 1977. 774 p.)
8. Logachev O. A., Yashchenko V. V., Denisenko M. P. Local affinity of Boolean mappings // Boolean functions in cryptology and information security: Proceedings of the NATO Advanced Study Institute. — IOS Press, 2008. P. 148–172.
9. Горшков С. П. Применение теории NP-полных задач для оценки сложности решения систем булевых уравнений // Обзорные прикладной и промышленной математики, 1995. Т. 2. Вып. 3. С. 325–398.
10. Dickson L. E. Linear groups: With an exposition of the Galois field theory. — Leipzig: B. G. Teubner, 1901. 322 p.

Поступила в редакцию 11.01.19

POLYNOMIAL ALGORITHMS FOR CONSTRUCTING LOCAL AFFINITIES OF QUADRATIC BOOLEAN FUNCTIONS

O. A. Logachev^{1,2}, A. A. Sukayev¹, and S. N. Fedorov¹

¹Information Security Institute, M. V. Lomonosov Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, 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: Due to the affine normal form, one can consider a Boolean function as affine on certain flats in its domain — so-called local affinities. This Boolean function representation — affine approximation — could be useful for solving systems of nonlinear equations over two-element field. The problem of solving these systems (of a special sort) arises, in particular, in some methods of the information security tools design and analysis. The paper describes an approach to finding local affinities for quadratic Boolean functions which is based on Dickson’s

theorem. By this, one obtains affine normal forms for such functions. Besides, the paper concerns the efficiency of corresponding algorithms. This approach can be profitable for constructing efficient methods of solving systems of quadratic Boolean equations via “approximation” of corresponding Boolean functions by their affine normal forms.

Keywords: Boolean function; system of quadratic Boolean equations; vector space partition; flat; local affinity; Dickson’s theorem; affine normal form (ANF) of Boolean function; algebraic cryptanalysis

DOI: 10.14357/19922264190110

Acknowledgments

The paper was partly supported by the Russian Foundation for Basic Research (project 18-29-03124 mk).

References

1. Garey, M. R., and D. S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W. H. Freeman and Co. 348 p.
2. Gorshkov, S. P., and A. V. Tarasov. 2017. *Slozhnost’ resheniya sistem bulevykh uravneniy* [Complexity of solving the systems of Boolean equations]. Moscow: Kurs. 192 p.
3. Smirnov, V. G. 2000. Nekotorye klassy effektivno reshaemykh sistem bulevykh uravneniy [Some classes of Boolean equation systems permitting effective solution]. *Trudy po diskretnoy matematike* [Proceedings on Discrete Mathematics] 3:269–282.
4. Bard, G. V. 2009. *Algebraic cryptanalysis*. Springer. 389 p.
5. Bard, G., N. Courtois, and C. Jefferson. 2007. Efficient methods for conversion and solution of sparse systems of low-degree multivariate polynomials over GF(2) via SAT-solvers. *Cryptology ePrint Archive*. Report 2007/024. Available at: <http://eprint.iacr.org/2007/024.pdf> (accessed August 30, 2018).
6. Logachev, O. A., A. A. Sal’nikov, S. V. Smyshlyaev, and V. V. Yashchenko. 2015. *Bulevy funktsii v teorii kodirovaniya i kriptologii* [Boolean functions in coding theory and cryptology]. Moscow: LENAND. 576 p.
7. MacWilliams, F. J., and N. J. A. Sloane. 1977. *The theory of error-correcting codes*. North-Holland mathematical library ser. North-Holland Publishing Co. 774 p.
8. Logachev, O. A., V. V. Yashchenko, and M. P. Denisenko. 2008. Local affinity of Boolean mappings. *Boolean functions in cryptology and information security: Proceedings of the NATO Advanced Study Institute*. IOS Press. 148–172.
9. Gorshkov, S. P. 1995. Primenenie teorii NP-polnykh zadach dlya otsenki slozhnosti resheniya sistem bulevykh uravneniy [Application of the NP-complete problem theory to assessment of complexity of solving the systems of Boolean equations]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Applied and Industrial Mathematics Review] 2(3):325–398.
10. Dickson, L. E. 1901. *Linear groups: With an exposition of the Galois field theory*. Leipzig: B. G. Teubner. 322 p.

Received January 11, 2019

Contributors

Logachev Oleg A. (b. 1950) — Candidate of Science (PhD) in physics and mathematics, head of department, Information Security Institute, M. V. Lomonosov Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, 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; logol@iisi.msu.ru

Sukayev Al’bert A. (b. 1994) — student, Information Security Institute, Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, Russian Federation; asukaev@gmail.com

Fedorov Sergey N. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Information Security Institute, M. V. Lomonosov Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, Russian Federation; s.n.feodorov@yandex.ru

ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ ПРЕДСКАЗАНИЯ ОСАДКОВ*

А. К. Горшенин¹, В. Ю. Кузьмин²

Аннотация: Рассмотрена процедура настройки гиперпараметров нейронной сети для анализа пространственных метеорологических данных с использованием инструментов гибридного высокопроизводительного вычислительного комплекса (ГВБК). Проведено сравнение точности прогнозирования осадков на основе таких методов, как поиск по решетке и случайный поиск. Продемонстрировано, что даже при сравнительно небольшом числе случайных выборов комбинаций гиперпараметров возможно получить точность, сопоставимую с полным перебором, при умеренных временных затратах. Данные результаты означают возможность автоматического построения архитектуры нейросети на основе базовой модели для решения конкретных прикладных задач.

Ключевые слова: нейронные сети; прогнозирование; глубокое обучение; гиперпараметры; высокопроизводительные вычисления; CUDA

DOI: 10.14357/19922264190111

1 Введение

Нарастающий объем данных, накапливаемых в различных предметных областях, ведет к необходимости оптимизации вычислительных методов и задействования наиболее современных устройств, значительно превосходящих возможности настольных решений. Высокопроизводительные вычисления позволяют получать новые научные результаты для широкого класса прикладных задач. Так, в работе [1] проводятся расчеты для различных конфигураций оборудования (от 36 до 200 ядер центральных процессоров) для моделирования пыльных бурь. В статье [2] рассмотрены высокопроизводительные решения, которые могут применяться для обработки данных дистанционного зондирования (см. также [3]), в том числе на основе видеокарт. Подобные гибридные решения, сочетающие эффективность расчетов как на центральных, так и на графических процессорах, являются наиболее востребованными. В частности, в работе [4] рассмотрены вопросы моделирования длинноволнового излучения с максимально возможным использованием всех доступных вычислительных ресурсов. Стоит отметить использование высокопроизводительных решений в задачах, ориентированных на анализ изменения климата [5], построение погодных моделей [6], гидродинамическое моделирова-

ние внезапных наводнений в горных водоразделах [7] и цунами [8].

Одним из наиболее популярных современных трендов обработки данных (в том числе и метеорологических) стало использование искусственных нейронных сетей. При этом на передний план помимо выбора типа архитектуры выходит задача корректного определения гиперпараметров [9], т. е. настроек, которые должны быть заданы до начала процесса обучения (в отличие от остальных параметров, которые определяются непосредственно в его процессе). Фактически, для каждой достаточно сложной задачи оптимальные величины должны подбираться индивидуально, что представляет собой нетривиальную вычислительную проблему.

В статье [10] авторами данной работы была осуществлена попытка повышения качества прогнозирования метеорологических данных с помощью паттернов путем оптимизации архитектуры за счет тонкой ручной настройки гиперпараметров нейронной сети прямого распространения. Выбранный набор (число скрытых слоев и нейронов в каждом; метод оптимизации, используемый для обучения; доля исключаемых элементов между слоями и др.) тестировался с помощью непосредственной проверки ряда вручную подобранных значений.

* Работа выполнена при частичной поддержке РФФИ (проекты 17-07-00851 и 18-29-03100) и Стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-538.2018.5).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, agorshenin@frcsc.ru

² ООО «Бай2Гео», shadesilent@yandex.ru

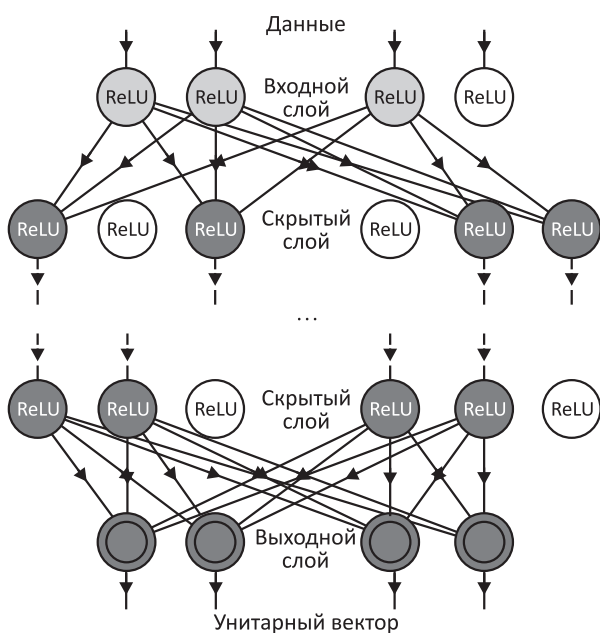
В настоящей статье на примере анализа данных 22 европейских метеорологических станций будет проверена адекватность и эффективность использования двух достаточно популярных методов подбора гиперпараметров — случайного выбора [11] и так называемого *поиска по решетке*, фактически реализующего полный перебор всех возможных комбинаций. Очевидна необходимость задействования для таких методов высокоскоростных современных вычислителей.

В данной статье для проведения расчетов были использованы инструменты ГВВК Федерального исследовательского центра «Информатика и управление» Российской академии наук (<http://hhpcc.frccsc.ru>).

2 Базовая архитектура нейронной сети

Анализ данных производится с помощью многослойной сети прямого распространения, состоящей из входного, одного или нескольких скрытых и выходного слоев (см. рисунок). На вход нейронной сети подаются квантиль-преобразованные данные: непосредственно наблюдаемые величины объемов осадков заменяются на M дискретных значений — индекс одного из заранее определенных классов, соответствующих диапазонам значений анализируемого ряда. Подробно данная методология описана, например, в статье [12].

В скрытых слоях (их число служит одним из гиперпараметров и может изменяться) использу-



Базовая архитектура нейронной сети

ются нейроны ReLU (*Rectified Linear Unit*) [13]. На выходе сети выводится унитарный код длины M (набор, состоящий из $M - 1$ нулей и единицы, расположенной в позиции с индексом, соответствующим номеру диапазона) — предсказание об индексе класса следующего наблюдения после анализируемого окна. Поскольку в такой постановке построение прогноза сводится к задаче классификации, для выходного слоя была выбрана функция активации *softmax*. Для повышения точности прогноза и предотвращения возможности переобучения (*overfitting*) использованы следующие методы:

- L^2 -регуляризация [14];
- исключение (*dropout*, *дропаут* [15]) для входного и скрытых слоев (на рисунке такие нейроны не закрашены и не имеют связей с другими элементами);
- изменение коэффициента скорости обучения нейронной сети при достижении условного «плато» точности (в ситуации, когда увеличение числа эпох обучения не приводит к заметному изменению данного показателя — величина отличается на сотые доли процента).

3 Методы выбора значений гиперпараметров

При исследовании данных с применением машинного обучения возникает задача подбора гиперпараметров — величин, которые не изменяются в процессе обучения нейронной сети. Их число может широко варьироваться в зависимости от сложности и архитектуры нейронной сети и включать в себя, например, такие параметры, как коэффициенты регуляризации, свойства топологии сети, критерии настройки и продолжительности обучения. Предполагается, что при обучении двух нейросетей с одинаковыми значениями гиперпараметров должна получаться одна и та же величина точности обучения, т. е. имеет место воспроизводимость. Таким образом, возникает задача максимизации точности обучения нейронной сети с помощью выбора вектора в пространстве доступных гиперпараметров.

Наиболее простой метод подобной оптимизации — ручной подбор соответствующих значений вектора. На основании собственного опыта и знаний о природе анализируемых данных исследователь подбирает вектор параметров и последовательно обучает нейросеть. При необходимости значения корректируются и процесс повторяется. Пример подобного выбора параметров можно найти, например, в статье [10].

Другой популярный метод определения оптимального вектора параметров — поиск по решетке (*grid search*). Для каждого параметра исследователь задает диапазон возможных значений, после чего выбираются узлы (например, равномерно), являющиеся подмножеством прямого произведения допустимых множеств для каждого из параметров. Затем с помощью полного перебора всех узлов выбирается комбинация с наибольшей точностью. У данного метода имеется ряд достоинств, в частности:

- простота реализации;
- возможность эффективной параллелизации;
- получаемые результаты точнее, чем в случае ручного перебора.

Однако в реальных задачах для корректного решения требуется значительное число гиперпараметров, а значит, размерность решетки быстро возрастает, что приводит к очень высокой вычислительной сложности данного метода.

Альтернативой методу поиска по решетке выступает метод случайного поиска. Выбираются N случайных векторов в пространстве гиперпараметров, для каждой комбинации проводится обучение нейронной сети и определяется наилучшая с точки зрения полученной точности. Как показано в статье [11], данный подход позволяет получить результаты как минимум не хуже, чем в случае поиска по решетке, при этом временные затраты снижаются значительным образом. Отметим, что данный метод достаточно популярен при решении различных прикладных задач, например можно упомянуть медицинские приложения [16, 17], кредитный скоринг [18] или тонкую настройку популярной архитектуры рекуррентных нейронных сетей LSTM (*Long Short-Term Memory*) для распознавания речи и почерка [19].

В следующем разделе на примере ряда европейских метеорологических станций будет проведено сравнение рассмотренных в этом разделе методов настройки гиперпараметров нейронной сети для прогнозирования осадков.

4 Обработка метеорологических данных с использованием высокопроизводительных решений

Пространственные геоклиматические данные собираются, как правило, с использованием значительного числа метеостанций, поэтому их обработка может быть весьма успешно оптимизирована с применением методов высокопроизводительных

вычислений. В качестве тестовых рядов были выбраны суточные данные об объемах осадков за период с 1904 по 1999 гг. на 22 европейских станциях, расположенных в таких странах, как Австрия, Германия, Голландия, Дания, Норвегия, Франция. В наблюдениях отсутствуют пропуски, поэтому нет необходимости их заполнения для корректного решения задач анализа и прогнозирования. Обучение нейронных сетей проводилось с использованием технологии NVIDIA CUDA (*Compute Unified Device Architecture*) с реализацией инструментов на языке программирования Python с помощью библиотек Keras и Tensorflow. Для выполнения расчетов был использован ГВБК в конфигурации: 2 центральных процессора IBM Power 9, 1 ТБ оперативной памяти, 2 графические карты NVIDIA Tesla V100 (каждая с объемом памяти 16 ГБ) с технологией NVLink.

Для рассмотренной в разд. 2 архитектуры нейронной сети для обработки метеорологических данных были выбраны следующие гиперпараметры: число скрытых слоев и нейронов в них, коэффициенты дропаута во входном и скрытых слоях, коэффициент изменения скорости обучения, методы оптимизации (Adam [20], AdaDelta [21], AdaMax [22]).

Точностью на решетке Acc_{Grid} будем называть лучшую полученную точность среди всех моделей, обученных на этой решетке Grid:

$$\text{Acc}_{\text{Grid}} = \max(\text{Acc}(X) | X \in \text{Grid}). \quad (1)$$

Для метода случайного поиска будем рассматривать наборы Set_K из $K = \overline{1, 25}$ случайно выбранных векторов гиперпараметров. Точностью набора $\text{Acc}_{\text{Set}_K}$ будем называть лучшую точность среди всех моделей, обученных на этом наборе:

$$\text{Acc}_{\text{Set}_K} = \max(\text{Acc}(X) | X \in \text{Set}_K), \quad K = \overline{1, 25}. \quad (2)$$

Для n наборов длины K , точность каждого из которых обозначена $\text{Acc}_{\text{Set}_K, i}$, $i = \overline{1, n}$, и определяется формулой (2), среднюю величину точности естественно определить как

$$\text{Acc}_{\text{Avg}_K} = \frac{1}{n} \sum_{i=1}^n \text{Acc}_{\text{Set}_K, i}. \quad (3)$$

На вход нейронной сети подается вектор из 180 преобразованных наблюдений, каждое из которых относится к одному из 10 интервалов. На выходе ожидается прогноз номера класса для следующего за окном наблюдения. Использование высокопроизводительного вычислительного комплекса позволило реализовать метод поиска по решетке из 3904 узлов для каждого из рассматриваемых 22 рядов (общее время обучения — 4 ч 20 мин). По сравнению с настольными решениями это позволило ускорить обучение как минимум в 5–8 раз без существенного изменения программного кода:

Таблица 1 Точность прогнозирования нейронной сети методами случайного выбора и поиска по решетке для станций в различных европейских странах (1 шаг, 10 интервалов)

Станция	Число случайных выборов								Поиск по решетке
	1	2	3	5	10	15	20	25	
Австрия	83,49%	87,04%	87,28%	87,55%	87,62%	87,63%	87,64%	87,64%	87,66%
Дания	76,26%	80,11%	80,80%	80,99%	81,01%	81,02%	81,03%	81,03%	81,05%
Франция	79,96%	83,49%	84,06%	84,18%	84,22%	84,23%	84,23%	84,23%	84,23%
Германия 1	80,31%	83,18%	83,41%	83,52%	83,58%	83,59%	83,59%	83,59%	83,60%
Германия 2	82,42%	85,42%	85,59%	85,73%	85,75%	85,75%	85,75%	85,75%	85,74%
Голландия	76,49%	79,40%	80,04%	80,16%	80,21%	80,21%	80,22%	80,22%	80,23%
Норвегия	78,48%	81,96%	82,27%	82,51%	82,58%	82,59%	82,60%	82,60%	82,60%

Таблица 2 Точность прогнозирования нейронной сети методами случайного выбора и поиска по решетке для станций в различных европейских странах (3 шага, 10 интервалов)

Станция	Число случайных выборов								Поиск по решетке
	1	2	3	5	10	15	20	25	
Австрия	70,34%	71,01%	71,02%	71,10%	71,12%	71,12%	71,12%	71,12%	71,14%
Дания	57,13%	58,76%	59,25%	59,31%	59,36%	59,36%	59,36%	59,36%	59,37%
Франция	63,45%	63,78%	63,79%	63,80%	63,80%	63,80%	63,80%	63,80%	63,80%
Германия 1	63,04%	63,36%	63,37%	63,38%	63,39%	63,39%	63,39%	63,39%	63,39%
Германия 2	63,90%	65,80%	65,95%	66,14%	66,14%	66,14%	66,14%	66,14%	66,14%
Голландия	56,11%	58,14%	58,63%	58,74%	58,84%	58,83%	58,84%	58,84%	58,85%
Норвегия	59,74%	61,51%	61,72%	61,74%	61,74%	61,74%	61,74%	61,74%	61,75%

скорость обчета узла решетки составила около 4 с по сравнению с 30–35 с до этого.

Для анализа поведения методов поиска по решетке и метода случайного поиска определены значения величин Acc_{Grid} (1) и Acc_{Avg_K} (3) для $K = 1, 2, 3, 5, 10, 15, 20$ и 25. В табл. 1 приведены примеры для некоторых метеостанций в различных европейских странах, расположенных как на уровне моря, так и на равнине и в горах. Исследуется прогноз на 1 шаг (день) вперед. В столбцах с указанным числом случайных выборов продемонстрирована средняя точность (3) для $n = 50$ построенных наборов соответствующей длины K .

В табл. 2 показана средняя точность при случайном поиске и точность поиска по решетке для прогнозов на 3 дня вперед.

Из приведенных табл. 1 и 2 следует, что средняя точность наборов из 10 элементов уже сопоставима с величиной точности для метода поиска по решетке. При этом скорость поиска оптимальных гиперпараметров может быть на 1–3 порядка выше (в зависимости от числа узлов решетки), чем полный перебор. Таким образом, для метеорологических данных и выбранной архитектуры подтверждается эффективность метода случайного поиска [11].

Еще один немаловажный результат заключается в успешной проверке применимости методологии на основе паттернов [12, 23] для наблюдений, полученных на различных европейских станциях, рас-

положенных на уровне моря, на равнинах, а также в горах — в предшествующих работах эксперименты проводились только для Потсдама и Элисты.

Рассмотренная архитектура нейронной сети обладает существенным недостатком — точность прогнозов уменьшается с увеличением их длительности и числа интервалов разбиения исходных наблюдений (см. табл. 2). При построении многодневных прогнозов резко возрастает вычислительная сложность обучения сети из-за экспоненциального роста числа нейронов в выходном слое. Задействование высокопроизводительных ресурсов и тонких методов настройки гиперпараметров позволяет обрабатывать данные и в таком режиме, однако представляется разумным рассмотреть альтернативные архитектуры.

5 Заключение

В работе рассмотрена эффективность использования метода случайного поиска в сравнении с классическим алгоритмом перебора всех возможных значений гиперпараметров на некоторой выбранной сетке. Продемонстрировано, что даже сравнительно небольшое число (порядка десяти) случайно выбранных комбинаций позволяет получить сопоставимую точность, при этом затраченное время оказывается весьма умеренным. Полу-

ченые результаты означают возможность реализовать предложенную в статьях [12, 23] методологию паттернов для нейронных сетей в виде исследовательского сервиса цифровой платформы. При этом от пользователей не требуется значительной квалификации в области работы с нейронными сетями — гиперпараметры могут подбираться в автоматическом режиме, а значит, тонкая настройка архитектуры возможна и без участия исследователя.

Дальнейшие исследования будут ориентированы на повышение скорости обсчета каждого набора гиперпараметров за счет оптимизации программного кода для ресурсов ГВБК, а также на переход к более сложным архитектурам нейронных сетей (рекуррентным, сверточным) для построения прогнозов исходных данных с использованием метрик типа среднеквадратичных ошибок для оценивания качества получаемых результатов.

Авторы выражают признательность член-корреспонденту РАН С. К. Гулеву и доктору физико-математических наук О. Г. Золиной за предоставленные для экспериментов данные европейских метеорологических станций.

Литература

1. Xie J., Yang C., Zhou B., Huang Q. High-performance computing for the simulation of dust storms // *Comput. Environ. Urban*, 2010. Vol. 34. P. 278–290.
2. Lee C. A., Gasster S. D., Plaza A., Chang C.-I., Huang B. Recent developments in high performance computing for remote sensing: A review // *IEEE J. Sel. Top. Appl.*, 2011. Vol. 4. Iss. 3. P. 508–527.
3. Xue Y., Palmer-Brown D., Guo H. The use of high-performance and high-throughput computing for the fertilization of Digital Earth and global change studies // *Int. J. Digit. Earth*, 2011. Vol. 4. Iss. 3. P. 185–210.
4. Lu F., Song J., Cao X., Zhu X. CPU/GPU computing for long-wave radiation physics on large GPU clusters // *Comput. Geosci.*, 2012. Vol. 41. P. 47–55.
5. Oubeidillah A. A., Kao S.-C., Ashfaq M., Naz B. S., Tootle G. A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US // *Hydrol. Earth Syst. Sc.*, 2014. Vol. 18. Iss. 1. P. 67–84.
6. Thompson G., Politovich M. K., Rasmussen R. M. A numerical weather model's ability to predict characteristics of aircraft icing environments // *Weather Forecast.*, 2017. Vol. 32. Iss. 1. P. 207–221.
7. Hu X., Song L. Hydrodynamic modeling of flash flood in mountain watersheds based on high-performance GPU computing // *Nat. Hazards*, 2018. Vol. 91. Iss. 2. P. 567–586.
8. Reguly I. Z., Giles D., Gopinathan D., Quivy L., Beck J. H., Giles M. B., Guillas S., Dias F. The VOLNA-OP2 tsunami code (version 1.5) // *Geosci. Model Dev.*, 2018. Vol. 11. Iss. 11. P. 4621–4635.
9. Zheng M., Tang W., Zhao X. Hyperparameter optimization of neural network-driven spatial models accelerated using cyber-enabled high-performance computing // *Int. J. Geogr. Inf. Sci.*, 2019. Vol. 33. Iss. 2. P. 314–345.
10. Gorshenin A. K., Kuzmin V. Yu. Improved architecture of feedforward neural networks to increase accuracy of predictions for moments of finite normal mixtures // *Pattern Recognition Image Anal.*, 2019. Vol. 29. No. 1. P. 68–77.
11. Bergstra J., Bengio Y. Random search for hyper-parameter optimization // *J. Mach. Learn. Res.*, 2012. Vol. 13. P. 281–305.
12. Gorshenin A. K., Kuzmin V. Yu. Neural network forecasting of precipitation volumes using patterns // *Pattern Recognition Image Anal.*, 2018. Vol. 28. No. 3. P. 450–461.
13. Glorot X., Bordes A., Bengio Y. Deep sparse rectifier neural networks // *J. Mach. Learn. Res.*, 2011. Vol. 15. P. 315–323.
14. Tikhonov A. N., Leonov A. S., Yagola A. G. Nonlinear ill-posed problems. — Heidelberg: Springer, 1998. 386 p.
15. 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.
16. Rojas-Dominguez A., Padierna L. C., Valadez J. M. C., Puga-Soberanes H. J., Fraire H. J. Optimal hyperparameter tuning of SVM classifiers with application to medical diagnosis // *IEEE Access*, 2018. Vol. 6. P. 7164–7176.
17. Uppu S., Krishna A. A deep hybrid model to detect multi-locus interacting SNPs in the presence of noise // *Int. J. Med. Inform.*, 2018. Vol. 119. P. 134–151.
18. Xia Y., Liu C., Li Y., Liu N. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring // *Expert Syst. Appl.*, 2017. Vol. 78. P. 225–241.
19. Greff K., Srivastava R. K., Koutnik J., Steunebrink B. R., Schmidhuber J. LSTM: A search space Odyssey // *IEEE T. Neur. Net. Lear.*, 2017. Vol. 28. Iss. 10. P. 2222–2232.
20. Kingma D., Ba J. Adam: A method for stochastic optimization // 3rd Conference (International) for Learning Representations. — San Diego, CA, USA, 2015. arXiv:1412.6980. 13 p.
21. Zeiler M. D. ADADELTA: An adaptive learning rate method. arXiv:1212.5701, 2012. 6 p.
22. Buduma N. Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. — Sebastopol, CA, USA: O'Reilly Media, 2017. 295 p.
23. Горшенин А. К. Анализ вероятностно-статистических характеристик осадков на основе паттернов // *Информатика и её применения*, 2017. Т. 11. Вып. 4. С. 38–46.

Поступила в редакцию 15.01.19

OPTIMIZATION OF HYPERPARAMETERS OF NEURAL NETWORKS USING HIGH-PERFORMANCE COMPUTING FOR PREDICTION OF PRECIPITATION

A. K. Gorshenin^{1,2}, and V. Yu. Kuzmin³

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

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

³“Wi2Geo LLC,” 3-1 Mira Prosp., Moscow 129090, Russian Federation

Abstract: The paper describes the procedure for tuning hyperparameters of the neural network for analyzing spatial meteorological data using the tools of the hybrid high-performance computing system. The comparison of precipitation forecasting accuracy has been carried out on the basis of such methods as grid and random searches. It has been demonstrated that even with a relatively small number of random choices of combinations of hyperparameters, it is possible to obtain an accuracy comparable to brute force, with moderate time costs. These results show the ability to automatically build a neural network architecture based on the general model for solving applied problems.

Keywords: artificial neural network; forecasting; deep learning; hyperparameters; high-performance computing; CUDA

DOI: 10.14357/19922264190111

Acknowledgments

The research is partially supported by the Russian Foundation for Basic Research (projects 17-07-00851 and 18-29-03100) and the RF Presidential scholarship program (project No. 538.2018.5). The calculations were performed by Hybrid high-performance computing cluster of FRC CSC RAS (<http://hhpcc.frccsc.ru>).

References

1. Xie, J., C. Yang, B. Zhou, and Q. Huang. 2010. High-performance computing for the simulation of dust storms. *Comput. Environ. Urban* 34:278–290.
2. Lee, C.A., S.D. Gasster, A. Plaza, C.-I. Chang, and B. Huang. 2011. Recent developments in high performance computing for remote sensing: A review. *IEEE J. Sel. Top. Appl.* 4(3):508–527.
3. Xue, Y., D. Palmer-Brown, and H. Guo. 2011. The use of high-performance and high-throughput computing for the fertilization of Digital Earth and global change studies. *Int. J. Digit. Earth* 4(3):185–210.
4. Lu, F., J. Song, X. Cao, and X. Zhu. 2012. CPU/GPU computing for long-wave radiation physics on large GPU clusters. *Comput. Geosci.* 41:47–55.
5. Oubeidillah, A. A., S.-C. Kao, M. Ashfaq, B. S. Naz, and G. Tootle. 2014. A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrol. Earth Syst. Sc.* 18(1):67–84.
6. Thompson, G., M. K. Politovich, and R. M. Rasmussen. 2017. A numerical weather model’s ability to predict characteristics of aircraft icing environments. *Weather Forecast.* 32(1):207–221.
7. Hu, X., and L. Song. 2018. Hydrodynamic modeling of flash flood in mountain watersheds based on high-performance GPU computing. *Nat. Hazards* 91(2):567–586.
8. Reguly, I. Z., D. Giles, D. Gopinathan, L. Quivy, J. H. Beck, M. B. Giles, S. Guillas, and F. Dias. 2018. The VOLNA-OP2 tsunami code (version 1.5). *Geosci. Model Dev.* 11(11):4621–4635.
9. Zheng, M., W. Tang, and X. Zhao. 2019. Hyperparameter optimization of neural network-driven spatial models accelerated using cyber-enabled high-performance computing. *Int. J. Geogr. Inf. Sci.* 33(2):314–345.
10. Gorshenin, A. K., and V. Yu. Kuzmin. 2019. Improved architecture of feedforward neural networks to increase accuracy of predictions for moments of finite normal mixtures. *Pattern Recognition Image Anal.* 29(1):68–77.
11. Bergstra, J., and Y. Bengio. 2012. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13:281–305.
12. Gorshenin, A. K., and V. Yu. Kuzmin. 2018. Neural network forecasting of precipitation volumes using patterns. *Pattern Recognition Image Anal.* 28(3):450–461.

13. Glorot, X., A. Bordes, and Y. Bengio. 2011. Deep sparse rectifier neural networks. *J. Mach. Learn. Res.* 15:315–323.
14. Tikhonov, A. N., A. S. Leonov, and A. G. Yagola. 1998. *Nonlinear Ill-posed problems*. Heidelberg: Springer. 386 p.
15. 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.
16. Rojas-Dominguez, A., L. C. Padierna, J. M. C. Valadez, H. J. Puga-Soberanes, and H. J. Fraire. 2018. Optimal hyper-parameter tuning of SVM classifiers with application to medical diagnosis. *IEEE Access* 6:7164–7176.
17. Uppu, S., and A. Krishna. 2018. A deep hybrid model to detect multi-locus interacting SNPs in the presence of noise. *Int. J. Med. Inform.* 119:134–151.
18. Xia, Y., C. Liu, Y. Li, and N. Liu. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst. Appl.* 78:225–241.
19. Greff, K., R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber. 2017. LSTM: A search space Odyssey. *IEEE T. Neur. Net. Lear.* 28(10):2222–2232.
20. Kingma, D., and J. Ba. 2015. Adam: A method for stochastic optimization. *3rd Conference (International) for Learning Representations*. San Diego, CA. arXiv:1412.6980. 13 p.
21. Zeiler, M. D. 2012. ADADELTA: An adaptive learning rate method. arXiv:1212.5701.
22. Buduma, N. 2017. *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. Sebastopol, CA: O'Reilly Media. 295 p.
23. Gorshenin, A. K. 2017. Analiz veroyatnostno-statisticheskikh kharakteristik osadkov na osnove patternov [Pattern-based analysis of probabilistic and statistical characteristics of precipitations]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):38–46.

Received January 15, 2019

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, 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; leading scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; agorshenin@frccsc.ru

Kuzmin Victor Yu. (b. 1986) — Head of Development, “Wi2Geo LLC”, 3-1 Mira Prosp., Moscow 129090, Russian Federation; shadesilent@yandex.ru

СИНТЕЗ ГЕОДАНЫХ В ПРОСТРАНСТВЕННЫХ ИНФРАСТРУКТУРАХ НА ОСНОВЕ СВЯЗАННЫХ ДАННЫХ*

С. К. Дулин¹, Н. Г. Дулина², О. С. Кожунова³

Аннотация: Синтез пространственных данных из различных источников, доступных в Web, — главная задача для современных приложений, использующих информационный поиск в сети и нацеленных на принятие решений на основе геоданных. Эта работа посвящена синтезу пространственных данных с акцентом на его приложения в пространственных инфраструктурах данных (Spatial Data Infrastructures — SDIs). Возможности интеграции SDIs и семантического контекста обсуждаются при условии согласованного описания и использования отношений характеристик объектов. Предложена классификация и декомпозиция процессов синтеза в сервис-ориентированной структуре для обслуживания широкого круга запросов.

Ключевые слова: синтез данных; пространственная инфраструктура данных; связанные данные; семантическая сеть

DOI: 10.14357/19922264190112

1 Введение

Быстрое развитие Web от варианта, ориентированного на данные, до структур для обслуживания широкого круга запросов наряду с широким распространением мобильных устройств с определением местоположения сильно повлияло на понимание, доступность и использование геоданных. В результате этого объем геоданных, доступных через сеть, непрерывно увеличивается. Для успешного анализа наборов геоданных необходимы методы установления связей и объединения геоданных, полученных из разнообразных источников. Становится очевидным, что как только веб-сервисы обеспечат синтез пространственной информации из произвольного числа источников геоданных, возникнет намного больший потенциал обработки, чем дают сегодняшние пространственные инфраструктуры данных (SDIs), которые действуют только как доступные через сеть платформы доставки пространственных данных.

Методики синтеза пространственных данных играют важную роль в создании интегрированного представления распределенных источников пространственных данных в сети. Поскольку гибкость и функциональная совместимость — ключевые факторы в такой интеграции геоданных, использование стандартов — неоспоримое требование. Поэтому в дополнение к геопространственным стандартам, установленным открытым геопространственным консорциумом (Open Geospatial Consortium — OGC), семантические стандарты сети, изданные консорциумом Web (World Wide Web Consortium — W3C) применительно к связанным геоданным, стали хорошим дополнением для формализации и управления отношениями характеристик объектов как части процесса их синтеза. Хотя архитектуры SDIs и Semantic Web сильно отличаются по линии применяемых технологических процессов и стандартов, они могут дополнять друг друга.

В этой работе исследуются подходы, требования и ограничивающие факторы для пространственного синтеза данных на основе веб-сервисов с акцентом на взаимодействие SDI и стандартов Semantic Web.

Основными целями исследования являются сервис-ориентированная декомпозиция широкого круга запросов процессов синтеза и стратегии управления отношениями характеристик объектов, использующие связанные данные в комбинации с SDIs.

Основными целями исследования являются сервис-ориентированная декомпозиция широкого круга запросов процессов синтеза и стратегии управления отношениями характеристик объектов, использующие связанные данные в комбинации с SDIs.

*Работа выполнена при частичной поддержке РФФИ (проект 17-20-02153 офи_м.РЖД).

¹Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте (ОАО НИИАС); Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, skdul@mail.ru

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, ngdulina@mail.ru

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, kozhunovka@mail.ru

Классификация и декомпозиция для обслуживания широкого круга запросов пространственных процессов синтеза данных, облегчающих выполнение взаимодействующих гибких технологических процессов синтеза, рассмотрены в разд. 2. Определение интеграции SDI и разработок Semantic Web со специфическим акцентом на формализацию и управление отношениями пространственных данных предложено в разд. 3. В разд. 4 обсуждаются классификация и структура операций низкого уровня процессов синтеза пространственных данных. Взаимодействие между SDI (интерфейс OWS) и компонентами Semantic Web (интерфейс RDF) рассмотрено в разд. 5.

2 Синтез геоданных — классификация и декомпозиция

Термин «синтез данных» (data fusion) широко используется в области электронной обработки данных и имеет множество различных определений и классификаций. В 2010 г. при исследовании стандартов информационный синтез данных был определен как «акт или процесс объединения или соединения данных или информации, касающихся одного или более объектов, которые рассматриваются в явной или неявной структуре знаний, чтобы улучшить возможность (или выявить новую возможность) для обнаружения, идентификации или характеристик рассматриваемых объектов» [1].

В этой работе синтез пространственных данных рассматривается применительно к множеству источников для извлечения значимой информации, касающейся определенного прикладного контекста. Поэтому понимание синтеза геоданных, пред-

ложенного здесь, включает то, что другие описывают как синтез [2], интеграцию геоданных [3] или конкатенацию геоданных [4].

Многие процессы синтеза могут быть реализованы в закрытой архитектуре при наличии единственного провайдера программного обеспечения и оборудования. Однако без использования открытых стандартов множество геоданных и сервисов трудно автоматизировать и масштабировать. Стандартизованные данные, приложения и сервисы обеспечивают автоматизированную и интероперабельную среду синтеза геоданных, поддерживая безопасное совместное использование геоданных и прозрачное многократное использование сервисов обработки больших объемов геоданных и непредсказуемых аналитических запросов.

Некоторые элементы предпочтительной открытой структуры геоинформационного синтеза на основе геостандартов уже распространяются среди пользователей. Например, обслуживание карты в сети (Web Map Service [5] — WMS) обеспечивает синтез карт. Сервис WMS представляет карты как иллюстрированные уровни изображений, полученных из различных источников, чтобы географическим оверлеем создавать агрегированные карты, удовлетворяющие потребностям пользователей.

Простым примером геоинформационного синтеза можно считать широко используемое агрегирование слоев карты (иногда называемых темами) в единое изображение (рис. 1).

Очень важным шагом в исследовании синтеза данных стало введение OGC термина *feature*, который не имеет однозначного перевода на русский язык, поскольку трактуется в английском очень широко: особенность, характеристика, свойство, признак, функция, феномен, элемент. В терминологической базе OGC (<http://www.opengeospatial.org/ogc/glossary/f>) этот термин определяется так: «Циф-

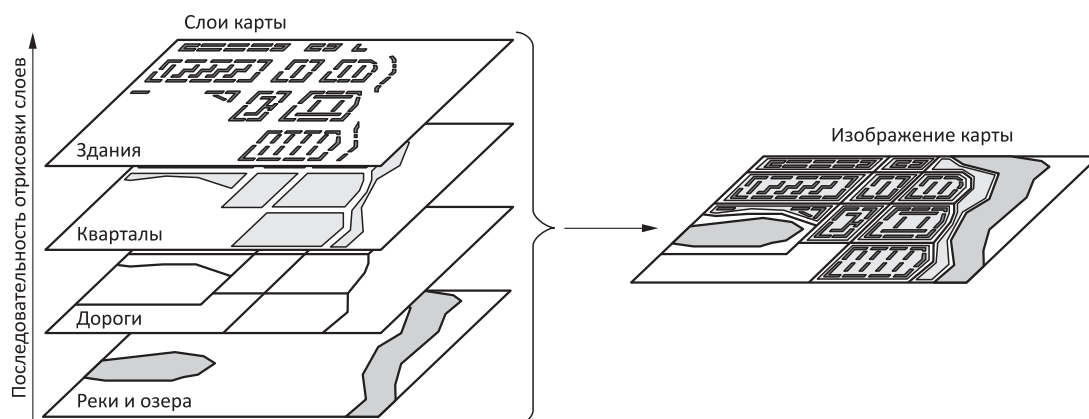


Рис. 1 Формирование карты на основе слоев

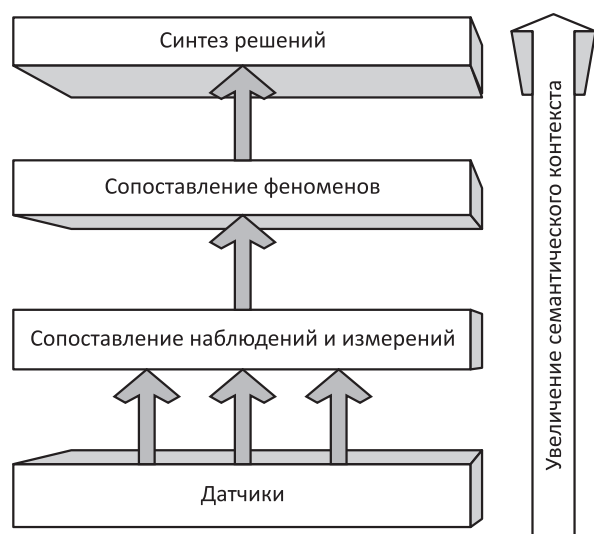


Рис. 2 Три категории, используемые в исследовании синтеза геоданных (расположены в соответствии с увеличением семантического контекста)

ровое представление объекта реального мира или абстракция реального мира. В качестве одного из его атрибутов оно может иметь пространственные, временные или пространственно-временные характеристики. Феномены обычно управляются группами в виде коллекций феноменов. Термины феномен и объект часто используются как синонимы. Термины феномен, набор феноменов и покрытие определяются в соответствии с принятыми в OpenGIS». В дальнейшем в этой работе термин *feature* будет обозначаться как феномен.

В исследовании синтеза геоданных принято рассматривать три категории (рис. 2): сопоставление наблюдений и измерений (*observation fusion*); сопоставление феноменов (*feature fusion*) и синтез решений (*decision fusion*) [5].

3 Сопоставление наблюдений и измерений

В результате синтеза наблюдений многократные измерения датчиками одних и тех же явлений объединяются в некое комбинированное наблюдение. Процессы сопоставления тем самым формируют комбинацию различных измерений датчиков в виде хорошо описываемого наблюдения, учитывая при этом присущие этому процессу неопределенности (например, анализ подписи клиента или распознавание его лица в различных ракурсах).

Основные требования для сопоставления наблюдений и измерений с помощью датчика включают [6]:

- (1) формирование системы датчиков, необходимых наблюдений и процессов наблюдения, которые удовлетворяют потребностям пользователей;
- (2) определение возможностей датчиков и качества измерений;
- (3) доступность параметров датчика, которые позволяют автоматически определять местонахождение наблюдений;
- (4) определение реального времени наблюдений в стандартных кодировках, включая кодирование неопределенности измерений, и параметров, необходимых для обработки измерений;
- (5) настройка датчиков на проведение наблюдений;
- (6) выработка и публикация предупреждений, которые будут выдаваться датчиками или сервисами датчиков;
- (7) идентификация и классификация объекта;
- (8) включение сопоставления наблюдений и измерений обеспечением доступа к механизмам обработки и необходимой информации об объекте.

Информация для проведения синтеза может быть получена как из наблюдений за датчиками, так и от людей. Информация как результат наблюдения может служить вводом в процессы сопоставления наблюдений и измерений или может использоваться для идентификации распознаваемых объектов, характеристики которых обрабатываются как вводимые для процесса синтеза. Стандарты для сопоставления наблюдений и измерений используются уже продолжительное время и доказали свою жизнеспособность.

4 Сопоставление феноменов и синтез решений

Сопоставление феноменов включает обработку наблюдений на более высоком уровне семантических свойств феноменов. Это улучшает понимание операционной ситуации и уточняет оценку потенциальных угроз и влияний, позволяя более корректно идентифицировать, классифицировать, связать и объединить объекты, представляющие интерес. Процессы сопоставления феноменов включают их обобщение и агрегирование. Технология агрегирования включает полезные опции, позволяя работать с несовершенными, гетерогенными, противоречивыми и дублированными геоданными.

Сервис-ориентированная архитектура (service-oriented architecture, SOA) [7] хорошо подходит для поддержки в распределенных сервисах правил агрегирования феноменов. Сопоставление феноменов обеспечивает работу с более мощными, гибкими и точными информационными ресурсами, чем с полученными из оригинальных источников.

Синтез решений инициирует процессы, поддерживающие способность человека принять решение, обеспечивая среду взаимодействующего сетевого обслуживания для оценки ситуации и поддержки принятия решений, используя информацию от различных датчиков и уже обработанную информацию.

Синтез решений обеспечивает аналитикам среду, где они, используя интерфейс отдельного клиента, могут обращаться к интероперабельным инструментальным средствам, чтобы находить, обрабатывать и использовать различные типы данных, полученных от разных датчиков и баз данных. Синтез решений включает использование информации от корпоративных геосообществ, которая позволяет оценить ситуацию в целом и воспользоваться общей операционной ситуацией. Поэтому можно сказать, что информация, доступная для осуществления синтеза решений, является коллекцией сведений интеллектуального корпоративного капитала. Источниками таких сведений выступают люди, документы, оборудование или технические датчики.

Развитие синтеза решений требует разработок стандартов для структурированной информации, таких как методы отображения схемы с соответствующей идентификацией правил отображения, и увеличения акцента на обработку ассоциаций, поскольку идентификация ассоциации между объектами лежит в основе синтеза. Самая эффективная среда для реализации указанных категорий синтеза — это архитектура с распределенными базами данных и сервисами, основанными на общем ядре форматов геоданных, алгоритмов и приложений [8].

Синтез решений представляет собой масштабную операцию, включающую как геоданные, получаемые от человека с мобильного устройства, так и геоданные общей обстановки из центра МЧС. Невяная коммуникация — это сотрудничество с другими организациями и ведомствами.

Существующие публикации предлагают разнообразие классификаций для синтеза пространственных данных. Эти подходы могут различаться (рис. 3) по области применения, уровню автоматизации, частоте операций, уровню соответствия, принятой модели данных и пространственно-временной привязке входных данных [2]. Соответ-

ствующие реализации процессов можно различать, основываясь на структурах геоданных ввода и вывода, которые эти процессы поддерживают, на соответствующих прикладных стратегиях, на вычислительных показателях и других качественных мерах.

Обеспечение процессов синтеза пространственных данных в сервис-ориентированной архитектуре (SOA) для обслуживания широкого круга запросов, такой как SDI, требует организации и определения соответствующих сервисных интерфейсов. Эти интерфейсы предназначены для декомпозиции и инкапсуляции соответствующих наборов взаимодействующих функций синтеза, которые будут использоваться для различных технологических процессов синтеза. Такая организация обеспечит интерфейсы для поддержки свободного связывания геоданных, многократного использования и компоновки сервисов. Можно определить три уровня детализации:

- (1) атомарные операции, которые обеспечивают прикладные функциональные возможности системы типа простых арифметических операций и определения географических размеров;
- (2) операции низкого уровня, которые могут использоваться приложением, но все-таки характеризуются как базисные и поэтому пригодны для использования множеством приложений. Они включают, например, геометрические преобразования, показывают степень подобия феноменов и операции перемещения феноменов;
- (3) операции высокого уровня, которые ориентированы на специфику определенного приложения, обладая ограниченной гибкостью и возможностью многократного использования, такого как повторяющиеся технологические процессы, привязанные к конкретным исходным данным и целевым показателям приложений.

Таким образом, любая операция может неоднократно использоваться или быть композицией операций низшего уровня. Однако на практике эти уровни не всегда строго разделяются. Чтобы реализовать синтез пространственных данных на основе сервисов гибким и практичным способом при условии неоднократного их применения, можно использовать комплект инструментов операций низкого уровня, и это, возможно, самый перспективный подход. Типичный технологический процесс синтеза и его логическая декомпозиция представлены на рис. 4. Этот рисунок служит формальной структурой для процессов синтеза низкого уровня, включающей следующие шаги:

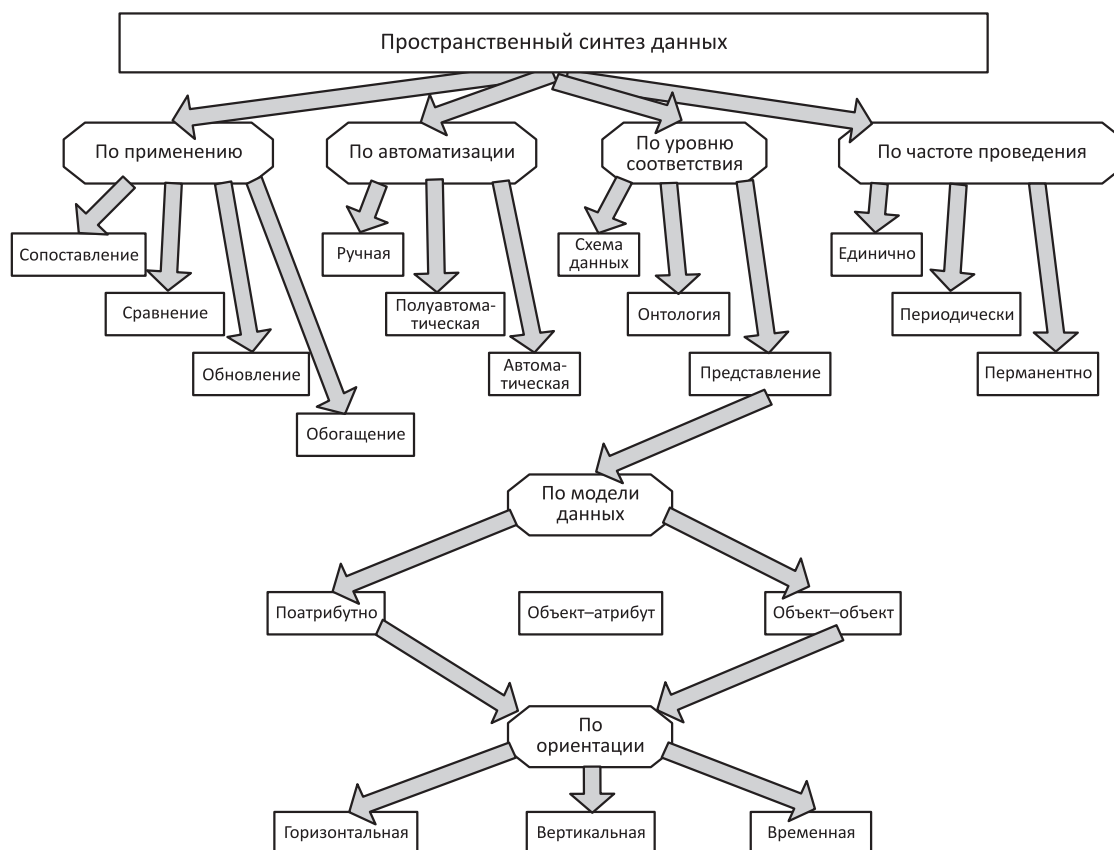


Рис. 3 Классификация процесса синтеза пространственных данных

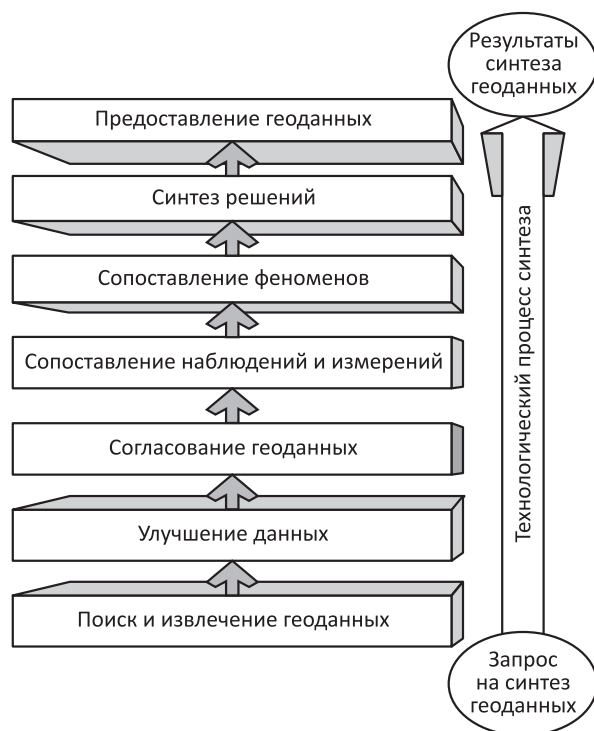


Рис. 4 Операции низкого уровня для проведения синтеза пространственных данных

- (1) поиск и извлечение геоданных, нацеленных на идентификацию и сбор входных геоданных для синтеза пространственных данных;
- (2) улучшение данных производится на основе анализа качественных характеристик индивидуальных входных источников, релевантных процессу синтеза. Оно включает задачи формирования и исправления [9, 10], предварительной обработки признаков и процессы классификации [11];
- (3) согласование геоданных нацелено на устранение несогласованности между вводами, чтобы обеспечить общую синтаксическую, структурную и семантическую основу для поддержки корректного синтеза геоданных. Согласование обычно выполняется на уровне набора геоданных и включает обработку позиционной перегруппировки [4], синхронизации времени, приведения форматов и координатного преобразования [12] и генерализации;
- (4) сопоставление наблюдений и измерений исследует квантификацию размерных отношений между входными источниками. Оно может быть выполнено в представлении

феноменов, схеме или на семантическом уровне и типично выражает определенный вид подобия;

- (5) сопоставление феноменов использует соотношения измерений для определения типов связей между феноменами в представлении и типов отношений между характеристиками в представлении, схеме или на семантическом уровне. Так же как и соотношение измерений, оно зависит от надежности входных источников и знаний об объекте [13];
- (6) синтез решений зависит от предыдущих результатов и определяется целевой программой. Стратегии разрешения противоречий могут быть применены для устранения любых несогласованностей, идентифицированных в процессе реализации синтеза. Фактически цель процесса синтеза может быть достигнута путем использования идентифицированных отношений для сопоставления, сравнения, обновления или улучшения данных. Этот этап включает передачу феноменов и признаков, которые не присутствуют во всех наборах геоданных (например, полученных при взаимных обновлениях), а также агрегирование связанных феноменов, чтобы получить расширенное или улучшенное представление феноменов;
- (7) предоставление геоданных связано с кодированием, хранением и регистрацией результатов, чтобы облегчить доступ и использование результатов для визуализации или дальнейшего анализа. Идеально, если результаты синтеза будут включать информацию о неопределенности и происхождении геоданных.

Шаги, описанные выше, не должны расцениваться как формирование строгой последовательности. Отдельные операции могут быть пропущены, повторены или объединены иначе. В частности, шаги 4 (соотношение измерений), 5 (сопоставление феноменов) и 6 (решение) часто объединяются в единый процесс идентификации и сопоставления феноменов.

5 Синтез пространственных данных на основе SDI и связанных данных

Принятие стандарта SDIs привело к публикации, поиску, доступу и обработке пространственных данных открытым и стандартизованным способом [14]. Semantic Web обеспечивает повсеместный доступ к связанным данным в Web. Па-

радигма связанных данных предполагает, что URI (Universal Resource Identifier) используется для однозначного определения объекта или понятия, которые могут быть соотнесены с HTTP-адресами, обеспечивая данные в стандартизованных форматах, связанных с другими источниками данных в Web.

Ориентируясь на интероперабельный, основанный на сервисах синтез пространственных данных, следует учитывать, что все этапы синтеза геоданных, описанные в предыдущем разделе, должны быть согласованы с текущими стандартами SDI и Semantic Web. Это может быть достигнуто с помощью стандартов OGC (Open Geospatial Consortium) для регистрации, кодирования, визуализации и обработки пространственных данных и использования стандартов W3C для связанных данных, в частности в форматах RDF (Resource Description Framework) и SPARQL (протокол SPARQL и язык запросов RDF). На рис. 5 показан пример возможной коммуникации между SDI и компонентами Semantic Web для синтеза пространственных данных в Web. В то время как компоненты SDI достижимы с помощью сервисов OWS (OGC Web Services), компоненты Semantic Web обеспечивают функциональные возможности, основанные на интерфейсах RDF.

Метод управления на основе уникальной идентификации в SDIs предопределяет использование связанных данных. Однако управление постоянными уникальными идентификаторами и реализация эффективной идентификации ресурса и стратегии

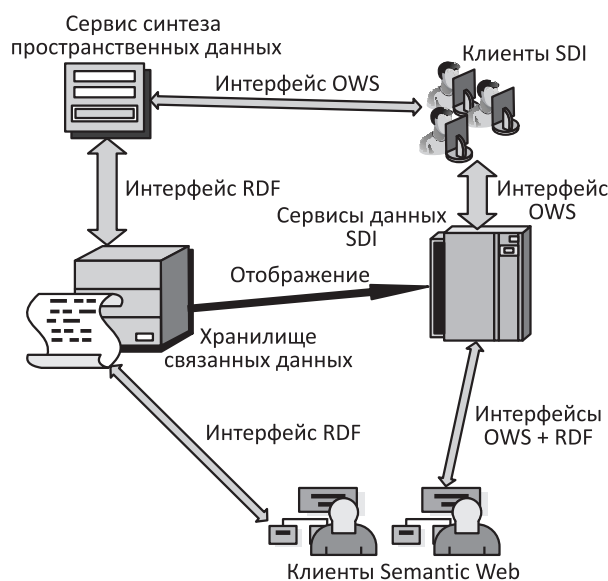


Рис. 5 Взаимодействие между SDI (интерфейс OWS) и компонентами Semantic Web (интерфейс RDF)

управления в большом масштабе требуют значительных усилий по согласованности на концептуальном уровне.

6 Заключение

Как показали вышеизложенные исследования, для обеспечения доступа и использования отношений феноменов из различных распределенных источников геоданных эти источники данных должны быть полностью определены. Схема отношений, которая описывает отношения феноменов, адекватные SDI и согласованные с принципами связанных данных, должна содержать следующие компоненты:

- ресурс, соответствующий фундаментальному понятию Semantic Web;
- отношение феноменов — основной класс, который содержит все феномены, участвующие в отношении, типы связанных отношений и измерения;
- тип отношения, который квалифицирует типы связей между входными феноменами;
- ресурс феномена, который представляет собой еще один тип феномена, связанный отношением. Он указывает на сервис данных SDI, обслуживающий тип феномена и уникально сопоставимый с ресурсом;
- соотношение измерений, обеспечивающее получение информации об основных измерениях, которые поддерживаются типами отношений и сами занесены в отношение, включая информацию о подобию и секретности.

Литература

1. OGC Engineering Reports. <http://www.opengeospatial.org/standards>.
2. Ruiz J. J., Ariza F. I., Ureña M. A., Blázquez E. B. Digital map conflation: A review of the process and a proposal for classification // *Int. J. Geogr. Inf. Sci.*, 2011. Vol. 25. Iss. 9. P. 1439–1466.
3. Schwinn A., Schelp J. Design patterns for data integration // *J. Enterprise Information Management*, 2005. Vol. 18. Iss. 4. P. 471–482.
4. López-Vázquez C., Callejo M. A. M. Point- and curve-based geometric conflation // *Int. J. Geogr. Inf. Sci.*, 2013. Vol. 27. Iss. 1. P. 192–207.
5. OGC Web Services Testbed, Phase 8 (OWS-8) Demonstrations. <http://www.opengeospatial.org/pub/www/ows8/index.html>.
6. Botts M., Percivall G., Carl Reed C., Davidson J. OGC Sensor Web Enablement (SWE): Overview and high level architecture. OGC White Paper. Document OGC 07-165r1, April 2, 2013.
7. Pravia M. Generation of a fundamental data set for hard/soft information fusion // 11th Conference (International) on Information Fusion. — Cologne: International Society of Information Fusion, 2008. P. 134–145.
8. OGC Fusion Standards Study: Phase 2 Engineering Report / Ed. G. Percivall. OGC Document 10-184, December 13, 2010. http://www.opengeospatial.org/files/?artifact_id=41573.
9. Butenuth M., von Goesseln G., Sester M. Integration of heterogeneous geospatial data in a federated database // *ISPRS J. Photogramm.*, 2007. Vol. 62. Iss. 5. P. 328–346.
10. Al-Bakri M., Fairbairn D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources // *Int. J. Geogr. Inf. Sci.*, 2012. Vol. 26. Iss. 8. P. 1437–1456.
11. Koukoletsos T., Haklay M., Ellul C. Assessing data completeness of VGI through an automated matching procedure for linear data // *T. GIS*, 2012. Vol. 16. Iss. 4. P. 477–498.
12. Stankute S., Asche H. A data fusion system for spatial data mining, analysis and improvement // *Computational science and its applications / Eds. B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A.A.C. Rocha, D. Taniar, B.O. Apduhan. — Lecture notes in computer science ser. — Springer-Verlag, 2012. Vol. 7334. P. 439–449.*
13. Cholvy L. Modelling information evaluation in fusion // 10th Conference (International) on Information Fusion Proceedings. — Quebec, Canada: IEEE, 2007. CD-ROM. P. 1–6. doi: 10.1109/ICIF.2007.4408060.
14. Bernard L., Kanellopoulos I., Annoni A., Smits P. The European Geoportal — one step towards the establishment of a European spatial data infrastructure // *Comput. Environ. Urban*, 2005. Vol. 29. Iss. 1. P. 15–31.

Поступила в редакцию 04.07.18

SYNTHESIS OF GEODATA IN SPATIAL INFRASTRUCTURES BASED ON RELATED DATA

S. K. Dulin^{1,2}, N. G. Dulina³, and O. S. Kozhunova¹

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

²Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation

³A. 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: Synthesis of spatial data from various sources available on the Web is the main task for modern applications that use information retrieval in the network and are aimed at making decisions based on geodata. This work is devoted to the synthesis of spatial data with an emphasis on its application in Spatial Data Infrastructures (SDIs). The possibilities for integrating SDIs and semantic contexts are discussed subject to an agreed description and use of object characteristics relations. Classification and decomposition of synthesis processes in a service-oriented structure for servicing a wide range of queries are proposed.

Keywords: data synthesis; spatial data infrastructure; related data; semantic network

DOI: 10.14357/19922264190112

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-20-02153 ofi_m. RZhD).

References

- OGC Engineering Reports. Available at: <http://www.opengeospatial.org/standards/> (accessed March 19, 2019).
- Ruiz, J. J., F. J. Ariza, M. A. Ureña, and E. B. Blázquez. 2011. Digital map conflation: A review of the process and a proposal for classification. *Int. J. Geogr. Inf. Sci.* 25(9):1439–1466.
- Schwinn, A., and J. Schelp. 2005. Design patterns for data integration. *J. Enterprise Information Management* 18(4):471–482.
- López-Vázquez, C., and M. A. M. Callejo. 2013. Point- and curve-based geometric conflation. *Int. J. Geogr. Inf. Sci.* 27(1):192–207.
- OGC Web Services Testbed, Phase 8 (OWS-8) Demonstrations, Available at: <http://www.opengeospatial.org/pub/www/ows8/index.html> (accessed March 2018).
- Botts, M., G. Percivall, C. Carl Reed, and J. Davidson. April 2, 2013. OGC Sensor Web Enablement (SWE): Overview and high level architecture. OGC White Paper. Document OGC 07-165r1.
- Pravia, M. 2008. Generation of a fundamental data set for hard/soft information fusion. *11th Conference (International) on Information Fusion*. Cologne: International Society of Information Fusion. 134–145.
- Percivall, G., ed. December 13, 2010. OGC Fusion Standards Study, Phase 2 Engineering Report. OGC Document 10-184. Available at: <http://www.opengeospatial.org/files/?artifact.id=41573> (accessed April 2018).
- Butenuth, M., G. von Goesseln, and M. Sester. 2007. Integration of heterogeneous geospatial data in a federated database. *ISPRS J. Photogramm.* 62(5):328–346.
- Al-Bakri, M., and D. Fairbairn. 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *Int. J. Geogr. Inf. Sci.* 26(8):1437–1456.
- Koukoletsos, T., M. Haklay, and C. Ellul. 2012. Assessing data completeness of VGI through an automated matching procedure for linear data. *T. GIS* 16(4):477–498.
- Stankute, S., and H. Asche. 2012. A data fusion system for spatial data mining, analysis and improvement. *Computational science and its applications*. Eds. B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. A. C. Rocha, D. Taniar, and B. O. Apduhan. Lecture notes in computer science ser. Springer-Verlag. 7334:439–449.
- Cholvy, L. 2007. Modelling information evaluation in fusion. *10th Conference (International) on Information Fusion Proceedings*. Quebec, Canada: IEEE. CD-ROM. 1–6. doi: 10.1109/ICIF.2007.4408060.
- Bernard, L., I. Kanellopoulos, A. Annoni, and P. Smits. 2005. The European Geoportal — one step towards the establishment of a European spatial data infrastructure. *Comput. Environ. Urban* 29(1):15–31.

Received July 4, 2018

Contributors

Dulin Sergey K. (b. 1950) — 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; principal scientist, Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation; skdulin@mail.ru

Dulina Natalia G. (b. 1947) — Candidate of Science (PhD) in technology, leading programmer, 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; ngdulina@mail.ru

Kozhunova Olga S. (b. 1982) — Candidate of Science (PhD) in technology, Head of Laboratory, 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; kozhunovka@mail.ru

ЦЕЛЕНАПРАВЛЕННОЕ РАЗВИТИЕ СИСТЕМ ЛИНГВИСТИЧЕСКИХ ЗНАНИЙ: ВЫЯВЛЕНИЕ И ЗАПОЛНЕНИЕ ЛАКУН*

И. М. Зацман¹

Аннотация: Дано описание процесса целенаправленного развития лингвистических типологий² как форм представления знаний об исследуемых языковых единицах (ЯЕ). Рассматриваемая задача заключается в обнаружении лакун в системе современного знания (ССЗ) об исследуемых ЯЕ, что предполагает выбор и использование некоторого эталона, отражающего современный уровень знания в соответствующей предметной области. Процесс обнаружения лакун представляет собой вид лингвистического аннотирования текстов, содержащих исследуемые ЯЕ, с применением методов и средств информатики. В качестве источника новых знаний для заполнения лакун используются параллельные тексты, которые включают оригинальные литературные произведения и их переводы. Цель статьи состоит в описании подхода к обнаружению лакун в процессе аннотирования параллельных текстов, содержащих исследуемые ЯЕ и их переводы.

Ключевые слова: лингвистическая типология; параллельные тексты; компьютерная лингвистика; извлечение новых знаний; информационная технология; целенаправленное развитие типологий

DOI: 10.14357/19922264190113

1 Введение

Разработка методов и средств информатики для обнаружения лакун в ССЗ о языке в процессе аннотирования параллельных текстов [1–4] — одна из задач проекта по гранту РФФИ, который в настоящее время выполняется в Институте проблем информатики ФИЦ ИУ РАН. Обнаружение лакун является основой для последующего целенаправленного пополнения существующих или формирования новых лингвистических типологий как форм представления лингвистами извлеченного ими знания об исследуемых ЯЕ [5–8].

Актуальность проблемы обнаружения лакун проявляется наиболее наглядно в процессе разработки и развития систем машинного перевода, когда имеющиеся типологии моделей перевода оказываются неполными для тех или иных классов ЯЕ и при этом отсутствуют средства информатики, обеспечивающие целенаправленное их пополнение. Более того, для некоторых классов ЯЕ такие типологии могут полностью отсутствовать. Тогда речь идет о необходимости формирования новых лингвистических типологий. Например, практически полностью отсутствует типология русских

конструкций с модальным значением как переводных эквивалентов немецких модальных глаголов [9] и сегодня нет средств информатики, обеспечивающих ее целенаправленное формирование. Именно в подобных ситуациях возникает необходимость в формировании новых и развитии существующих типологий, чем и обусловлена актуальность создания обеспечивающих их формирование методов и средств информатики.

Таким образом, речь идет не о представлении в виде типологий конвенциональных лингвистических знаний, эксплицированных в справочниках и словарях. Предлагается подход к выявлению лакун в ССЗ о языке, представленного в справочниках и словарях, на основе автоматизированной обработки параллельных текстов. Описание предлагаемого подхода иллюстрируется примером, в котором цель исследования состоит в уточнении типологии значений немецких модальных глаголов и их переводных эквивалентов на русском языке. Ее планируется пополнять в процессе обработки немецко-русских параллельных текстов [9].

Процесс пополнения существующих или формирования новых лингвистических типологий

*Работа выполнена в Институте проблем информатики ФИЦ ИУ РАН при частичной поддержке РФФИ (проект 18-07-00192).

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

²В лингвистике типологии используются, как правило, для описания сходства и различий между языками. В статье этот термин используется в другом значении: для структурированного описания связей исследуемых языковых единиц с их значениями и переводными эквивалентами.

Таблица 1 Примеры объектов интерпретации из трех произведений

Sollte jetzt etwa eine Predigt stattfinden? [Franz Kafka. Der Prozess (1914)]	Неужели сейчас кто-то будет читать проповедь? [Франц Кафка. Процесс (Р. Райт-Ковалева, 1965)]
Warum mußte er diese Demoiselle Stüwing heiraten und den. . . Laden. . . [Thomas Mann. Buddenbrooks (1896-1900)]	Зачем ему понадобилось жениться на этой мадемуазель Штювинг с ее. . . лавкой? [Томас Манн. Будденброки (Н. Ман, 1953)]
Er durfte nun eine Weile lang guten Gewissens ruhen. [Patrick Süskind. Das Parfum: Die Geschichte eines Mörders (1985)]	Теперь он имел право некоторое время отдыхать. [Патрик Зюскинд. Парфюмер: История одного убийцы (Э. Венгерова, 1992)]

включает несколько итерационно повторяемых стадий, которые в совокупности образуют одну итерацию целенаправленного извлечения знаний и представления их в виде новых типологий или новых рубрик, включаемых в существующие типологии. Каждая итерация состоит из следующих стадий (их описание является развитием результатов работы [10]).

Стадия 1. Каждая итерация начинается с определения одной или нескольких ЯЕ, исследуемых на текущей итерации, и поиска в параллельных текстах тех предложений, которые содержат исследуемую ЯЕ, и соответствующих им переводов. Одно или несколько найденных предложений и их переводы будем называть объектами интерпретации. Три их примера из работы [10] с тремя разными немецкими модальными глаголами, состоящие из одного предложения и его перевода, приведены в табл. 1.

Массив найденных объектов интерпретации может включать тысячи, а иногда и десятки тысяч предложений и их переводов [6, 7]. В таких случаях на первой стадии формируется подмассив объектов интерпретации для семантического их анализа именно на этой итерации.

В настоящее время наиболее представительный массив немецко-русских параллельных текстов в электронной форме хранится в Параллельном немецком корпусе (ПНК), который находится в открытом доступе [11]. Примеры в табл. 1 скопированы из этого корпуса. Его параллельные тексты являются выровненными (оригинальным предложениям поставлены в соответствие их переводы), т. е. необходимая их фрагментация на объекты интерпретации (т. е. на пары немецких и русских предложений) уже выполнена в процессе формирования этого корпуса.

Стадия 2. Выполняется лингвистическое аннотирование объектов интерпретации с исследуемыми ЯЕ, найденных в параллельных текстах на стадии 1. Метод аннотирования с использованием динамических фасетных классификаций описан в работах [1–3, 10, 12]. Он отличается от традиционных методов аннотирования [13] тем, что допускает возможность незавершенности процесса аннотирования объектов интерпретации. В этом случае анализируются причины незавершенности и в аннотации объектов интерпретации включаются специальные рубрики-теги, которые указывают на ту или иную причину. Например, в объекте интерпретации может встретиться новое значение некоторого немецкого модального глагола, которое не отражено в ССЗ, или в процессе аннотирования не удалось определить значение, в котором употреблен модальный глагол [10]. Возможна ситуация, когда не удастся определить смысл всего аннотируемого объекта интерпретации.

Для решения задач проекта интерес представляют в первую очередь те рубрики аннотаций, которые в качестве причины указывают на лакуны в ССЗ о значениях немецких модальных глаголов. В качестве авторитетного описания ССЗ об их значениях лингвистами был выбран немецко-русский словарь [14].

Стадия 3. Выполняется семантический анализ незавершенных аннотаций, помеченных рубриками, которые указывают на причины незавершенности. Предметом анализа являются те незавершенные аннотации, в которых указано, что не удалось определить значение немецкого модального глагола или смысл всего объекта интерпретации.

Стадия 4. Выполняется пополнение существующей типологии, которое основано на процессе доработки незавершенных аннотаций, включающих рубрики, которые в качестве причины указывают на лакуны в ССЗ о значениях.

Стадия 5. Вычисляются числовые параметры, характеризующие состояние процесса пополнения типологии на момент завершения итерации. После этого осуществляется переход к следующей итерации, начиная с первой стадии, если не достигнута желаемая степень заполнения обнаруженных лакун.

Цель статьи состоит, главным образом, в описании первых двух из пяти перечисленных стадий,

включая описание подхода к обнаружению лакун в процессе аннотирования параллельных текстов, содержащих исследуемые ЯЕ, в интересах пополнения существующей типологии. Кратко рассматриваются последние три стадии. Вопросы формирования новой типологии с нуля в этой статье не рассматриваются.

2 Поиск объектов интерпретации в параллельных текстах

В интересах решения задач проекта был создан макет базы знаний (МБЗ), который обеспечивает выполнение лингвистами пяти перечисленных стадий. Для проверки реализуемости предлагаемого подхода к обнаружению лакун в ССЗ о немецких модальных глаголах в МБЗ были загружены выровненные тексты ПНК общим объемом 2,6 млн словоупотреблений, включая 1,4 млн словоупотреблений в оригинальных текстах на немецком языке и 1,2 млн словоупотреблений в их переводах на русский язык. Загруженные немецко-русские па-

раллельные тексты состоят из пар немецких и русских предложений, всего — 83 190 пар (см. табл. 1 с тремя парами). В среднем одна пара содержит 31,25 словоупотреблений (16,83 в левой немецкой части и 14,42 в русской).

Предположим, что сначала для исследования лингвистами был выбран модальный глагол *sollen* и его значения [9]. По предварительной оценке, для уточнения типологии значений немецких модальных глаголов уже загруженный массив немецко-русских параллельных текстов дает возможность лингвистам найти в нем более 16 тыс. пар, включая более 2 тыс. пар, содержащих словоформы леммы *sollen* [10]. Для извлечения объектов интерпретации в МБЗ реализован поиск пар как по словоформам, так и по леммам. Поиск можно вести как по тексту одного произведения, так и по всем загруженным текстам.

Например, если задать поиск пар только в произведении Ф. Кафки «Процесс» по словоформе *sollte*, то будет найдена 61 пара с исследуемой ЯЕ, четыре из которых приведены в табл. 2 (включая пару из табл. 1).

Таблица 2 Примеры результатов поиска пар предложений по словоформе *sollte*

Номер пары	Оригинальный текст	Перевод
3433	Sollte jetzt etwa eine Predigt stattfinden?	Неужели сейчас кто-то будет читать проповедь?
3435	K sah an der Treppe hinab, die an die Säule sich anschmiegend zur Kanzel führte und so schmal war, als sollte sie nicht für Menschen, sondern nur zum Schmuck der Säule dienen.	К поглядел на лесенку, которая вела на кафедру, лепясь к самой колонне; она была настолько узкой, что, казалось, служила не людям, а просто украшению колонны.
3439	Sollte wirklich eine Predigt beginnen?	Неужели сейчас начнется проповедь?
3442	Und wenn es schon eine Predigt sein sollte, warum wurde sie nicht von der Orgel eingeleitet?	А если уж собираются начинать проповедь, почему перед этим не вступает орган?

Таблица 3 Объединение соседних пар в один объект интерпретации

Номер пары	Оригинальный текст	Перевод
3431	Das Ganze war wie zur Qual des Predigers bestimmt, es war unverständlich, wozu man diese Kanzel benötigte, da man doch die andere, große und so kunstvoll geschmückte zur Verfügung hatte.	Казалось, все было задумано нарочно для мучений проповедника, и нельзя было понять, зачем нужна эта кафедра, когда можно располагать главной, большой, столь искусно разукрашенной.
3432	K wäre auch diese kleine Kanzel gewiß nicht aufgefallen, wenn nicht oben eine Lampe befestigt gewesen wäre, wie man sie kurz vor einer Predigt bereitzustellen pflegt.	К, наверно, не заметил бы эту маленькую кафедру, если бы в ней не горела лампа, какие обычно зажигают для проповедника перед проповедью.
3433	Sollte jetzt etwa eine Predigt stattfinden?	Неужели сейчас кто-то будет читать проповедь?
3434	In der leeren Kirche?	Тут, в пустом соборе?
3435	K sah an der Treppe hinab, die an die Säule sich anschmiegend zur Kanzel führte und so schmal war, als sollte sie nicht für Menschen, sondern nur zum Schmuck der Säule dienen.	К поглядел на лесенку, которая вела на кафедру, лепясь к самой колонне; она была настолько узкой, что, казалось, служила не людям, а просто украшению колонны.

По умолчанию каждая пара, найденная по запросу лингвиста, является для него объектом семантической интерпретации. Однако лингвист может переформировать объект интерпретации, объединяя несколько пар. Например, если для интерпретации выбрана найденная пара № 3433, то она может быть объединена с соседними по тексту книги парами в единый объект интерпретации (табл. 3), что обеспечивает тот уровень его локализации в параллельных текстах произведений, который необходим для проведения исследования и извлечения новых знаний. При этом объектом исследования является модальный глагол в паре № 3433, а остальные пары образуют только контекст, необходимый для формирования аннотации на второй стадии итерации. Отметим, что в паре № 3435, которая входит и в поисковую выдачу (см. табл. 2), и в объект интерпретации (см. табл. 3), есть тот же самый модальный глагол, но это словопотребление будет рассматриваться и аннотироваться в рамках уже другого объекта интерпретации и в результате его семантического анализа будет сформирована другая аннотация.

3 Аннотирование объектов интерпретации

Для каждого объекта интерпретации с исследуемыми ЯЕ, найденного на стадии 1 (включая возможное объединение соседних пар), формируется аннотация. Как уже отмечалось выше, метод аннотирования, используемый в проекте, отличается тем, что допускает возможность незавершенности процесса аннотирования. Именно незавершенные аннотации позволяют зафиксировать неполноту (лакуны) в ССЗ о языке и о модальных глаголах, в частности. Ключевым элементом второй стадии является метод описания лакун, рассматриваемый ниже.

В процессе формирования аннотации происходит сопоставление значения глагола в каждом объекте интерпретации со списком значений в немецко-русском словаре [14], который был выбран как авторитетное описание ССЗ о значениях немецких модальных глаголов (как исследуемых ЯЕ) и их переводов на русский язык.

В настоящее время в словаре описано 13 значений исследуемого модального глагола *sollen*, которые обозначены в МБЗ следующими рубриками: *sollen-01*, ..., *sollen-13* (ряд значений рассматривается в работе [9]). Начальное состояние типологии (до аннотирования) включает рубрики только тех значений модальных глаголов, которые опи-

саны в словаре. Основная цель лингвистического аннотирования состоит в пополнении типологии и словаря с помощью МБЗ.

Если значение глагола в объекте интерпретации совпадает с одним из этих 13 значений, то в аннотации этого объекта интерпретации проставляется рубрика с соответствующим номером. Если значение глагола не совпадает ни с одним из этих 13 значений, то в аннотацию этого объекта интерпретации проставляется специальная рубрика *sollen-x*, которая говорит о незавершенности процесса аннотирования, а также о ее причине: отсутствие в типологии рубрики для того значения исследуемой ЯЕ, которое было найдено в объекте интерпретации.

Таким образом, хранимые в МБЗ незавершенные аннотации, соответствующие им объекты интерпретации и состояние типологии на каждой итерации позволяют зафиксировать текущую неполноту ССЗ о модальных глаголах, а также причину неполноты (отсутствие найденного нового значения), если проставлена специальная рубрика с литерой «x».

Пример незавершенной аннотации приведен в [10, табл. 3]. Последующий семантический анализ незавершенных аннотаций с одним и тем же тегом *sollen-x* и соответствующих им объектов интерпретации может привести к тому, что лингвисты извлекут из них несколько новых значений глагола *sollen* (= не описанных в словаре и в текущем состоянии пополняемой типологии). Таким образом, одной и той же специальной рубрикой-тегом *sollen-x* могут быть отмечены аннотации с несколькими новыми значениями модального глагола.

В процессе формирования аннотации возможна такая ситуация, когда значение модального глагола в некотором объекте интерпретации не удается определить без проведения дополнительного исследования на следующей, третьей, стадии, выполняемой после аннотирования. В этом случае в незавершенной аннотации ставится специальная рубрика *sollen-TBD* (to be defined), которая помечает аннотации и соответствующие объекты интерпретации, требующие проведения дополнительного исследования.

Аналогичная ситуация возникает, когда в процессе аннотирования не удастся определить смысл объекта интерпретации в целом. В этом случае ставится специальная рубрика *unknown*.

Предметом анализа на третьей стадии являются те незавершенные аннотации, в которых указано, что не удалось определить значение немецкого модального глагола, т. е. была проставлена рубрика с аббревиатурой *TBD*, или смысл всего объекта интерпретации, т. е. была проставлена рубрика *unknown*.

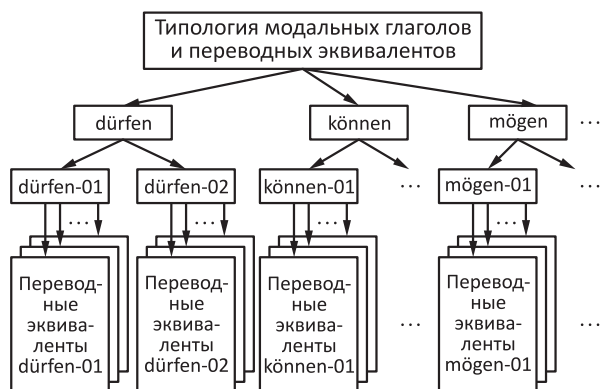
4 Пополнение типологии

Для пополнения типологии значений немецких модальных глаголов была разработана иерархическая трехуровневая структура (см. рисунок). Эта типология может обновляться лингвистами в двух ситуациях:

- (1) описано новое значение модального глагола и тогда типология обновляется на среднем уровне ее структуры;
- (2) зафиксирован новый переводной эквивалент на русском языке для некоторого значения модального глагола и тогда типология обновляется на нижнем уровне, где описываются эквиваленты.

На рисунке показаны три уровня типологии. Верхний (постоянный) уровень включает немецкие модальные глаголы *dürfen*, *können*, *mögen*, *lassen*, *müssen*, *sollen* и *wollen* (показаны только первые три глагола). Средний уровень включает номера значений немецких модальных глаголов. Число значений может увеличиваться в процессе анализа незавершенных аннотаций. Нижний уровень включает для каждого значения глагола те его переводные эквиваленты на русском языке, которые описаны лингвистами в процессе аннотирования.

Пополнение типологии, т.е. выполнение четвертой стадии итерации, основано на процессе доработки незавершенных аннотаций, имеющих в МБЗ специальную рубрику с литерой «х». Для каждой такой аннотации сначала создается описание того нового значения модального глагола, которое было извлечено из незавершенной аннотации. Например, если было извлечено новое значение глагола *sollen*, а на среднем уровне текущего состояния типологии было описано 13 его значений, то тогда в нее добавляется новая рубрика для 14-го зна-



Структура пополняемой типологии немецких модальных глаголов и их переводных эквивалентов на русском языке

чения на этом уровне. При этом в незавершенной аннотации специальная рубрика *sollen-x* заменяется на *sollen-14*. Таким образом, в случае описания на четвертой стадии нового значения незавершенная аннотация становится завершенной благодаря появлению в типологии новой рубрики.

Возможны ситуации, когда лингвисты не вводят в типологию рубрику с новым номером для обнаруженного нового значения немецкого модального глагола, а меняют дефиницию уже существующей рубрики, т.е. используют существующий номер для нового значения. В этом случае необходимо провести поиск по рубрике с этим номером и отредактировать все найденные аннотации (кроме случаев увеличения объема значения соответствующей рубрики).

На пятой стадии вычисляются следующие числовые параметры, характеризующие состояние процесса пополнения типологии на момент завершения итерации:

- число объектов интерпретации, содержащих исследуемые ЯЕ, хранящихся в МБЗ (оно может увеличиться, если в процессе исследования загружаются новые тексты);
- число аннотированных объектов интерпретации, в том числе включающих специальные рубрики;
- число новых значений исследуемых ЯЕ, извлеченных из объектов интерпретации и включенных в типологию;
- число новых переводных эквивалентов, включенных в типологию.

5 Заключение

Выполнение лингвистами итерационно повторяемых стадий с использованием МБЗ основано на новом подходе к уточнению лингвистических типологий. Он предполагает определение и применение специальных рубрик, фиксирующих причины, по которым процесс аннотирования отдельных объектов интерпретации не был завершен.

Как было показано на рисунке, конструкции русского языка с модальным значением, возникающие в переводе конструкций немецкого языка с модальными глаголами, группируются по значениям этих глаголов, так как цель исследования состоит в уточнения типологии их значений. Однако в упомянутой ранее работе [9] одновременно ставится еще одна цель: построение типологии русских конструкций с модальным значением для описания соответствующего фрагмента грамматики русского

языка. Эта цель уже относится только к русскому языку и не связана со средствами выражения модальности в немецком языке. Однако именно переводные эквиваленты, включаемые в нижний уровень типологии на рисунке, являются исходной информацией для создания новой типологии конструкций с модальным значением для описания соответствующего фрагмента грамматики русского языка.

Предлагаемый подход дает возможность не только уточнять уже существующие типологии, но и формировать новые типологии с нуля, что планируется продемонстрировать при продолжении проекта на примере построения типологии русских конструкций с модальным значением с использованием МБЗ. В существующих методах уточнения лингвистических типологий в процессе аннотирования не допускается никаких изменений в типологии. После завершения работ по аннотированию в типологию могут быть внесены изменения, но в процессе аннотирования она остается неизменной до его окончания [15]. Такие подходы не позволяют формировать типологии с нуля в процессе аннотирования, так как для его обеспечения уже требуется иметь некоторый начальный вариант типологии.

Литература

1. Дурново А. А., Зацман И. М., Лоцилова Е. Ю. Кросс-лингвистическая база данных для аннотирования логико-семантических отношений в тексте // Системы и средства информатики, 2016. Т. 26. № 4. С. 124–137.
2. Зацман И. М., Мамонова О. С., Шурова А. Ю. Обратимость и альтернативность генерализации моделей перевода коннекторов в параллельных текстах // Системы и средства информатики, 2017. Т. 27. № 2. С. 125–142.
3. Зацман И. М., Кружков М. Г., Лоцилова Е. Ю. Методы анализа частотности моделей перевода коннекторов и обратимость генерализации статистических данных // Системы и средства информатики, 2017. Т. 27. № 4. С. 164–176.
4. Kruzchkov M. G. Approaches to annotation of discourse relations in linguistic corpora // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 118–125.
5. Zatsman I., Buntman N., Kruzchkov M., Nuriev V., Zalizniak 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.
6. 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.
7. 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.
8. Zatsman I. Goal-oriented creation of individual knowledge: Model and information technology // 19th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2018. Vol. 2. P. 947–956.
9. Добровольский Д. О., Зализняк Анна А. Немецкие конструкции с модальными глаголами и их русские соответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. Т. 17(24). С. 172–184.
10. Зацман И. М. Стадии целенаправленного извлечения знаний, имплицитированных в параллельных текстах // Системы и средства информатики, 2018. Т. 28. № 3. С. 175–188.
11. Параллельный немецкий корпус. <http://www.ruscorpora.ru/search-para-de.html>.
12. Зализняк Анна А., Зацман И. М., Инькова О. Ю. Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–106.
13. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1468 p.
14. Немецко-русский словарь: актуальная лексика / Под ред. Д. О. Добровольского. — М.: Лексрус, 2019 (в печати).
15. Zufferey S., Degand L. Annotating the meaning of discourse connectives in multilingual corpora // Corpus Linguist. Ling., 2013. Vol. 13. Iss. 2. P. 399–423. doi: 10.1515/cllt-2013-0022.

Поступила в редакцию 15.01.19

GOAL-ORIENTED DEVELOPMENT OF LINGUISTIC KNOWLEDGE SYSTEMS: IDENTIFYING AND FILLING OF LACUNAE

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 description of the process of goal-oriented development of linguistic typologies as forms of knowledge representation about linguistic units in question is given. The task is to identify and fill the lacunae in the system of modern knowledge about the linguistic units. For identifying the lacunae, the selection of a reliable authority that reflects the current level of knowledge in the relevant subject area is necessary. The process of identifying the lacunae is a type of annotation of texts containing the linguistic units, using methods and means of informatics. As the source of new knowledge, parallel texts are used to fill the lacunae. They include original literary works and their translations. The objective of the paper is to describe the approach to identifying the lacunae in the process of annotating parallel texts containing the linguistic units and their translations.

Keywords: linguistic typology; parallel texts; computational linguistics; discovering new knowledge; information technology; goal-oriented development of typologies

DOI: 10.14357/19922264190113

Acknowledgments

The work was fulfilled at the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences with partial support of the Russian Foundation for Basic Research (project 18-07-00192).

References

1. Durnovo, A., I. Zatsman, and E. Loshchilova. 2016. Krosslingvisticheskaya baza dannykh dlya annotirovaniya logiko-semanticheskikh otnosheniy v tekste [Cross-linguistic database for annotating logical-semantic relations in the text]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(4):124–137.
2. Zatsman, I., O. Mamonova, and A. Shchurova. 2017. Obratimost' i al'ternativnost' generalizatsii modeley perevoda konnektorov v parallel'nykh tekstakh [Reversibility and alternativeness of generalization of connective translation models in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):125–142.
3. Zatsman, I., M. Kruzhkov, and E. Loshchilova. 2017. Metody analiza chastotnosti modeley perevoda konnektorov i obratimost' generalizatsii statisticheskikh dannykh [Methods of frequency analysis of connectives translations and reversibility of statistical data generalization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):164–176.
4. Kruzhkov, M. 2017. Approaches to annotation of discourse relations in linguistic corpora. *Informatika i ee Primeniya — Inform. Appl.* 11(4):118–125.
5. Zatsman, I., N. Buntman, M. Kruzhkov, V. Nuriev, and Anna A. Zalizniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 3:1063–1071.
6. Zatsman, I., and N. Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery. *16th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 851–860.
7. 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 Limited. 1:976–983.
8. Zatsman, I. 2018. Goal-oriented creation of individual knowledge: Model and information technology. *19th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 2:947–956.
9. Dobrovol'skij, D.O., and Anna A. Zalizniak. 2018. Nemetskiye konstruksii s modal'nymi glagolami i ikh russkie sootvetstviya: proekt nadkorporusnoy bazy dannykh [German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Computer Linguistics and Intellectual Technologies: Conference (International) “Dialog” Proceedings*. Moscow: RGGU. 17(24):172–184.
10. Zatsman, I. 2018. Stadii tselenapravlennoy izvlecheniya znaniy, implitsirovannykh v parallel'nykh tekstakh [Stages of goal-oriented discovery of knowledge implied in par-

- allel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):175–188.
11. *Parallelnyy nemetskiy korpus* [Parallel German corpus]. Available at: <http://www.ruscorpora.ru/search-para-de.html> (accessed January 18, 2019).
 12. Zaliznyak, Anna A., I. Zatsman, and O. Inkova. 2017. Nadkorpusnaya baza dannyykh konnektorov: postroenie sistemy terminov [Supracorpora database on connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–106.
 13. Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1468 p.
 14. Dobrovolskiy, D. O., ed. 2019 (in press). *Nemetsko-russkiy slovar': aktual'naya leksika* [German-Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
 15. Zufferey, S., and L. Degand. 2013. Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguist. Ling.* 13(2):399–423. doi: 10.1515/cllt-2013-0022.

Received January 15, 2019

Contributor

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

РЕСУРСНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПРОИЗВОЛЬНЫМ ОБСЛУЖИВАНИЕМ*

А. В. Горбунова¹, В. А. Наумов², Ю. В. Гайдамака³, К. Е. Самуйлов⁴

Аннотация: Представлен обзор ресурсных систем массового обслуживания (РСМО). Одному из методов их исследования посвящен целый раздел. Ценной особенностью предложенного метода является значительное упрощение анализа системы и при этом сохранение высокой точности оценки, а в отдельных случаях и отсутствие потери точности в принципе. Так, в частности, вместо исходной ресурсной модели было предложено рассматривать упрощенную, в которой после ухода заявки освобождается не тот объем ресурсов, который заявка занимала, а его некоторая случайная величина, что позволяет значительно упростить случайный процесс, описывающий поведение системы. Впоследствии для случая пуассоновского входящего потока и экспоненциального времени обслуживания была строго доказана эквивалентность результатов для исходной и упрощенной моделей. Отдельный практический интерес представляют работы с рекуррентной дисциплиной обслуживания. Результатам их анализа посвящена значительная часть работы.

Ключевые слова: ресурсная система массового обслуживания; непрерывный ресурс; дискретный ресурс; ограниченный ресурс; рекуррентное обслуживание; гетерогенная сеть; стационарное распределение; полумарковский процесс

DOI: 10.14357/19922264190114

1 Введение

Описание модели РСМО достаточно подробно представлено в [1]. Напомним лишь, что ее отличительной чертой является то, что заявкам помимо приборов и мест ожидания требуются дополнительные ресурсы. В случае нехватки какого-либо из них заявка теряется. Случайный объем дополнительного ресурса, занимаемого заявкой в подавляющем большинстве исследований на все время ее пребывания в системе, может носить как дискретный, так и непрерывный характер. Случаи детерминированных требований заявок к ресурсам подробно описаны в [2–4] и здесь не рассматриваются.

Если в ресурсной модели имеется единственный тип конечного ресурса, то, как правило, таким образом моделируется ограниченность памяти некоторого технического устройства или целой информационно-вычислительной системы [5–8]. Системы с множественными ресурсами моделируют беспроводные сети, в которых разного рода радиоресурсы должны распределяться между активными абонентами и освобождаться после завершения их обслуживания [9]. Системы с мно-

жественными ресурсами, которые выделяются не одновременно, а последовательно, могут также использоваться для моделирования беспроводных гетерогенных сетей, в которых обслуживание пользователей происходит последовательно вследствие необходимости выделения ресурсов на нескольких станциях, имеющих различные каналы передачи данных. В этом случае речь уже идет о модели многофазной РСМО [10, 11].

Сеть РСМО с заявками нескольких классов и сигналами, инициирующими переходы заявок между узлами, в которой объем ресурсов, запрашиваемый клиентами, может быть как положительным, так и отрицательным, применима для моделирования автономных точек доступа, которые имеют возможность динамически увеличивать свою пропускную способность, или для моделирования гетерогенной сети, развернутой в какой-либо области [12–15]. Положительные заявки в таком случае будут представлять запросы на выделение некоторого случайного объема работ, тогда как отрицательные заявки будут соответствовать увеличению объема ресурсов, доступных положительным заявкам. Таким образом, становится возможным

* Публикация подготовлена при финансовой поддержке Минобрнауки России (проект 2.882.2017/4.6).

¹ Российский университет дружбы народов, gorbunova_av@rudn.university

² Исследовательский институт инноваций, Хельсинки, Финляндия, valeriy.naumov@pfi.fi

³ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gaydamaka_yuv@rudn.university

⁴ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samouylov_ke@rudn.university

учитывать динамику пользователей, когда новые «объекты» могут предоставлять свои сетевые возможности соседним потребителям, а также возвращать их, когда они покидают территорию.

Статья организована следующим образом: в разд. 2 описываются основные типы РСМО с рекуррентным обслуживанием, методы их исследования и полученные результаты. Третий раздел посвящен новому подходу к исследованию РСМО — методу упрощения и результатам его применения к анализу ресурсных моделей. В заключении кратко подведены итоги работы.

2 Ресурсная система массового обслуживания с произвольным обслуживанием

Стоит отметить, что одним из первых в числе ученых, обративших внимание на актуальность исследования РСМО и много сделавших для ее изучения, является О. М. Тихоненко. Его авторству начиная с 1985 г. [7] принадлежит множество статей, посвященных этой тематике. В большинстве из них рассматриваются модели ресурсных систем с дисциплиной разделения процессора (Egalitarian Processor Sharing, EPS). Так, работа [16] посвящена анализу СМО $M|G|1|m$ ($m \leq \infty$) с разделением процессора, одним типом ресурса и существующей зависимостью между временем обслуживания заявки и объемом требуемого ей ресурса. Для этой системы определяются стационарные вероятности состояний и вероятность потери заявок.

В статье [17] исследуются однолинейные СМО вида $M|G|1|m$ ($m \leq \infty$) с единственным типом ограниченного объема ресурса, в которых поступающие заявки также характеризуются зависимостью между временем обслуживания заявки и случайной величиной (с.в.) требуемого ей объема ресурса. Для первой исследуемой СМО объем ресурса в системе не ограничен, в отличие от числа мест для ожидания; в результате ее анализа получено выражение для преобразования Лапласа—Стилтьеса суммарного объема ресурсов, занятого заявками. Для второй СМО, в которой объем ресурса ограничен, а накопитель — бесконечной емкости, получена оценка вероятности потери заявки.

В статьях [18, 19] анализируется СМО $M|G|1|\infty$ с разделением процессора и одним типом ресурса неограниченного объема $R = \infty$, при этом заявке помимо ресурса некоторого случайного объема требуется некоторое число приборов для ее обслуживания, и эти величины зависимы, т.е. задана их совместная функция распределения. В результа-

те было получено выражение для преобразования Лапласа—Стилтьеса суммарного объема ресурсов, занятых находящимися в системе заявками, в стационарном режиме. В [19] результаты были обобщены на нестационарный режим и, кроме того, было предложено использовать полученные в [18] результаты для оценки вероятности потери заявок в системе с ограниченным объемом ресурса, однако такая оценка далеко не всегда приемлема для использования [19]. Поэтому в работе [20] анализируется более общий вид СМО с ограничением на объем ресурса $R < \infty$: $M_L|G|1|m$ с разделением процессора и одним типом ресурса, т.е. в систему поступает L пуассоновских потоков заявок, причем заявка класса l , $l = \overline{1, L}$, характеризуется не только необходимым ей объемом ресурса, но и константой $n_l \leq N$ — числом приборов, необходимых для ее обслуживания. Объем требуемого заявке ресурса и ее длина, под которой понимается время пребывания заявки в системе при условии отсутствия в ней других заявок (т.е. фактически время обслуживания заявки), являются зависимыми величинами. Их совместное распределение также зависит от класса заявки. Было получено стационарное распределение числа заявок в системе и вероятности потери заявок каждого класса и, кроме того, аналогичные характеристики для некоторых частных случаев.

В [21] фактически исследуется СМО $M|G|N|0$ с единственным типом ограниченного ресурса $R \leq \infty$, причем каждая заявка характеризуется тремя случайными признаками: числом приборов, необходимых для обслуживания, объемом ресурса и временем обслуживания, т.е. помимо некоторого случайного объема ресурсов заявке требуется еще и случайное число приборов для обслуживания. Всего имеется N классов таких требований, для каждого класса заявок задана своя совместная функция распределения времени ее обслуживания и объема требуемого ей ресурса. Получено стационарное распределение числа заявок в системе и вероятности потери заявки каждого из имеющихся классов. Также в [21] проанализированы некоторые частные случаи, например: когда объем требуемого заявке ресурса и время ее обслуживания являются независимыми с.в., когда время обслуживания заявки пропорционально объему ее ресурса, а также, как и в работе [5], рассматривается частный случай, когда для обслуживания любой заявки требуется один прибор и функция распределения объема требуемых ресурсов одна и та же для всех классов заявок.

В серии работ другого авторства [8, 22–24] исследуются ресурсные системы, особенностью которых является инверсионный порядок обслужива-

ния. По мнению авторов, в случае с дисциплиной LIFO (last in, first out) можно получить рекуррентные алгоритмы, пригодные для численных методов вычисления стационарных вероятностей состояний и стационарного распределения времени пребывания заявки в системе. Однако стоит заметить, что предложенные рекуррентные алгоритмы в случае с непрерывной функцией распределения объема ресурса [22, 23] имеют большую вычислительную сложность в силу необходимости решения довольно сложных интегральных уравнений, а условие дискретности с.в. требуемого заявке объема ресурса [8, 24] позволяет получить более простые и эффективные алгоритмы расчета основных стационарных характеристик СМО.

В работе [22] рассматривается СМО типа $M_k|G|1|m$, $0 \leq m < \infty$, с инверсионным порядком обслуживания без прерывания и единственным типом ресурса объема R . Иными словами, на систему поступает пуассоновский входящий поток второго рода: его интенсивность λ_k зависит от числа заявок k , находящихся в системе. Как и в предыдущих работах, упомянутых в этом разделе, здесь тоже существует зависимость между временем обслуживания и требуемым заявке объемом ресурса, задаваемая совместной функцией распределения. Под инверсионным порядком обслуживания без прерывания подразумевается, что каждая принятая в систему заявка становится на первое место в очереди, причем если поступающая в систему заявка застает в очереди уже m заявок, но требуемый ей для обслуживания объем ресурса суммарно с необходимыми объемами ресурсов для имеющихся в системе заявок меньше R , то она вытесняет заявку, стоящую на приборе, и занимает ее место. Получены выражения для стационарных вероятностей различных состояний, в том числе распределение числа заявок в системе и суммарного объема занятого этими заявками ресурса, а также вероятность принятия заявки в систему, вероятность ее обслуживания; стационарное распределение времени пребывания заявки в системе. Также рассмотрены некоторые частные случаи: независимость времени обслуживания и объема необходимого ресурса; отсутствие ограничения на суммарный объем ($R = \infty$); экспоненциальное время обслуживания.

В статье [23] исследуется СМО с марковским входящим потоком вида $MAR|G|1|m$, так же как и в [22], с инверсионным порядком обслуживания, но уже с прерыванием, а также с единственным типом ограниченного ресурса объема R и зависимыми временем обслуживания и требуемым объемом ресурса для заявок. Получены формулы для стационарного распределения числа заявок в системе, для времени пребывания заявки в системе, а так-

же некоторые характеристики для частных случаев: независимые время обслуживания и объем ресурса; $m = \infty$; $R = \infty$.

В работе [24] исследуется однолинейная СМО с ограничением на суммарный объем занятых ресурсов одного типа (R) и геометрическим второго рода входящим потоком заявок $Geo_k|G|1|m$, функционирующая в дискретном времени с инверсионным порядком обслуживания без прерывания обслуживания и с существующей зависимостью для объема ресурса, т.е. фактически система, аналогичная СМО $M_k|G|1|m$ с непрерывным временем. Получены соотношения, которые позволяют вычислять основные стационарные характеристики (стационарные вероятности состояний и стационарное распределение времени пребывания заявки в системе).

В [8] рассматривается система $Geo_k|G|1|\infty$, аналогичная СМО из [24], но объем требуемого заявке ресурса имеет дискретное распределение, потери происходят только в том случае, когда суммарный объем находящихся в системе заявок превышает величину R . Получены соотношения, позволяющие вычислять основные стационарные характеристики СМО.

В статьях [10, 11, 25–27] исследуются многофазные РСМО с непуассоновским входящим потоком, неэкспоненциальным обслуживанием, бесконечным числом приборов и неограниченным объемом выделяемого ресурса на каждой из фаз. Так, в [27] проведен анализ РСМО типа $MMPP|G|\infty$, в [25] — $MAR|G|\infty$ и в [26] — $G|G|\infty$, во всех случаях речь идет об однофазных СМО. В статьях [10, 11] исследуются двухфазные СМО. Перечисленные статьи объединяет общий метод исследования систем: метод асимптотического анализа в условиях растущей интенсивности входящего потока с некоторыми особенностями в зависимости от типа входящего потока. В результате анализа получены асимптотические выражения для стационарного распределения вероятностей числа заявок в системе и суммарного объема занятых ресурсов (на каждой из фаз). Проведенные исследования позволили сделать вывод о том, что во всех случаях асимптотическое распределение суммарного объема занятых ресурсов является гауссовским с соответствующими параметрами и размерностью, согласующейся со случайным процессом, описывающим поведение изучаемой системы.

В статье [28] для РСМО с пуассоновским входящим потоком интенсивности λ , M типами ограниченных ресурсов и N приборами получены выражения для стационарных вероятностей процесса $X(t) = (\xi(t); \delta(t))$ ($\xi(t)$ — число заявок в системе; $\delta(t)$ — суммарный объем занятых ресурсов каждого

типа) в случае произвольного закона распределения времени обслуживания $B(x)$:

$$p_0 = \lim_{t \rightarrow \infty} P\{\xi(t) = 0\} = \left(1 + \sum_{k=1}^N G^{(k)}(\mathbf{R}) \frac{\rho^k}{k!}\right)^{-1};$$

$$Q_k(\mathbf{x}) = \lim_{t \rightarrow \infty} P\{\xi(t) = k; \delta(t) \leq \mathbf{x}\} =$$

$$= p_0 G^{(k)}(\mathbf{x}) \frac{\rho^k}{k!}, \quad \mathbf{0} \leq \mathbf{x} \leq \mathbf{R}, \quad 1 \leq k \leq N.$$

Здесь $G^{(k)}(\mathbf{x})$ — k -кратная свертка функции распределения:

$$G(\mathbf{x}) = \frac{1}{b} \int_0^\infty t H(dt, \mathbf{x}),$$

где $H(t, \mathbf{x})$ — совместная функция распределения длительности обслуживания и вектора объемов ресурсов, необходимых поступившей заявке; $\rho = \lambda b$, $b = \int_0^\infty t dB(t)$. Заметим, что в формуле (11) статьи [28] символ D следует читать как F .

3 Метод упрощения

Многолинейные СМО с множественными ресурсами, рекуррентным входящим потоком и произвольными функциями распределения длительностей обслуживания и объемов ресурсов впервые исследовались в [29]. В этой статье был предложен новый подход к анализу РСМО, позволяющий значительно упростить вычисления, что принципиально важно и заслуживает отдельного внимания. Итак, поскольку анализ РСМО довольно сложен, было предложено вместо исходной СМО исследовать ее упрощенный аналог, который функционирует аналогично исходной системе за исключением того, что объемы ресурсов, освобождаемые по завершении обслуживания, могут отличаться от тех, которые были выделены заявке в начале ее обслуживания. Случайный процесс, описывающий поведение упрощенной системы, легче поддается анализу, поскольку отпадает необходимость запоминания объемов ресурсов, занимаемых каждой обслуживаемой заявкой, и при анализе надо помнить лишь суммарные объемы занятых ресурсов. В этой же работе были продемонстрированы результаты моделирования исходных и упрощенных систем при пуассоновском входящем потоке и экспоненциальном обслуживании, которое показало, что средние значения объемов занимаемых ресурсов в стационарном режиме для них оказались очень близкими. Позднее в [30], также с помощью имитационного моделирования, было показано, что близки не только средние значения,

но и стационарные распределения суммарных объемов занимаемых ресурсов.

И, наконец, в [9] было строго обосновано, что переход к анализу упрощенной системы не меняет стационарного распределения суммарных объемов занятых ресурсов для случая пуассоновского входящего потока и экспоненциального времени обслуживания. Данный результат был доказан для случая непрерывной функции распределения объема требуемых заявок ресурсов. Вообще говоря, анализируемый здесь процесс $Y(t) = (\xi(t); \delta(t))$ не является марковским, поскольку в момент ухода заявки из системы должны быть освобождены ресурсы в объемах, равных объемам ресурсов, выделенных этой заявке при ее поступлении в систему. Однако при небольшом изменении исходной системы можно получить более простую систему с множественными ресурсами, для которой составной случайный процесс, компонентами которого являются два процесса (число заявок в системе и суммарный объем занимаемых ими ресурсов), образуют полумарковский процесс.

Итак, рассмотрим многолинейную СМО с пуассоновским входящим потоком, экспоненциальным временем обслуживания и M типами ресурсов, функционирование которой подчиняется классическому набору правил [1], за исключением того, что условие:

в момент окончания обслуживания заявки суммарный объем занятого ресурса каждого типа уменьшается на величину ресурса, выделенного этой заявке,

заменяется следующим:

* в момент окончания обслуживания заявки вектор суммарных объемов занятых ресурсов уменьшается на случайную величину, которая при заданном числе заявок в системе и векторе суммарных объемов занятых ресурсов не зависит от поведения системы до упомянутого момента времени.

Тогда процесс $Y^*(t) = (\xi^*(t); \delta^*(t))$, где $\delta^*(t)$ — это те же, что и для исходной системы, векторы суммарных объемов ресурсов, необходимых для обслуживания заявок, а $\xi^*(t)$ — число заявок в системе, является полумарковским и его предельное распределение определяется как

$$q_0^* = \lim_{t \rightarrow \infty} P\{\xi^*(t) = 0\};$$

$$Q_k^*(\mathbf{x}) = \lim_{t \rightarrow \infty} P\{\xi^*(t) = k; \delta^*(t) \leq \mathbf{x}\}, \quad 1 \leq k \leq N.$$

В результате решения соответствующей системы уравнений получаем, что предельные распределе-

ния суммарных объемов занятых ресурсов в исходной и упрощенной СМО совпадают: $q_0^* = p_0$; $Q_k^*(\mathbf{x}) = Q_k(\mathbf{x})$ [9].

В работе [31] рассматривается РСМО с единственным типом пуассоновского входящего потока и M типами ограниченных ресурсов. Здесь доказана инвариантность стационарного процесса $X(t) = (\xi(t); \delta(t))$ относительно функции распределения для упрощенной СМО, т.е. в случае произвольного закона распределения времени обслуживания $B(x)$ стационарные вероятности упрощенной СМО p_0^* и $Q_k^*(\mathbf{x})$ будут иметь тот же вид, что и в случае экспоненциального распределения времени обслуживания. А стационарные вероятности случайного процесса $(\xi(t); \delta(t), \beta(t))$, где $\beta(t) = (\beta_1(t), \dots, \beta_{\xi(t)}(t))$ — время, прошедшее с момента поступления заявок на обслуживание, примут вид:

$$\begin{aligned} Q_k(\mathbf{x}, \tau_1, \dots, \tau_k) &= \\ &= \lim_{t \rightarrow \infty} P \{ \xi(t) = k; \delta(t) \leq \mathbf{x}; \beta_1(t) \leq \tau_1, \dots \\ &\dots, \beta_k(t) \leq \tau_k \} = p_0 F^{(k)}(\mathbf{x}) \frac{\rho^k}{k!}, \\ &\mathbf{0} \leq \mathbf{x} \leq \mathbf{R}, \quad 1 \leq k \leq N, \end{aligned}$$

где $\rho = \lambda b$, $b = \int_0^\infty t dB(t)$.

Для СМО с N приборами, одним типом дискретного ресурса общим объемом R , пуассоновским входящим потоком и экспоненциальным временем обслуживания в [32, 33] благодаря методу упрощения получены выражения для вероятности блокировки системы, стационарные вероятности простоя системы, маргинальные вероятности числа заявок в системе $q_{k,\cdot}$, среднее число заявок в системе, а также $q_{k,j}$ — стационарные вероятности того, что число заявок в системе равно k , а суммарный объем занятого ресурса равен j :

$$\begin{aligned} q_{k,\cdot} &= P \lim_{t \rightarrow \infty} P \{ \xi(t) = k \} = p_0 \frac{\rho^k}{k!} \sum_{i=0}^R p_i^{(k)}, & 0 < k \leq N; \\ q_{k,j} &= P \lim_{t \rightarrow \infty} P \{ \xi(t) = k; \delta(t) = j \} = p_0 \frac{\rho^k}{k!} p_j^{(k)}, & 0 \leq j \leq R, \quad 0 < k \leq N; \end{aligned}$$

где

$$p_0 = \left(1 + \sum_{k=1}^N \frac{\rho^k}{k!} \sum_{i=0}^R p_i^{(k)} \right)^{-1}, \quad 0 < k \leq N,$$

а $p_i^{(k)}$ — k -кратная свертка вероятности p_i с.в. $\sum_{i=1}^k r_i$, r_i — с.в. объема ресурса, необходимого поступившей в систему i -й заявке.

В [34] с помощью метода упрощения была рассмотрена РСМО с марковским входящим потоком типа $MAP|M|N|0$ и единственным типом ограниченного ресурса объема R . Система уравнений равновесия выведена в векторной форме и решена численно. Наряду со стационарным распределением вероятностей в статье представлены формулы для математического ожидания и дисперсии числа занятых ресурсов, а также вероятности блокировки. Результаты проиллюстрированы численным примером.

4 Заключение

В настоящем обзоре кратко представлены основные разновидности РСМО с рекуррентным обслуживанием, существующие методы их анализа, выражения для оценки основных вероятностно-временных характеристик.

Литература

1. Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 48–55.
2. Kelly F. P. Loss networks // Ann. Appl. Probab., 1991. No. 1. P. 319–378.
3. Ross K. W. Multiservice loss models for broadband telecommunication networks. — London: Springer-Verlag, 1995. 343 p.
4. Башарин Г. П., Самуйлов К. Е., Яркина Н. В., Гудкова И. А. Новый этап развития математической теории телеграфика // Автоматика и телемеханика, 2009. № 12. С. 16–28.
5. Ромм Э. Л., Скитович В. В. Об одном обобщении задачи Эрланга // Автоматика и телемеханика, 1971. № 6. С. 164–168.
6. Кац Б. А. Об обслуживании сообщений случайной длины // Теория массового обслуживания: Труды 3-й Всесоюзн. школы-совещания по теории массового обслуживания. — М.: МГУ, 1976. С. 157–168.
7. Тихоненко О. М. Распределение суммарного объема сообщений в однолинейной системе массового обслуживания с групповым поступлением // Автоматика и телемеханика, 1985. № 11. С. 78–83.
8. Печинкин А. В., Соколов И. А., Шоргин С. Я. Ограничение на суммарный объем заявок в дискретной системе $Geo|G|1|\infty$ // Информатика и её применения, 2012. Т. 6. Вып. 3. С. 107–113.
9. Наумов В. А., Самуйлов К. Е., Самуйлов А. К. О суммарном объеме ресурсов, занимаемых обслуживаемыми заявками // Автоматика и телемеханика, 2016. № 8. С. 105–110.

10. Lisovskaya E., Moiseeva S., Pagano M. Infinite-server tandem queue with renewal arrivals and random capacity of customers // Comm. Com. Inf. Sc., 2017. Vol. 700. P. 201–216.
11. Moiseev A., Moiseeva S., Lisovskaya E. Infinite-server queueing tandem with MMPP arrivals and random capacity of customers // 31st European Conference on Modelling and Simulation Proceedings. — Budapest, Hungary, 2017. P. 673–679.
12. Samouylov K., Sopin E., Vikhrova O. Analysis of queueing system with resources and signals // Comm. Com. Inf. Sc., 2017. Vol. 800. P. 358–369.
13. Sopin E., Vikhrova O., Samouylov K. LTE network model with signals and random resource requirement // 9th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Munich, Germany: IEEE, 2017. P. 101–106.
14. Naumov V., Samouylov K. Analysis of multi-resource loss system with state dependent arrival and service rates // Probab. Eng. Inform. Sc., 2017. Vol. 31. No. 4. P. 413–419.
15. Наумов В. А., Самуйлов К. Е. Анализ сетей ресурсных систем массового обслуживания // Автоматика и телемеханика, 2018. № 5. С. 59–68.
16. Тихоненко О. М., Климович К. Г. Анализ систем обслуживания требований случайной длины при ограниченном суммарном объеме // Проблемы передачи информации, 2001. Т. 37. Вып. 1. С. 78–88.
17. Позняк Р. И., Ревинский В. В., Старовойтов А. М., Тихоненко О. М. Определение характеристик суммарного объема требований в однолинейных системах обслуживания с ограничениями // Автоматика и телемеханика, 1990. № 11. С. 182–186.
18. Sengupta B. The spatial requirement of an $M|G|1$ queue, or: How to design for buffer space // Modelling and performance evaluation methodology / Eds. F. Baccelli, G. Fayolle. — Lecture notes in control and information sciences book ser. — Springer, 1984. Vol. 60. P. 545–562.
19. Тихоненко О. М. Анализ системы обслуживания неоднородных требований с дисциплиной разделения процессора // Известия Национальной академии наук Беларуси. Сер. физико-математических наук, 2002. № 2. С. 105–111.
20. Тихоненко О. М. Система обслуживания с разделением процессора и ограниченными ресурсами // Автоматика и телемеханика, 2010. № 5. С. 84–98.
21. Тихоненко О. М. Обобщенная задача Эрланга для систем обслуживания с ограниченным суммарным объемом // Проблемы передачи информации, 2005. Т. 41. Вып. 3. С. 64–75.
22. Печинкин А. В. Система $M_i|G|1|n$ с дисциплиной LIFO и ограничением на суммарный объем требований // Автоматика и телемеханика, 1998. № 4. С. 106–116.
23. Печинкин А. В. Система $MAP|G|1|n$ с дисциплиной LIFO с прерыванием и ограничением на суммарный объем требований // Автоматика и телемеханика, 1999. № 12. С. 114–120.
24. Касконе А., Манзо Р., Печинкин А. В., Шоргин С. Я. Система $Geo_m|G|1|n$ с дисциплиной LIFO без прерывания обслуживания и ограничением на суммарный объем заявок // Автоматика и телемеханика, 2011. № 1. С. 107–120.
25. Кононов И. А., Лисовская Е. Ю. Исследование бесконечнолинейной $SMO MAP|GI|\infty$ с заявками случайного объема // Информационные технологии и математическое моделирование: Мат-лы XV Междунар. конф. имени А. Ф. Терпугова. — Томск: ТГУ, 2016. Ч. 1. С. 67–71.
26. Лисовская Е. Ю., Моисеева С. П. Асимптотический анализ немарковской бесконечнолинейной системы обслуживания требований случайного объема с входящим рекуррентным потоком // Вестник Томского государственного университета. Управление, вычислительная техника и информатика, 2017. № 39. С. 30–38.
27. Lisovskaya E., Moiseeva S., Pagano M., Potatueva V. Study of the MMPP $|GI|\infty$ queueing system with random customers' capacities // Информатика и её применения, 2017. Т. 11. № 4. С. 111–119.
28. Наумов В. А., Самуйлов А. К. О связи ресурсных систем массового обслуживания с сетями Эрланга // Информатика и её применения, 2016. Т. 10. Вып. 3. С. 9–14.
29. Наумов В. А., Самуйлов К. Е. О моделировании систем массового обслуживания с множественными ресурсами // Вестник РУДН. Сер. Математика, информатика, физика, 2014. № 3. С. 60–64.
30. Naumov V., Samouylov K., Sopin E., Andreev S. Two approaches to analyzing dynamic cellular networks with limited resources // 6th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — St. Petersburg: IEEE, 2014. P. 485–488.
31. Samouylov K., Gaidamaka Yu., Sopin E. Simplified analysis of queueing systems with random requirements // Statistics and simulation / Eds. J. Pilsz, D. Rasch, V. B. Melas, K. Moder. — Springer proceedings in mathematics & statistics book ser. — Springer, 2018. Vol. 231. P. 381–390.
32. Samouylov K., Sopin E., Vikhrova O. Analyzing blocking probability in LTE wireless network via queueing system with finite amount of resources // Comm. Com. Inf. Sc., 2015. Vol. 564. P. 393–403.
33. Вихрова О. Г., Самуйлов К. Е., Сопин Э. С., Шоргин С. Я. К анализу показателей качества обслуживания в современных беспроводных сетях // Информатика и её применения, 2015. Т. 9. Вып. 4. С. 48–55.
34. Sopin E., Samouylov K. On the analysis of the limited resources queueing system under MAP arrivals // Conference (International) on Applied Mathematics, Computational Science and Systems Engineering Proceedings. — Athens, Greece: EDP Sciences, 2018. Vol. 16. Art. No. 01008. 4 p.

Поступила в редакцию 15.01.19

RESOURCE QUEUING SYSTEMS WITH GENERAL SERVICE DISCIPLINE

A. V. Gorbunova¹, V. A. Naumov², Yu. V. Gaidamaka^{1,3}, and K. E. Samouylov^{1,3}

¹Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

²Service Innovation Research Institute, 8A Annankatu, Helsinki 00120, Finland

³Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article gives an overview of resource queuing systems with the concentration on the methods of their investigation. A valuable part of the article is devoted to the method, which leads to a significant simplification of the system analysis while maintaining high accuracy of the estimate, and in some cases without any loss of accuracy. Simplification is to consider a system with random resource amount release at the instant of a customer departure instead of a system with the exact resource amount release equal to the occupied by the customer at the beginning of service. Subsequently, for the case of a Poisson flow of arrivals and exponential service time, the equivalence of the results for the initial and the simplified models was rigorously proved. In addition, a significant part of the paper is devoted to the overview of publications on the recurrent service discipline.

Keywords: resource queueing systems; continuous resource; discrete resource; limited resource; recurrent service; heterogeneous network; stationary distribution; semi-Markov process

DOI: 10.14357/19922264190114

Acknowledgments

The publication was supported by the Ministry of Education and Science of the Russian Federation (project No. 2.882.2017/4.6).

References

- Gorbunova, A. V., V. A. Naumov, Yu. V. Gaidamaka, and K. E. Samouylov. 2018. Resursnye sistemy massovogo obsluzhivaniya kak modeli besprovodnykh sistem svyazi [Resource queuing systems as models of wireless communication systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):48–55.
- Kelly, F. P. 1991. Loss networks. *Ann. Appl. Probab.* 1:319–378.
- Ross, K. W. 1995. *Multiservice loss models for broadband telecommunication networks*. London: Springer-Verlag. 343 p.
- Basharin, G. P., K. E. Samouylov, N. V. Yarkina, and I. A. Gudkova. 2009. A new stage in mathematical teletraffic theory. *Automat. Rem. Contr.* 70(12):1954–1964.
- Romm, E. L., and V. V. Skitovitch. 1971. Ob odnom obobshchenii zadachi Ehrlanga [On a generalization of the Erlang problem]. *Automat. Rem. Contr.* 6:164–168.
- Kats, B. A. 1976. Ob obsluzhivanii soobshcheniy sluchaynoy dliny [On serving messages of random length]. *Teoriya massovogo obsluzhivaniya: Trudy 3-y Vsesoyuzn. shkoly-soveshchaniya po teorii massovogo obsluzhivaniya* [3rd All-Union School-seminar on Queuing Theory Proceedings]. 157–168.
- Tikhonenko, O. M. 1985. Raspredelenie summarnogo ob"ema soobshcheniy v odnolineynoy sisteme massovogo obsluzhivaniya s gruppovym postupleniem [Distribution of the total message flow in a single-line service system]. *Automat. Rem. Contr.* 11:78–83.
- Pechinkin, A. V., I. A. Sokolov, and S. Ya. Shorgin. 2012. Ogranichenie na summarnyy ob"em zayavok v diskretnoy sisteme $Geo|G|1|\infty$ [A restriction on the total volume of demands in the discrete-time system $Geo|G|1|\infty$]. *Informatika i ee Primeneniya — Inform. Appl.* 6(3):107–113.
- Naumov, V. A., K. E. Samuilov, and A. K. Samuilov. 2016. On the total amount of resources occupied by serviced customers. *Automat. Rem. Contr.* 77(8):1419–1427.
- Lisovskaya, E., S. Moiseeva, and M. Pagano. 2017. Infinite-server tandem queue with renewal arrivals and random capacity of customers. *Comm. Com. Inf. Sc.* 700:201–216.
- Moiseev, A., S. Moiseeva, and E. Lisovskaya. 2017. Infinite-server queueing tandem with MMPP arrivals and random capacity of customers. *31st European Conference on Modelling and Simulation Proceedings*. Budapest. 673–679.
- Samouylov, K., E. Sopin, and O. Vikhrova. 2017. Analysis of queueing system with resources and signals. *Comm. Com. Inf. Sc.* 800:358–369.
- Sopin, E., O. Vikhrova, and K. Samouylov. 2017. LTE network model with signals and random resource requirement. *9th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops*. Munich, Germany: IEEE. 101–106.

14. Naumov, V., and K. Samouylov. 2017. Analysis of multi-resource loss system with state dependent arrival and service rates. *Probab. Eng. Inform. Sc.* 31(4):413–419.
15. Naumov, V. A., and K. E. Samuilov. 2018. Analysis of networks of the resource queuing systems. *Automat. Rem. Contr.* 79(5):822–829.
16. Tikhonenko, O. M., and K. G. Klimovich. 2001. Analiz sistem obsluzhivaniya trebovaniy sluchaynoy dliny pri ogranichenom summarnom ob”eme [Analysis of queuing systems for random-length arrivals with limited cumulative volume]. *Probl. Inform. Transm.* 37(1):78–88.
17. Poznyak, R. I., V. V. Revinskiy, A. M. Starovoytov, and O. M. Tikhonenko. 1990. Opredelenie kharakteristik summarnogo ob”ema trebovaniy v odnolineynykh sistemakh obsluzhivaniya s ogranicheniyami [Calculation of characteristics of the total amount of customers in single-server queueing systems with constraints]. *Automat. Rem. Contr.* 11:182–186.
18. Sengupta, B. 1984. The spatial requirement of an $M|G|1$ queue, or: How to design for buffer space. *Modelling and performance evaluation methodology*. Eds. F. Baccelli and G. Fayolle. Lecture notes in control and information sciences book ser. Springer. 60:545–562.
19. Tikhonenko, O. M. 2002. Analiz sistemy obsluzhivaniya neodnorodnykh trebovaniy s distsiplinoy razdeleniya protsessora [Analysis of the queueing system with heterogeneous customers and processor sharing discipline]. *Izvestiya Natsional’noy akademii nauk Belarusi. Ser. fiziko-matematicheskikh nauk* [Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics Ser.] 2:105–111.
20. Tikhonenko, O. M. 2010. Queuing system with processor sharing and limited resources. *Automat. Rem. Contr.* 71(5):803–815.
21. Tikhonenko, O. M. 2005. Obobshchennaya zadacha Ehrlanga dlya sistem obsluzhivaniya s ogranichenym summarnym ob”emom [Generalized Erlang problem for queueing systems with bounded total size]. *Probl. Inform. Transm.* 41(3):64–75.
22. Pechinkin, A. V. 1998. Sistema $M_i|G|1|n$ s distsiplinoy LIFO i ogranicheniem na summarnyy ob”em trebovaniy [$M_i|G|1|n$ system with LIFO discipline and constrained total amount of items]. *Automat. Rem. Contr.* 4:106–116.
23. Pechinkin, A. V. 1999. Sistema $MAP|G|1|n$ s distsiplinoy LIFO s preryvaniem i ogranicheniem na summarnyy ob”em trebovaniy [The $MAP|G|1|n$ system with LIFO service discipline with interruptions and limitations on the total amount of requests]. *Automat. Rem. Contr.* 12:114–120.
24. Cascone, A., R. Manzo, A. V. Pechinkin, and S. Ya. Shorgin. 2011. $Geo_m|G|1|n$ system with LIFO discipline without interrupts and constrained total amount of customers. *Automat. Rem. Contr.* 72(1):99–110.
25. Kononov, I. A., and E. Yu. Lisovskaya. 2016. Issledovanie beskonechnolinyey SMO $MAP|GI|∞$ s zayavkami sluchaynogo ob”ema [Study of the infinite server queue $MAP|GI|∞$ with customers of random volume]. *15th Conference (International) named after A. F. Terpugov “Information Technologies and Mathematical Modelling” Proceedings*. Tomsk: Tomsk State University. 67–71.
26. Lisovskaya, E. Yu., and S. P. Moiseeva. 2017. Asimptoticheskiy analiz nemarkovskoy beskonechnolinyey sistemy obsluzhivaniya trebovaniy sluchaynogo ob”ema s vkhodyashchim rekurrentnym potokom [Asymptotical analysis of a non-Markovian queueing system with renewal input process and random capacity of customers]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel’naya tekhnika i informatika* [Tomsk State University J. Control Computer Science] 39:30–38.
27. Lisovskaya, E., S. Moiseeva, M. Pagano, and V. Potatueva. 2017. Study of the MMPP $|GI|∞$ queueing system with random customers’ capacities. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):111–119.
28. Naumov, V. A., and K. E. Samouylov. 2016. O svyazi resursnykh sistem massovogo obsluzhivaniya s setyami Ehrlanga [On relationship between queueing systems with resources and Erlang networks]. *Informatika i ee Primeneniya — Inform. Appl.* 10(3):9–14.
29. Naumov, V. A., and K. E. Samouylov. 2014. O modelirovani sistemy massovogo obsluzhivaniya s mnozhestvennymi resursami [On the modeling of queueing systems with multiple resources]. *Vestnik RUDN. Ser. Matematika, informatika, fizika* [RUDN J. Mathematics Information Sciences Physics] 3:60–64.
30. Naumov, V., K. Samouylov, E. Sopin, and S. Andreev. 2014. Two approaches to analyzing dynamic cellular networks with limited resources. *6th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) Proceedings*. St. Petersburg. 485–488.
31. Samouylov, K., Yu. Gaidamaka, and E. Sopin. 2018. Simplified analysis of queueing systems with random requirements. *Statistics and simulation*. Eds. J. Pilz, D. Rasch, V. B. Melas, and K. Moder. Springer proceedings in mathematics & statistics book ser. Springer. 231:381–390.
32. Samouylov, K., E. Sopin, and O. Vikhrova. 2015. Analyzing blocking probability in LTE wireless network via queueing system with finite amount of resources. *Comm. Com. Inf. Sc.* 564:393–403.
33. Vikhrova, O. G., K. E. Samouylov, E. S. Sopin, and S. Ya. Shorgin. 2015. K analizu pokazateley kachestva obsluzhivaniya v sovremennykh besprovodnykh setyakh [On performance analysis of modern wireless networks]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):48–55.
34. Sopin, E., and K. Samouylov. 2018. On the analysis of the limited resources queueing system under MAP arrivals. *Conference (International) on Applied Mathematics, Computational Science and Systems Engineering Proceedings*. 16:01008. 4 p.

Received January 15, 2019

Contributors

Gorbunova Anastasiya V. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; gorbunova_av@rudn.university

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

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gaidamaka-yuv@rudn.university

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; samouylov-ke@rudn.university

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ МОДЕЛИ БЕСПРОВОДНОЙ СЕТИ МЕЖМАШИННОГО ВЗАИМОДЕЙСТВИЯ, РАБОТАЮЩЕЙ В РАМКАХ ДВУХ ПОЛИТИК РАЗДЕЛЕНИЯ РАДИОРЕСУРСОВ*

Е. В. Маркова¹, А. А. Гольская², И. Л. Дзантиев³, И. А. Гудкова⁴, С. Я. Шоргин⁵

Аннотация: В настоящее время информационно-коммуникационные технологии (ИКТ) все глубже проникают во многие области современной жизни. Например, концепция интеграции ИКТ и Интернета вещей (Internet of Things, IoT) для управления городской инфраструктурой «Умный город» позволяет городской власти следить за изменениями и ситуацией в городе с помощью датчиков. При этом специализированные системы осуществляют сбор данных в автоматическом режиме без участия человека. Важным параметром при определении показателей эффективности беспроводных сетей межмашинного взаимодействия (Machine-to-Machine, M2M) — скорости передачи данных, вероятности блокировки — служит удаленность устройства (датчика) от радиопередающей аппаратуры (базовой станции, БС). Поэтому при описании такой сети в виде системы массового обслуживания с потоковым (гарантированная скорость передачи данных) или эластичным (негарантированная скорость) трафиком необходимо рассматривать входящий поток запросов на передачу данных от M2M-устройств таким образом, чтобы учесть расположение устройств относительно БС. Представлена модель соты беспроводной сети со стационарными M2M-устройствами, находящимися либо в пассивном, либо в активном состоянии. Устройства описываются точками, случайно возникающими на плоскости, и генерируют потоковый трафик, скорость передачи которого зависит от расположения устройства в соте, его мощности и уровня шума. Состояние системы описывает вектор переменной длины, компонентами которого служат расстояния от каждого активного устройства до БС. Рассмотрены две политики управления радиоресурсами — *round robin* (RR) и *full power* (FP), отличающиеся друг от друга распределением временного интервала обслуживания M2M-устройства и предоставляемой скоростью передачи данных. Проведен сравнительный анализ значений вероятности блокировки запроса на передачу данных.

Ключевые слова: беспроводная сеть; LTE; устройство межмашинного взаимодействия; индикатор качества канала; формула Шеннона; равномерное распределение; политика циклического обслуживания; политика обслуживания на максимальной мощности; вероятность блокировки

DOI: 10.14357/19922264190115

1 Введение

В связи со стремительным ростом числа используемых M2M-устройств [1] в широком спектре отраслей они становятся все более распространенными. Согласно прогнозу Cisco [2], количество M2M-соединений составит более половины от общего числа подключенных устройств и соединений (51%) и достигнет значения 14,6 млрд к 2022 г.

Для обслуживания M2M-приложений, генерирующих в совокупности большой объем данных с необходимым уровнем QoS (Quality of Service),

требуется развитие существующих и разработка новых технологий. Отметим, что для различных типов M2M-приложений эти технологии значительно отличаются [3].

Первый тип приложений включает в себя смарт-измерения, мобильный трекинг, электронное здравоохранение и т. д. и характеризуется малыми объемами данных и длительными задержками, а также низкой стоимостью и энергопотреблением.

Второй тип приложений относится к связи между подвижными объектами и характеризуется не-

* Публикация подготовлена при финансовой поддержке Минобрнауки России (проект 2.3397.2017/4.6).

¹ Российский университет дружбы народов, markova_ev@rudn.university

² Российский университет дружбы народов, golskaya_aa@rudn.university

³ Российский университет дружбы народов, dzantiev_il@rudn.university

⁴ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gudkova_ia@rudn.university

⁵ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

большими объемами данных, короткими задержками и высокой надежностью.

Третий тип относится к видеонаблюдению, для которого характерны большие объемы данных и высокая скорость их передачи.

Стремительный рост объема трафика, генерируемого М2М-приложениями [4, 5], в конечном итоге может привести к нехватке частотного диапазона [6]. Одним из возможных решений этой проблемы может стать более эффективное использование частотного ресурса, а именно использование более высоких диапазонов частот [7]. Наиболее эффективно использовать частотный диапазон помогают различные политики управления частотно-временными ресурсами, например политика, основанная на равномерном распределении временного ресурса между всеми обслуживаемыми М2М-устройствами (RR policy), или политика, основанная на передаче данных с максимальной мощностью (FP policy).

В статье предложена модель соты беспроводной сети LTE со стационарными М2М-устройствами, генерирующими потоковый трафик и способными находиться в двух состояниях: пассивном или активном. Устройства описываются точками, случайно возникающими на плоскости и получающими различные значения индикатора качества канала CQI (Channel Quality Indicator) в соответствии с их удаленностью от БС [8]. Параметр CQI может принимать значения от 0 до 15. Чем выше значение, тем лучше — тем выше скорость, которую может выделить БС оператора LTE. В связи с этим все устройства соты можно разделить по значению соответствующего CQI на 15 групп.

В работе рассмотрены две политики управления радиоресурсами, учитывающие значение CQI. Проведено построение математической модели для случая конечной мощности. Для обеих политик проведен анализ вероятности блокировки запросов устройств на передачу данных.

2 Описание модели

Рассмотрим соту сети беспроводной передачи данных радиуса R с равномерно распределенными по территории стационарными М2М-устройствами. С интенсивностью λ устройства переходят из пассивного в активное состояние и передают данные в восходящем канале (uplink channel) со средним временем передачи μ^{-1} . Пусть случайная величина (СВ) $\eta = 1, \dots, L$ определяет значение CQI, соответствующее каждому М2М-устройству. Тогда расстояние от устройства до БС определяется как $\xi_d(\eta) = RL^{-1}\eta$. Обозначим максимальную

мощность передачи сигнала М2М-устройством через p_{\max} . Текущая мощность передачи представляет собой СВ ξ_p , которая удовлетворяет условию $\xi_p \leq p_{\max}$. Предположим, что устройства передают данные с гарантированной скоростью r_0 . Будем считать, что максимальная мощность сигнала одинакова для всех устройств. Отметим, что рассматриваются идеальные условия, так называемая среда free space [9].

Предположим, что М2М-устройства распределены в соте равномерно, тогда функция распределения (ФР) $F_{\xi_d(\eta)}(d) = P\{\xi_d(\eta) \leq d\}$ СВ $\xi_d(\eta)$ равна 0 при $d < 0$, d^2/R^2 при $0 < d < R$ и 1 при $d > R$, а соответствующая плотность распределения $f_{\xi_d(\eta)}(d) = 2d/R^2$, $0 \leq d \leq R$.

Согласно теореме Шеннона достижимая М2М-устройством скорость передачи данных $r(\xi_d(\eta), \xi_p)$ зависит от ширины полосы частот ω восходящего канала, мощности передачи сигнала ξ_p , удаленности $\xi_d(\eta)$ от БС и определяется следующим образом:

$$\begin{aligned} r(\xi_d(\eta), \xi_p) &= \omega \ln \left(1 + \frac{G\xi_p}{\xi_d^\kappa(\eta)N_0} \right) = \\ &= \omega \ln \left(1 + \frac{G\xi_p}{(R\eta/L)^\kappa N_0} \right), \end{aligned}$$

где ω — ширина полосы пропускания; N_0 — мощность шума; G — коэффициент затухания сигнала; κ — степень затухания сигнала.

Рассмотрим две политики управления радиоресурсами, позволяющие учесть удаленность устройства от БС. Первая — RR policy — характеризуется разбиением временного ресурса на равные интервалы. Вторая — FP policy — передачей сигнала с максимальной мощностью. Проиллюстрируем функционирование обеих политик в зависимости от расположения устройств [10–12], а следовательно, и значений CQI.

На рис. 1 первое устройство расположено на расстоянии $\xi_d(\eta_1) = d_1$, второе — на расстоянии $\xi_d(\eta_2) = d_2$ от БС. А третье устройство может быть расположено либо близко к БС, на расстоянии $\xi_d(\eta_1^I) = d_3^I$ (случай I), либо далеко — $\xi_d(\eta_1^{II}) = d_3^{II}$ (случай II).

Рассмотрим политику RR. Временной ресурс делится поровну между всеми пользовательскими устройствами, которые, в свою очередь, регулируют мощность передачи для обеспечения гарантированной скорости r_0 , т. е. при переходе М2М-устройства в активное состояние все остальные устройства, находящиеся также в активном состоянии и передающие данные, в связи с уменьшением временного интервала, отведенного им на обслуживание, вынуждены увеличить мощность передачи данных для

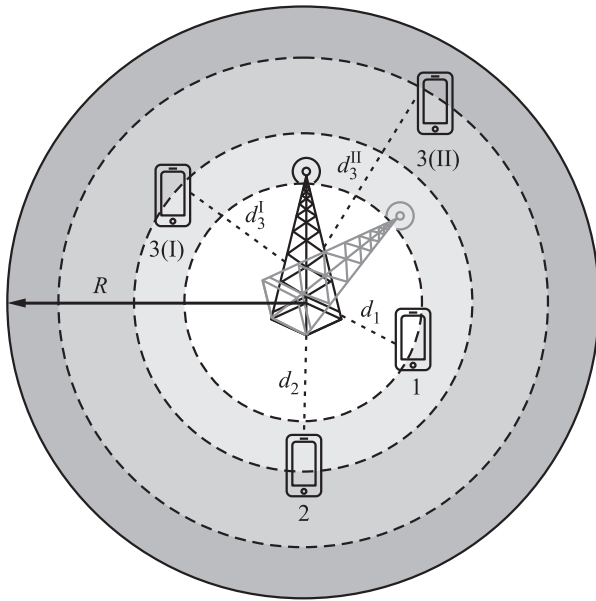


Рис. 1 Пример расположения М2М-устройства в сети

достижения гарантированной скорости r_0 . В противном случае запрос на обслуживание блокируется и М2М-устройство остается в пассивном состоянии.

Рассмотрим случай I, когда новое устройство находится близко к БС (рис. 2, а). Близкое рас-

положение позволяет увеличить мощность других устройств, не превышая ее максимального значения p_{max} , и принять запрос на обслуживание нового устройства. В случае сильной удаленности от БС (случай II) увеличения мощности М2М-устройств до максимального значения p_{max} недостаточно для достижения гарантированной скорости и запрос на обслуживание устройства блокируется (рис. 2, б).

Далее рассмотрим поведение системы при использовании политики *full power*, при которой все М2М-устройства всегда работают на максимальной мощности, а временной ресурс распределяется пропорционально достижимой каждым устройством скорости r_i , где i — номер устройства. В случае I временного интервала достаточно для достижения гарантированной скорости, поэтому возможно обслуживание нового устройства (рис. 3, а). В случае II из-за удаленности от БС устройству недостаточно отведенного временного интервала для достижения скорости r_0 и происходит блокировка запроса на обслуживание (рис. 3, б).

3 Построение математической модели

Введем следующие обозначения: $\xi(t)$ — число М2М-устройств, находящихся в активном со-

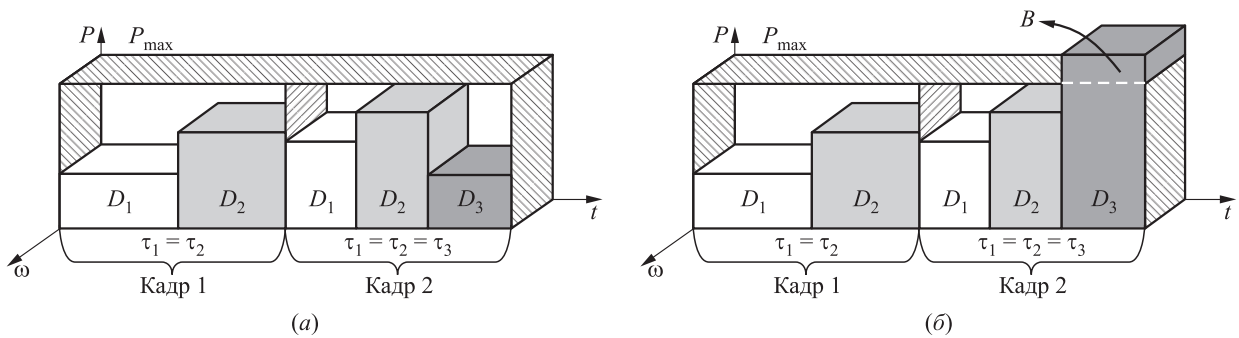


Рис. 2 Случаи I (а) и II (б) для политики RR

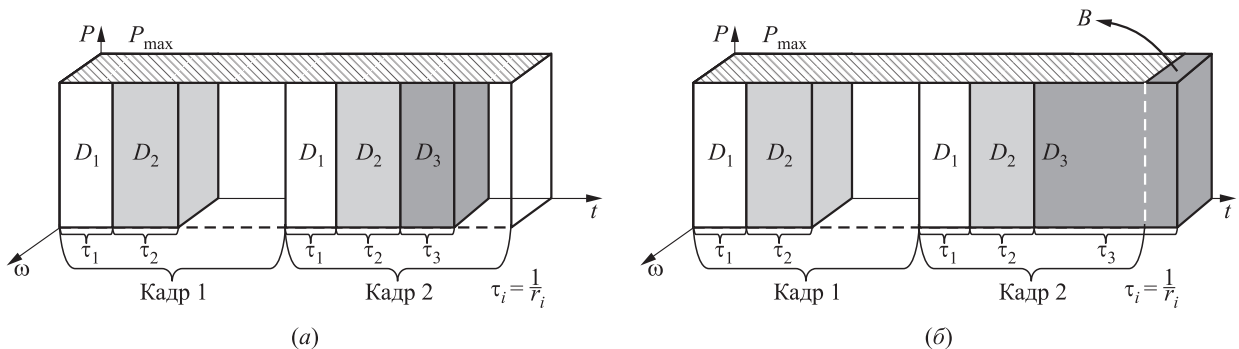


Рис. 3 Случаи I (а) и II (б) для политики FP

стоянии в момент времени $t > 0$; $\eta_i(t)$, $i = 1, \dots, \xi(t)$, — значение параметра CQI для i -го M2M-устройства в момент времени $t > 0$; $\{\xi(t), \eta_1(t), \dots, \eta_{\xi(t)}(t), t \geq 0\}$ — марковский случайный процесс (СП), описывающий поведение системы; (k, l_1, \dots, l_k) , $l_i = 1, \dots, L$, $k = 0, 1, \dots$, — состояние системы.

Для адекватного построения модели сначала рассмотрим случай, когда мощность передачи сигнала активными M2M-устройствами не имеет ограничений. Тогда пространство состояний системы L^\sim имеет следующий вид:

$$\begin{aligned} L^\sim &= \\ &= \{(0), (1, 1), \dots, (1, L), (2, 1, 1), \dots, (2, L, L), \dots \\ &\dots, \underbrace{\left(k, \underbrace{1, \dots, 1}_k\right), \dots, \left(k, \underbrace{L, \dots, L}_k\right), \dots\} = \\ &= \{(0), (k, l_1, l_2, \dots, l_k), l_i \in \{1, \dots, L\}, \\ &\quad i = 1, \dots, k, k = 1, 2, \dots\}. \end{aligned}$$

Предположим, что СВ η принимает значение, равное l , с вероятностью q_l , $l \in \{1, \dots, L\}$, которая с учетом равномерного распределения M2M-устройств в соте определяется как

$$q_l = \frac{2L - 2l - 1}{L^2}, \quad l = 1, \dots, L.$$

Пусть в системе k активных M2M-устройств, каждому из которых соответствует некоторое значение параметра CQI $16 - l_i$. Для упрощения расчетов агрегируем состояния системы по числу обслуживаемых M2M-устройств k [13]. Тогда пространство состояний L^\sim разбивается следующим образом:

$$L^\sim = \bigcup_{n=0}^{\infty} L^\sim(k), \quad L^\sim(k) = \{(k, l_1, l_2, \dots, l_k) \in L^\sim\}.$$

4 Сужение случайного процесса в соответствии с политикой управления радиоресурсами

В действительности мощность передачи сигнала не может быть бесконечной, а следовательно, ограничено и число обслуживаемых в соте M2M-устройств. Эти ограничения определяются в соответствии с выбранной политикой управления радиоресурсами. В связи с этим рассмотрим сужение СП $\{\xi(t), \eta_1(t), \dots, \eta_{\xi(t)}(t), t \geq 0\}$ над множеством $L \subset L^\sim$. Для корректного определения множества L введем функции доступа, соответствующие каждой

из рассматриваемых политик. Для политики RR функция доступа имеет вид:

$$g_{\xi_d(\eta)}(k, l_1, l_2, \dots, l_k) = \begin{cases} 1, & \frac{r_0}{r(\xi_d(\eta), p_{\max})} \leq \frac{1}{k+1}, \quad i = 1, \dots, k+1; \\ 0 & \text{в противном случае;} \end{cases} \quad (1)$$

для политики FP

$$g_{\xi_d(\eta)}(k, l_1, l_2, \dots, l_k) = \begin{cases} 1, & \sum_{i=1}^n \frac{r_0}{r(\xi_d(\eta), p_{\max})} \leq 1; \\ 0 & \text{в противном случае.} \end{cases} \quad (2)$$

Тогда пространство состояний $L_{RR}(k)$ или $L_{FP}(k)$, в зависимости от использования политики RR или FP, определяется в соответствии с формулой (1) или (2) следующим образом:

$$\begin{aligned} L_{RR}(k) &= \left\{ 0 \leq d_1 \leq R, \dots, 0 \leq d_k \leq R : \right. \\ &\quad \left. \frac{r_0}{\omega \ln(1 + Gp_{\max}/(d_i^k N_0))} \leq \frac{1}{k}, \quad i = 1, \dots, k \right\}, \\ L_{RR} &\subset L^\sim; \quad (3) \end{aligned}$$

$$\begin{aligned} L_{FP}(k) &= \left\{ 0 \leq d_1 \leq R, \dots, 0 \leq d_k \leq R : \right. \\ &\quad \left. \sum_{i=1}^k \frac{r_0}{\omega \ln(1 + Gp_{\max}/(d_i^k N_0))} \leq 1, \right\}, \\ L_{FP} &\subset L^\sim. \quad (4) \end{aligned}$$

Обозначим через

$$P(k-1) = P\{(k, l_1, l_2, \dots, l_k) \in L(k) | (k, l_1, l_2, \dots, l_{k-1}) \in L(k-1)\}$$

условную вероятность того, что k -е M2M-устройство с соответствующим значением CQI l_k , $l_k \in \{1, \dots, L\}$, будет обслужено при условии, что в системе уже обслуживается $k-1$ M2M-устройств, каждое из которых имеет соответствующее значение CQI l_1, \dots, l_{k-1} .

Тогда стационарное распределение вероятностей p_k , $k = 1, 2, \dots$, того, что в системе ровно k активных M2M-устройств, рассчитывается по формуле:

$$\begin{aligned} p_k &= \left(\frac{\sum_{j=0}^k \left((\lambda/\mu)^j \prod_{i=0}^{j-1} P(i) \right)}{j!} \right)^{-1} \times \\ &\quad \times \frac{(\lambda/\mu)^k \prod_{i=0}^{k-1} P(i)}{k!}, \quad k = 1, 2, \dots \end{aligned}$$

Основной характеристикой рассматриваемой модели служит вероятность B блокировки передачи данных от М2М-устройства, которая рассчитывается по формуле:

$$B = \sum_{k=1}^{\infty} (1 - P(k-1)) p_k. \quad (5)$$

5 Расчет условной вероятности в соответствии с выбранной политикой управления радиоресурсами

С учетом функции доступа (1) и пространства состояний (3) условная вероятность $P(k-1)$ при использовании политики RR имеет вид:

$$P(k-1) = \left(P \left\{ \frac{r_0}{r(\xi_d(\eta), p_{\max})} \leq \frac{1}{k} \right\} \right)^k \times \left(P \left\{ \frac{r_0}{r(\xi_d(\eta), p_{\max})} \leq \frac{1}{k-1} \right\} \right)^{-(k-1)}.$$

Так как СВ $\xi_d(\eta)$ является равномерно распределенной, то

$$P(k) = \frac{1}{R^2} \left(\frac{Gp_{\max}}{N_0} \right)^{2/\kappa} \left(e^{r_0 k/\omega} - 1 \right)^{2k/\kappa} \times \left(e^{(r_0 k+1)/\omega} - 1 \right)^{-2(k+1)/\kappa}.$$

Перейдем к расчету условной вероятности $P(k-1)$ для политики FR, опираясь на формулы (2) и (4). Для случая, когда в системе нет активных М2М-устройств, т. е. $k = 1$, получим:

$$P(0) = F_{\xi_d(\eta)} \left(\min \left\{ R, \left(\frac{Gp_{\max}}{(e^{r_0/\omega} - 1)N_0} \right)^{1/\kappa} \right\} \right).$$

Для вывода формул расчета остальных условных вероятностей $P(k-1)$, $k > 1$, воспользуемся определением условной вероятности и функцией доступа (4):

$$P(k-1) = P \left\{ \sum_{i=1}^k \frac{1}{r(l_i, p_{\max})} \leq \frac{1}{r_0}, \sum_{i=1}^{k-1} \frac{1}{r(l_i, p_{\max})} \leq \frac{1}{r_0} \right\} / P \left\{ \sum_{i=1}^{k-1} \frac{1}{r(l_i, p_{\max})} \leq \frac{1}{r_0} \right\},$$

где согласно центральной предельной теореме

$$P \left\{ \sum_{i=1}^k \frac{1}{r(l_i, p_{\max})} \leq \frac{1}{r_0} \right\} = \Phi \left(\frac{1 - \theta k r_0}{r_0 \sigma \sqrt{k}} \right).$$

Здесь θ и σ — математическое ожидание и среднеквадратичное отклонение независимых и одинаково распределенных СВ $1/r(l_i, p_{\max})$; $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x e^{-t^2/2} dt$ — стандартное нормальное распределение.

Для дальнейшего упрощения записи введем обозначения:

$$m_k = k r_0 E \frac{1}{r(l_i, p_{\max})};$$

$$\tau_k^2 = k r_0 \left(E \left(\frac{1}{r(l_i, p_{\max})} \right)^2 - \left(E \frac{1}{r(l_i, p_{\max})} \right)^2 \right).$$

Тогда условная вероятность при использовании политики FR примет вид:

$$P(k-1) = \frac{\Phi((1 - m_k)/\tau_k)}{\Phi((1 - m_{k-1})/\tau_{k-1})}.$$

Расчет математического ожидания случайных величин $1/r(l_i, p_{\max})$ и $(1/r(l_i, p_{\max}))^2$ представлен в [13].

6 Численный анализ

Для проведения численного анализа вероятности блокировки, рассчитываемой по формуле (5) с учетом используемой политики управления радиоресурсами, рассмотрим три различных сценария, в которых динамически меняются:

- (1) радиус соты;
- (2) мощность передачи;
- (3) интенсивность перехода устройств в активное состояние.

Сведем исходные данные в табл. 1, характеризующую динамические значения параметров, и табл. 2 с постоянными значениями.

Построим графики поведения вероятности блокировки для каждого из рассматриваемых сценариев и в рамках используемой политики распределения радиоресурсов (рис. 4).

Очевидно, что вероятность блокировки ниже при большей мощности передачи. При одинаковых значениях мощности вероятность блокировки принимает различные значения в зависимости

Таблица 1 Динамически меняющиеся данные

Сценарий	R	p_{\max}	λ
1	200 → 400 м	23 дБм (0,2 Вт) 42 дБм (15,85 Вт)	10
2	200, 400 м	23 → 42 дБм	10
3	200, 400 м	23 дБм (0,2 Вт) 42 дБм (15,85 Вт)	2 → 10

Таблица 2 Постоянные данные

ω	L	d_0	r_0	μ^{-1}	N_0	G	κ
10 МГц	15	$R/15$	1 Мбит/с	0,1 с	-60 дБм	197,43	5

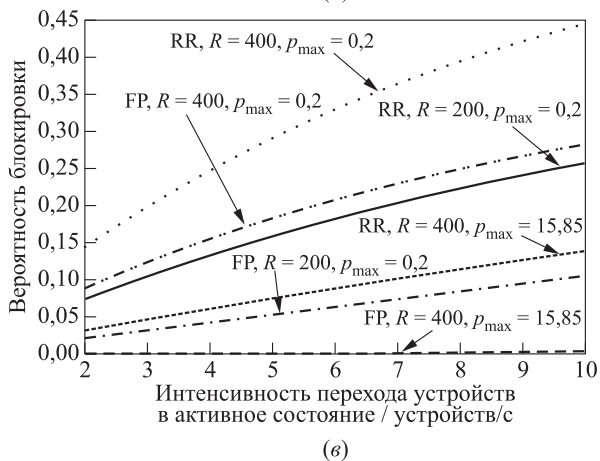
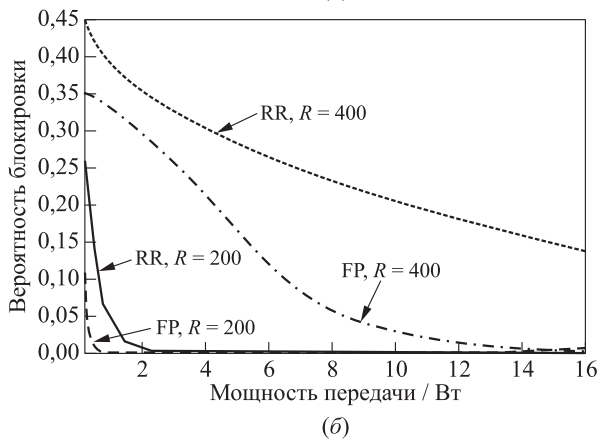
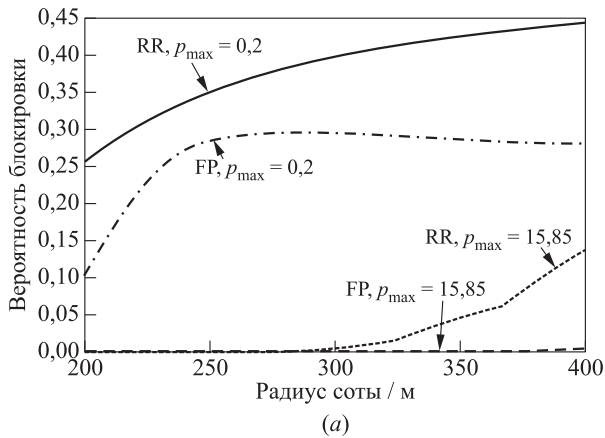


Рис. 4 Вероятность блокировки в рамках реализации сценариев 1 (а), 2 (б) и 3 (в)

от используемой политики распределения радиоресурсов. На рис. 4, а видно, что при использовании политики RR вероятность блокировки выше.

При увеличении мощности передачи данных вероятность блокировки снижается (рис. 4, б).

С увеличением частоты поступления новых запросов на обслуживание устройств, вероятность блокировки увеличивается (рис. 4, в).

7 Заключение

В статье построена модель соты беспроводной сети LTE с M2M-устройствами, равномерно распределенными в соте и сгруппированными по расстоянию от БС, определяемому с учетом соответствующего устройству значения CQI. Устройства генерируют потоковый трафик, передаваемый в восходящем канале с гарантированной скоростью. Рассмотрены две политики управления радиоресурсами, основанные на различных вариациях таких ресурсов, как ширина полосы частот, мощность передачи сигнала и время обслуживания M2M-запроса.

В рамках текущего исследования моделей сети с двумя различными политиками распределения (FP и RP) результаты численного анализа показали, что политика FP является преимущественной в использовании и наиболее эффективной, так как в этом случае вероятность блокировки ниже.

Литература

1. Laya A., Alonso L., Alonso-Zarate J. Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives // IEEE Commun. Surv. Tut., 2014. Vol. 16. Iss. 1. P. 4–16. doi: 10.1109/SURV.2013.111313.00244.
2. Cisco Visual Networking Index: Forecast and trends, 2017–2022. White Papers, November 26, 2018. Document ID: 1543280537836565.
3. Future technology trends of terrestrial IMT systems. ITU-R Reports. Report M.2320, November 2014.
4. Aijaz A., Tshangini M., Nakhai M.R., Chu X., Aghvami A.-H. Energy-efficient uplink resource allocation in LTE networks with M2M/H2N co-existence under statistical QoS guarantees // IEEE T. Commun., 2014. Vol. 62. Iss. 7. P. 2353–2365. doi: 10.1109/TCOMM.2014.2328338.
5. Hamdoun S., Rachedi A., Ghamri-Doudane Y. A flexible M2M radio resource sharing scheme in LTE networks

- within an H2H/M2M coexistence scenario // Conference (International) on Communications. — IEEE, 2016. P. 1–7. doi: 10.1109/ICC.2016.7511237.
6. Ericsson mobility report: On the pulse of the networked society. — Ericsson, June 2016.
 7. Requirements, evaluation criteria and submission templates for the development of IMT-2020 // ITU-R Report M.2411, November 2017.
 8. Samouylov K., Gudkova I., Markova E., Dzantiev I. On analyzing the blocking probability of M2M transmissions for a CQI-based RRM scheme model in 3GPP LTE // Information technologies and mathematical modelling — queueing theory and applications / Eds. A. Dudin, A. Gortsev, A. Nazarov, R. Yakupov. — Communications in computer and information science ser. — Springer, 2016. Vol. 638. P. 327–340. doi: 10.1007/978-3-319-44615-8_29.
 9. Markova E., Dzantiev I., Gudkova I., Shorgin S. Analyzing impact of path loss models on probability characteristics of wireless network with licensed shared access framework // 9th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2017. P. 20–25. doi: 10.1109/ICUMT.2017.8255189.
 10. Begishev V., Kovalchukov R., Samuylov A., Ometov A., Moltchanov D., Gaidamaka Y., Andreev S. An analytical approach to SINR estimation in adjacent rectangular cells // Internet of things, smart spaces, and next generation networks and systems / Eds. S. I. Balandin, S. D. Andreev, Y. Koucheryavy. — Lecture notes in computer science ser. — Springer, 2015. Vol. 9247. P. 446–458. doi: 10.1007/978-3-319-23126-6_39.
 11. Samuylov A., Moltchanov D., Gaidamaka Y., Andreev S., Koucheryavy Y. Random triangle: A baseline model for interference analysis in heterogeneous networks // IEEE T. Veh. Technol., 2016. Vol. 65. Iss. 8. P. 6778–6782. doi: 10.1109/TVT.2015.2481795.
 12. Naumov V., Samouylov K. Analysis of multi-resource loss system with state-dependent arrival and service rates // Probab. Eng. Inform. Sc., 2017. Vol. 31. Iss. 4. P. 413–419. doi: 10.1017/S0269964817000079.
 13. Markova E., Gudkova I., Ometov A., Dzantiev I., Andreev S., Koucheryavy Ye., Samouylov K. Flexible spectrum management in a smart city within licensed shared access framework // IEEE Access, 2017. Vol. 5. P. 22252–22261. doi: 10.1109/ACCESS.2017.2758840.

Поступила в редакцию 15.01.19

COMPARATIVE ANALYSIS OF PERFORMANCE MEASURES FOR A WIRELESS MACHINE-TO-MACHINE NETWORK MODEL OPERATING WITHIN TWO RADIO RESOURCE MANAGEMENT POLICIES

E. V. Markova¹, A. A. Golskaia¹, I. L. Dzantiev¹, I. A. Gudkova^{1,2}, and S. Ya. Shorgin^{1,2}

¹Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, 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: Currently, information and communication technologies (ICT) deeply penetrate into many areas of modern life. For example, the concept of integrating ICT and the Internet of Things for managing Smart City infrastructure allows city authorities to monitor changes and the situation in the city using sensors. Thus, specialized systems collect data automatically without human intervention. An important parameter in determining the performance of wireless networks of machine-to-machine (M2M) interaction — data transfer rates, blocking probabilities, becomes the distance between the device (sensor) and the radio transmitting equipment (base station, BS). Therefore, describing such a network in the form of a queueing system with streaming (guaranteed data transfer rate) or elastic traffic (nonguaranteed speed), it is necessary to consider the incoming stream of requests for data transmission of M2M devices in such a way as to take into account the distance between devices and BS. In this paper, there is built a cell model of a wireless network with stationary M2M devices that are in a passive or active state, shown by points that appear randomly on a plane. The devices generate streaming traffic which depends on the distance from the BS, the device transmit power, and the noise level. The state of the system is described by the vector of variable length, the components of which are the distance of each active device to the BS. Two different disciplines of radio resource separation are considered — “round robin” and “full power,” which differ

in the distribution of the time interval for servicing an M2M device and the data transfer rate provided. There is built a random process with states enlarged by the number of users and a formula for calculating the probability of blocking a data transfer request is proposed.

Keywords: wireless network; LTE; machine-to-machine communication; channel quality indicator; Shannon's formula; uniform distribution; round robin policy; full power policy; blocking probability

DOI: 10.14357/19922264190115

Acknowledgments

The publication was supported by the Ministry of Education and Science of the Russian Federation (project No. 2.3397.2017/4.6).

References

1. Laya, A., L. Alonso, and J. Alonso-Zarate. 2014. Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives. *IEEE Commun. Surv. Tut.* 16(1):4–16. doi: 10.1109/SURV.2013.111313.00244.
2. Cisco Visual Networking Index: Forecast and trends, 2017–2022. November 26, 2018. White Paper. Available at: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html> (accessed January 15, 2019).
3. Future technology trends of terrestrial IMT systems. November 2014. ITU-R Report M.2320. Available at: <https://www.itu.int/pub/R-REP-M.2320-2014> (accessed January 15, 2019).
4. Aijaz, A., M. Tshangini, M. R. Nakhai, X. Chu, and A.-H. Aghvami. 2014. Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees. *IEEE T. Commun.* 62(7):2353–2365. doi: 10.1109/TCOMM.2014.2328338.
5. Hamdoun, S., A. Rachedi, and Y. Ghamri-Doudane. 2016. A flexible M2M radio resource sharing scheme in LTE networks within an H2H/M2M coexistence scenario. *Conference (International) on Communications*. IEEE. 1–7. doi: 10.1109/ICC.2016.7511237.
6. On the pulse of the networked society. June 2016. Ericsson mobility report. Available at: <https://www.ericsson.com/assets/local/mobility-report/documents/2016/ericsson-mobility-report-june-2016.pdf> (accessed January 15, 2019).
7. Requirements, evaluation criteria and submission templates for the development of IMT-2020. November 2017. ITU-R Report M.2411. Available at: <https://www.itu.int/pub/R-REP-M.2411-2017> (accessed January 15, 2019).
8. Samouylov, K., I. Gudkova, E. Markova, and I. Dzantiev. 2016. On analyzing the blocking probability of M2M transmissions for a CQI-based RRM scheme model in 3GPP LTE. *Information technologies and mathematical modelling — queueing theory and applications*. Eds. A. Dudin, A. Gortsev, A. Nazarov, and R. Yakupov. Communications in computer and information science ser. Springer. 638:327–340. doi: 10.1007/978-3-319-44615-8_29.
9. Markova, E., I. Dzantiev, I. Gudkova, and S. Shorgin. 2017. Analyzing impact of path loss models on probability characteristics of wireless network with licensed shared access framework. *9th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings*. Piscataway, NJ: IEEE. 20–25. doi: 10.1109/ICUMT.2017.8255189.
10. Begishev, V., R. Kovalchukov, A. Samuylov, A. Ometov, D. Moltchanov, Y. Gaidamaka, and S. Andreev. 2015. An analytical approach to SINR estimation in adjacent rectangular cells. *Internet of things, smart spaces, and next generation networks and systems*. Eds. S. I. Balandin, S. D. Andreev, and Y. Koucheryavy. Lecture notes in computer science ser. Springer. 9247:446–458. doi: 10.1007/978-3-319-23126-6_39.
11. Samuylov, A., D. Moltchanov, Y. Gaidamaka, S. Andreev, and Y. Koucheryavy. 2016. Random triangle: A baseline model for interference analysis in heterogeneous networks. *IEEE T. Veh. Technol.* 65(8):6778–6782. doi: 10.1109/TVT.2015.2481795.
12. Naumov, V., and K. Samouylov. 2017. Analysis of multi-resource loss system with state-dependent arrival and service rates. *Probab. Eng. Inform. Sc.* 31(4):413–419. doi: 10.1017/S0269964817000079.
13. Markova, E., I. Gudkova, A. Ometov, I. Dzantiev, S. Andreev, Ye. Koucheryavy, and K. Samouylov. 2017. Flexible spectrum management in a smart city within licensed shared access framework. *IEEE Access* 5:22252–22261. doi: 10.1109/ACCESS.2017.2758840.

Received January 15, 2019

Contributors

Markova Ekaterina V. (b. 1987) — Candidate of Science (PhD) in physics and mathematics; associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; markova_ev@rudn.university

Golskaia Anastasia A. (b. 1994) — PhD student, Department of Applied Informatics and Probability Theory, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; golskaya_aa@rudn.university

Dzantiev Iliya L. (b. 1991) — PhD student, Department of Applied Informatics and Probability Theory, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; dzantiev_il@rudn.university

Gudkova Irina A. (b. 1985) — Candidate of Science (PhD) in physics and mathematics; associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gudkova_ia@rudn.university

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor; principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Аникеев Дмитрий Александрович (р. 1995) — студент Московского физико-технического института

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гольская Анастасия Андреевна (р. 1994) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Горбунова Анастасия Владимировна (р. 1986) — кандидат физико-математических наук, старший преподаватель кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Гудкова Ирина Андреевна (р. 1985) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Рос-

сийского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дзантиев Илья Леонидович (р. 1991) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Дулин Сергей Константинович (р. 1950) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Научно-исследовательского и проектно-конструкторского института информатизации, автоматизации и связи на железнодорожном транспорте (ОАО «НИИАС»)

Дулина Наталья Георгиевна (р. 1947) — кандидат технических наук, ведущий программист Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дюкова Елена Всеволодовна (р. 1945) — доктор физико-математических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кожунова Ольга Сергеевна (р. 1982) — кандидат технических наук, заведующая сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры

математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Кузьмин Виктор Юрьевич (р. 1986) — руководитель Департамента разработки ООО «Вай2Гео»

Логачев Олег Алексеевич (р. 1950) — кандидат физико-математических наук, заведующий отделом Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Малашенко Юрий Евгеньевич (р. 1946) — доктор физико-математических наук, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Маркова Екатерина Викторовна (р. 1987) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Масляков Глеб Олегович (р. 1996) — студент факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Назарова Ирина Александровна (р. 1966) — кандидат физико-математических наук, научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Наумов Валерий Арсентьевич (р. 1950) — кандидат физико-математических наук, научный руководитель Исследовательского института инноваций (Финляндия, Хельсинки)

Новикова Наталья Михайловна (р. 1953) — доктор физико-математических наук, профессор, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Палионная Софья Игоревна (р. 1995) — студентка факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Пенкин Григорий Олегович (р. 1993) — аспирант Московского физико-технического института

Прокофьев Петр Александрович (р. 1982) — кандидат физико-математических наук, научный сотрудник Института машиноведения им. А. А. Благонравова Российской академии наук

Рыбаков Константин Александрович (р. 1979) — кандидат физико-математических наук, доцент Московского авиационного института (национального исследовательского университета)

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Стефанович Алексей Игоревич (р. 1983) — главный специалист Института проблем информатики 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

Сукаев Альберт Атласович (р. 1994) — студент Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова

Ушаков Владимир Георгиевич (р. 1952) — 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

Федоров Сергей Николаевич (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова

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 исследова-

тельского центра «Информатика и управление» Российской академии наук

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 15 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 resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (Timoshenko S. P., Young D. H., Weaver W. *Vibration problems in engineering*. — 4th ed. — 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:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 15 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 resursosberegayushchie 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 slovary 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 Prof. Rustem Seyful-Mulyukov)
<http://www.ipiran.ru/english/journal.asp>