

Информатика и её применения

Том 12 Выпуск 3 Год 2018

СОДЕРЖАНИЕ

Управление случайным блужданием с эталонным стационарным распределением М. Г. Коновалов, Р. В. Разумчик	2
Среднеквадратичный риск пороговой обработки при случайном объеме выборки О. В. Шестаков	14
Байесовские модели баланса А. А. Кудрявцев	18
Зашумление данных конечными смесями нормальных и гамма-распределений с применением к задаче округления наблюдений А. К. Горшенин	28
Анализ разрезных повреждений в многополюсных сетях Ю. Е. Малашенко, И. А. Назарова, Н. М. Новикова	35
Об инвариантности стационарного распределения системы массового обслуживания с ограниченными ресурсами и с интенсивностями поступления и обслуживания, зависящими от состояния системы Э. С. Сопин, В. А. Наумов, К. Е. Самуйлов	42
Ресурсные системы массового обслуживания как модели беспроводных систем связи А. В. Горбунова, В. А. Наумов, Ю. В. Гайдамака, К. Е. Самуйлов	48
Обучаемая классификация данных с учетом анализа главных компонент М. П. Кривенко	56
Параметризация в прикладных задачах поиска эмпирических причин А. А. Грушо, Н. А. Грушо, М. И. Забежайло, Д. В. Смирнов, Е. Е. Тимонина	62
Методы и средства обнаружения нештатных ситуаций, возникающих на элементах жилищно-коммунальной инфраструктуры И. А. Шанин, С. А. Ступников, В. Н. Захаров	67
Имплицитные знания: основания и технологии извлечения И. М. Зацман	74
Statistical analysis of language specificity of connectives based on parallel texts О. Ю. Inkova and M. G. KruzHKov	83
Семантическая обработка неструктурированных текстовых данных на основе лингвистического процессора PullEnti Е. Б. Козеренко, К. И. Кузнецов, Д. А. Романов	91
Управление выходом стохастической дифференциальной системы по квадратичному критерию. I. Оптимальное решение методом динамического программирования А. В. Босов, А. И. Стефанович	99
Модель передвижения поездов и маневровых локомотивов на железнодорожной станции в приложении к оценке и анализу вероятности бокового столкновения А. В. Босов, А. Н. Игнатов, А. В. Наумов	107
Фильтрация состояний марковских скачкообразных процессов по дискретизованным наблюдениям А. В. Борисов	115
Об авторах	122
Правила подготовки рукописей	124
Requirements for manuscripts	127

Технический редактор *Л. Кокушкина* Художественный редактор *М. Седакова*

Сдано в набор 27.08.18. Подписано в печать 24.09.18. Формат 60 x 84 / 8

Бумага офсетная. Печать цифровая. Усл.-печ. л. 14,88. Уч.-изд. л. 13,5. Тираж 100 экз.

Заказ № 1044

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43

Отпечатано в НИПКЦ «Восход-А» с готовых файлов

Москва 109052, ул. Смирновская, д. 25, стр. 3

УПРАВЛЕНИЕ СЛУЧАЙНЫМ БЛУЖДЕНИЕМ С ЭТАЛОННЫМ СТАЦИОНАРНЫМ РАСПРЕДЕЛЕНИЕМ*

М. Г. Коновалов¹, Р. В. Разумчик²

Аннотация: Рассматривается случайное блуждание на отрезке, допускающее управление в форме выбора направления для очередного шага. Задано множество стратегий управления, параметризованных конечными векторами. Требуется найти из этого множества такую стратегию, при которой плотность стационарного распределения марковской цепи, описывающей блуждание, максимально приближена к заданной эталонной плотности распределения. Постановка задачи отличается от классической схемы марковского процесса принятия решений тем, что отсутствует одношаговый доход. Содержательная трактовка задачи появляется в психологии, робототехнике, генетике. Предложен квазиградиентный алгоритм определения оптимальных значений параметров, основанный на оценках частных производных целевой функции по наблюдениям за фазовой траекторией. Приведены численные результаты работы алгоритма в примерах с различными классами стратегий и различными эталонными плотностями распределения.

Ключевые слова: управление марковской цепью с непрерывным множеством состояний; квазиградиентные алгоритмы; оценки производных по наблюдениям

DOI: 10.14357/19922264180301

1 Введение

В ряде прикладных областей возникает следующая модель управляемого случайного блуждания. Пусть $n = 0, 1, 2, \dots$ и пусть $x_n \in [0, 1]$ — положение блуждающего объекта в момент n . В этом положении выбирается направление движения, т. е. один из двух доступных сегментов $[0, x_n]$ или $[x_n, 1]$ соответственно с вероятностями $s(x_n)$ и $1 - s(x_n)$. Затем, если был выбран отрезок $[0, x_n]$, объект переходит в точку $f(x_n) \in [0, x_n]$, иначе — в точку $g(x_n) \in [x_n, 1]$. Функции $f(x_n)$ и $g(x_n)$ могут быть как детерминированными, так и рандомизированными, и их вид обусловлен спецификой задачи. Очевидно, процесс x_n является управляемой марковской цепью с множеством состояний $[0, 1]$.

По-видимому, впервые подобные цепи Маркова начали появляться в связи с математическим моделированием процессов, обладающих или наделенных свойствами обучаемости. В этой связи надо отметить основополагающую работу [1], где для модели предлагается интерпретация из области психологии: если предыдущий результат действия испытуемого есть x_n , то $f(x_n)$ и $g(x_n)$ — результат следующего действия, выбираемого из двух альтернатив с вероятностями $s(x_n)$ и $1 - s(x_n)$.

В другой интерпретации [2] значения x_n могут рассматриваться как текущий уровень интеллекта испытуемого, значения $g(x_n)$ — как уровень интеллекта испытуемого после неправильного (правильного) ответа на очередной вопрос, а $s(x_n)$ — вероятность неправильного ответа при условии, что текущий уровень интеллекта равен x_n . Задачей является определение стационарного уровня интеллекта.

В работах [3, 4] описан пример из робототехники, в котором изменение каждой из координат робота при механическом перемещении по двумерной области моделируется в точности по схеме, описанной в первом абзаце. Другие примеры применения рассматриваемой модели случайного блуждания, в том числе в области генетики, можно найти в [5–7].

Основной вопрос, который возникает в работах, использующих марковскую цепь x_n , — это условия существования стационарного распределения и нахождение для него замкнутых формул или численного алгоритма расчета [3, 4, 8–10]. Например, в [9] показано, что если $s(x_n) = 1/2$, $f(x_n) = 1/x_n$, $g(x_n) = 1/(1 - x_n)$, то цепь Маркова $\{x_n\}$ ³ имеет единственное стационарное абсолютно непрерывное распределение с арксинус-плотностью. Если

* Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научного проекта № 18-07-00692.

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

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, rrazumchik@ipiran.ru

³ Эта цепь известна в зарубежной литературе как цепь Диакониса–Фридмана [9].

же $s(x_n) = 1/2$ заменить на $s(x_n) = s$, $0 < s < 1$, то стационарное распределение также единственно и имеет бета-плотность. Как показано в [4], бета-плотность является предельной также и в более общем случае, когда длина «шага влево» $f(x_n)$ и длина «шага вправо» $g(x_n)$ имеют бета-распределения с разными параметрами¹, а $s(x_n)$ — линейная функция. Для произвольной функции $s(x_n)$ предложен метод численного расчета стационарного распределения, а также доказываются достаточные условия для эргодичности цепи. Отдельно стоит отметить пример из [11], который показывает, что как только вероятность выбора направления начинает зависеть от текущего состояния цепи, ситуация заметно усложняется: если $f(x_n) = x_n/3$ и $g(x_n) = (x_n + 2)/3$, то существует стратегия $s(x_n)$, $0 < s(x_n) < 1$, при которой стационарное распределение не единственно. Известно несколько простых достаточных условий эргодичности цепи, которые с теоретической точки зрения, быть может, и являются ограничительными, но с практической точки зрения предлагают удобное средство для исключения патологических ситуаций [3, 4]. К примеру, если длины шагов влево и вправо распределены равномерно, то достаточным условием является одновременное выполнение двух неравенств: $s(0) < 1$ и $s(1) > 0$.

Даже беглый обзор результатов свидетельствует о том, что задача анализа, т. е. задача нахождения стационарного распределения цепи при фиксированных $s(x_n)$, $f(x_n)$ и $g(x_n)$, изучена хорошо. Примечательно, что стационарное распределение цепи редко удается выписать в явном виде. В связи с этим возникает вопрос: как решать задачу синтеза, т. е. задачу нахождения таких $s(x_n)$, $f(x_n)$ и $g(x_n)$, которые приводят к заданному стационарному распределению цепи?

Прикладным мотивом к рассмотрению подобной «обратной задачи» может являться упомянутый пример из [3]. Пусть задачей робота является регулярное посещение «каждого участка» некоторой области, причем определенные заранее заданные части этой области должны посещаться чаще, чем другие. Какая стратегия обеспечивает решение поставленной задачи? В [3] предложено решение этой задачи в случае, когда известен явный вид стационарного распределения вектора и когда целевая плотность распределения вектора унимодальная.

Эта статья посвящена задаче синтеза управления случайным блужданием x_n . В разд. 2 формулируется постановка задачи и конкретизируются случайные функции $f(x_n)$ и $g(x_n)$. Цель управления — из-

данного множества стратегий, параметризованных конечномерными наборами числовых параметров, найти ту, при которой стационарное распределение цепи максимально приближено к заданному эталонному виду. В разд. 3 и 4 конструируется алгоритм решения, использующий идеи стохастической градиентной оптимизации на марковских цепях. В разд. 5 представлены результаты численных экспериментов, которые позволяют оценить эффективность предложенного алгоритма.

В заключение этой вводной части сделаем еще одно замечание.

На сформулированную задачу приближения эталонного стационарного распределения на отрезке $[0, 1]$ можно посмотреть и выйдя за рамки случайного блуждания x_n . Плодотворным в этом случае является подход, основанный на системах «итерационных случайных функций» [9, 12, 13]. При таком подходе приближение осуществляется с помощью специально подобранного набора случайных функций, каждая из которых, выбираемая с некоторым вероятностным распределением, отображает отрезок $[0, 1]$ в себя. Однако для цепи x_n такой метод не позволяет получить удовлетворительное решение из-за необходимости оперировать «небольшой» системой случайных функций (лишь $f(x_n)$ и $g(x_n)$).

2 Постановка задачи

Пусть управляемое случайное блуждание задается рекуррентным соотношением

$$x_{n+1} = x_n + \frac{\xi_n}{\theta} [-(1 - \sigma_n)x_n + \sigma_n(1 - x_n)],$$

$$n = 0, 1, 2, \dots$$

В этой формуле σ_n — взаимно условно-независимые бинарные случайные величины, принимающие значения 0 с вероятностью $s(x_n)$, и 1 с вероятностью $1 - s(x_n)$; $s(x)$ — функция на отрезке $[0, 1]$, $0 < s(x) < 1$; ξ_n — взаимно независимые случайные величины, равномерно распределенные на отрезке $[0, 1]$; $0 < \theta \leq 1$; $x_0 \in [0, 1]$.

Значения величины σ_n трактуются как выбор направления движения на каждом шаге: 0 — сдвиг влево; 1 — сдвиг вправо. Подлежащую выбору функцию $s(x)$ будем называть правилом управления. Размер сдвига на каждом шаге определяется значением случайных величин ξ_n и числовым параметром θ . Начальное значение процесса не существенно, но известно.

¹Но один из параметров бета-распределения должен равняться единице.

Последовательность правил $\{s_n(x), n \geq 0\}$ будем называть стратегией управления, если правило $s_n(x)$ определяет выбор направления на шаге n . Если все правила одинаковы, т. е. если $s_n(x) = s(x)$ для всех n , то стратегию будем называть однородной. Таким образом, имеется взаимно однозначное соответствие между правилами и однородными стратегиями.

При фиксированной однородной стратегии $s(x)$ и фиксированном параметре θ последовательность x_n является марковской цепью с множеством состояний $[0, 1]$ и переходной вероятностью с плотностью

$$p(x, y) = sq_0(x, y) + (1 - s)q_1(x, y), \quad 0 < x < 1, \quad (1)$$

где

$$q_0(x, y) = \begin{cases} \frac{1}{x\theta}, & y \in [x - x\theta, x]; \\ 0, & y \notin [x - x\theta, x]; \end{cases}$$

$$q_1(x, y) = \begin{cases} \frac{1}{(1-x)\theta}, & y \in [x, x + (1-x)\theta]; \\ 0, & y \notin [x, x + (1-x)\theta]. \end{cases}$$

В граничных точках $x = 0$ и $x = 1$ переходная плотность доопределяется по непрерывности. Относительно этой цепи предположим, что она при любых $s(x)$ и θ имеет абсолютно непрерывное стационарное распределение с плотностью $\pi(x) = \pi(x, s(x), \theta), x \in [0, 1]$.

Пусть $S = \{s(x, a), a \in A \subset \mathbb{R}^k\}$ — некоторое заданное множество однородных стратегий (или, что эквивалентно, множество правил управления), параметризованных векторами $a = (a^{(1)}, \dots, a^{(k)})$ из множества $A = A(S)$. Стационарное распределение $\pi(x)$, соответствующее стратегии $s(x, a)$, зависит от набора параметров a . Пусть также $\rho(x)$ — заданная плотность вероятностного распределения на отрезке $[0, 1]$, которую будем называть эталонной плотностью.

Цель управления заключается в том, чтобы найти такой набор параметров $a \in A(S)$, который минимизирует функцию

$$W = W(A) = \int_0^1 (p(x) - \rho(x))^2 dx. \quad (2)$$

Таким образом, требуется отыскать стратегию из заданного параметризованного множества S , при которой стационарное распределение цепи x_n наиболее приближено в смысле критерия (2) к заданному распределению ρ .

3 Производная целевой функции по параметру

Плотность переходной вероятности марковской цепи за n шагов определяется как

$$p^{(n)}(x, y) = \int_0^1 p^{(n-1)}(x, z)p(z, y) dz, \quad n = 1, 2, \dots;$$

$$p^{(0)}(x, z) \equiv p_0(z).$$

Стационарная плотность распределения $\pi(x)$ удовлетворяет следующим условиям:

$$\pi(y) = \int_0^1 \pi(x)p(x, y) dx; \quad (3)$$

$$\int_0^1 \pi(x) dx = 1. \quad (4)$$

Выберем произвольный параметр из набора a и будем обозначать дифференцирование по этому параметру штрихом. Продифференцируем функцию (2) и равенство (3) в предположении, что производные p' и π' существуют:

$$W' = 2 \int_0^1 \pi'(x) (\pi(x) - \rho(x)) dx;$$

$$\pi'(y) = \int_0^1 \pi'(x)p(x, y) dx + \int_0^1 \pi(x)p'(x, y) dx. \quad (5)$$

Воспользуемся номенклатурой теории обобщенных функций [14]. Согласно этой теории существует взаимно однозначное соответствие между локально суммируемыми функциями $r(x)$ на отрезке $[0, 1]$ и регулярными обобщенными функциями r (линейными непрерывными функционалами на пространстве функций, непрерывных на отрезке $[0, 1]$). Интегралу $\int \varphi(x)r(x) dx$ соответствует функциональное обозначение $\langle \varphi, r \rangle$ для результата действия функционала r на функцию φ .

Обозначая обобщенные функции переходной плотности $p(x, y)$, а также ее производной $p'(x, y)$ соответственно через p_y и p'_y и используя обозначение δ_y для обобщенной плотности, сосредоточенной в точке x , равенство (5) перепишем в виде:

$$\langle \pi', \delta_y - p_y \rangle = \langle \pi, p'_y \rangle. \quad (6)$$

Здесь использовано характеристическое свойство δ -функции, согласно которому $\langle \pi', \delta_y \rangle = \pi'(y)$.

Для фиксированного $N > 0$ определим функционал

$$P_x(N) = \delta_x + p_x + p_x^{(2)} + \dots + p_x^{(N)} = \sum_{n=0}^N p_x^{(n)},$$

где $p_x^{(n)}$ — обобщенные функции, соответствующие переходным плотностям $p^{(n)}(y, x)$, $n > 0$, $p^{(0)}(x, y) = \delta(x, y)$. (Заметим, что из приведенных определений следует, что $p_x^{(n+1)} = p_x^{(n)} p_x = p_x p_x^{(n)}$.) Применим этот функционал к обеим частям равенства (6), рассматривая их как функции y :

$$\begin{aligned} \langle \langle \pi', \delta_y - p_y \rangle, P_x(N) \rangle &= \pi'(x) - \\ - \langle \pi', p_x \rangle + \langle \pi', p_x \rangle - \langle \pi', p_x^{(2)} \rangle &+ \langle \pi', p_x^{(2)} \rangle - \dots \\ \dots - \langle \pi', p_x^{(N+1)} \rangle &= \langle \langle \pi, p'_y \rangle, P_x(N) \rangle. \end{aligned}$$

Отсюда получаем, что

$$\pi'(x) = \langle \langle \pi, p'_y \rangle, P_x(N) \rangle + \langle \pi', p_x^{(N+1)} \rangle.$$

Перейдем к пределу при $N \rightarrow \infty$. Второе слагаемое в правой части стремится к нулю:

$$\begin{aligned} \lim_{N \rightarrow \infty} \langle \pi', p_x^{(N)} \rangle &= \lim_{N \rightarrow \infty} \int_0^1 \pi'(y) p^{(N)}(y, x) dy = \\ &= \int_0^1 \pi'(y) \pi(x) dy = \pi(x) \int_0^1 \pi'(y) dy = 0 \end{aligned}$$

(последнее равенство следует из дифференцирования равенства (4)), поэтому

$$\pi'(x) = \langle \langle \pi, p'_y \rangle, P_x \rangle,$$

где

$$P_x = \sum_{m=0}^{\infty} p_x^{(m)}. \quad (7)$$

После подстановки полученного выражения для $\pi'(x)$ в формулу (4) получим:

$$W' = 2 \langle \langle \langle \pi, p'_y \rangle, P_x \rangle, \gamma \rangle, \quad (8)$$

где через γ обозначена обобщенная функция, соответствующая разности функций $\pi(x) - \rho(x)$.

Дифференцирование плотности (1) по выбранному параметру приводит к выражению:

$$p'(x, y) = s'(x) (q_0(x, y) - q_1(x, y)),$$

которое подставим в формулу (8). Правая часть (8) представляет собой ряд из-за наличия функционала (7). Рассмотрим вначале произвольное слагаемое

с $m > 0$. Поскольку входящие в него обобщенные функции регулярны, то можно эквивалентным образом перейти к обычным функциям и записать такое слагаемое в виде: $\int_0^1 \int_0^1 \int_0^1 \pi(x) s'(x) (q_0(x, y) - q_1(x, y)) p^{(m)}(y, z) \gamma(z) dz dy dx$.

Эта запись интерпретируется следующим образом. Положим

$$G_m^{(i)} = \lim_{t \rightarrow \infty} M \left(s'(x_t) M_{x_t}^{(i)} (\gamma_{x_{t+m}}) \right), \quad (9)$$

где M — безусловное математическое ожидание; $M_{x_t}^{(i)}$ — условное математическое ожидание при условии, что в состоянии x_t был совершен шаг влево ($i = 0$) или вправо ($i = 1$); $\gamma = \pi(\cdot) - \rho(\cdot)$. Тогда рассматриваемое слагаемое записывается как $G_{m+1}^{(0)} - G_{m+1}^{(1)}$.

Для слагаемого с $m = 0$, где фигурирует сингулярная обобщенная функция δ_x , непосредственная интегральная запись неправомерна, однако интерпретация этого слагаемого совершенно аналогична. Поэтому окончательно

$$W' = \sum_{m=0}^{\infty} (G_m^{(0)} - G_m^{(1)}).$$

4 Алгоритм оптимизации параметров стратегии

Пусть по-прежнему S — заданное множество стратегий, параметризованных k -мерными векторами из некоторого множества $A = A(S) \subset \mathbb{R}^k$, и пусть $a_0 \in A$ — некоторый начальный набор параметров. Будем корректировать значения параметров для приближенной минимизации целевой функции (2), используя стохастический вариант алгоритма проекции градиента. Обозначим через a_n вектор параметров, который применяется для выбора направления движения на n -м шаге случайного блуждания x_n . Алгоритм коррекции имеет вид:

$$a_{n+1} = \prod_A (a_n - \alpha_n w_n), \quad (10)$$

где \prod_A — оператор проектирования на множество A ; α_n — подходящим образом подобранная числовая последовательность; w_n — последовательность случайных величин, являющихся оценками градиента $\nabla W(a_n)$. Управление случайным блужданием осуществляется таким образом, что на n -м шаге выбор направления «сдвига» происходит с помощью правила $s_n(x) = s(x, a_n) \in S$. (Заметим, что стратегия $\mathbf{s} = \{s_n(x), n \geq 0\}$ является неоднородной и множеству S не принадлежит.) Алгоритмы

типа (10) являются широко распространенным инструментом оптимизации и предметом изучения в огромном числе публикаций.

Реализация схемы (10) предполагает построение оценок градиента целевой функции. Согласно (9) каждая частная производная функции W представляет собой ряд из слагаемых сложной структуры в виде предела повторного математического ожидания. При этом внешнее усреднение происходит по предельному распределению цепи, которое соответствует текущему значению набора параметров, задающих стратегию. Основная проблема заключается в том, чтобы совместить пошаговое изменение параметров, соответствующее изменению индекса n , с необходимостью «зафиксировать» значения параметров для оценки предельного математического ожидания. Для решения используется прием «оценивания с забыванием», который поясним на примере оценки величины (9).

Пусть для определенности $i = 0$ и пусть $\tau_1, \tau_2, \dots, \tau_l, \dots$ — последовательные моменты выбора действия 0 («сдвиг влево») при случайном блуждании, управляемом согласно стратегии s . Выберем числовую последовательность $\beta_l \uparrow 1$, положим $z_0 = 0, b_0 = 1$ и зададим рекуррентные соотношения:

$$z_{l+1} = \beta_l z_l + s'_{\tau_l}(x_{\tau_l}) \gamma_{\tau_l+m}, \quad b_{l+1} = \beta_l b_l + 1.$$

Оценкой величины $G_m^{(0)}$ на n -м шаге случайного блуждания является отношение $g_{m,n}^{(0)} = z_l/b_l$. Аналогично строятся оценки $g_{m,n}^{(1)} = z_l/b_l$ величин $G_m^{(1)}$.

«Скользящие» суммы были использованы в [15] для построения оценок градиента предельного среднего дохода в задаче управления дискретной марковской цепью с доходами с помощью алгоритма типа (10). Там же было доказано, что для сходимости алгоритма необходимо выполнение дополнительных условий, в частности на последовательности α_n и β_l . В данной работе исследование алгоритма ограничивается численными экспериментами.

5 Экспериментальный анализ

В этом разделе приведены результаты экспериментов с управляемым случайным блужданием x_n с коэффициентом θ , задающим максимальный размер сдвига на одном шаге, равным 0,5. Была выбрана упрощенная модификация управляющего алгоритма (10), для которой последовательности α_n и β_l суть константы: $\alpha_n \equiv \alpha, \beta_l \equiv \beta$. В качестве оценки частных производных целевой функции на n -м шаге было взято выражение $\sum_{m=0}^M (g_{m,n}^{(0)} - g_{m,n}^{(1)})$.

Константы α, β и M варьировались в диапазонах $[10^{-8}; 10^{-7}]$, $[0,999; 0,9999]$ и $\{3, 4, \dots, 10\}$ соответственно.

Определим четыре множества правил управления, образованных многочленами степеней от 0 до 3:

$$S_k = \left\{ s(x, a) = \max \left(0, \min \left(1, \sum_{i=0}^k a^{(i)} x^i \right) \right), \right. \\ \left. a^{(i)} \in \mathbb{R} \right\}, \quad k = 0, 1, 2, 3.$$

Стратегии из класса S_k параметризуются набором коэффициентов $a = (a^{(0)}, \dots, a^{(k)}) \in A_k = \mathbb{R}^{k+1}$.

Кроме того, определим множество правил управления синусоидального типа:

$$S_t = \\ = \{s(x, a) = \max(0, \min(1, a_0 \sin(a_1 + a_2 x) + a_3)); \\ a_0, a_1, a_2, a_3 \in \mathbb{R}\},$$

которые параметризуются набором коэффициентов $(a^{(0)}, \dots, a^{(3)}) \in A_t = \mathbb{R}^4$.

Для выбранного класса правил управления S и установленной эталонной плотности $c(x)$ эксперимент заключался в имитации траектории случайного блуждания x_n , управляемого согласно стратегии $s = \{s_n(x) = s(x, a_n) \in S, n \geq 0\}$. При этом последовательность a_n порождалась алгоритмом (10), в котором $A = A(S)$, а последовательности σ_n и w_n определены в начале текущего раздела. Продолжительность эксперимента составляла $N \sim 10^7$ тактов. Результатами эксперимента стали финальное правило управления $s_N(x)$, а также оценки значений целевой функции W_N и стационарной плотности $p_N(x)$, соответствующих финальному правилу $s_N(x)$.

Пример 1. Эталонная плотность $c(x)$ — линейная. Она приведена в табл. 1. В этой же таблице указано начальное правило управления $s_0(x)$, а также полученные экспериментально значение целевой функции W_0 и график плотности стационарного распределения $p_0(x)$, соответствующие правилу управления $s_0(x)$. Результаты эксперимента приведены в табл. 2. Они показывают, что алгоритм существенно улучшает начальное правило как по значениям целевой функции, так и по форме графика стационарной плотности. Заметим, впрочем, что приближение графика стационарной плотности к эталонному виду не ставилось задачей управления. Это побочный эффект оптимизации по метрике (2). Дополнительно следует отметить,

Таблица 1 Исходные данные для эксперимента с линейной эталонной плотностью

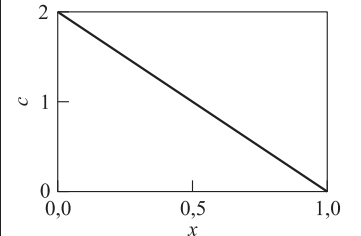
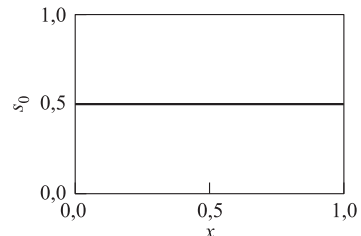
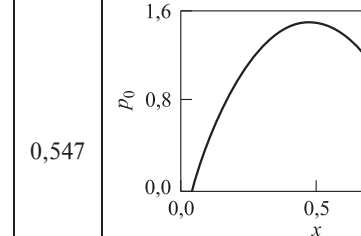
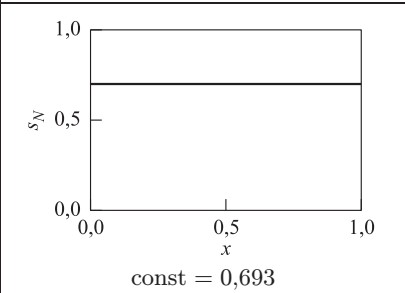
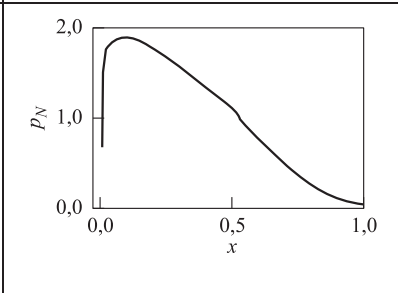
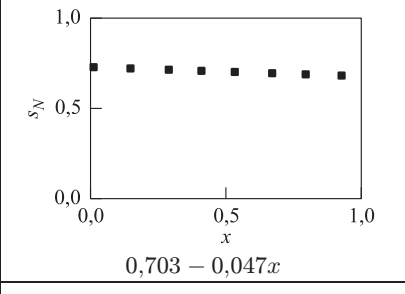
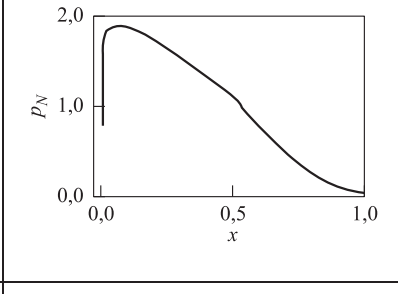
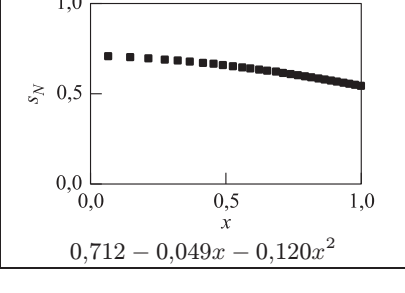
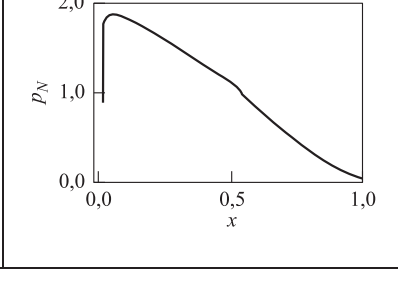
$c(x) = 2(1 - x)$	$s_0(x) \equiv 0,5$	W_0	$p_0(x)$
		0,547	

Таблица 2 Результаты эксперимента с линейной эталонной плотностью

S	$s_N(x)$	W_N	$p_N(x)$
S_0	 const = 0,693	0,033	
S_1	 $0,703 - 0,047x$	0,024	
S_2	 $0,712 - 0,049x - 0,120x^2$	0,013	

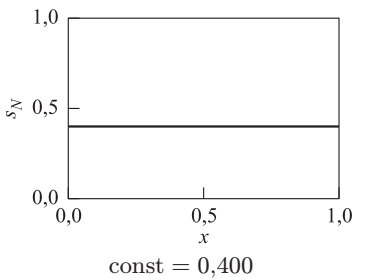
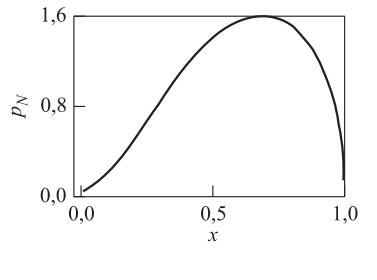
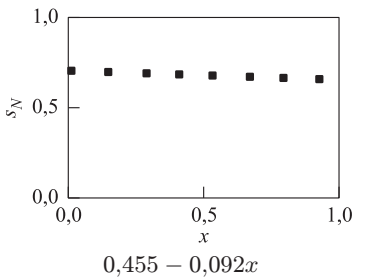
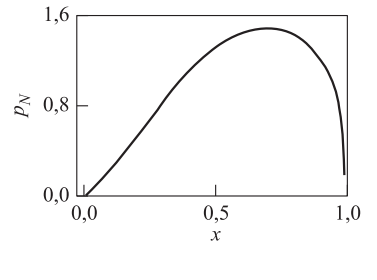
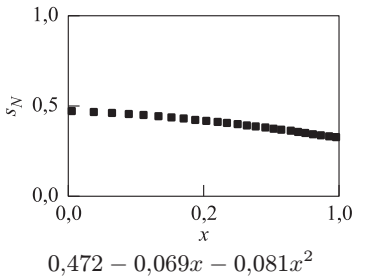
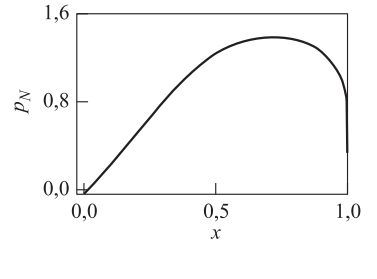
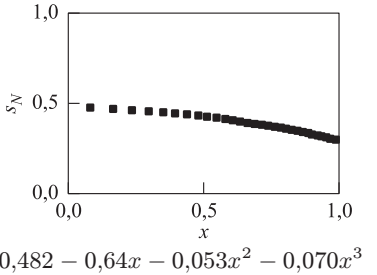
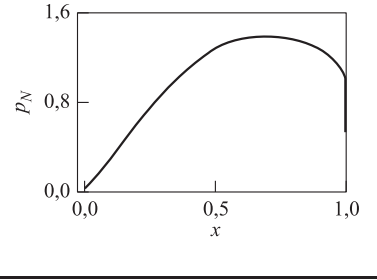
что качество оптимизации повышается с ростом степени многочленов — «базисных» правил.

Пример 2. Эталонная плотность — квадратичная: $c(x) = 2,727(1,4x - x^2)$. Начальное правило такое же, как в примере 1. Ему соответствует значение целевой функции $W_0 = 0,163$. Результаты эксперимента приведены в табл. 3. Качественные выводы

относительно результатов такие же, как в примере 1. Плотность $c(x)$ для наглядного сравнения с результирующей плотностью изображена на рис. 1.

Пример 3. Эталонная плотность — кубическая: $c(x) = 1,622(0,8 + 0,4x - 2,5x^2 + 1,8x^3)$ (рис. 2). Начальное правило такое же, как в предыдущих примерах. Ему соответствует значение целевой

Таблица 3 Результаты эксперимента с квадратичной эталонной плотностью

S	$s_N(x)$	W_N	$p_N(x)$
S_0	 <p>const = 0,400</p>	0,042	
S_1	 <p>$0,455 - 0,092x$</p>	0,023	
S_2	 <p>$0,472 - 0,069x - 0,081x^2$</p>	0,010	
S_3	 <p>$0,482 - 0,64x - 0,053x^2 - 0,070x^3$</p>	0,004	

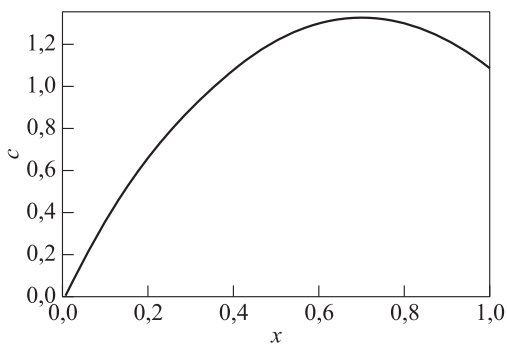


Рис. 1 Квадратичная эталонная плотность

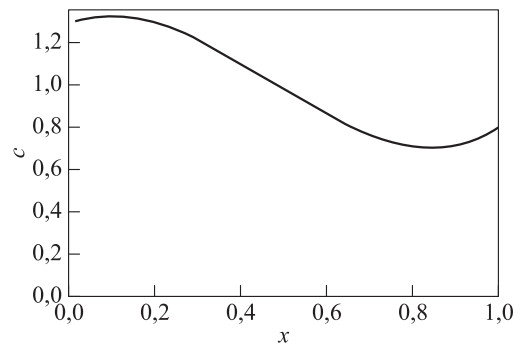
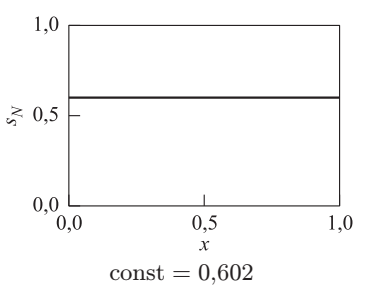
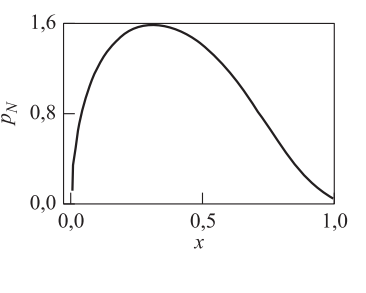
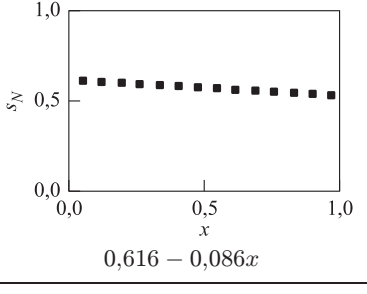
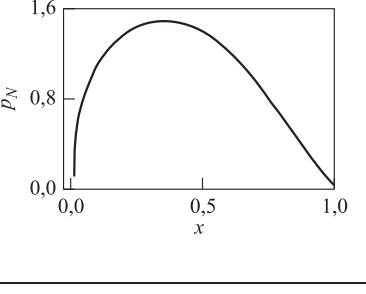
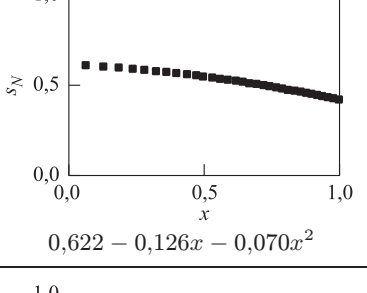
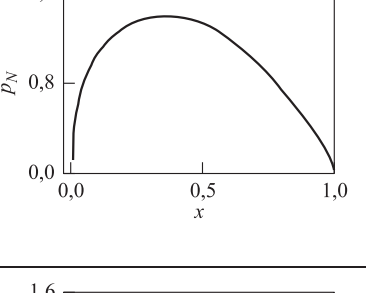
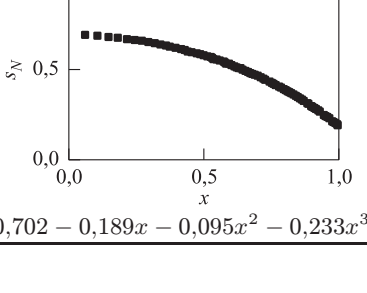
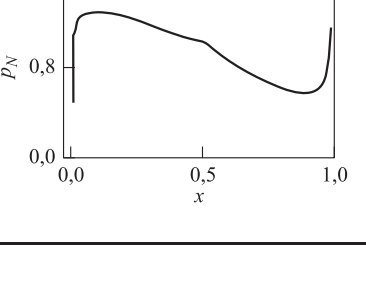


Рис. 2 Кубическая эталонная плотность

Таблица 4 Результаты эксперимента с кубической эталонной плотностью

S	$s_N(x)$	W_N	$p_N(x)$
S_0	 <p style="text-align: center;">const = 0,602</p>	0,042	
S_1	 <p style="text-align: center;">$0,616 - 0,086x$</p>	0,023	
S_2	 <p style="text-align: center;">$0,622 - 0,126x - 0,070x^2$</p>	0,010	
S_3	 <p style="text-align: center;">$0,702 - 0,189x - 0,095x^2 - 0,233x^3$</p>	0,004	

функции $W_0 = 0,284$. Результаты эксперимента приведены в табл. 4. Как и в предыдущих примерах, оптимизация на любом из множеств S_k дает существенный выигрыш в целевой функции по сравнению с начальным приближением. Вновь качество оптимизации повышается с ростом значения k , т. е. с увеличением мощности множества «базисных» правил.

В отличие от первых двух примеров, форма стационарной плотности становится похожей на эталонную плотность только при использовании

правил управления — многочленов третьей степени.

Пример 4. Эталонная плотность — синусоидальная. Исходные данные указаны в табл. 5. Результаты приведены в табл. 6. Этот пример особенно подчеркивает важность выбора множества «базисных» правил. «Тригонометрическое» множество S_t , несмотря на плохое начальное правило, позволяет добиться гораздо лучшего приближения к эталонной плотности, чем множество полиномиальных правил.

Таблица 5 Исходные данные для эксперимента с синусоидальной эталонной плотностью

$c(x) = 0,798 \sin(0,2 + 12x) + 0,977$	$s_0(x)$	W_0	$p_0(x)$
	<p>для базовых классов S_0-S_2: $s_0(x) \equiv 0,5$</p>	0,533	
	<p>для базового класса S_t: $s(x) = \min(1, \sin 10x + 1)$</p>	143,9	

6 Заключение

Сформулирована задача поиска стратегии управления случайным блужданием с целью минимизировать функцию, оценивающую отклонение плотности стационарного распределения от заданной эталонной плотности. Задача относится к теории управления марковскими цепями с не-дискретным множеством состояний, однако отличается от классической постановки марковского процесса принятия решений тем, что отсутствует одношаговый доход. Оптимальная стратегия ищется в заданном множестве стратегий, параметризованных конечномерными наборами числовых параметров.

Предложенное приближенное численное решение представляет собой градиентный алгоритм коррекции параметров стратегии, причем оценки производных целевой функции строятся по результатам наблюдения за имитируемой траекторией процесса. Проведен экспериментальный анализ алгоритма для ряда параметризованных классов стратегий и эталонных плотностей, который показал хорошие результаты в плане минимизации целевой функции. Основной вывод на основании работы заключается в констатации эффективности градиентного подхода к оптимизации на марковских цепях с непрерывным множеством состояний.

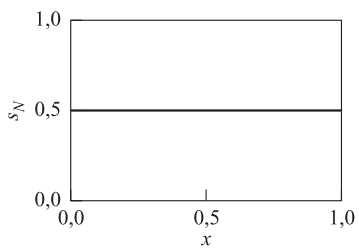
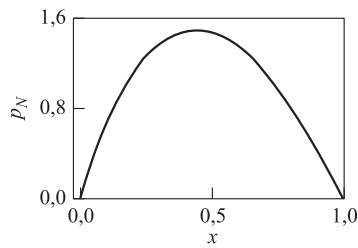
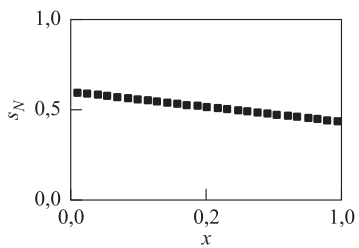
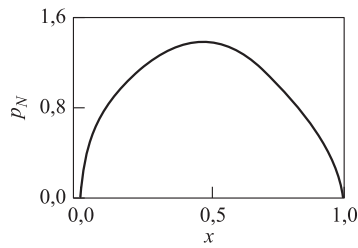
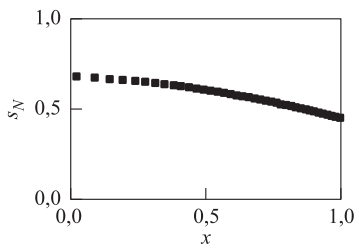
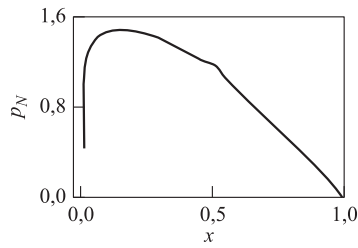
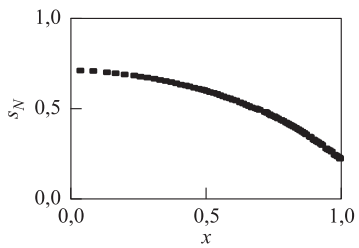
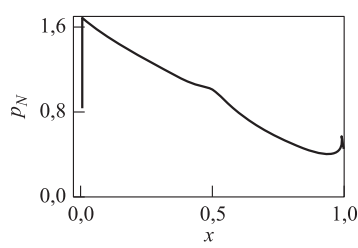
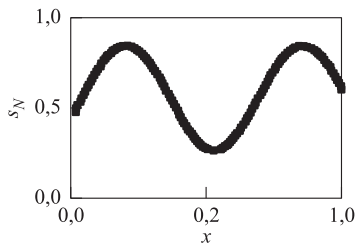
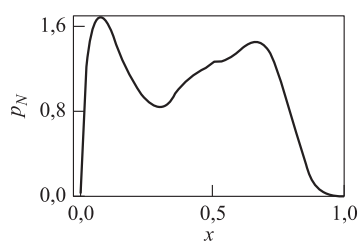
Представляются интересными следующие направления дальнейших исследований: теоретический анализ сходимости алгоритма; распространение и обоснование метода на произвольные

распределения без предположения о существовании плотностей; использование различных метрик для оценки близости распределений; изучение связи между исходным множеством заданных стратегий и точностью приближения эталонного распределения.

Литература

1. *Karlin S.* Some random walks arising in learning models. I // *Pac. J. Math.*, 1953. Vol. 3. No. 4. P. 725–756.
2. *Kajser T.* On a theorem of Karlin // *Acta Appl. Math.*, 1994. Vol. 34. P. 51–69.
3. *Ramli M. A., Leng G.* The stationary probability density of a class of bounded Markov processes // *Adv. Appl. Probab.*, 2010. Vol. 42. P. 986–993.
4. *McKinlay S., Borovkov K.* On explicit form of the stationary distributions for a class of bounded Markov chains // *J. Appl. Probab.*, 2016. Vol. 53. Iss. 1. P. 231–243.
5. *Li C.* Human genetics. — New York, NY, USA: McGraw-Hill, 1961. 218 p.
6. *DeGroot M. H., Rao M. M.* Stochastic give-and-take // *J. Math. Anal. Appl.*, 1963. Vol. 7. P. 489–498.
7. *McKinlay S.* A characterization of transient random walks on stochastic matrices with Dirichlet distributed limits // *J. Appl. Probab.*, 2014. Vol. 51. P. 542–555.
8. *Peigné M.* Iterated function systems and spectral decomposition of the associated Markov operator // *Publications mathématiques et informatique de Rennes*, 1993. No. 2. P. 1–28.

Таблица 6 Результаты эксперимента с синусоидальной эталонной плотностью

S	$s_N(x)$	W	$p_N(x)$
S_0	 <p>$\text{const} = 0,533$</p>	0,505	
S_1	 <p>$0,602 - 0,165x$</p>	0,425	
S_2	 <p>$0,686 - 0,074x - 0,162x^2$</p>	0,372	
S_3	 <p>$0,724 - 0,093x - 0,182x^2 - 0,226x^3$</p>	0,321	
S_π	 <p>$0,288 \sin(-0,355 + 9,531x) + 0,566$</p>	0,152	

9. Diaconis P., Freedman D. Iterated random functions // SIAM Rev., 1999. Vol. 41. Iss. 1. P. 45–76.
10. Ladjimi F., Peigné M. Iterated function systems with place dependent probabilities and application to the Diaconis–Friedman’s chain on $[0, 1]$. <https://hal.archives-ouvertes.fr/LMPT/hal-01567392v1>.
11. Stenflo Ö. A note on a theorem of Karlin // Stat. Probabil. Lett., 2001. Vol. 54. Iss. 2. P. 183–187.
12. Jacquin A. A fractal theory of iterated Markov operators with applications to digital image coding. — Atlanta, GA, USA: Georgia Institute of Technology, 1989. Ph.D. Thesis.
13. Forte B., Vrscay E. R. Solving the inverse problem for measures using iterated function systems: A new approach // Adv. Appl. Probab., 1995. Vol. 27. Iss. 3. P. 800–820.
14. Владимиров В. С. Обобщенные функции в математической физике. — М.: Наука, 1976. 280 с.
15. Конавалов М. Г. Методы адаптивной обработки информации и их приложения. — М.: ИПИ РАН, 2007. 212 с.

Поступила в редакцию 28.04.18

FINDING CONTROL POLICY FOR ONE DISCRETE-TIME MARKOV CHAIN ON $[0, 1]$ WITH A GIVEN INVARIANT MEASURE

M. G. Konovalov¹ and R. V. Razumchik^{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

²Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

Abstract: A discrete-time Markov chain on the interval $[0, 1]$ with two possible transitions (left or right) at each step has been considered. The probability of transition towards 0 (and towards 1) is a function of the current value of the chain. Having chosen the direction, the chain moves to the randomly chosen point from the appropriate interval. The authors assume that the transition probabilities depend on the current value of the chain only through a finite number of real-valued numbers. Under this assumption, they seek the transition probabilities, which guarantee the L_2 distance between the stationary density of the Markov chain and the given invariant measure on $[0, 1]$ is minimal. Since there is no reward function in this problem, it does not fit in the MDP (Markov decision process) framework. The authors follow the sensitivity-based approach and propose the gradient- and simulation-based method for estimating the parameters of the transition probabilities. Numerical results are presented which show the performance of the method for various transition probabilities and invariant measures on $[0, 1]$.

Keywords: Markov chain; control; continuous state space; sensitivity-based approach; derivative estimation

DOI: 10.14357/19922264180301

Acknowledgments

The reported study was partly supported by the Russian Foundation for Basic Research according to the research project No. 18-07-00692.

References

1. Karlin, S. 1953. Some random walks arising in learning models. I. *Pac. J. Math.* 3(4):725–756.
2. Kaijser, T. 1994. On a theorem of Karlin. *Acta Appl. Math.* 34:51–69.
3. Ramli, M. A., and G. Leng. 2010. The stationary probability density of a class of bounded Markov processes. *Adv. Appl. Probab.* 42:986–993.
4. McKinlay, S., and K. Borovkov. 2016. On explicit form of the stationary distributions for a class of bounded Markov chains. *J. Appl. Probab.* 53(1):231–243.
5. Li, C. 1961. *Human genetics*. New York, NY: McGraw-Hill. 218 p.
6. DeGroot, M. H., and M. M. Rao. 1963. Stochastic give-and-take. *J. Math. Anal. Appl.* 7:489–498.
7. McKinlay, S. 2014. A characterization of transient random walks on stochastic matrices with Dirichlet distributed limits. *J. Appl. Probab.* 51:542–555.
8. Peigné, M. 1993. Iterated function systems and spectral decomposition of the associated Markov operator. *Publications mathématiques et informatique de Rennes. 2:* 1–28.
9. Diaconis, P., and D. Freedman. 1999. Iterated random functions. *SIAM Rev.* 41(1):45–76.
10. Ladjimi, F., and M. Peigné. Iterated function systems with place dependent probabilities and appli-

- cation to the Diaconis–Friedman’s chain on $[0, 1]$. Available at: <https://hal.archives-ouvertes.fr/LMPT/hal-01567392v1/> (accessed April 4, 2018).
11. Stenflo, Ö. 2001. A note on a theorem of Karlin. *Stat. Probabil. Lett.* 54(2):183–187.
 12. Jacquin, A. 1989. A fractal theory of iterated Markov operators with applications to digital image coding. Atlanta, GA: Georgia Institute of Technology. Ph.D. Thesis.
 13. Forte, B., and E. R. Vrscay. 1995. Solving the inverse problem for measures using iterated function systems: A new approach. *Adv. Appl. Probab.* 27(3):800–820.
 14. Vladimirov, V. S. 1976. *Obobshchennye funktsii v matematicheskoy fizike* [Generalized functions in mathematical physics]. Moscow: Nauka. 280 p.
 15. Kononov, M. G. 2007. *Metody adaptivnoy obrabotki informatsii i ikh prilozheniya* [Methods of adaptive information processing and their applications]. Moscow: IPI RAN. 212 p.

Received April 28, 2018

Contributors

Kononov Mikhail G. (b. 1950) — 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; mkononov@ipiran.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ipiran.ru

СРЕДНЕКВАДРАТИЧНЫЙ РИСК ПОРОГОВОЙ ОБРАБОТКИ ПРИ СЛУЧАЙНОМ ОБЪЕМЕ ВЫБОРКИ*

О. В. Шестаков¹

Аннотация: Нелинейные методы удаления шума из сигналов, основанные на пороговой обработке вейвлет-коэффициентов, широко используются в различных прикладных областях. Свою популярность эти методы приобрели за счет скорости алгоритмов построения оценок и возможности лучшей, чем линейные методы, адаптации к функциям, принадлежащим различным классам регулярности. При использовании методов пороговой обработки обычно предполагается, что число вейвлет-коэффициентов фиксировано, а распределение шума является гауссовым. Эта модель хорошо изучена в литературе, и для разных классов сигналов вычислены оптимальные значения порогов. Однако в некоторых ситуациях объем выборки заранее не известен и моделируется случайной величиной. В данной работе рассматривается модель со случайным числом наблюдений, содержащих гауссов шум и оценивается порядок среднеквадратичного риска при растущем объеме выборки.

Ключевые слова: пороговая обработка; случайный объем выборки; среднеквадратичный риск

DOI: 10.14357/19922264180302

1 Введение

Во многих областях, таких как физика плазмы, медицина, геофизика, астрономия, радиолокация, системы связи и передачи информации, компьютерная графика и т. д., возникает потребность в анализе и обработке сигналов самых разных видов и происхождения. В некоторых случаях объем данных, доступных для анализа (объем выборки), заранее не известен. Такие ситуации могут возникать, например, в случае пропуска данных (по техническим причинам), ограниченности времени сбора данных при случайных временах регистрации наблюдений или недостатке информации о характеристиках используемого оборудования (например, используемый прибор может принадлежать некоторой партии, внутри которой определенные технические характеристики могут быть жестко не фиксированы — это может приводить к тому, что «разрешение» сигнала заранее не известно). В таком случае предполагается, что объем выборки данных представляет собой случайную величину с некоторым распределением вероятностей.

На одном из первых этапов обработки данных обычно осуществляется преобразование, приводящее к их «экономному» (разреженному) представлению, т. е. фактически разложение функции, описывающей наблюдения (сигнал) по некоторому

базису, зависящему от конкретного класса рассматриваемых сигналов. Примерами подобных базисов могут служить различные классы вейвлетов. В данной работе рассматриваются именно такие вейвлет-базисы. Затем осуществляется пороговая обработка коэффициентов; в результате часть коэффициентов разложения, которая считается незначительной, обнуляется. Такие процедуры, во-первых, позволяют подавить большую часть шума, возникающего из-за дискретизации исходной информации, несовершенства оборудования, случайных помех, наличия фонового излучения и других причин. Во-вторых, таким образом осуществляется сжатие данных с незначительной потерей информации, что позволяет более экономно хранить информацию и быстрее передавать ее по каналам цифровой связи.

В моделях с фиксированным объемом выборки статистические свойства данных процедур хорошо изучены и получены выражения для «оптимальных» порогов, ориентированных на различные классы функций сигналов и различные распределения шума (см., например, [1–6]).

В данной работе рассматривается модель со случайным числом вейвлет-коэффициентов функции сигнала, «загрязненных» белым гауссовым шумом, и оценивается порядок среднеквадратичного риска пороговой обработки.

* Работа выполнена при финансовой поддержке Российского научного фонда (проект 18-11-00155).

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

2 Пороговая обработка вейвлет-коэффициентов

Пусть функция наблюдаемого сигнала принадлежит классу функций, регулярных по Липшицу с показателем $\gamma > 0$. После вейвлет-преобразования сигнала получается набор эмпирических вейвлет-коэффициентов, имеющих следующий вид:

$$Y_{j,k} = \mu_{j,k} + z_{j,k}, \quad j = 0, \dots, J-1, \quad k = 0, \dots, 2^j - 1, \quad (1)$$

где $\mu_{j,k}$ — вейвлет-коэффициенты «чистого» сигнала, а $z_{j,k}$ — «шумовые» коэффициенты, относительно которых предполагается, что они независимы и имеют нормальное распределение с нулевым средним и дисперсией σ^2 . Если базисные вейвлет-функции удовлетворяют некоторым дополнительным условиям гладкости [7], то, поскольку функция сигнала регулярна по Липшицу, существует такая положительная константа A , что

$$|\mu_{j,k}| \leq \frac{A \cdot 2^{J/2}}{2^{j(\gamma+1/2)}}. \quad (2)$$

Через $\text{Lip}(A, \gamma)$ обозначим класс регулярных по Липшицу функций, вейвлет-коэффициенты которых удовлетворяют (2).

Популярным методом удаления шума является пороговая обработка эмпирических вейвлет-коэффициентов, смысл которой заключается в обнулении коэффициентов, чьи абсолютные значения не превышают заданного порога. Оценка $\hat{Y}_{j,k}$ вычисляется с помощью пороговой функции $\rho_T(Y_{j,k})$ с порогом T . Наиболее распространены функции жесткой пороговой обработки $\rho_T^{(h)}(x) = x \mathbf{1}(|x| > T)$ и мягкой пороговой обработки $\rho_T^{(s)}(x) = \text{sign}(x)(|x| - T)_+$. Смысл такой обработки заключается в том, что, поскольку большинство «чистых» коэффициентов малы по абсолютному значению, обнуление эмпирических коэффициентов должно удалить шум, не сильно затронув полезный сигнал.

Основным критерием качества методов пороговой обработки является среднеквадратичный риск

$$r_J(f) = \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\hat{Y}_{j,k} - \mu_{j,k} \right)^2. \quad (3)$$

В ситуации, когда число эмпирических вейвлет-коэффициентов не случайно, вычислены оптимальные значения порогов и оценен порядок среднеквадратичного риска для различных классов

функций сигналов. В частности, в работе [2] доказано следующее утверждение, оценивающее минимальный порядок (3).

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

$$\sup_{f \in \text{Lip}(A, \gamma)} r_J(f) \leq C_M 2^{-(2\gamma/(2\gamma+1))J} J^{2\gamma/(2\gamma+1)},$$

где C_M — некоторая положительная константа.

Асимптотически оптимальный порог в теореме 1 при $J \rightarrow \infty$ удовлетворяет соотношению [4]:

$$T \simeq \sigma \sqrt{\frac{4\gamma}{2\gamma+1} \ln 2^J}.$$

В следующем разделе оценивается порядок среднеквадратичного риска пороговой обработки в модели со случайным числом эмпирических вейвлет-коэффициентов.

3 Случайное число вейвлет-коэффициентов

Пусть M — положительная целочисленная случайная величина и $N = 2^M$. Тогда среднеквадратичный риск для модели (1) принимает вид:

$$r(f) = \sum_{J=0}^{\infty} \mathbb{P}(N = 2^J) \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\hat{Y}_{j,k} - \mu_{j,k} \right)^2 \quad (4)$$

и его асимптотический порядок в значительной степени зависит от распределения N . Чтобы получить осмысленные оценки порядка риска (4), величина N должна быть «большой». Рассмотрим последовательность $N_n, n = 1, 2, \dots$, и предположим, что существует неслучайная возрастающая последовательность натуральных чисел $J_n, n = 1, 2, \dots$, такая, что $N_n/2^{J_n}$ имеет некоторый предел (в смысле равномерной сходимости по распределению) при $n \rightarrow \infty$, т. е.

$$\sup_{x \geq 0} |H_n(x) - H(x)| < \frac{\varepsilon_n}{2} \rightarrow 0, \quad n \rightarrow \infty, \quad (5)$$

где

$$H_n(x) = \mathbb{P} \left(\frac{N_n}{2^{J_n}} < x \right),$$

а $H(x)$ — предельная функция распределения. Предположим, что $H(x)$ не имеет атома в нуле и исследуем поведение

$$r_n(f) = \sum_{J=0}^{\infty} \mathbb{P}(N_n = 2^J) \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\hat{Y}_{j,k} - \mu_{j,k} \right)^2$$

при $n \rightarrow \infty$.

Пусть $\delta_n \rightarrow 0$, $\alpha_n \rightarrow 0$ при $n \rightarrow \infty$ так, что $J_n + \log_2 \delta_n \rightarrow \infty$ и $H(\delta_n) + 1 - H(\delta_n^{-1}) < \alpha_n$ при всех $n = 1, 2, \dots$. Тогда

$$\begin{aligned} r_n(f) &= \\ &= \sum_{J=0}^{[J_n + \log_2 \delta_n]} P(N_n = 2^J) \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} E(\hat{Y}_{j,k} - \mu_{j,k})^2 + \\ &+ \sum_{J=[J_n + \log_2 \delta_n] + 1}^{[J_n - \log_2 \delta_n]} P(N_n = 2^J) \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} E(\hat{Y}_{j,k} - \mu_{j,k})^2 + \\ &+ \sum_{J=[J_n - \log_2 \delta_n] + 1}^{\infty} P(N_n = 2^J) \frac{1}{2^J} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} E(\hat{Y}_{j,k} - \mu_{j,k})^2 \equiv \\ &\equiv S_1 + S_2 + S_3. \end{aligned}$$

Учитывая (5), для $S_1 + S_3$ имеем

$$S_1 + S_3 \leq H_n(\delta_n) + 1 - H_n(\delta_n^{-1}) \leq \alpha_n + \varepsilon_n.$$

Для S_2 с помощью теоремы 1 можно получить оценку:

$$S_2 \leq C_1 2^{-(2\gamma/(2\gamma+1))(J_n + \log_2 \delta_n)} (J_n + \log_2 \delta_n)^{2\gamma/(2\gamma+1)},$$

справедливую для всех $f \in \text{Lip}(A, \gamma)$ (здесь C_1 — некоторая положительная константа). Таким образом, справедливо следующее утверждение.

Теорема 2. В модели со случайным числом вейвлет-коэффициентов при выборе асимптотически оптимального порога начиная с некоторого n справедлива оценка:

$$\begin{aligned} \sup_{f \in \text{Lip}(A, \gamma)} r_n(f) &\leq \alpha_n + \varepsilon_n + \\ &+ C_1 2^{-(2\gamma/(2\gamma+1))(J_n + \log_2 \delta_n)} (J_n + \log_2 \delta_n)^{2\gamma/(2\gamma+1)}. \end{aligned}$$

Сам асимптотически оптимальный порог при $n \rightarrow \infty$ удовлетворяет соотношению:

$$T_n \simeq \sigma \sqrt{\frac{4\gamma}{2\gamma+1} \ln 2^{J_n + \log_2 \delta_n}}.$$

Вид α_n , ε_n и δ_n в теореме 2 существенно зависит от поведения последовательности $N_n/2^{J_n}$ и предельной функции распределения $H(x)$. Так, ε_n характеризует скорость сходимости $H_n(x)$ к $H(x)$, а α_n и δ_n зависят от поведения $H(x)$ в окрестности нуля и бесконечности.

Следствие 1. Если предельное распределение $N_n/2^{J_n}$ вырожденно: $N_n/2^{J_n} \xrightarrow{P} 1$ при $n \rightarrow \infty$, тогда начиная с некоторого n

$$\sup_{f \in \text{Lip}(A, \gamma)} r_n(f) \leq \varepsilon_n + C_2 2^{-(2\gamma/(2\gamma+1))J_n} J_n^{2\gamma/(2\gamma+1)},$$

где C_2 — некоторая положительная константа.

Если ε_n убывает достаточно быстро, то эта оценка совпадает с оценкой для среднеквадратичного риска в модели с неслучайным числом вейвлет-коэффициентов.

Следствие 2. Пусть $H(x)$ дифференцируема в окрестности нуля и для некоторых положительных констант b и B в этой окрестности выполнено $b \leq H'(x) \leq B$. Пусть $\delta_n = 2^{-(2\gamma/(4\gamma+1))J_n}$. Тогда начиная с некоторого n справедлива оценка:

$$\begin{aligned} \sup_{f \in \text{Lip}(A, \gamma)} r_n(f) &\leq \\ &\leq \varepsilon_n + C_3 2^{-(2\gamma/(4\gamma+1))J_n} J_n^{2\gamma/(2\gamma+1)}, \end{aligned} \quad (6)$$

где C_3 — некоторая положительная константа.

Заметим, что $2\gamma/(4\gamma+1) < 1/2$ при $\gamma > 0$, и если $\varepsilon_n = O(2^{-J_n/2})$ (что является распространенной оценкой для скорости сходимости по распределению), то второе слагаемое в (6) определяет скорость убывания среднеквадратичного риска. Таким образом, среднеквадратичный риск для случайного числа вейвлет-коэффициентов может стремиться к нулю значительно медленнее, чем среднеквадратичный риск для неслучайного числа вейвлет-коэффициентов.

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Donoho D., Johnstone I. M., Kerkycharian G., Picard D. Wavelet shrinkage: Asymptopia? // *J. Roy. Stat. Soc. B*, 1995. Vol. 57. No. 2. P. 301–369.
3. Donoho D., Johnstone I. M. Minimax estimation via wavelet shrinkage // *Ann. Stat.*, 1998. Vol. 26. No. 3. P. 879–921.
4. Jansen M. Noise reduction by wavelet thresholding. — *Lecture notes in statistics ser.* — New York, NY, USA: Springer Verlag, 2001. Vol. 161. 196 p.
5. Jansen M. Minimum risk thresholds for data with heavy noise // *IEEE Signal Proc. Let.*, 2006. Vol. 13. No. 5. P. 296–299.
6. Шестаков О. В. Минимаксный среднеквадратичный риск пороговой обработки в моделях с негауссовым распределением шума // *Вестн. Моск. ун-та. Сер. 15: Вычисл. матем. и киберн.*, 2017. № 4. С. 35–40.
7. Mallat S. A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.

Поступила в редакцию 10.05.18

MEAN-SQUARE THRESHOLDING RISK WITH A RANDOM SAMPLE SIZE

O. V. Shestakov^{1,2}

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

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

Abstract: Nonlinear methods of signal de-noising based on the threshold processing of wavelet coefficients are widely used in various application areas. These methods have gained their popularity due to the speed of the algorithms for constructing estimates and the possibility of adapting to functions belonging to different classes of regularity better than linear methods. When applying thresholding techniques, it is usually assumed that the number of wavelet coefficients is fixed and the noise distribution is Gaussian. This model has been well studied in the literature, and optimal threshold values have been calculated for different classes of signals. However, in some situations, the sample size is not known in advance and is modeled by a random variable. The present author considers a model with a random number of observations containing Gaussian noise and estimates the order of the mean-square risk with increasing sample size.

Keywords: thresholding; random sample size; mean-square risk

DOI: 10.14357/19922264180302

Acknowledgments

This research is supported by the Russian Science Foundation (project No. 18-11-00155).

References

1. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455.
2. Donoho, D., I. M. Johnstone, G. Kerkycharian, and D. Picard. 1995. Wavelet shrinkage: Asymptopia? *J. Roy. Stat. Soc. B* 57(3):301–369.
3. Donoho, D., and I. M. Johnstone. 1998. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 26(3):879–921.
4. Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. New York, NY: Springer Verlag. Vol. 161. 196 p.
5. Jansen, M. 2006. Minimum risk thresholds for data with heavy noise. *IEEE Signal Proc. Let.* 13(5):296–299.
6. Shestakov, O. V. 2017. Minimax mean-square thresholding risk in models with non-Gaussian noise distribution. *Mosc. Univ. Comput. Math. Cybern.* 41(4):187–192.
7. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.

Received May 10, 2018

Contributor

Shestakov Oleg V. (b. 1976) — Doctor of Science 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; 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

БАЙЕСОВСКИЕ МОДЕЛИ БАЛАНСА*

А. А. Кудрявцев¹

Аннотация: Ряд предыдущих работ автора был посвящен применению байесовского подхода к задачам массового обслуживания и надежности. В данной статье метод распространяется на широкий круг задач из различных областей знания: демографии, физики, политологии, моделирования чрезвычайных ситуаций, медицины и др. В основе метода лежит разделение факторов, влияющих на исследуемую систему, на способствующие функционированию (позитивные, или p -факторы) и препятствующие функционированию (негативные, или n -факторы). Рассматривается индекс баланса системы, равный отношению n -фактора к p -фактору, и индекс преимущества, равный отношению p -фактора к сумме n - и p -факторов. Предполагается, что факторы, влияющие на систему, меняются со временем, причем точные значения факторов невозможно определить ввиду несовершенства измерительного оборудования, излишне высокой стоимости досконального изучения, нехватки временных и материальных ресурсов и т. п. Такие предпосылки обуславливают применение к описанным задачам байесовского метода, заключающегося в рандомизации исходных параметров (факторов) и, как следствие, индексов баланса и преимущества. Основной целью исследования является изучение вероятностных характеристик индексов баланса и преимущества в предположении, что априорные распределения факторов известны. В случае, когда n - и p -факторы являются независимыми случайными величинами, задача сводится к исследованию свойств смесей распределений. В отличие от популярных в настоящее время смесей нормальных законов в байесовских моделях баланса смешиваемые распределения имеют положительные носители. Особое внимание уделяется априорным распределениям гамма-типа, поскольку эти распределения являются адекватными асимптотическими аппроксимациями широкого класса вероятностных распределений. Ранее рассматривались смеси показательного, эрланговского и вейбулловского априорных распределений. В данной статье особое внимание уделено случаю, когда n - и p -факторы имеют m -распределение Накагами и его частные виды (распределение Рэлея, Максвелла–Больцмана, хи-распределение и др.). Получены явные виды плотности, функции распределения и моментов индекса баланса для различных комбинаций описанных априорных распределений. Результаты статьи могут применяться в задачах исследования разного рода индексов, рейтингов и показателей.

Ключевые слова: байесовский метод; смешанные распределения; индекс баланса; индекс преимущества; процесс баланса; m -распределение Накагами

DOI: 10.14357/19922264180303

1 Введение

подавляющее большинство аспектов современной жизни — от бытовых приборов до государственного управления — усложнилось настолько, что определение критериев эффективности путем детерминированного анализа стало практически невозможным. По этой причине все чаще можно встретить разного рода индексы и рейтинги, дающие возможность быстро принимать решения в ситуациях, на исследование которых могли бы уйти годы и значительные финансовые и материальные ресурсы.

Об эффективности и адекватности рейтингов как инструмента анализа можно спорить [1], однако сложно представить современный мир без их использования.

2 Примеры моделей баланса

В основе построения рейтингов и индексов обычно лежит разделение параметров модели на два класса. Первый класс включает параметры, способствующие функционированию целевого объекта и позитивно влияющие на исследуемый процесс (p -факторы); второй класс включает параметры, препятствующие и негативно влияющие (n -факторы). Такое разделение достаточно условно. Так, при катализе в зависимости от постановки задачи ингибитор может выступать и в роли n -фактора (если требуется увеличить скорость реакции), и в роли p -фактора (при замедлении нежелательных реакций).

Разделение параметров на негативные и позитивные факторы свойственно моделям из очень

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00577).

¹ Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, nubigena@mail.ru

разных областей знания. Так, в модели Ричардсона гонки вооружений [2] для достижения баланса сил необходимо рассматривать коэффициент роста вооружений и «прошлые обиды» как условные р-факторы и бремя расходов и коэффициент доброй воли как условные п-факторы; при определении степени пожарной опасности принято использовать комплексный показатель Нестерова [3], в котором р-фактором можно считать температуру точки росы, а п-фактором — температуру воздуха; математическая модель инфекционного заболевания строится на основе соотношений баланса [4] для каждой из зависимых переменных: концентрация размножающихся антигенов, доля разрушенных антигеном клеток (п-факторы), концентрация носителей антител, концентрация антител (р-факторы); при анализе экономического развития государства легальные и теневые базовые параметры рынка можно рассматривать и как р-факторы, и как п-факторы [5]. Можно привести множество других примеров.

Вполне естественно, что функционирование исследуемой системы в итоге зависит не столько от значений п- и р-факторов, сколько от их соотношения. При этом большое расхождение между величинами факторов обычно свидетельствует либо о чрезмерных затратах на «борьбу с негативным влиянием», либо о недооценке негативного воздействия. Таким образом, для того чтобы система была сбалансированной, имеет смысл стремиться приблизить к единице отношение п-фактора к р-фактору.

Однако существуют постановки задач, в которых преобладания р-фактора над п-фактором имеет смысл добиваться, невзирая на «цену вопроса». Такие постановки имеют место, когда речь идет, например, о безопасности или надежности. В этом случае отношение негативного к позитивному фактору стремится к нулю и для лучшего понимания близости к решению поставленной задачи рассматривается отношение р-фактора к сумме р- и п-факторов и его близость к единице. Наилучшим образом различие между двумя постановками задач можно проиллюстрировать современными геополитическими реалиями, в которых наряду с «международными отношениями, построенными на балансе взаимодействия и конкуренции» присутствует идея глобального доминирования, при которой «свои интересы продавливаются любой ценой» [6].

Обозначим через λ и μ соответственно п- и р-факторы модели. Рассмотрим *индекс баланса* $\rho = \lambda/\mu$ и *индекс преимущества* $\pi = \mu/(\mu + \lambda) = 1/(1 + \rho)$. Заметим, что зачастую в качестве индекса баланса удобно рассматривать величину ρ^{-1} ,

а наряду с индексом преимущества рассматривать *индекс недостатка* $1 - \pi$.

Приведем ряд примеров индексов баланса и преимущества и соответствующих им п- и р-факторов из различных областей знания. Так, в демографии уровень младенческой смертности определяется как отношение числа умерших в возрасте до года к числу родившихся за период [7], а индекс разводимости — как отношение коэффициента суммарной разводимости к коэффициенту суммарной брачности [8]; в физике коэффициент трансформации — это отношение выходного напряжения к входному, а универсальная функция Кирхгофа — отношение излучательной способности тела к поглощательной [9]; в теории массового обслуживания отношение интенсивности входящего потока к интенсивности обслуживания определяет коэффициент загрузки системы [10]; при моделировании чрезвычайных ситуаций пожароопасность объекта определяется отношением угрозы возникновения пожара к фактору пожарозащиты [3]; в теории надежности ожидаемое время безотказной работы представимо в виде отношения среднего времени безотказной работы к среднему времени восстановления [11].

Примерами индексов преимущества служат: в физике коэффициент полезного действия источника тока (п-фактор — сопротивление источника тока, р-фактор — сопротивление нагрузки) [9]; в политологии предельная доля мобилизованного населения (п-фактор — коэффициент выбытия, р-фактор — коэффициент агитируемости) [2]; в теории надежности коэффициент готовности, вероятность пребывания в работоспособном состоянии (п-фактор — среднее время восстановления, р-фактор — средняя наработка на отказ) [11–13] и предельная надежность (п-фактор — средний параметр дефективности, р-фактор — средний параметр эффективности) [14]; в теории массового обслуживания вероятность того, что вызов не будет потерян (п-фактор — интенсивность входящего потока, р-фактор — интенсивность обслуживания) [15]. Индексы баланса и преимущества также можно использовать при анализе рынка ценных бумаг для исследования процесса дисбаланса интенсивности [16], спроса при помощи функций Торнквиста и оптимальных запасов по формуле Харриса [2]. Ряд примеров можно продолжить.

3 Байесовские процессы баланса

С течением времени п- и р-факторы, а следовательно, и индексы баланса/преимущества претерпевают изменения. Это связано с неустойчивостью среды, в которой происходит функциони-

рование, — изменяются экономическое развитие, политическая система, технологии производства, пристрастия населения и т. д. По этой причине имеет смысл рассматривать не только мгновенные значения факторов и индексов, но и соответствующие функции от времени: *n*-процесс $\lambda(t)$, *p*-процесс $\mu(t)$, процесс баланса $\rho(t) = \lambda(t)/\mu(t)$ и процесс преимущества $\pi(t) = \mu(t)/(\mu(t) + \lambda(t))$. Кроме того, в подавляющем большинстве случаев перечисленные процессы отслеживаются не постоянно, а в некоторые отсчеты времени $t_1 < t_2 < \dots$. Моменты времени t_i могут распределяться не однородно (минуты, дни, годы), а быть привязаны к некоторым внешним событиям (публикация доклада, внеплановая проверка, стихийное бедствие и т. п.). Обозначим мгновенное значение фактора или индекса через $\xi_i = \xi(t_i)$, где $\xi(t)$ — один из рассматриваемых процессов.

Заметим, что с течением времени изменяются «состояния природы», которые в теории вероятностей принято называть элементарными исходами. Таким образом, невозможно предугадать частное значение фактора из-за невозможности досконального изучения «состояния природы». Измерения при помощи приборов также не могут дать абсолютно точного значения из-за неизбежно вносимых погрешностей, имеющих в большинстве случаев изменчивый (случайный) характер: «наблюдения в различные моменты времени учитывают влияние изменения условий окружающей среды и перекалибровки оборудования между наблюдениями» [17]. Это дает предпосылки для рассмотрения факторов, а следовательно, и индексов как случайных величин. При этом стоит учитывать, что глобальные изменения окружающей среды происходят достаточно редко, поэтому законы, влияющие на значения факторов, можно считать (в рамках конкретной модели) неизменными. Из этого следует, что распределения рассматриваемых случайных величин надо полагать заданными априорно.

Приведенные выше рассуждения обуславливают применение к моделям баланса байесовского метода.

В настоящее время в прикладной статистике, медицине, социологии, экономике и других науках активно применяется байесовский подход [18–21], основанный на вычислении апостериорных вероятностей и формуле Байеса [22]; метод, исследовавшийся Байесом [23] и основанный на рандомизации параметров при помощи известных априорных распределений, было предложено применять к задачам массового обслуживания и надежности [24–26].

Везде далее будем предполагать, что даны два случайных процесса $\lambda(t)$ и $\mu(t)$ с известными апри-

орными распределениями. Основной задачей исследования является вычисление (или оценивание) распределения процесса баланса $\rho(t)$ и процесса преимущества $\pi(t)$. Зная эти распределения, можно делать выводы о стабильности системы, строить прогнозы, исследовать средние и экстремальные значения и т. п. При вычислении таких важных характеристик распределений процессов, как моментные функции и ковариационные матрицы, необходимо делать предположения о взаимной зависимости проекций процесса. При этом следует учитывать, что зачастую *n*-процесс зависит от *p*-процесса и наоборот. Так, при подготовке обновления программного обеспечения, необходимого для устранения сбоев, улучшения показателей производительности и/или других характеристик продукта [27], *p*-фактор меняется в зависимости от *n*-фактора, а мутацию вируса гриппа под влиянием вакцинации и эффект первородного антигенного греха [28–30] можно рассматривать как влияние *p*-фактора на *n*-фактор.

Строить предположения о зависимости факторов следует исходя из конкретной модели. Однако, если взять за основу зависимость λ_i только от μ_{i-1} , а μ_i — только от λ_{i-1} , $i = 1, 2, \dots$, а также независимость λ_0 и μ_0 в начальный момент времени, получаем для каждого i независимость пары случайных величин (λ_i, μ_i) .

Исследование процессов баланса и преимущества и их конечномерных распределений, таким образом, имеет смысл осуществлять при помощи метода анализа масштабных смесей априорных распределений негативных и позитивных факторов. При этом в байесовских моделях баланса, в отличие от популярных моделей, описываемых в терминах сдвиговых и масштабных смесей нормальных законов [31], оба смешиваемых распределения в большинстве случаев имеют положительные носители.

4 Примеры вычисления одномерных распределений процессов баланса

Основная сложность при исследовании проекций процессов баланса заключается в том, что характеристики распределений индексов баланса и преимущества зачастую выражаются в терминах специальных функций: интегральной показательной функции Эйлера [32], обобщенной гипергеометрической функции Гаусса [33], преобразований Лапласа специального вида [34], гамма-экспоненциальной функции [35], бета- и гамма-функций и др. Данное обстоятельство не только вынуждает

применять компьютерные средства для вычисления характеристик модели, но и искать новые аналитические подходы [35].

Рассмотрим ряд утверждений, обоснования которых приводятся в следующем разделе, с примерами характеристик баланса.

Введем следующие обозначения. Пусть $f_\lambda(x)$ и $F_\lambda(x)$ — соответственно плотность и функция распределения случайной величины λ . Через $B(q, p)$ и $\Gamma(q)$, $q, p > 0$, будем соответственно обозначать бета- и гамма-функции. Пусть

$$(\alpha)_j = \alpha(\alpha + 1) \cdots (\alpha + j - 1), \quad (\alpha)_0 = 1.$$

Рассмотрим вырожденную гипергеометрическую функцию

$$G_0(\alpha, \beta; x) = \sum_{j=0}^{\infty} \frac{(\alpha)_j x^j}{(\beta)_j j!},$$

$$\alpha \in \mathbb{R}, \beta \neq 0, -1, \dots, x \in \mathbb{R},$$

неполную гамма-функцию

$$\gamma(p, y) = \int_0^y x^{p-1} e^{-x} dx, \quad p > 0, y > 0,$$

и неполную бета-функцию

$$\beta(q, p, y) = \int_0^y x^{q-1} (1-x)^{p-1} dx, \quad q, p > 0, y > 0.$$

В дальнейшем изложении речь будет идти о случайных величинах λ , имеющих бета-распределение:

$$f_\lambda(x) = \frac{x^{q-1} (1-x)^{\theta-1}}{B(q, \theta)}, \quad q, \theta > 0, x \in (0, 1),$$

гамма-распределение:

$$f_\lambda(x) = \frac{\theta^q x^{q-1} e^{-\theta x}}{\Gamma(q)}, \quad q, \theta > 0, x > 0,$$

m-распределение Накагами $m(q, \theta)$ [36]:

$$f_\lambda(x) = \frac{2q^q x^{2q-1}}{\theta^q \Gamma(q)} \exp\left\{-\frac{qx^2}{\theta}\right\},$$

$$q \geq 1/2, \theta > 0, x > 0, \quad (1)$$

распределение максимума процесса броуновского движения $Vm(\theta)$ [37]:

$$f_\lambda(x) = \frac{2}{\sqrt{2\pi\theta}} \exp\left\{-\frac{x^2}{2\theta}\right\}, \quad \theta > 0, x > 0, \quad (2)$$

распределение Рэлея Ray (θ) [38]:

$$f_\lambda(x) = \frac{2x}{\theta^2} \exp\left\{-\frac{x^2}{\theta^2}\right\}, \quad \theta > 0, x > 0, \quad (3)$$

хи-распределение $\chi(q)$:

$$f_\lambda(x) = \frac{x^{q-1} e^{-x^2/2}}{2^{q/2-1} \Gamma(q/2)}, \quad q > 0, x > 0, \quad (4)$$

распределение Максвелла–Больцмана MB (θ) [39]:

$$f_\lambda(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\theta^{3/2}} \exp\left\{-\frac{x^2}{2\theta}\right\}, \quad \theta > 0, x > 0. \quad (5)$$

Плотности $f_\mu(x)$ случайных величин μ будут иметь аналогичный вид с заменой параметров (q, θ) парой параметров (p, α) .

Теорема 1. Пусть негативный фактор λ имеет гамма-распределение с параметрами q и θ , а позитивный фактор μ имеет бета-распределение с параметрами p и α , причем λ и μ независимы. Тогда индекс баланса $\rho = \lambda/\mu$ имеет плотность

$$f_\rho(x) = \frac{\theta^q x^{q-1} B(q+p, \alpha)}{\Gamma(q) B(p, \alpha)} G_0(q+p, q+p+\alpha; -\theta x),$$

$$x > 0, \quad (6)$$

и функцию распределения

$$F_\rho(x) = \frac{\theta^q B(q+p, \alpha)}{\Gamma(q) B(p, \alpha)} \sum_{j=0}^{\infty} \frac{(q+p)_j (-\theta)^j}{(q+p+\alpha)_j j!} \frac{x^{q+j}}{q+j},$$

$$x > 0.$$

Выражение плотности индекса баланса ρ через специальную функцию G_0 затрудняет исследование свойств распределения ρ стандартными методами теории вероятностей и математической статистики. Однако для приложений существенно большую роль играют моментные характеристики ρ .

Следующее утверждение и предположение о независимости p - и r -факторов дает возможность исследовать средние значения индекса баланса.

Теорема 2. Пусть негативный фактор λ имеет вырожденное в $\theta > 0$ распределение, а позитивный фактор μ имеет бета-распределение с параметрами p и α . Тогда вероятностные характеристики индекса баланса ρ имеют вид:

$$f_\rho(x) = \frac{x^{-1}}{B(p, \alpha)} \left(\frac{\theta}{x}\right)^p \left(1 - \frac{\theta}{x}\right)^{\alpha-1}, \quad x > \theta;$$

$$E \rho^l = \frac{\theta^l B(p-l, \alpha)}{B(p, \alpha)}, \quad l < p.$$

Следствие 1. Пусть негативный фактор λ имеет гамма-распределение с параметрами q и θ , а позитив-

ный фактор μ имеет бета-распределение с параметрами p и α , причем λ и μ независимы. Тогда для моментов индекса баланса ρ справедливо

$$E \rho^l = \frac{\Gamma(q+l)B(p-l, \alpha)}{\theta^l \Gamma(q)B(p, \alpha)}, \quad l < p.$$

Отдельное место при исследовании индексов баланса занимают модели, в которых распределения n - и p -факторов являются различными обобщениями гамма-распределения. Это обусловлено тем, что распределения гамма-типа встречаются во многих прикладных задачах. Так, m -распределение Накагами используется для моделирования замираний сигналов в беспроводных каналах связи, распределение Рэлея было предложено для исследования гармонических колебаний со случайными фазами и применяется для описания амплитудных флуктуаций радиосигнала, распределение Максвелла–Больцмана лежит в основании кинетической теории газов и описывает электронные процессы переноса и т. д. Кроме того, распределения гамма-типа выступают в роли адекватных асимптотических аппроксимаций широкого круга вероятностных распределений [40]. В свою очередь, это дает возможность рассматривать факторы с распределениями гамма-типа в качестве агрегирующих факторов.

Ранее рассматривались модели, в которых факторы имели показательное [41], эрланговское [32] и вейбулловское распределения [34, 35]. Приведем ряд примеров применения аналогичных методов получения характеристик индекса баланса для распределения Накагами и некоторых его частных случаев.

Теорема 3. Пусть негативный и позитивный факторы λ и μ имеют m -распределение Накагами с параметрами (q, θ) и (p, α) соответственно, причем λ и μ независимы. Тогда индекс баланса ρ имеет плотность распределения

$$f_\rho(x) = \frac{2(q/\theta)^q (p/\alpha)^p x^{2q-1}}{B(q, p)(qx^2/\theta + p/\alpha)^{q+p}}, \quad x > 0, \quad (7)$$

и функцию распределения

$$F_\rho(x) = \frac{\beta(q, p, \alpha qx^2 / (\alpha qx^2 + \theta p))}{B(q, p)}, \quad x > 0.$$

Для нахождения моментов, соответствующих плотности (7), воспользуемся следующим утверждением.

Теорема 4. Пусть негативный фактор λ имеет вырожденное в $\theta > 0$ распределение, а позитивный фактор μ имеет распределение Накагами с параметрами p и α . Тогда вероятностные характеристики индекса баланса ρ имеют вид:

$$F_\rho(x) = 1 - \frac{\gamma(p, \theta^2 p / (\alpha x^2))}{\Gamma(p)}, \quad x > 0;$$

$$f_\rho(x) = \frac{2\theta^{2p} (p/\alpha)^p}{\Gamma(p)x^{2p+1}} \exp\left\{-\frac{\theta^2 p}{\alpha x^2}\right\}, \quad x > 0;$$

$$E \rho^l = \frac{\theta^l (p/\alpha)^{l/2}}{\Gamma(p)} \Gamma\left(p - \frac{l}{2}\right), \quad l < 2p.$$

Следствие 2. Пусть негативный и позитивный факторы λ и μ имеют распределение Накагами с параметрами (q, θ) и (p, α) соответственно, причем λ и μ независимы. Тогда для моментов индекса баланса ρ справедливо

$$E \rho^l = \frac{(p/\alpha)^{l/2} \Gamma(q+l/2)}{(q/\theta)^{l/2} \Gamma(q) \Gamma(p)} \Gamma\left(p - \frac{l}{2}\right), \quad l < 2p.$$

Заметим, что плотность распределения Накагами (1) обобщает плотности распределений (2)–(5), что дает возможность сформулировать ряд следствий для всевозможных комбинаций априорных распределений (1)–(5) n - и p -факторов.

Следствие 3. Пусть негативный и позитивный факторы λ и μ имеют распределения, определяемые плотностями (1)–(5) соответственно с параметрами (q, θ) и (p, α) , причем λ и μ независимы. Тогда вероятностные характеристики индекса баланса ρ имеют вид:

$$f_\rho(x) = \frac{2R^Q S^P x^{2Q-1}}{B(Q, P)(Rx^2 + S)^{Q+P}}, \quad x > 0;$$

$$F_\rho(x) = \frac{\beta(Q, P, Rx^2 / (Rx^2 + S))}{B(Q, P)}, \quad x > 0;$$

$$E \rho^l = \frac{S^{l/2} \Gamma(Q+l/2)}{R^{l/2} \Gamma(Q) \Gamma(P)} \Gamma\left(P - \frac{l}{2}\right), \quad l < 2P,$$

где вектор констант (Q, R, P, S) определяется из таблицы.

Замечание 1. Условие $l < 2P$ в следствии 3 ограничивает число конечных моментов. Например, в случае, когда p -фактор имеет распределение Рэлея, у индекса баланса ρ не существует дисперсии.

Замечание 2. Для получения вероятностных характеристик индекса преимущества $\pi = 1/(1 + \rho)$ можно использовать соотношения:

$$f_\pi(x) = \frac{1}{x^2} f_\rho\left(\frac{1-x}{x}\right),$$

$$F_\pi(x) = 1 - F_\rho\left(\frac{1-x}{x}\right), \quad x \in (0, 1).$$

Однако вычисление моментов $E \pi^l$ представляет отдельную задачу, в том числе и по причине представления плотности ρ в терминах специальных функций.

Значения констант (Q, R, P, S) для частных случаев m -распределения Накагами

λ	μ				
	$m(p, \alpha)$	$Bm(\alpha)$	$Ray(\alpha)$	$\chi(p)$	$MB(\alpha)$
$m(q, \theta)$	$\left(\frac{q}{\theta}, \frac{q}{\theta}, p, \frac{p}{\alpha}\right)$	$\left(\frac{q}{\theta}, \frac{q}{\theta}, \frac{1}{2}, \frac{1}{2\alpha}\right)$	$\left(\frac{q}{\theta}, \frac{q}{\theta}, 1, \frac{1}{\alpha^2}\right)$	$\left(\frac{q}{\theta}, \frac{q}{\theta}, \frac{p}{2}, \frac{1}{2}\right)$	$\left(\frac{q}{\theta}, \frac{q}{\theta}, \frac{3}{2}, \frac{1}{2\alpha}\right)$
$Bm(\theta)$	$\left(\frac{1}{2}, \frac{1}{2\theta}, p, \frac{p}{\alpha}\right)$	$\left(\frac{1}{2}, \frac{1}{2\theta}, \frac{1}{2}, \frac{1}{2\alpha}\right)$	$\left(\frac{1}{2}, \frac{1}{2\theta}, 1, \frac{1}{\alpha^2}\right)$	$\left(\frac{1}{2}, \frac{1}{2\theta}, \frac{p}{2}, \frac{1}{2}\right)$	$\left(\frac{1}{2}, \frac{1}{2\theta}, \frac{3}{2}, \frac{1}{2\alpha}\right)$
$Ray(\theta)$	$\left(1, \frac{1}{\theta^2}, p, \frac{p}{\alpha}\right)$	$\left(1, \frac{1}{\theta^2}, \frac{1}{2}, \frac{1}{2\alpha}\right)$	$\left(1, \frac{1}{\theta^2}, 1, \frac{1}{\alpha^2}\right)$	$\left(1, \frac{1}{\theta^2}, \frac{p}{2}, \frac{1}{2}\right)$	$\left(1, \frac{1}{\theta^2}, \frac{3}{2}, \frac{1}{2\alpha}\right)$
$\chi(q)$	$\left(\frac{q}{2}, \frac{1}{2}, p, \frac{p}{\alpha}\right)$	$\left(\frac{q}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2\alpha}\right)$	$\left(\frac{q}{2}, \frac{1}{2}, 1, \frac{1}{\alpha^2}\right)$	$\left(\frac{q}{2}, \frac{1}{2}, \frac{p}{2}, \frac{1}{2}\right)$	$\left(\frac{q}{2}, \frac{1}{2}, \frac{3}{2}, \frac{1}{2\alpha}\right)$
$MB(\theta)$	$\left(\frac{3}{2}, \frac{1}{2\theta}, p, \frac{p}{\alpha}\right)$	$\left(\frac{3}{2}, \frac{1}{2\theta}, \frac{1}{2}, \frac{1}{2\alpha}\right)$	$\left(\frac{3}{2}, \frac{1}{2\theta}, 1, \frac{1}{\alpha^2}\right)$	$\left(\frac{3}{2}, \frac{1}{2\theta}, \frac{p}{2}, \frac{1}{2}\right)$	$\left(\frac{3}{2}, \frac{1}{2\theta}, \frac{3}{2}, \frac{1}{2\alpha}\right)$

5 Доказательства утверждений

Приведем обоснования утверждений из разд. 4. Доказательство **теоремы 1**. Поскольку при $x > 0$

$$f_\rho(x) = \int_0^1 y f_\lambda(xy) f_\mu(y) dy = \frac{\theta^q x^{q-1}}{\Gamma(q) B(p, \alpha)} \int_0^1 y^{q+p-1} (1-y)^{\alpha-1} e^{-\theta xy} dy,$$

применив [42, формула 3.383], получаем (6). Выражение для функции распределения ρ получаем, проинтегрировав (6).

Доказательство **теоремы 2** непосредственно вытекает из равенства

$$f_\rho(x) = \frac{\theta}{x^2} f_\mu\left(\frac{\theta}{x}\right), \quad x > \theta.$$

Доказательство **теоремы 3**. Для плотности индекса баланса при $x > 0$ имеем

$$\begin{aligned} f_\rho(x) &= \frac{4(q/\theta)^q (p/\alpha)^p}{\Gamma(q)\Gamma(p)} x^{2q-1} \int_0^\infty y^{2q+2p-1} e^{-(qx^2/\theta+p/\alpha)y^2} dy = \\ &= \frac{2(q/\theta)^q (p/\alpha)^p}{\Gamma(q)\Gamma(p)} x^{2q-1} \int_0^\infty t^{q+p-1} e^{-(qx^2/\theta+p/\alpha)t} dt = \\ &= \frac{2(q/\theta)^q (p/\alpha)^p x^{2q-1}}{B(q, p)(qx^2/\theta + p/\alpha)^{q+p}}, \end{aligned}$$

откуда получаем:

$$\begin{aligned} F_\rho(x) &= \frac{2(q/\theta)^q (p/\alpha)^p}{B(q, p)} \int_0^x \frac{u^{2q-1} du}{(qu^2/\theta + p/\alpha)^{q+p}} = \\ &= \frac{(p/\alpha)^p}{B(q, p)} \int_{p/\alpha}^{qx^2/\theta+p/\alpha} \frac{(t-p/\alpha)^{q-1}}{t^{q+p}} dt = \\ &= \frac{1}{B(q, p)} \int_0^{qx^2/(qx^2+\theta p/\alpha)} z^{q-1} (1-z)^{p-1} dz. \end{aligned}$$

Доказательство **теоремы 4**. Поскольку при $x > 0$

$$F_\rho(x) = 1 - \frac{(p/\alpha)^p}{\Gamma(p)} \int_0^{\theta^2/x^2} y^{p-1} e^{-py/\alpha} dy,$$

продифференцировав по x , получаем $f_\rho(x)$ и выражение для моментов при $l < 2p$:

$$\begin{aligned} E\rho^l &= \frac{\theta^{2p} (p/\alpha)^p}{\Gamma(p)} \int_0^\infty y^{p-l/2-1} e^{-\theta^2 py/\alpha} dy = \\ &= \frac{\theta^l (p/\alpha)^{l/2}}{\Gamma(p)} \Gamma\left(p - \frac{l}{2}\right). \end{aligned}$$

Последнее равенство наряду с представлением моментов распределения Накагами

$$E\lambda^l = \frac{\Gamma(q+l/2)}{(q/\theta)^{l/2} \Gamma(q)}$$

обосновывает следствие 2.

6 Заключение

В данной работе было существенно расширено множество возможных областей применения байесовских постановок задач определения значений

индексов баланса и преимущества, целесообразность которых ранее обуславливалась только применительно к теории массового обслуживания и надежности [24–26]. Решение задачи нахождения вероятностных характеристик частного двух случайных факторов имеет не только практическую значимость, но и представляет теоретический интерес.

Литература

1. Юрасова М. В. Рейтинг как инструмент измерения успеха: «за» и «против» // Вестник Московского ун-та. Сер. 18: Социология и политология, 2017. Т. 23. Вып. 2. С. 137–164.
2. Шикин Е. В., Чхартишвили А. Г. Математические методы и модели в управлении. — 3-е изд. — М.: Дело, 2004. 440 с.
3. Шаптала В. Г., Радоуцкий В. Ю., Шаптала В. В. Основы моделирования чрезвычайных ситуаций. — Белгород: БГТУ, 2010. 166 с.
4. Болодурина И. П., Луговскова Ю. П. Оптимальное управление динамикой взаимодействия иммунной системы человека с инфекционными заболеваниями // Вестник СамГУ. Естественнонаучная сер., 2009. № 8(74). С. 138–153.
5. Пескова Д. Р. Теневой сектор: ингибитор или катализатор экономического развития? // Вестник Башкирского ун-та. Раздел экономика, 2006. Т. 11. Вып. 3. С. 141–143.
6. Мир будущего: через столкновение к гармонии: Итоговая пленарная сессия XIV ежегодного заседания Международного дискуссионного клуба «Валдай». — Сочи, 2017. <http://kremlin.ru/events/president/news/55882>.
7. Борисов В. А. Демография. — М.: NOTABENE, 2001. 272 с.
8. Волгин Н. А., Рыбаковский Л. Л., Калмыкова Н. М., Архангельский В. Н., Иванова Е. И., Захарова О. Д., Иванова А. Е., Денисенко М. Б., Тихомиров Н. П., Тихомирова Т. М. Демография / Под ред. Н. А. Волгина, Л. Л. Рыбаковского. — М.: Логос, 2005. 280 с.
9. Кузнецов С. И., Рогозин К. И. Справочник по физике. — Томск: ТПУ, 2012. 224 с.
10. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
11. Здоровцов И. А., Королев В. Ю. Основы теории надежности волоконно-оптических линий передачи железнодорожного транспорта. — М.: МАКС Пресс, 2004. 308 с.
12. Kozlov B. A., Ushakov I. A. Reliability handbook. — New York, NY, USA: Holt, Rinehart & Winston. 1970. 391 p.
13. ГОСТ 27.002-89 Надежность в технике. Основные понятия. Термины и определения. — М.: Госкомстандарт, 1989. 24 с.
14. Королев В. Ю., Соколов И. А. Основы математической теории надежности модифицируемых систем. — М.: ИПИ РАН, 2006. 102 с.
15. Ивченко Г. И., Каштанов В. А., Коваленко И. Н. Теория массового обслуживания. — М.: Высшая школа, 1982. 256 с.
16. Korolev V. Yu., Chertok A. V., Korchagin A. Yu., Zeifman A. I. Modeling high-frequency order flow imbalance by functional limit theorems for two-sided risk processes // Appl. Math. Comput., 2015. Vol. 253. P. 224–241.
17. ГОСТ Р ИСО 5725-1-2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения. — М.: Стандартинформ, 2009. 24 с.
18. Congdon P. Bayesian statistical modelling. — 2nd ed. — Chichester, U.K.: John Wiley & Sons, 2006. 596 p.
19. Hamada M. S., Wilson A., Reese C. S., Martz H. Bayesian reliability. — New York, NY, USA: Springer, 2008. 436 p.
20. Carlin B. P., Louis T. A. Bayesian methods for data analysis. — 3rd ed. — New York, NY, USA: Chapman & Hall, 2008. 552 p.
21. Albert J. Bayesian computation with R. — New York, NY, USA: Springer, 2009. 300 p.
22. Laplace P.-S. A philosophical essay on probabilities / Transl. from the French by F. W. Truscott, F. L. Emory. — New York, NY, USA: John Wiley & Sons, 1902. 223 p.
23. Bayes T., Price R. An essay towards solving a problem in the doctrine of chances // Phil. Trans., 1763. Vol. 53. P. 370–418.
24. Шоргин С. Я. О байесовских моделях массового обслуживания // II Научная сессия Института проблем информатики РАН: Тезисы докладов. — М.: ИПИ РАН, 2005. С. 120–121.
25. Бенинг В. Е., Королев В. Ю., Соколов И. А., Шоргин С. Я. Рандомизационные модели и методы теории надежности информационных и технических систем. — М.: ТОРУС ПРЕСС, 2007. 256 с.
26. Кудрявцев А. А., Шоргин С. Я. Байесовские модели в теории массового обслуживания и надежности. — М.: ФИЦ ИУ РАН, 2015. 76 с.
27. Орлик С. Введение в программную инженерию и управление жизненным циклом ПО. Программная инженерия. Сопровождение программного обеспечения, 2004–2005. http://www.software-testing.ru/files/se/3-5-software_engineering_maintenance.pdf.
28. Boni M. F. Vaccination and antigenic drift in influenza // Vaccine, 2008. Vol. 26. Suppl. 3. P. 8–14.
29. Kim J. H., Skountzou I., Compans R., Jacob J. Original antigenic sin responses to influenza viruses // J. Immunology, 2009. Vol. 183. Iss. 5. P. 3294–3301.
30. Amantonio. Разбираемся с прививками. Часть 20. Грипп (1) // Livejournal, 09.11.2017. <https://amantonio.livejournal.com/29621.html>. Продолжение: Часть 21. Грипп (2) // Livejournal, 10.11.2017. <https://amantonio.livejournal.com/29886.html>.

31. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 510 с.
32. *Кудрявцев А. А., Шоргин В. С., Шоргин С. Я.* Байесовские модели массового обслуживания и надежности: общий эрланговский случай // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 30–34.
33. *Жаворонкова Ю. В., Кудрявцев А. А., Шоргин С. Я.* Байесовская рекуррентная модель роста надежности: бета-распределение параметров // Информатика и её применения, 2014. Т. 8. Вып. 2. С. 48–54.
34. *Кудрявцев А. А., Титова А. И.* Байесовские модели массового обслуживания и надежности: вырожденно-вейбулловский случай // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 68–71.
35. *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.
36. *Nakagami M.* The m-distribution, a general formula of intensity of rapid fading // Statistical Methods in Radio Wave Propagation Symposium Proceedings / Ed. W. C. Hoffman. — New York, NY, USA: Pergamon Press, 1960. P. 3–36.
37. *Круглов В. М.* Случайные процессы. Ч. 1. Основы общей теории. — 2-е изд., перераб. и доп. — М.: Юрайт, 2016. 276 с.
38. *Siddiqui M. M.* Statistical inference for Rayleigh distributions // J. Res. NBS D Rad. Sci., 1964. Vol. 68D. No. 9. P. 1005–1010.
39. *Mandl F.* Statistical physics. — 2nd ed. — Chichester, U.K.: John Wiley & Sons, 1988. 385 p.
40. *Закс Л. М., Королев В. Ю.* Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 105–115.
41. *Кудрявцев А. А., Шоргин С. Я.* Байесовские модели массового обслуживания и надежности: экспоненциально-эрланговский случай // Информатика и её применения, 2009. Т. 3. Вып. 1. С. 44–48.
42. *Градштейн И. С., Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. — М.: Наука, 1971. 1108 с.

Поступила в редакцию 03.02.18

BAYESIAN BALANCE MODELS

A. A. Kudryavtsev

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

Abstract: A number of previous author's works were devoted to the Bayesian approach to queuing and reliability. In this paper, the method application is extended to a wide range of problems, such as demography, physics, political science, modeling of emergencies, medicine, etc. The method is based on separation of system factors into two groups: those that support functioning of the system (positive, or p-factors) and those that inhibit system's functioning (negative, or n-factors). In the paper, system's balance index, which equals to the ratio of n- and p-factors, and the advantage index, which equals to the ratio of p-factor to the sum of n- and p-factors, are considered. It is assumed that the factors, which affect the system, change over time, and besides their exact values are impossible to determine due to the measuring equipment's imperfections, excessively high expenses on thorough research, lack of time and resources, and so on. Such prerequisites lead to usage of the Bayesian method in application to the problems described. The method implies randomization of the initial parameters (factors) and, as a consequence, randomization of the balance and advantage indices. The main goal of the research is to study probabilistic characteristics of the balance and advantage indices assuming that the apriori distributions of the system's factors are known. In the case of independently distributed n- and p-factors, which are random variables, the problem is reduced to studying properties of the distributions' mixtures. As opposed to popular normal mixtures, in Bayesian balance models, the distribution being mixed has a positive support. Special attention is paid to apriori gamma-type distributions, since these distributions are adequate asymptotic approximations of a wide range of probability distributions. The mixtures of exponential, Erlang, and Weibull apriori distributions were considered earlier. In this paper, special attention is paid to the case of Nakagami m-distribution of n- and p-factors (with its particular cases of Rayleigh, Maxwell–Boltzmann, chi-, and other distributions). The explicit formulas for density, distribution functions, and moments of the balance index for different combinations of distributions are obtained. The results provided in this paper can be applied to many different tasks concerning indices, ratings, and indicators.

Keywords: Bayesian method; mixed distributions; balance index; advantage index; balance process; Nakagami m-distribution

DOI: 10.14357/19922264180303

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project No. 17-07-00577).

References

1. Yurasova, M. V. 2017. Reyting kak instrument izmereniya uspekha: “za” i “protiv” [Rating as a tool for measuring success: “Pro” and “contra”]. *Moscow State University Bull. Ser. 18. Sociology Political Sci.* 23(2):137–164.
2. Shikin, E. V., and A. G. Chkhartishvili. 2004. *Matematicheskie metody i modeli v upravlenii* [Mathematical methods and models in management]. 3rd ed. Moscow: Delo. 440 p.
3. Shaptala, V. G., V. Yu. Radoutskiy, and V. V. Shaptala. 2010. *Osnovy modelirovaniya chrezvychaynykh situatsiy* [Basics of modeling of emergency situations]. Belgorod: BGTU. 166 p.
4. Bolodurina, I. P., and Yu. P. Lugovskova. 2009. Optimal’noe upravlenie dinamikoy vzaimodeystviya immunnogo sistema cheloveka s infektsionnymi zabolevaniyami [Optimum control of dynamics of interaction of the human immune system with infectious diseases]. *Vestnik SamGU* 74(8):138–153.
5. Peskova, D. R. 2006. Tenevoy sektor: ingibitor ili katalizator ekonomicheskogo razvitiya? [The informal sector: An inhibitor or catalyst for economic development?]. *Vestnik Bashkirskogo universiteta* [Bull. Bashkir University. Section of the economy] 11(3):141–143.
6. Mezhdunarodnyy diskussionnyy klub “Valday” [International Discussion Club “Valdai”]. 2017. *Mir budushchego: cherez stolknovenie k garmonii. Itogovaya plenarnaya sessiya XIV ezhegodnogo zasedaniya* [The world of the future: Through a clash to harmony: The Final Plenary Session of the 14th Annual Meeting]. Sochi. Available at: <http://kremlin.ru/events/president/news/55882> (accessed February 13, 2018).
7. Borisov, V. A. 2001. *Demografiya* [Demography]. Moscow: NOTABENE. 272 p.
8. Volgin, N. A., L. L. Rybakovskiy, N. M. Kalmykova, et al. 2005. *Demografiya* [Demography]. Eds. N. A. Volgin and L. L. Rybakovskiy. Moscow: Logos. 280 p.
9. Kuznetsov, S. I., and K. I. Rogozin. 2012. *Spravochnik po fizike* [Handbook of physics]. Tomsk: TPU. 224 p.
10. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
11. Zdorovtsov, I. A., and V. Yu. Korolev. 2004. *Osnovy teorii nadezhnosti volokonno-opticheskikh liniy peredachi zheleznodorozhnogo transporta* [Fundamentals of reliability theory of fiber optic transmission lines for railway transport]. Moscow: MAKS Press. 308 p.
12. Kozlov, B. A., and I. A. Ushakov. 1970. *Reliability handbook*. New York, NY: Holt, Rinehart & Winston. 391 p.
13. GOST 27.002-89. 1989. *Nadezhnost’ v tekhnike. Osnovnyye ponyatiya. Terminy i opredeleniya* [Industrial product dependability. General concepts. Terms and definitions]. Moscow. 24 p.
14. Korolev, V. Yu., and I. A. Sokolov. 2006. *Osnovy matematicheskoy teorii nadezhnosti modifitsiruemykh sistem* [Fundamentals of mathematical theory of modified systems reliability]. Moscow: IPI RAN. 102 p.
15. Ivchenko, G. I., V. A. Kashtanov, and I. N. Kovalenko. 1982. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: Higher school. 256 p.
16. Korolev, V. Yu., A. V. Chertok, A. Yu. Korchagin, and A. I. Zeifman. 2015. Modeling high-frequency order flow imbalance by functional limit theorems for two-sided risk processes. *Appl. Math. Comput.* 253:224–241.
17. GOST R ISO 5725-1-2002. 2009. *Tochnost’ (pravil’nost’ i pretzionnost’) metodov i rezul’tatov izmereniy. Chast’ 1. Osnovnyye polozheniya i opredeleniya* [Accuracy (trueness and precision) of measurement methods and results. Part 1. General principles and definitions]. Moscow: Standardinform Publ. 24 p.
18. Congdon, P. 2006. *Bayesian statistical modelling*. 2nd ed. Chichester, U.K.: John Wiley & Sons. 596 p.
19. Hamada, M. S., A. Wilson, C. S. Reese, and H. Martz. 2008. *Bayesian reliability*. New York, NY: Springer. 436 p.
20. Carlin, B. P., and T. A. Louis. 2008. *Bayesian methods for data analysis*. 3rd ed. New York, NY: Chapman & Hall. 552 p.
21. Albert, J. 2009. *Bayesian computation with R*. New York, NY: Springer. 300 p.
22. Laplace, P.-S. 1902. *A philosophical essay on probabilities*. Transl. from the French by F. W. Truscott and F. L. Emory. New York, NY: John Wiley & Sons. 223 p.
23. Bayes, T., and R. Price. 1763. An essay towards solving a problem in the doctrine of chances. *Phil. Trans.* 53:370–418.
24. Shorgin, S. Ya. 2005. O bayesovskikh modelyakh massovogo obsluzhivaniya [On Bayesian queueing models]. *II Nauchnaya Sessiya instituta problem informatiki RAN: Tezisy dokladov* [2nd Scientific Session of the Institute of Informatics Problems of the Russian Academy of Sciences: Abstracts]. Moscow: IPI RAN. 120–121.
25. Bening, V. E., V. Yu. Korolev, I. A. Sokolov, and S. Ya. Shorgin. 2007. *Randomizatsionnye modeli i metody teorii nadezhnosti informatsionnykh i tekhnicheskikh sistem* [Randomization models and methods of reliability theory for information and technical systems]. Moscow: TORUS PRESS. 256 p.
26. Kudryavtsev, A. A., and S. Ya. Shorgin. 2015. *Bayesovskie modeli v teorii massovogo obsluzhivaniya i nadezhnosti* [Bayesian models in mass service and reliability theories]. Moscow: FIC IU RAN. 76 p.
27. Orlik, S. 2004–2005. *Vvedenie v programmnyy inzheneriyu i upravlenie zhiznennym tsiklom PO. Pro-*

- grammnaya inzheneriya. Soprovozhdenie programmno obespecheniya* [Introduction to software engineering and software lifecycle management. Software engineering. Maintenance of software]. Available at: http://www.software-testing.ru/files/se/3-5-software_engineering_maintenance.pdf (accessed February 13, 2018).
28. Boni, M. F. 2008. Vaccination and antigenic drift in influenza. *Vaccine* 26(Suppl. 3):8–14.
 29. Kim, J. H., I. Skountzou, R. Compans, and J. Jacob. 2009. Original antigenic sin responses to influenza viruses. *J. Immunology* 183(5):3294–3301.
 30. Amantonio. 09.11.2017. Razbiraemysya s privivkami. Chast' 20. Gripp (1) [We deal with vaccinations. Part 20. Influenza (1)]. *Livejournal*. Available at: <https://amantonio.livejournal.com/29621.html> (accessed February 13, 2018). 10.11.2017. Prodolzhenie: Chast' 21. Gripp (2) [Continuation: Part 21. Influenza (2)]. *Livejournal*. Available at: <https://amantonio.livejournal.com/29886.html> (accessed February 13, 2018).
 31. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: MSU Publ. 510 p.
 32. Kudryavtsev, A. A., V. S. Shorgin, and S. Ya. Shorgin. 2009. Bayesovskie modeli massovogo obsluzhivaniya i nadezhnosti: obshchiy erlangovskiy sluchay [Bayesian queuing and reliability models: General Erlang case]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):30–34.
 33. Zhavoronkova, Iu. V., A. A. Kudryavtsev, and S. Ya. Shorgin. 2014. Bayesovskaya rekurrentnaya model' rosta nadezhnosti: beta-raspredelenie parametrov [Bayesian recurrent model of reliability growth: Beta-distribution of parameters]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):48–54.
 34. Kudryavtsev, A. A., and A. I. Titova. 2016. Bayesovskie modeli massovogo obsluzhivaniya i nadezhnosti: vyrozhdennno-veybullovskiy sluchay [Bayesian queuing and reliability models: Degenerate-Weibull case]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):68–71.
 35. 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.
 36. Nakagami, M. 1960. The m-distribution, a general formula of intensity of rapid fading. *Statistical Methods in Radio Wave Propagation Symposium Proceedings*. Ed. W. C. Hoffman. New York, NY: Pergamon Press. 3–36.
 37. Kruglov, V. M. 2016. *Sluchaynye protsessy. Ch. 1. Osnovy obshchey teorii* [Stochastic processes. Part 1. Bases of general theory]. 2nd ed. Moscow: Yurayt. 276 p.
 38. Siddiqui, M. M. 1964. Statistical inference for Rayleigh distributions. *J. Res. NBS D Rad. Sci.* 68D(9):1005–1010.
 39. Mandl, F. 1988. *Statistical physics*. 2nd ed. Chichester, U.K.: John Wiley & Sons. 385 p.
 40. Zaks, L. M., and V. Yu. Korolev. 2013. Obobshchennyye dispersionnyye gamma-raspredeleniya kak predel'nye dlya sluchaynykh sum [Generalized dispersion gamma distributions as limiting for random sums]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):105–115.
 41. Kudryavtsev, A. A., and S. Ya. Shorgin. 2009. Bayesovskie modeli massovogo obsluzhivaniya i nadezhnosti: eksponentsial'no-erlangovskiy sluchay [Bayesian queuing and reliability models: An exponential-Erlang case]. *Informatika i ee Primeneniya — Inform. Appl.* 3(1):44–48.
 42. Gradshteyn, I. S., and I. M. Ryzhik. 1971. *Tablitsy integralov, summ, ryadov i proizvedeniy* [Tables of integrals, sums, series, and products]. Moscow: Nauka. 1108 p.

Received February 3, 2018

Contributor

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

ЗАШУМЛЕНИЕ ДАННЫХ КОНЕЧНЫМИ СМЕСЯМИ НОРМАЛЬНЫХ И ГАММА-РАСПРЕДЕЛЕНИЙ С ПРИМЕНЕНИЕМ К ЗАДАЧЕ ОКРУГЛЕНИЯ НАБЛЮДЕНИЙ*

А. К. Горшенин¹

Аннотация: Во многих реальных задачах проводится статистический анализ данных, содержащих дополнительные ошибки измерения, в том числе в виде округления, что в ряде ситуаций может приводить к достаточно существенным искажениям. В настоящей статье для одной из возможных моделей округления получены оценки для неизвестного математического ожидания наблюдений в предположении, что исходные данные дополнительно зашумлены с помощью случайных величин, имеющих распределения типа конечных смесей нормальных и гамма-законов. Построены доверительные интервалы для неизвестного математического ожидания с использованием уточненной оценки для дисперсии целой части случайной величины. Обсуждается алгоритм определения значения параметра для искусственного шума, добавление которого к исходным данным способствует повышению качества работы метода скользящего разделения смесей.

Ключевые слова: зашумленные данные; округленные наблюдения; конечные смеси нормальных распределений; конечные смеси гамма-распределений; доверительные интервалы; метод скользящего разделения смесей

DOI: 10.14357/19922264180304

1 Введение

Во многих реальных задачах данные, являющиеся непрерывными по своей сути, регистрируются с помощью инструментов, вносящих дополнительные ошибки измерения, в том числе в виде округления. Таким образом, статистический анализ проводится не для исходных, а для преобразованных некоторым случайным образом наблюдений, что в ряде ситуаций может приводить к достаточно существенным искажениям.

Для преодоления данной проблемы развивались различные подходы, в том числе на основе смешанных моделей (см., например, статью [1], в которой различные компоненты используются для представления уровней округления). В работе [2] приводятся результаты для моделей авторегрессии и скользящего среднего для округленных данных, а в статье [3] эти результаты развиваются и исследуются их асимптотические свойства. В статье [4] исследован метод оценивания параметров конечных смесей вероятностных распределений (в том числе, и многомерных) на основе использования EM (expectation-maximization) алгоритма [5] с целью получения состоятельных и асимптотически нормальных оценок.

В настоящей статье развиваются результаты для моделей округления, описанных в работах [6–8].

В их рамках будут получены оценки для неизвестного математического ожидания наблюдений в предположении, что исходные данные зашумлены с помощью случайных величин, имеющих распределения типа конечных смесей нормальных и гамма-законов. Это позволяет учесть большее количество случайных факторов, влияющих на величину «дополнительной» ошибки. Также будут построены доверительные интервалы для неизвестного математического ожидания. Выражения для гамма-распределений получены впервые. Также обсуждается алгоритм определения значения параметра для искусственного шума, добавление которого к исходным данным способствует повышению качества работы метода скользящего разделения смесей [9].

2 Предположения и базовые отношения

Для сокращения формулировок теорем в следующих разделах сделаем ряд предположений, на которые будем ссылаться в дальнейшем. Итак, пусть:

(A) X_1, X_2, \dots — независимые одинаково распределенные случайные величины с неизвестным математическим ожиданием $E_X < +\infty$;

*Работа выполнена при поддержке РФФ (проект 18-71-00156).

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

- (B) $\varepsilon_1, \varepsilon_2, \dots$ — независимые одинаково распределенные случайные величины с математическим ожиданием $E_\varepsilon < +\infty$;
- (C) X_1, X_2, \dots и $\varepsilon_1, \varepsilon_2, \dots$ являются независимыми;
- (D) $Y_j = [X_j + \varepsilon_j + 1/2]$ для всех $j = 1, 2, \dots$ представляют собой округление значения суммы случайных величин $X_j + \varepsilon_j$ до ближайшего целого сверху (при этом запись $[\cdot]$ соответствует целой части выражения).

В рамках данных предположений в статье будут рассмотрены вопросы качества приближения неизвестного математического ожидания E_X для исходных данных в ситуации, когда наблюдения для анализа получены с аддитивной ошибкой с известными распределениями (см. предположение (B)) и дополнительно округляются до ближайшего целого (см. предположение (D)).

Заметим, что в силу усиленного закона больших чисел справедливы следующие выражения:

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n Y_j &\xrightarrow[n \rightarrow \infty]{\text{п.н.}} E_Y \equiv \mathbb{E} \left[X_1 + \varepsilon_1 + \frac{1}{2} \right] = \\ &= \mathbb{E} \left(X_j + \varepsilon_j + \frac{1}{2} \right) - \mathbb{E} \left\{ X_j + \varepsilon_j + \frac{1}{2} \right\} = \\ &= E_X + E_\varepsilon + \frac{1}{2} - \mathbb{E} \left\{ X_j + \varepsilon_j + \frac{1}{2} \right\}. \end{aligned} \quad (1)$$

Запись $\{\cdot\}$ в формуле (1) соответствует дробной части выражения, а п.н. обозначает сходимость в смысле почти наверное.

Для доказательства результатов в дальнейшем потребуются следующие представления для дробной части абсолютно непрерывной случайной величины Z с абсолютно интегрируемой характеристической функцией $\varphi_Z(t)$ (см., например, Лемму 4 в работе [8]):

$$\mathbb{E}\{Z\} = \frac{1}{2} - \sum_{n=1}^{\infty} \frac{\text{Im}(\varphi_Z(2\pi n))}{\pi n}. \quad (2)$$

Через $\text{Im}(\cdot)$ в формуле (2) обозначена мнимая часть соответствующей функции.

При построении доверительных интервалов в дальнейшем будет использована следующая оценка, справедливая для любой случайной величины Z :

$$\mathbb{D}[Z] \leq \left(\sqrt{\mathbb{D}Z} + \frac{1}{2} \right)^2. \quad (3)$$

Она может быть проверена непосредственно с учетом представления $\mathbb{D}[Z] = \mathbb{D}(Z - \{Z\})$, неравенства Коши–Буняковского для ковариации и соотношения $\mathbb{D}\{Z\} \leq 1/4$, справедливого для любой

случайной величины Z (см., например, статью [8]). Отметим, что данная оценка является более точной по сравнению с использованным для аналогичных целей в работе [8] соотношением $\mathbb{D}[Z] \leq 2\mathbb{D}Z + 1/2$. Действительно,

$$2\mathbb{D}Z + \frac{1}{2} - \left(\sqrt{\mathbb{D}Z} + \frac{1}{2} \right)^2 = \left(\sqrt{\mathbb{D}Z} - \frac{1}{2} \right)^2 \geq 0,$$

причем для всех $\sqrt{\mathbb{D}Z} \neq 1/2$ данное неравенство является строгим.

3 Конечные смеси нормальных законов

Для случайной величины X , имеющей распределение типа конечной смеси нормальных законов [5] с параметрами $\mathbf{a} = (a_1, \dots, a_k)$, $a_j \in \mathbb{R}$, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$, $\sigma_j > 0$, $\mathbf{p} = (p_1, \dots, p_k)$, $p_j \geq 0$, $\sum_{j=1}^k p_j = 1$, плотность которого задается выражением

$$f_X(x) = \sum_{j=1}^k \frac{p_j}{\sigma_j \sqrt{2\pi}} e^{-(x-a_j)^2/(2\sigma_j^2)}, \quad (4)$$

характеристическая функция имеет вид:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx = \sum_{j=1}^k p_j e^{ita_j - \sigma_j^2 t^2/2}. \quad (5)$$

Абсолютная интегрируемость $\varphi_X(t)$ вытекает из свойств характеристической функции нормального распределения. Заметим, что в точке $t = 2\pi n$ выражение (5) принимает следующий вид:

$$\varphi_X(2\pi n) = \sum_{j=1}^k p_j e^{-2\pi^2 \sigma_j^2 n^2}. \quad (6)$$

Рассмотрим вопрос точности оценивания неизвестного математического ожидания E_X при добавлении зашумления.

Теорема 1. Пусть выполнены предположения (A)–(D), причем случайные величины ε_j , $j = 1, 2, \dots$, имеют распределение типа конечной k -компонентной смеси нормальных законов вида (4) с параметрами \mathbf{a} , $\boldsymbol{\sigma}$ и \mathbf{p} . Тогда

$$|E_Y - E_X| \leq A + \frac{1}{\pi} \left(1 + \frac{1}{4\pi^2 \sigma^2} \right) e^{-2\pi^2 \sigma^2}, \quad (7)$$

где $A = \max(|a_1|, \dots, |a_k|)$, $\sigma = \min(\sigma_1, \dots, \sigma_k)$.

Доказательство. С учетом представлений (1), (2) и (6), ограниченности модуля характеристической функции, а также независимости случайных величин X_j и ε_j имеем:

$$\begin{aligned} |E_Y - E_X| &= \left| E_\varepsilon + \frac{1}{2} - \mathbb{E} \left\{ X_j + \varepsilon_j + \frac{1}{2} \right\} \right| = \\ &= \left| E_\varepsilon + \sum_{n=1}^{\infty} \frac{\text{Im}(\varphi_{X_j}(2\pi n)\varphi_{\varepsilon_j}(2\pi n)\varphi_{1/2}(2\pi n))}{\pi n} \right| = \\ &= \left| E_\varepsilon + \sum_{n=1}^{\infty} \frac{\text{Im}(\varphi_{X_j}(2\pi n) \sum_{j=1}^k p_j e^{-2\pi^2 \sigma_j^2 n^2} e^{\pi n})}{\pi n} \right| = \\ &= \left| E_\varepsilon + \sum_{n=1}^{\infty} \frac{(-1)^n \sum_{j=1}^k p_j e^{-2\pi^2 \sigma_j^2 n^2} \text{Im}(\varphi_{X_j}(2\pi n))}{\pi n} \right| \leq \\ &\leq |E_\varepsilon| + \left| \sum_{j=1}^k p_j \sum_{n=1}^{\infty} \frac{1}{\pi n} e^{-2\pi^2 \sigma_j^2 n^2} \right| \leq \\ &\leq \max(|a_1|, \dots, |a_k|) + \\ &+ \sum_{j=1}^k \frac{p_j}{\pi} \left(1 + \frac{1}{4\pi^2 \sigma_j^2} \right) e^{-2\pi^2 \sigma_j^2} \leq \\ &\leq A + \frac{1}{\pi} \left(1 + \frac{1}{4\pi^2 \sigma^2} \right) e^{-2\pi^2 \sigma^2}. \end{aligned}$$

Справедливость использованной оценки

$$\sum_{n=1}^{\infty} \frac{e^{-2\pi^2 \sigma_j^2 n^2}}{n} \leq \left(1 + \frac{1}{4\pi^2 \sigma_j^2} \right) e^{-2\pi^2 \sigma_j^2}$$

может быть проверена непосредственно (например, см. доказательство Теоремы 6 в статье [8]). \square

Замечание 1. В случае, если зашумление производится нормально распределенными случайными величинами с нулевыми средними (т.е. в формуле (7) необходимо считать $A = 0, k = 1$), то Теорема 1 совпадает с результатом, полученным в работе [8].

Рассмотрим вопросы построения доверительного интервала для неизвестного математического ожидания E_X в предположении, что случайные величины X_j не содержат ошибок измерения, а все

погрешности учтены исключительно в зашумляющих элементах ε_j .

Теорема 2. Пусть выполнены предположения (A)–(D), причем случайные величины $\varepsilon_j, j = 1, 2, \dots$, имеют распределение типа конечной k -компонентной смеси нормальных законов вида (4) с параметрами $\mathbf{a}, \boldsymbol{\sigma}$ и \mathbf{p} , а случайные величины $X_j \stackrel{\text{н.н.}}{=} E_X, j = 1, 2, \dots$. Тогда доверительный интервал для E_X при условии $0 < \alpha < 1$ имеет вид:

$$\hat{E}_X - f(\mathbf{a}, \boldsymbol{\sigma}, \alpha, n) \leq E_X \leq \hat{E}_X + f(\mathbf{a}, \boldsymbol{\sigma}, \alpha, n), \quad (8)$$

где

$$\hat{E}_X = \frac{1}{n} \sum_{j=1}^n Y_j; \quad (9)$$

$$\begin{aligned} f(\mathbf{a}, \boldsymbol{\sigma}, \alpha, n) &= \frac{z_{1-\alpha/2}}{\sqrt{n}} \left(\sqrt{A^2 + \Sigma^2} + \frac{1}{2} \right) + \\ &+ A + \frac{1}{\pi} \left(1 + \frac{1}{4\pi^2 \sigma^2} \right) e^{-2\pi^2 \sigma^2}; \quad (10) \end{aligned}$$

$z_{1-\alpha/2}$ — $(1 - \alpha/2)$ -квантиль стандартного нормального распределения; $A = \max(|a_1|, \dots, |a_k|)$; $\Sigma = \max(\sigma_1, \dots, \sigma_k)$; $\sigma = \min(\sigma_1, \dots, \sigma_k)$.

Доказательство. Из центральной предельной теоремы с учетом условия (A) следует, что величина \hat{E}_X (9) асимптотически нормальна с математическим ожиданием

$$E_Y \equiv \mathbb{E} \left[E_X + \varepsilon_1 + \frac{1}{2} \right] \quad (11)$$

и дисперсией

$$\frac{1}{n} D_Y \equiv \frac{1}{n} \mathbb{D} \left[E_X + \varepsilon_1 + \frac{1}{2} \right]. \quad (12)$$

Воспользовавшись оценкой (3), получим:

$$\begin{aligned} D_Y &\leq \left(\sqrt{\mathbb{D} \left(E_X + \varepsilon_1 + \frac{1}{2} \right)} + \frac{1}{2} \right)^2 = \\ &= \left(\sqrt{\mathbb{D}\varepsilon_1} + \frac{1}{2} \right)^2 = \\ &= \left(\sqrt{\sum_{j=1}^k p_j \left(\left(a_j - \sum_{t=1}^k p_t a_t \right)^2 + \sigma_j^2 \right)} + \frac{1}{2} \right)^2 \leq \\ &\leq \left(\sqrt{A^2 + \Sigma^2} + \frac{1}{2} \right)^2. \end{aligned}$$

Тогда доверительный интервал уровня $1 - \alpha$ для математического ожидания E_Y имеет вид:

$$\mathbb{P} \left(\left| \hat{E}_X - E_Y \right| \leq \frac{z_{1-\alpha/2}}{\sqrt{n}} \left(\sqrt{A^2 + \Sigma^2} + \frac{1}{2} \right) \right) \geq 1 - \alpha.$$

Откуда следует справедливость соотношения (8) с учетом очевидного неравенства

$$\left| \hat{E}_X - E_X \right| \leq \left| \hat{E}_X - E_Y \right| + |E_Y - E_X|$$

и оценки (7) из Теоремы 1. \square

Замечание 2. В работе [9] было продемонстрировано повышение точности работы метода скользящего разделения конечных нормальных смесей за счет введения дополнительной компоненты, имеющей нормальное распределение $\mathcal{N}(0, \sigma^2)$ с математическим ожиданием, равным 0, и стандартным отклонением σ . При этом была отмечена сложность выбора параметра σ для сохранения структуры выборки, близкой к исходной. Результат Теоремы 2 может быть использован с данной целью, если положить $k = 1$, $a_j = 0$ для всех $j = 1, 2, \dots$ и выбирать величину σ как минимизирующую длину доверительного интервала (8). Для этого необходимо найти производную функции $f(0, \sigma, \alpha, n)$ (10) и численно решить уравнение

$$f'_\sigma(0, \sigma, \alpha, n) \equiv \frac{z_{1-\alpha/2}}{\sqrt{n}} - e^{-2\pi^2\sigma^2} \left(4\pi\sigma + \frac{1}{2\pi^3\sigma^3} + \frac{1}{\pi\sigma} \right) = 0 \quad (13)$$

относительно неизвестного параметра σ при выбранных значениях величин n и α . В качестве альтернативы можно использовать вид доверительного интервала из статьи [8], полученный с помощью неравенства $\mathbb{D}[Z] \leq 2\mathbb{D}Z + 1/2$, и искать решение уравнения вида:

$$\frac{2\sigma z_{1-\alpha/2}}{\sqrt{n(2\sigma^2 + 1/2)}} - e^{-2\pi^2\sigma^2} \left(4\pi\sigma + \frac{1}{2\pi^3\sigma^3} + \frac{1}{\pi\sigma} \right) = 0. \quad (14)$$

Примеры найденных значений σ для типичных размеров выборок в методе скользящего разделения смесей (учитываются как возможная ширина окна, так и общее количество наблюдений в анализируемом ряде) приведены в таблице (использован метод оптимизации Trust-Region Dogleg пакета MATLAB с настройками по умолчанию), в которой через σ_1 обозначено приближенное решение уравнения (13), а через σ_2 — уравнения (14).

Численные решения уравнений (13) и (14) относительно параметра σ для некоторых значений n и α

Размер выборки n	Уровень $\alpha = 0,1$		Уровень $\alpha = 0,05$		Уровень $\alpha = 0,01$	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
100	0,4302	0,435	0,419	0,425	0,4002	0,408
200	0,452	0,455	0,441	0,445	0,424	0,429
1000	0,499	0,499	0,489	0,489	0,473	0,475
10000	0,558	0,556	0,549	0,547	0,536	0,534
100000	0,611	0,607	0,603	0,599	0,591	0,588

Проверка практической эффективности данного подхода в качестве критерия выбора параметров зашумляющего распределения для повышения точности работы метода скользящего разделения смесей может быть отмечена как задача для дальнейших исследований.

4 Конечные смеси гамма-распределений

Для случайной величины X , имеющей распределение типа конечной смеси гамма-распределений с параметрами $\mathbf{r} = (r_1, \dots, r_k)$, $r_j > 0$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$, $\lambda_j > 0$, $\mathbf{p} = (p_1, \dots, p_k)$, $p_j \geq 0$, $\sum_{j=1}^k p_j = 1$, плотность которого задается выражением

$$f_X(x) = \sum_{j=1}^k p_j \frac{\lambda_j^{r_j} e^{-\lambda_j x}}{\Gamma(r_j)} x^{r_j-1}, \quad (15)$$

характеристическая функция имеет следующий вид:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx = \sum_{j=1}^k p_j \left(1 - \frac{it}{\lambda_j} \right)^{-r_j}. \quad (16)$$

Отметим, что подобные модели зашумления разумно использовать в случае, если известно, что данные сосредоточены на положительной полуоси, например при анализе различных информационных потоков (см., в частности, работу [10]).

Проверим абсолютную интегрируемость функции $\varphi_X(t)$ (16). Имеем:

$$\begin{aligned} \int_{-\infty}^{+\infty} |\varphi_X(t)| dt &\leq \sum_{j=1}^k p_j \int_{-\infty}^{+\infty} \left| \left(1 - \frac{it}{\lambda_j} \right)^{-r_j} \right| dt = \\ &= \sum_{j=1}^k p_j \int_{-\infty}^{+\infty} \left| \left(\frac{\lambda_j(\lambda_j + it)}{\lambda_j^2 + t^2} \right)^{r_j} \right| dt \leq \\ &\leq \sum_{j=1}^k p_j \lambda_j \int_{-\infty}^{+\infty} (1 + y^2)^{-r_j/2} dy. \end{aligned}$$

Подынтегральное выражение при $r_j \geq 2$ может быть оценено сверху функцией $1/(1+y^2)$, при этом соответствующий интеграл равен π , что влечет абсолютную интегрируемость характеристической функции для конечной смеси гамма-распределений. Поэтому в дальнейшем будем предполагать, что $r_j \geq 2$ для всех возможных значений $j = 1, 2, \dots$.

Рассмотрим вопрос точности оценивания неизвестного математического ожидания $E_X > 0$ при добавлении зашумления.

Теорема 3. Пусть выполнены предположения (А)–(D), причем случайные величины $\varepsilon_j, j = 1, 2, \dots$, имеют распределение типа конечной k -компонентной смеси гамма-распределений вида (15) с параметрами \mathbf{r}, λ и \mathbf{p} . Тогда

$$|E_Y - E_X| \leq \frac{R}{\lambda} + \frac{\Lambda^R}{2^r \pi^{r+1}} \left(1 + \frac{1}{r}\right), \quad (17)$$

где $r = \min(r_1, \dots, r_k); R = \max(r_1, \dots, r_k); \lambda = \max(\lambda_1, \dots, \lambda_k); \Lambda = \max(\lambda_1, \dots, \lambda_k)$.

Доказательство. С учетом представлений (1) и (2), ограниченности модуля характеристической функции, перехода от тригонометрической к показательной записи комплексных чисел, а также независимости случайных величин X_j и ε_j имеем:

$$\begin{aligned} |E_Y - E_X| &\leq |E_\varepsilon| + \left| \sum_{n=1}^{\infty} \left((-1)^n \operatorname{Im} \left(\sum_{j=1}^k p_j \varphi_{X_j}(2\pi n) \left(1 - i \left(\frac{2\pi n}{\lambda_j} \right)^{-r_j} \right) / (\pi n) \right) \right) \right| = \\ &= |E_\varepsilon| + \left| \sum_{n=1}^{\infty} \left((-1)^n \operatorname{Im} \left(\sum_{j=1}^k p_j \left(1 + \frac{4\pi^2 n^2}{\lambda_j^2} \right)^{-r_j/2} \times \right. \right. \right. \\ &\quad \left. \left. \left. \times \varphi_{X_j}(2\pi n) e^{-ir_j \arctan(t/\lambda_j)} \right) / (\pi n) \right) \right| \leq \\ &\leq |E_\varepsilon| + \sum_{j=1}^k p_j \sum_{n=1}^{\infty} \frac{1}{\pi n} \left(1 + \frac{4\pi^2 n^2}{\lambda_j^2} \right)^{-r_j/2} \leq \\ &\leq \frac{R}{\lambda} + \sum_{j=1}^k p_j \sum_{n=1}^{\infty} \left(\frac{1}{\pi n} \frac{\lambda_j^{r_j}}{(2\pi)^{r_j} n^{r_j}} \right) \leq \\ &\leq \frac{R}{\lambda} + \sum_{j=1}^k p_j \frac{\lambda_j^{r_j}}{2^{r_j} \pi^{r_j+1}} \left(1 + \int_1^{\infty} \frac{1}{x^{r_j+1}} dx \right) \leq \\ &\leq \frac{R}{\lambda} + \frac{\Lambda^R}{2^r \pi^{r+1}} \left(1 + \frac{1}{r} \right). \end{aligned}$$

При переходе от суммы к интегралу используется факт убывания функции как переменной n (или x). \square

Замечание 3. Теорема 3 описывает соответствующий результат для гамма-распределенных зашумляющих случайных величин, если положить $k = 1$ в выражении (17). При этом, очевидно, $r \equiv R$ и $\lambda \equiv \Lambda$.

Рассмотрим вопросы построения доверительного интервала для неизвестного математического ожидания $E_X > 0$ в предположении, что случайные величины X_j не содержат ошибок измерения, а все погрешности учтены исключительно в зашумляющих элементах ε_j .

Теорема 4. Пусть выполнены предположения (А)–(D), причем случайные величины $\varepsilon_j, j = 1, 2, \dots$, имеют распределение типа конечной k -компонентной смеси гамма-распределений вида (15) с параметрами \mathbf{r}, λ и \mathbf{p} , а случайные величины $X_j \stackrel{\text{н.н.}}{\equiv} E_X, j = 1, 2, \dots$. Тогда доверительный интервал для E_X при условии $0 < \alpha < 1$ имеет вид:

$$|E_X - \hat{E}_X| \leq f(\mathbf{r}, \lambda, \alpha, n), \quad (18)$$

где

$$\hat{E}_X = \frac{1}{n} \sum_{j=1}^n Y_j; \quad (19)$$

$$f(\mathbf{r}, \lambda, \alpha, n) = \frac{z_{1-\alpha/2}}{\sqrt{n}} \left(\sqrt{\frac{R(R+1)}{\lambda^2} - \frac{r^2}{\Lambda^2} + \frac{1}{2}} \right) + \frac{R}{\lambda} + \frac{\Lambda^R}{2^r \pi^{r+1}} \left(1 + \frac{1}{r} \right);$$

$z_{1-\alpha/2}$ — $(1-\alpha/2)$ -квантиль стандартного нормального распределения; $r = \min(r_1, \dots, r_k); R = \max(r_1, \dots, r_k); \lambda = \max(\lambda_1, \dots, \lambda_k); \Lambda = \max(\lambda_1, \dots, \lambda_k)$.

Доказательство. Из центральной предельной теоремы с учетом условия (А) следует, что величина \hat{E}_X (19) асимптотически нормальна с математическим ожиданием E_Y (11) и дисперсией $(1/n)D_Y$ (12). Пользуясь определением и свойствами гамма-функции, а также оценкой (3) получим:

$$\begin{aligned} D_Y &\leq \left(\sqrt{\sum_{j=1}^k p_j \frac{\lambda_j^{r_j}}{\Gamma(r_j)} \int_0^{\infty} e^{\lambda_j x} x^{r_j+1} dx} + \frac{1}{2} \right)^2 = \\ &= \left(\sqrt{\sum_{j=1}^k p_j \frac{r_j(r_j+1)}{\lambda_j^2} - \left(\sum_{j=1}^k p_j \frac{r_j}{\lambda_j} \right)^2} + \frac{1}{2} \right)^2 \leq \\ &\leq \left(\sqrt{\frac{R(R+1)}{\lambda^2} - \frac{r^2}{\Lambda^2} + \frac{1}{2}} \right)^2. \end{aligned}$$

Аналогично доказательству Теоремы 2 с учетом оценки (17) отсюда следует справедливость соотношения (18). \square

5 Заключение

Итак, в работе получены оценки для математического ожидания наблюдений в предположении зашумления конечными смесями нормальных (Теорема 1) и гамма-распределений (Теорема 3). Построены доверительные интервалы для неизвестного математического ожидания в этих случаях с использованием уточненной оценки (3) (Теоремы 2 и 4 соответственно). Отметим, что соответствующие соотношения зависят только от «экстремальных» значений параметров смесей, но не от числа компонент и весов в распределении зашумляющих наблюдений. Замечание 2 предлагает подход, который может быть использован для определения неизвестного параметра искусственно добавляемого к исходным данным шума для улучшения качества работы метода скользящего разделения смесей.

Автор выражает признательность доктору физико-математических наук, профессору Виктору Юрьевичу Королеву за идею использования оценки дисперсии вида (3) и другие полезные обсуждения в рамках работы над данной статьей.

Литература

1. *Wright D. E., Bray I.* A mixture model for rounded data // *J. Roy. Stat. Soc. D Sta.*, 2003. Vol. 52. P. 3–13.
2. *Bai Z., Zheng S., Zhang B., Hu G.* Statistical analysis for rounded data // *J. Stat. Plan. Infer.*, 2009. Vol. 139. Iss. 8. P. 2526–2542.
3. *Zhang B., Liu T., Bai Z. D.* Analysis of rounded data from dependent sequences // *Ann. I. Stat. Math.*, 2010. Vol. 62. Iss. 6. P. 1143–1173.
4. *Zhao N., Bai Z.* Analysis of rounded data in mixture normal model // *Stat. Pap.*, 2012. Vol. 53. P. 895–914.
5. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011. 512 с.
6. *Ушаков В. Г., Ушаков Н. Г.* Об усреднении округленных данных // *Информатика и её применения*, 2015. Т. 9. Вып. 4. С. 106–109.
7. *Ушаков В. Г., Ушаков Н. Г.* Границы точности восстановления информации, теряемой при округлении результатов наблюдений // *Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика*, 2017. № 2. С. 26–30.
8. *Ushakov N. G., Ushakov V. G.* Statistical analysis of rounded data: Recovering of information lost due to rounding // *J. Korean Stat. Soc.*, 2017. Vol. 46. No. 3. P. 426–437.
9. *Gorshenin A. K., Korolev V. Yu.* A noising method for the identification of the stochastic structure of information flows // *Comm. Com. Inf. Sc.*, 2017. Vol. 678. P. 279–289.
10. *Gorshenin A., Korolev V.* Modelling of statistical fluctuations of information flows by mixtures of gamma distributions // *27th European Conference on Modelling and Simulation Proceedings*. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2013. P. 569–572.

Поступила в редакцию 03.08.18

DATA NOISING BY FINITE NORMAL AND GAMMA MIXTURES WITH APPLICATION TO THE PROBLEM OF ROUNDED OBSERVATIONS

A. K. Gorshenin

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: In many real problems, statistical analysis of data containing additional measurement errors, including rounding, is performed, which in some situations can lead to sufficiently significant distortions. In this paper, estimates for an unknown expectation of observations are obtained for one of the possible rounding models under the assumption that the original data are additionally noised with random variables having distributions of the type of finite mixtures of normal and gamma laws. Confidence intervals for an unknown expectation are constructed using the refined estimate for the variance of the integer part of the random variable. An algorithm for determining the value of the parameter of artificial noise, which can be added to the initial data to improve the quality of the method of moving separation of mixtures, is discussed.

Keywords: noisy data; rounded data; finite normal mixtures; finite gamma mixtures; confidence intervals; moving separation of mixtures

DOI: 10.14357/19922264180304

Acknowledgments

The research was supported by the Russian Science Foundation (project 18-71-00156).

References

1. Wright, D. E., and I. Bray. 2003. A mixture model for rounded data. *J. Roy. Stat. Soc. D Sta.* 52:3–13.
2. Bai, Z., S. Zheng, B. Zhang, and G. Hu. 2009. Statistical analysis for rounded data. *J. Stat. Plan. Inference* 139(8):2526–2542.
3. Zhang, B., T. Liu, and Z. D. Bai. 2010. Analysis of rounded data from dependent sequences. *Ann. I. Stat. Math.* 62(6):1143–1173.
4. Zhao, N., and Z. Bai. 2012. Analysis of rounded data in mixture normal model. *Stat. Pap.* 53:895–914.
5. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Moscow University Publishing House. 512 p.
6. Ushakov, V. G., and N. G. Ushakov. 2015. Ob usrednenii okruglennykh dannykh [On averaging of rounded data]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):106–109.
7. Ushakov, V. G., and N. G. Ushakov. 2017. Boundaries of the precision of restoring information lost after rounding the results from observations. *Moscow University Computational Math. Cybernetics* 41(2):76–80.
8. Ushakov, N. G., and V. G. Ushakov. 2017. Statistical analysis of rounded data: Recovering of information lost due to rounding. *J. Korean Stat. Soc.* 46(3):426–437.
9. Gorshenin, A. K., and V. Yu. Korolev. 2016. A noising method for the identification of the stochastic structure of information flows. *Comm. Com. Inf. Sc.* 678:279–289.
10. Gorshenin, A., and V. Korolev. 2013. Modelling of statistical fluctuations of information flows by mixtures of gamma distributions. *27th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 569–572.

Received August 3, 2018

Contributor

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; agorshenin@frccsc.ru

АНАЛИЗ РАЗРЕЗНЫХ ПОВРЕЖДЕНИЙ В МНОГОПОЛЮСНЫХ СЕТЯХ

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

Аннотация: Предложен метод оценки изменения функциональных возможностей многополюсной потоковой сетевой системы после повреждающего воздействия. Для каждой стоковой дуги вычисляется максимальный поток, не зависящий от величины потока по остальным стоковым дугам. Разрезным структурным повреждением считается удаление из сети всех дуг, образующих минимальный разрез, соответствующий максимальному потоку по некоторой стоковой дуге. Среди найденных структурных повреждений по введенному критерию выбираются критически опасные. Для каждой дуги, принадлежащей хотя бы одному разрезному структурному повреждению, проводится количественная оценка последствий ее разрушения. Описанный подход предлагается использовать при исследовании уязвимости территориально распределенных многопользовательских систем передачи одного вида продукта, имеющих сетевую структуру связей.

Ключевые слова: однопродуктовая потоковая сеть; структурная уязвимость сети; многополюсная потоковая модель

DOI: 10.14357/19922264180305

1 Введение

В [1] на примере многополюсной однопродуктовой модели изучались возможности передачи пользователям запрашиваемых величин потоков при полном разрушении отдельных элементов исходной сетевой системы. Последовательно определялись гарантированные оценки величин потоков, которые можно передать в сети после аварии. В [2] был описан метод вычисления векторов ущерба для пользователей сети при различных повреждающих воздействиях. Предлагалось одновременно анализировать уязвимость и живучесть сети, а также вычислять величины потоков, которые можно гарантированно доставить всем пользователям.

В настоящей работе в рамках того же формализма однопродуктового потока и многостокковой модели рассматриваются предельные функциональные возможности сети и их изменение после аварий. Для оценки функциональных характеристик системы для каждой стоковой дуги вычисляется максимальный поток, не зависящий от величины потока по остальным стоковым дугам, и определяется соответствующий ему минимальный разрез [3]. Для сети вводится понятие критически опасного повреждения как повреждения, в результате которого передача потока из источни-

ка хотя бы в одну стоковую вершину оказывается невозможной. Разрезным структурным повреждением считается удаление из сети всех дуг, соответствующих минимальному разрезу для некоторой стоковой вершины. Для каждой стоковой вершины выбирается свой минимальный разрез. После моделирования такого повреждения сети определяются значения максимального потока по всем стоковым дугам и подсчитывается число нулевых значений. На основании проведенных расчетов выявляются стоковые вершины, передача потоков для которых чаще других оказывается в зоне критически опасных повреждений, вычисляются показатели, позволяющие оценить уязвимость системы в случае разрезных повреждений, строится таблица сравнительного «вклада» в уязвимость сети каждой из ее дуг, относящихся к множеству образующих критические разрезы.

2 Многополюсная потоковая модель

Сеть передачи единственного вида продукта в многополюсной системе будем описывать ориентированным графом $\bar{G} = \langle \bar{V}, \bar{L} \rangle$ без петель, который определяется множеством вершин (узлов)

¹Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, malash09@ccas.ru

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

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

$\bar{\mathcal{V}} = \{v_1, v_2, \dots, v_N\}$, где $|\bar{\mathcal{V}}| = N$, \mathcal{N} — множество индексов вершин, и множеством направленных дуг $\bar{\mathcal{L}} = \{l_{ij} \mid i \in \mathcal{N}, j \in \mathcal{N}, i \neq j\}$, соединяющих вершины. Здесь $l_{ij} = (v_i, v_j)$ — дуга, ведущая из вершины v_i в вершину v_j , $|\bar{\mathcal{L}}| = L$.

Обозначим через $\mathcal{V}_{\mathfrak{S}}$ и $\mathcal{V}_{\mathfrak{R}}$ множества вершин графа $\bar{\mathcal{G}}$, являющихся соответственно источниками и стоками для потока, который передается по многополюсной сети; $\mathcal{N}_{\mathfrak{S}}$ и $\mathcal{N}_{\mathfrak{R}}$ — множества индексов вершин-источников и вершин-стоков: $\mathcal{V}_{\mathfrak{S}} = \{v_i \mid i \in \mathcal{N}_{\mathfrak{S}}\}$, $|\mathcal{V}_{\mathfrak{S}}| = S$, $\mathcal{V}_{\mathfrak{S}} \subset \bar{\mathcal{V}}$, $\mathcal{N}_{\mathfrak{S}} \subset \mathcal{N}$, $S \geq 1$, $\mathcal{V}_{\mathfrak{R}} = \{v_i \mid i \in \mathcal{N}_{\mathfrak{R}}\}$, $|\mathcal{V}_{\mathfrak{R}}| = K$, $\mathcal{V}_{\mathfrak{R}} \subset \bar{\mathcal{V}}$, $\mathcal{N}_{\mathfrak{R}} \subset \mathcal{N}$, $K \geq 1$, $\mathcal{V}_{\mathfrak{S}} \cap \mathcal{V}_{\mathfrak{R}} = \emptyset$.

Считается, что на дугах графа $\bar{\mathcal{G}}$ заданы веса — значения d_{ij} пропускной способности дуг l_{ij} . Вектор d определяет максимально допустимую величину потока по дугам, $d = \{d_{ij} \mid d_{ij} \geq 0, l_{ij} \in \bar{\mathcal{L}}\}$. К графу $\bar{\mathcal{G}}$ добавим:

v_0 — единственный источник потока бесконечной мощности и дуги (v_0, v_j) , $j \in \mathcal{N}_{\mathfrak{S}}$, соединяющие v_0 с каждым узлом-источником. Для каждой дуги l_{0j} формально определим верхнее ограничение d_{0j} , которое соответствует величине максимального потока из узла-источника v_j в систему. Будем считать, что дуги (v_0, v_j) , $j \in \mathcal{N}_{\mathfrak{S}}$, являются дугами-источниками, и обозначим их множество через $\hat{\mathcal{L}} = \{l_{0j} \mid j \in \mathcal{N}_{\mathfrak{S}}\}$, $|\hat{\mathcal{L}}| = S$;

v_{N+1} — единственный узел-сток бесконечного объема и дуги (v_i, v_{N+1}) , $i \in \mathcal{N}_{\mathfrak{R}}$, соединяющие каждый узел-сток с v_{N+1} . Назовем направленные дуги (v_i, v_{N+1}) , $i \in \mathcal{N}_{\mathfrak{R}}$, дугами-стоками, или стоковыми дугами, и обозначим их множество через $\tilde{\mathcal{L}} = \{l_{iN+1} \mid i \in \mathcal{N}_{\mathfrak{R}}\}$, $|\tilde{\mathcal{L}}| = K$. Для каждой дуги l_{iN+1} формально определим верхнее ограничение d_{iN+1} , которое соответствует величине максимального потока из системы в v_i . Значение d_{iN+1} задает верхний предел для величины потока, покидающего систему через узел v_i . Ориентированный граф, который определяется множествами вершин $\mathcal{V} = \bar{\mathcal{V}} \cup \{v_0, v_{N+1}\}$ и дуг $\mathcal{L} = \bar{\mathcal{L}} \cup \hat{\mathcal{L}} \cup \tilde{\mathcal{L}}$, 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, $\mathcal{N}_j^+ \subset \mathcal{N} \cup \{N+1\}$. Вектор потоков $x = \langle x_{0j}, \dots, x_{ij}, \dots, x_{i(N+1)} \rangle$ по дугам $l_{ij} \in \mathcal{L}$, где

$$i \in \mathcal{N} \cup \{0\}, j \in \mathcal{N} \cup \{N+1\}, i \neq j, l_{ij} \in \mathcal{L},$$

и если $i = 0$, то $j \neq N+1$, (1)

должен удовлетворять условию сохранения потока в транзитных узлах

$$\sum_{i \in \mathcal{N}_j^-} x_{ij} = \sum_{i \in \mathcal{N}_j^+} x_{ji}, j \in \mathcal{N}, (2)$$

и ограничению на пропускную способность соответствующих дуг

$$0 \leq x_{ij} \leq d_{ij}, l_{ij} \in \mathcal{L}, \text{ для } i, j \text{ выполняется (1)}. (3)$$

Множество допустимых потоков в сети $\mathcal{X} = \{x \mid \text{выполняются (1)–(3)}\}$.

При анализе функциональных возможностей системы будем рассматривать потоки по стоковым дугам. Перенумеруем стоковые дуги по некоторому правилу натуральными числами от 1 до K , т.е. установим взаимно однозначное соответствие $l_k = l_{j(N+1)}$, $k = \overline{1, K}$, $j \in \mathcal{N}_{\mathfrak{R}}$. Пусть \bar{x}_k — величина потока по стоковой дуге l_k , $\bar{x}_k = x_{j(N+1)}$, $x_{j(N+1)} \geq 0$, $k = \overline{1, K}$, $j \in \mathcal{N}_{\mathfrak{R}}$. Таким образом, вектор $\bar{x} = \langle \bar{x}_1, \dots, \bar{x}_k, \dots, \bar{x}_K \rangle$, $k = \overline{1, K}$, покомпонентно определяет величины потока, которые передаются по каждой стоковой дуге сети в соответствии с некоторым допустимым потоком $x \in \mathcal{X}$. Обозначим через $\bar{\mathcal{X}}$ множество всех допустимых векторов \bar{x} :

$$\bar{\mathcal{X}} = \{\bar{x} \mid \bar{x}_k = x_{j(N+1)},$$

$k = \overline{1, K}, j \in \mathcal{N}_{\mathfrak{R}}, x \in \mathcal{X}\}. (4)$

3 Предельные функциональные характеристики

Рассмотрим множество $\bar{\mathcal{X}}$ векторов-потоков по стоковым дугам (4). Каждый допустимый вектор $\bar{x} \in \bar{\mathcal{X}}$ характеризует возможности системы по передаче потоков, покидающих сеть по всем стоковым дугам одновременно. Множество $\bar{\mathcal{X}}$ является выпуклым и многогранным. Точки, лежащие на границе множества $\bar{\mathcal{X}}$, будем называть предельными функциональными характеристиками системы (ПФ-характеристиками).

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 такой режим функционирования системы, при котором потоки по всем остальным стоковым дугам полагаются равными нулю.

Для анализа множества $\bar{\mathcal{X}}$ рассмотрим некоторую фиксированную стоковую дугу l_a , по которой

поток \bar{x}_a из вершины $v_a \in \mathcal{V}_{\mathcal{R}}$ покидает систему. Вычислим максимальную величину потока \bar{x}_a при работе системы в монопольном режиме, т. е. решим следующую задачу о максимальном потоке.

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

$$\bar{x}_a^* = \max_{\bar{x} \in \mathcal{X}} \bar{x}_a$$

при дополнительном условии $\bar{x}_i = 0$ для всех $i \neq a$, $i \in \mathcal{N}_{\mathcal{R}}$.

Оптимальное решение задачи 1 — максимально допустимый поток \bar{x}_a^* по дуге l_a из стоковой вершины $v_a \in \mathcal{V}_{\mathcal{R}}$ при передаче в монопольном режиме в вершину v_{N+1} . Значение \bar{x}_a^* для стоковой вершины v_a будем называть максимальным потоком, передаваемым в монопольном режиме, или МРМ-потоком. Вектор $\langle 0, 0, \dots, \bar{x}_a^*, 0, 0, \dots, 0 \rangle$ является угловой точкой множества \mathcal{X} и лежит на соответствующей координатной оси. Величина МРМ-потока \bar{x}_a^* определяет значение a -й ПФ-характеристики сети.

Последовательно решим задачу 1 и вычислим МРМ-потоки для каждой стоковой дуги l_k , $k = \overline{1, K}$. С помощью полученных решений сформируем вектор МРМ-потоков по стоковым дугам $\bar{x}^* = \langle \bar{x}_1^*, \dots, \bar{x}_k^*, \dots, \bar{x}_K^* \rangle$, значение k -й компоненты которого равно величине потока по стоковой дуге l_k в фиктивный узел-сток v_{N+1} при монопольном режиме, т. е. без учета потоков по другим стоковым дугам. Для вычисления предельных значений потоков при одновременной передаче их в сети решим следующую задачу.

Задача 2. Найти

$$\alpha^* = \max_{\alpha, \bar{x} \in \mathcal{X}} \alpha$$

при условиях $\alpha \bar{x}_k^* \leq \bar{x}_k$, $k = \overline{1, K}$, $\alpha \geq 0$.

Вектор потоков, отвечающий оптимальному решению $\alpha^* > 0$ задачи 2, обозначим через f^* , $f^* = \alpha^* \bar{x}^* = \langle \alpha^* \bar{x}_1^*, \dots, \alpha^* \bar{x}_j^*, \dots, \alpha^* \bar{x}_K^* \rangle$.

Точка с координатами $(\alpha^* \bar{x}_1^*, \dots, \alpha^* \bar{x}_j^*, \dots, \alpha^* \bar{x}_K^*)$ лежит на слейтеровской [4, 5] грани множества \mathcal{X} по критерию максимизации вектора потоков по всем стоковым дугам. В рамках формализма модели можно утверждать, что вектор f^* определяет передачу в равных долях максимальных потоков \bar{x}_j^* , достижимых только при монопольных режимах управления. Такое равнодолевое распределение обеспечивает допустимый поток при одновременной передаче из фиктивного источника бесконечной мощности $v_{\mathcal{S}}$ по всем стоковым дугам в фиктивный сток бесконечной емкости $v_{\mathcal{R}}$. Поскольку хотя бы для одной стоковой дуги компонента вектора f^* задает ее предельные функ-

циональные возможности при одновременной передаче потока в условиях равно долевого распределения, далее вектор f^* будем называть ПФР-характеристикой многополюсной потоковой сетевой системы (предельной функциональной равно долевой характеристикой).

4 Критически опасные повреждения сети

Пусть повреждение сети задается подмножеством w дуг исходной сети, пропускная способность которых становится равной нулю. Предполагается, что повреждены могут быть любые дуги сети, кроме стоковых. Положим $\mathcal{L}' = \overline{\mathcal{L}} \cup \hat{\mathcal{L}}$, тогда $w \subseteq \mathcal{L}'$. Обозначим через $d_{ij}(w)$ пропускную способность дуги l_{ij} поврежденной сети и положим

$$d_{ij}(w) = \begin{cases} 0, & \text{если } l_{ij} \in w; \\ d_{ij}, & \text{если } l_{ij} \in \mathcal{L} \setminus w. \end{cases} \quad (5)$$

Для допустимого после повреждения сети потока $x(w)$, где распределение потоков по дугам описывается компонентами вектора $x(w) = \langle x_{0j}(w), \dots, x_{ij}(w), \dots, x_{i(N+1)}(w) \rangle$, должно выполняться (1). Для любых допустимых потоков $x(w)$ в поврежденной сети выполняются стандартные ограничения

$$0 \leq x_{ij}(w) \leq d_{ij}(w), \quad \text{где } d_{ij}(w) \text{ определяется (5), } l_{ij} \in \mathcal{L}, \quad (6)$$

и закон сохранения потока в транзитных узлах

$$\sum_{i \in \mathcal{N}_j^-} x_{ij}(w) = \sum_{i \in \mathcal{N}_j^+} x_{ji}(w), \quad j \in \mathcal{N}. \quad (7)$$

Множество $\mathcal{X}(w)$ допустимых потоков в поврежденной сети определяется условиями (1), (6) и (7): $\mathcal{X}(w) = \{x(w) \mid \text{выполняется (1), (6), (7)}\}$.

Пусть $\bar{x}_k(w)$ — величина потока по k -й стоковой дуге после разрушающего воздействия, тогда $\bar{x}(w)$ — вектор величин потоков по всем стоковым дугам, $\bar{x}(w) = \langle \bar{x}_1(w), \dots, \bar{x}_k(w), \dots, \bar{x}_K(w) \rangle$. Аналогично (4) множество

$$\bar{\mathcal{X}}(w) = \{\bar{x}(w) \mid \bar{x}_k(w) = x_{j(N+1)}(w), \\ k = \overline{1, K}, j \in \mathcal{N}_{\mathcal{R}}, x(w) \in \mathcal{X}(w)\}$$

описывает достижимые потоки по стоковым дугам после повреждения.

Определение 2. Критически опасным структурным повреждением (КС-повреждением) w будем называть разрушающее воздействие на дуги сети, при

котором поток хотя бы к одной стоковой вершине оказывается равным нулю. Формально в случае КС-повреждения сети w у вектора потоков $\bar{x}(w)$ всегда найдется компонента, равная нулю, т. е.

$$\forall \bar{x}(w) \in \bar{\mathcal{X}}(w) \exists k : \bar{x}_k(w) = 0.$$

Рассмотрим влияние КС-повреждений на ПФ- и ПФР-характеристики многополюсной потоковой сетевой системы. Согласно классической теории о передаче однопродуктового потока [3] максимальному потоку в сети соответствует минимальный разрез, пропускная способность которого равна величине максимального потока. При разрушении дуг, определяющих минимальный разрез, передача потока в сети становится невозможной. Таким образом, по определению 2, все разрезы, соответствующие МРМ-потокам, образуются КС-повреждениями. Действительно, для того чтобы применить [3] в задаче 1, формально положим пропускную способность всех стоковых дуг, кроме l_a , равной нулю и решим задачу на максимум потока без дополнительного условия, имевшегося в задаче 1. Получим МРМ-поток для дуги l_a . Для МРМ-потока \bar{x}_a^* в этой задаче уже найдем соответствующее минимальному разрезу (далее — МРМ-разрез) подмножество $\mathcal{R}(a)$ дуг $l_{ij} \in \mathcal{L}'$, суммарная пропускная способность которых равна \bar{x}_a^* , и такое, что в случае повреждения $\mathcal{R}(a)$ при монопольном для l_a режиме функционирования передача потока из единственного источника v_0 в единственный узел-сток v_{N+1} через стоковую дугу l_a окажется невозможна, т. е.

$$\sum_{l_{ij} \in \mathcal{R}(a)} d_{ij} = \bar{x}_a^*, \quad \bar{x}_a = 0 \quad \forall x(\mathcal{R}(a)) \in \mathcal{X}(\mathcal{R}(a)).$$

Далее для сохранения простоты изложения будем считать, что каждому МРМ-потоку соответствует единственный МРМ-разрез и единственное КС-повреждение $\mathcal{R}(a)$. Все такие повреждения объединим в одно множество, т. е. введем $\bar{\mathcal{R}} = \{\mathcal{R}(a) \mid a \in \mathcal{N}_{\bar{\mathcal{R}}}\}$. Если МРМ-потоку \bar{x}_a^* соответствует несколько КС-повреждений, то любые (все) из них поместим в $\bar{\mathcal{R}}$. Повреждения, задаваемые МРМ-разрезами, будем называть разрезными. Каждое $\mathcal{R}(a) \in \bar{\mathcal{R}}$ является оптимальным в игровой задаче за условного противника, разрушающего сеть по критерию отделения одной из стоковых вершин (v_a) с минимальными затратами на уничтожение пропускной способности дуг. Когда же выбираемая «разрушителем» стоковая вершина не известна заранее (как в модели «оборона против нападения» [4]), можно ожидать любого повреждения из $\bar{\mathcal{R}}$. Поэтому из большого набора КС-повреждений для анализа уязвимости сети выберем

именно разрезные и получение оценок уязвимости проведем с помощью множества $\bar{\mathcal{R}}$.

Обозначим через R общее число разрезных КС-повреждений в множестве $\bar{\mathcal{R}}$. Для каждой стоковой дуги l_k вычислим долю случаев ρ_k , когда при повреждении того или иного множества дуг $\mathcal{R}(a)$, $a \in \mathcal{N}_{\bar{\mathcal{R}}}$, поток по выбранной стоковой дуге l_k оказывается равным нулю. Для этого число разрезных повреждений $m(l_k)$, «обнуляющих» поток по выбранной стоковой дуге l_k , разделим на общее число R повреждений в множестве $\bar{\mathcal{R}}$, т. е. положим

$$\rho_k = \frac{m(l_k)}{R}.$$

Передача потока через стоковые дуги l_k , для которых величина ρ_k больше чем у других, представляется более рискованной, так как пути в соответствующие стоковые вершины v_k чаще проходят в зоне критических повреждений сети. Следовательно, можно считать, что пользователи, стоящие за подобными стоковыми вершинами v_k , наиболее уязвимы в случае разрезных повреждений. Получили просто вычисляемый показатель относительной уязвимости стоковых вершин многополюсной сети.

5 Эффективные повреждения сети

Теперь предложим способ ранжирования разрезных КС-повреждений из $\bar{\mathcal{R}}$. Обозначим разрушение всех дуг некоторого $\mathcal{R}(a) \in \bar{\mathcal{R}}$ через w_a . Напомним, что в этом случае поток по стоковой дуге l_a оказывается равным нулю. Для оценки изменений относительно исходной ПФР-характеристики f^* решим на сети после повреждения $w_a = \mathcal{R}(a)$ следующую задачу.

Задача 3. Найти

$$\beta^*(w_a) = \min_{x(w_a) \in \mathcal{X}(w_a)} \sum_{k=1, K} \left(\frac{f_k^* - \bar{x}_k(w_a)}{f_k^*} \right)^2.$$

Вектор, отвечающий решению задачи 3, обозначим через $x^*(w_a)$. Вектору $x^*(w_a)$ соответствует вектор потоков по стоковым дугам $\bar{x}^*(w_a)$ при повреждении w_a : $\bar{x}^*(w_a) = \langle \bar{x}_1^*(w_a), \bar{x}_2^*(w_a), \dots, \bar{x}_K^*(w_a) \rangle$. Компоненты вектора $\bar{x}^*(w_a)$ определяют максимальные (но не выше исходных значений у ПФР-характеристики) потоки по стоковым дугам при условии, что пропускная способность дуг из выбранного множества $\mathcal{R}(a)$ равна нулю. Заметим, что a -я компонента вектора $\bar{x}^*(w_a)$ гарантированно равна нулю.

Для решения $\bar{x}^*(w_a)$ задачи 3 введем множество номеров стоковых дуг, величина потока по которым равна нулю, $\mathcal{K}^-(w_a) = \{k = \overline{1, K} \mid \bar{x}_k^*(w_a) = 0\}$, и обозначим через $\mathcal{K}^+(w_a)$ множество номеров стоковых дуг, величина потока по которым больше нуля, $\mathcal{K}^+(w_a) = \{k = \overline{1, K} \mid \bar{x}_k^*(w_a) > 0\}$. При этом $\mathcal{K}^-(w_a) \cup \mathcal{K}^+(w_a) = \hat{\mathcal{L}}$. Вычислим долю стоковых дуг с нулевым потоком и обозначим ее

$$\kappa(w_a) = \frac{|\mathcal{K}^-(w_a)|}{K}.$$

Найдем значение суммарных относительных потерь для сохранившихся потоков по стоковым дугам:

$$\Delta(w_a) = \sum_{k \in \mathcal{K}^+(w_a)} \frac{f_k^* - \bar{x}_k^*(w_a)}{f_k^*}.$$

Процедуру построения векторов $\bar{x}^*(w_k)$ и определения величин $\kappa(w_k)$ и $\Delta(w_k)$ проведем для каждого из КС-повреждений $\mathcal{R}(k)$, $k \in \mathcal{N}_{\bar{\mathcal{R}}}$, входящих в $\bar{\mathcal{R}}$. Выберем из $\bar{\mathcal{R}}$ подмножество $\bar{\mathcal{R}}^*$ разрезных повреждений, недоминируемых сразу по двум показателям (κ, Δ) :

$$\bar{\mathcal{R}}^* = \{\mathcal{R}(k) \in \bar{\mathcal{R}} \mid \forall a \in \mathcal{N}_{\bar{\mathcal{R}}} : \text{если } \kappa(w_a) > \kappa(w_k), \text{ то } \Delta(w_k) > \Delta(w_a)\}.$$

Критически опасные структурные повреждения из $\bar{\mathcal{R}}^*$ будем называть эффективными (с точки зрения разрушителя сети — по векторному критерию (κ, Δ) на максимум). Они задают конфигурации дуг сети, образующие ее «слабые места», защита которых требуется в первую очередь. Остальные элементы $\bar{\mathcal{R}}$ — доминируемые повреждения, но тоже могут быть полезны для целей аналитики.

Опираясь на полученные результаты, определим, полное разрушение каких дуг из КС-повреждений, входящих в множество $\bar{\mathcal{R}}$, оказывает наибольшее влияние на уязвимость сети. Для этого построим таблицу, организованную следующим образом. В первый столбец таблицы внесем дуги сети l_{ij} , которые могут быть подвержены разрушению, во второй — общее число разрезных повреждений $n(l_{ij})$, в которые входит дуга l_{ij} , в третий —

общее число эффективных разрезных повреждений $n^*(l_{ij})$ с дугой l_{ij} .

Чем больше значение $n(l_{ij})$, тем важнее сохранить дугу l_{ij} от разрушения, поскольку удаление l_{ij} снижает пропускную способность минимального разреза (и МРМ-потоки) для большего числа стоковых вершин. Заметим, что ПФР-характеристики КС-повреждений для указанных вершин не обязательно хуже, чем при удалении иной дуги. Для дуг с большими значениями как $n(l_{ij})$, так и $n^*(l_{ij})$ их влияние на уязвимость сети по отношению к разрезным повреждениям, по-видимому, еще существенней.

6 Заключение

Сценарии повреждения дуг, соответствующих минимальным разрезам, не являются единственно возможными. При исследовании сетевых систем приходится обращаться к теории графов и методам комбинаторной оптимизации [6, 7]. Проблема состоит в том, что попытки решения многих практических задач требуют применения ресурсоемких переборных методов анализа структуры сети с последующим специальным изучением функциональных потоковых характеристик [8, 9].

Предложенная в настоящей работе схема выделяет на дереве перебора содержательно важные точки ветвления — конфигурации дуг сети, образующие МРМ-разрезы. Подобным образом опытный шахматист или хорошая шахматная программа заранее тщательно рассматривает некоторые стандартные часто встречающиеся позиции, чтобы сэкономить время в реальной партии и/или лучше понять внутреннюю логику возможных последующих ходов в игре.

Разработанный подход опирается на методологию исследования операций [4]. Определение допустимых потоков через все стоковые дуги позволяет анализировать сложные взаимосвязи, свойственные большим территориально распределенным системам (см., например, [10]). Изучение разрезных повреждений помогает выявить наиболее уязвимые стоковые вершины. Следует подчеркнуть, что методы потокового программирования [3], положенные в основу анализа, дают способ быстро и эффективно проводить необходимые вычисления и получать значимые результаты.

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.

Формирование разрезных и эффективных повреждений

Дуги сети	Разрезные повреждения	Эффективные повреждения
l_{0j}	$n(l_{0j})$	$n^*(l_{0j})$
\vdots	\vdots	\vdots
l_{ij}	$n(l_{ij})$	$n^*(l_{ij})$
\vdots	\vdots	\vdots
l_{iN}	$n(l_{iN})$	$n^*(l_{iN})$

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. Форд Л., Фалкерсон Д. Потоки в сетях / Пер. с англ. — М.: Мир, 1966. 277 с. (Ford L. R., Fulkerson D. R. Flows in networks. — Princeton, NJ, USA: Princeton University Press, 1962. 332 p.)
4. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971. 384 с.
5. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — 2-е изд. — М.: Физматлит, 2007. 256 с.
6. Харари Ф. Теория графов / Пер. с англ. — М.: Мир, 1973. 302 с. (Harari F. Graph theory. — Boston, MA, USA: Addison-Wesley, 1969. 300 p.)
7. Сигал И. Х., Иванова А. П. Введение в прикладное дискретное программирование. — М.: Физматлит, 2002. 240 с.
8. Rosas-Casals M., Valverde S., Sole R. V. Topological vulnerability of the European power grid under errors and attacks // Int. J. Bifurcat. Chaos, 2007. Vol. 17. Iss. 7. P. 2465–2475.
9. Murray A. T. An overview of network vulnerability modeling approaches // GeoJournal, 2013. Vol. 78. P. 209–221.
10. 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. ТИСУ, 2017. № 6. С. 50–73.

Поступила в редакцию 28.06.18

ANALYSIS OF CUTTING DAMAGES TO MULTIPOLAR NETWORKS

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

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

Abstract: The method of estimating changes in the functional capabilities of a multipolar flow network system after a damage is proposed. For each sink arc, the maximal flow is calculated, independent of the flow value across the remaining sink arcs. The authors consider cutting structural damages that correspond to removing all arcs forming a minimal cut. The capacity of the cut is equal to the maximal flow along some sink arc. Among the structural damages, the critically dangerous ones are selected with an introduced criterion. For each arc belonging to at least one cutting structural damage, a quantitative characteristic is computed to estimate consequences of its destruction. The described approach is proposed to be used in studying vulnerability of territorially distributed multiuser systems with the network structure in the case of a single-product transfer.

Keywords: single-product flow network; structural vulnerability of network; multipolar flow model

DOI: 10.14357/19922264180305

References

1. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2017. Metod analiza funktsional'noy uyazvimosti potokovykh setevykh system [Method of the analysis of the functional vulnerability of flow network systems]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):50–73.
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. Ford, L. R., and D. R. Fulkerson. 1962. *Flows in networks*. Princeton, NJ: Princeton University Press. 332 p.
4. Germeier, Yu. B. 1971. *Vvedenie v teoriyu issledovaniya operatsiy* [An introduction to operations research theory]. Moscow: Nauka. 384 p.
5. Podinovskij, V. V., and V. D. Nogin. 2007. *Pareto-optimal'nye resheniya mnogokriterial'nykh zadach* [Pareto-optimal solutions of multicriteria tasks]. 2nd ed. Moscow: Fizmatlit. 256 p.
6. Harari, F. 1969. *Graph theory*. Boston, MA: Addison-Wesley Publ. 300 p.
7. Sigal, I. H., and A. P. Ivanova. 2002. *Vvedenie v prikladnoe diskretnoe programmirovaniye* [An introduction to applied discrete programming]. Moscow: Fizmatlit. 240 p.
8. Rosas-Casals, M., S. Valverde, and R. V. Sole. 2007. Topological vulnerability of the European power grid under errors and attacks. *Int. J. Bifurcat. Chaos* 17(7):2465–2475.
9. Murray, A. T. 2013. An overview of network vulnerability modeling approaches. *GeoJournal* 78:209–221.
10. 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.

Received June 28, 2018

Contributors

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

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

ОБ ИНВАРИАНТНОСТИ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ОГРАНИЧЕННЫМИ РЕСУРСАМИ И С ИНТЕНСИВНОСТЯМИ ПОСТУПЛЕНИЯ И ОБСЛУЖИВАНИЯ, ЗАВИСЯЩИМИ ОТ СОСТОЯНИЯ СИСТЕМЫ*

Э. С. Сопин¹, В. А. Наумов², К. Е. Самуйлов³

Аннотация: Рассматривается дальнейшее обобщение систем массового обслуживания (СМО), в которых заявкам для обслуживания требуется не только прибор, но и некоторый объем ресурсов, суммарный объем которых ограничен. В рассматриваемой системе интенсивности поступления и обслуживания заявок зависят от состояния системы, при этом объемы работ, приносимых заявками, имеют произвольное распределение с конечным средним. Сформулирована и доказана теорема о мультипликативном виде стационарного распределения вероятностей при пуассоновском входящем потоке, зависящем от числа заявок в системе. Показано, что стационарные вероятности числа заявок в системе и объемов занятых ими ресурсов, аналогично системе Эрланга с потерями, не зависят от вида функции распределения (ФР) объема работ, а зависят только от математического ожидания.

Ключевые слова: ограниченные ресурсы; система массового обслуживания; инвариантность; время обслуживания

DOI: 10.14357/19922264180306

1 Введение

Обобщение многолинейной модели Эрланга с потерями в виде СМО, в которых заявка на все время обслуживания занимает прибор и случайный объем ограниченных ресурсов, все чаще применяется для анализа современных вычислительных и телекоммуникационных систем. К примеру, при построении и анализе моделей перспективных беспроводных сетей основная сложность заключается в том, что объем используемых частотно-временных ресурсов одной пользовательской сессией зависит не только от типа услуги, но и массы других факторов, влияющих на распространение радиосигнала, таких как расстояние между мобильным устройством и базовой станцией, наличие дополнительных препятствий между ними, используемая кодово-модуляционная схема и др. Системы массового обслуживания с ограниченными ресурсами позволяют учесть все эти факторы и описывают функционирование таких систем с высокой степенью адекватности [1]. Первые результаты для систем с ограниченными ресурсами были получены

в работе [2]. Рассматриваемая в настоящей статье ресурсная СМО, в которой интенсивности поступления и обслуживания заявок зависят от состояния системы, была изучена в [3] в случае экспоненциального распределения времени обслуживания и пуассоновского входящего потока.

Впервые свойство инвариантности было доказано для системы Эрланга с потерями [4, 5]. В дальнейшем было показано, что свойство инвариантности справедливо для класса «симметричных» СМО, примерами которых служат однолинейная СМО с дисциплиной разделения процессора и бесконечно-линейная СМО. Для более полного обзора инвариантных систем читателю предлагается ознакомиться с работой [6].

Для СМО с ограниченными ресурсами и постоянными интенсивностями поступления и обслуживания заявок свойство инвариантности было доказано в [7]. В данной работе доказывается свойство инвариантности системы из [3] относительно вида распределения времени обслуживания. За основу доказательства свойства инвариантности более общей системы была взята идея доказательства из [6].

* Публикация подготовлена при финансовой поддержке Минобрнауки России (проект 2.882.2017/4.6).

¹ Российский университет дружбы народов, Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sopin_es@rudn.university

² Исследовательский институт инноваций, г. Хельсинки, Финляндия, valeriy.naumov@pfu.fi

³ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samouylov_ke@rudn.university

2 Модель системы в случае экспоненциального распределения времени обслуживания

Рассмотрим многолинейную СМО с $N < \infty$ приборами. Предположим, что поступающий поток является пуассоновским с параметром λ_k , зависящим от числа k заявок в системе, а объемы работ, которые необходимо выполнить для обслуживания заявок, независимы между собой и от поступающего потока и экспоненциально распределены с параметром $\mu = 1/b$. Предположим также, что каждый прибор обслуживает заявки с постоянной, зависящей от числа k заявок в системе скоростью σ_k . Система располагает ограниченным объемом ресурсов M типов и функционирует следующим образом:

1. Каждой находящейся в системе заявке требуется один прибор и некоторый объем ресурса каждого типа.
2. Поступившая заявка теряется, если в момент поступления объем требуемого ей ресурса превышает объем свободного ресурса этого типа либо в системе нет свободных приборов.
3. В момент поступления заявки объем свободного ресурса каждого типа уменьшается на величину ресурса, выделенного этой заявке.
4. В момент ухода заявки объем свободного ресурса каждого типа увеличивается на величину ресурса, выделенного этой заявке.

Обозначим R_m общий объем ресурса типа m и $\mathbf{r}_j = (r_{j1}, r_{j2}, \dots, r_{jM})$ — вектор объемов ресурсов, необходимых j -й поступившей заявке, $j \geq 1$. Будем считать, что случайные векторы \mathbf{r}_j не зависят от процесса поступления и обслуживания заявок, независимы в совокупности и одинаково распределены с ФР $F(\mathbf{x})$.

Будем считать, что поступившие заявки располагаются в очереди в порядке поступления и опишем состояние системы в момент t процессом $X(t) = (\xi(t), \gamma(t))$. Здесь $\xi(t)$ — число заявок в системе и $\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_{\xi(t)}(t))$, где $\gamma_i(t)$ — вектор объемов ресурсов, занимаемых i -й заявкой. Состояние системы может измениться только в моменты t_i , когда либо в систему поступает, либо ее покидает заявка.

Введем обозначения для стационарных вероятностей процесса $X(t)$:

$$p_0 = \lim_{t \rightarrow \infty} P(\xi(t) = 0); \quad (1)$$

$$p_k(\mathbf{r}_1, \dots, \mathbf{r}_k) = \lim_{t \rightarrow \infty} P(\xi(t) = k, \gamma_1(t) \leq \mathbf{r}_1, \dots, \gamma_k(t) \leq \mathbf{r}_k), \quad 1 \leq k \leq N. \quad (2)$$

Выражения для стационарного распределения (1), (2) имеют следующий вид:

$$p_0 = \left(1 + \sum_{k=1}^N F^{(k)}(\mathbf{R}) \frac{b^k}{k!} \prod_{i=1}^k \frac{\lambda_{i-1}}{\sigma_i} \right)^{-1};$$

$$p_k(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k) = p_0 F(\mathbf{r}_1) F(\mathbf{r}_2) \cdots F(\mathbf{r}_k) \frac{b^k}{k!} \prod_{i=1}^k \frac{\lambda_{i-1}}{\sigma_i},$$

$$1 \leq k \leq N, \mathbf{r}_1 \geq \mathbf{0}, \dots, \mathbf{r}_k \geq \mathbf{0}, \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R}, \quad (3)$$

где $F^{(k)}(\mathbf{x})$ есть k -кратная свертка ФР $F(\mathbf{x})$.

3 Свойство инвариантности относительно вида распределения времени обслуживания

Рассмотрим поведение системы в случае, когда распределение объемов работ не является экспоненциальным, а распределено в соответствии с ФР $B(x)$ с таким же математическим ожиданием b . Будем считать, что распределение времени обслуживания имеет плотность $b(x)$ на полуоси $[0, \infty)$, однако заметим, что опираясь на свойства слабой непрерывности, как в [8], полученные результаты можно доказать и в более общем случае. Состояние системы в момент t описывается случайным процессом $\tilde{X}(t) = (\xi(t), \gamma(t), \beta(t))$. Здесь, как и прежде, $\xi(t)$ — число заявок в системе, $\gamma(t)$ описывает объемы ресурсов, занимаемых каждой заявкой, а третья компонента $\beta(t) = (\beta_1(t), \beta_2(t), \dots, \beta_{\xi(t)}(t))$, где $\beta_i(t)$ — объем обслуженной работы i -й заявки. При поступлении заявки с вектором требований ресурсов \mathbf{r} система переходит из состояния $(k, (\mathbf{r}_1, \dots, \mathbf{r}_k), (x_1, \dots, x_k))$ в одно из состояний $(k+1, (\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}, \mathbf{r}_i, \dots, \mathbf{r}_k), (x_1, \dots, x_{i-1}, 0, x_i, \dots, x_k))$, $0 \leq i \leq k$, с вероятностью $1/(k+1)$. Обозначим

$$q_0(t) = P(\xi(t) = 0); \quad (4)$$

$$Q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t) = P(\xi(t) = k, \gamma_1(t) \leq \mathbf{r}_1, \dots, \gamma_k(t) \leq \mathbf{r}_k, \beta_1(t) < x_1, \dots, \beta_k(t) < x_k), \quad 1 \leq k \leq N, \quad (5)$$

вероятности состояний системы в момент времени t и $q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t)$ — плотности вероятностей $Q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t)$, $k > 0$.

Выпишем уравнения изменения k -мерных плотностей $q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t)$, в точке $\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k$ за малое время Δt . Для вероятностей пустой системы справедливо равенство:

$$q_0(t + \Delta t) = q_0(t) (1 - \lambda_0 F(\mathbf{R}) \Delta t) + \int_{0 \leq \mathbf{r} \leq \mathbf{R}} \int_0^\infty q_1(d\mathbf{r}, x; t) \frac{b(x) \sigma_1 \Delta t}{1 - B(x)} dx + o(\Delta t). \quad (6)$$

Первое слагаемое в правой части уравнения (6) соответствует случаю, при котором в интервале $(t, t + \Delta t)$ не было поступлений заявок, а второе — окончанию обслуживания заявки на данном интервале. Множитель $b(x) \sigma_1 \Delta t / (1 - B(x))$ имеет здесь смысл вероятности того, что на интервале $(t, t + \Delta t)$ заявка будет обслужена при условии, что объем работы, принесенной заявкой, превышает x . Выведем теперь уравнения для случая, когда в системе находится $0 < n < N$ заявок:

$$q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t + \Delta t) = \int_{\substack{0 \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1, 2, \dots, k}} q_k(d\mathbf{s}_1, \dots, d\mathbf{s}_k, x_1 - \sigma_k \Delta t, \dots, x_k - \sigma_k \Delta t; t) \times \left[1 - \left(\lambda_k F\left(\mathbf{R} - \sum_{i=1}^k \mathbf{s}_i\right) + \sum_{i=1}^k \frac{b(x_i) \sigma_k}{1 - B(x_i)} \right) \Delta t \right] + \sum_{j=1}^{k+1} \int_{\substack{0 \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1, 2, \dots, k, \\ \mathbf{s} \leq \mathbf{R} - \mathbf{s}_1 - \dots - \mathbf{s}_k}} \int_0^\infty q_{k+1}(d\mathbf{s}_1, \dots, d\mathbf{s}_{j-1}, d\mathbf{s}, d\mathbf{s}_j, \dots, d\mathbf{s}_k, x_1 - \sigma_{k+1} \Delta t, \dots, x_{j-1} - \sigma_{k+1} \Delta t, x, x_j - \sigma_{k+1} \Delta t, x_k - \sigma_{k+1} \Delta t; t) \frac{b(x) \sigma_{k+1} \Delta t}{1 - B(x)} dx + o(\Delta t), \quad \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R}, \quad 1 \leq k \leq N - 1. \quad (7)$$

Аналогично уравнению (6) первое слагаемое в правой части (7) соответствует случаю, в котором за интервал $(t, t + \Delta t)$ не происходит поступлений новых заявок и ни одна заявка не покидает систему, а второе — окончанию обслуживания одной заявки. Если в системе уже находится N заявок, то новых поступлений быть не может и невозможен переход из состояний с $N + 1$ заявками. Тогда уравнение упрощается:

$$q_N(\mathbf{r}_1, \mathbf{r}_N, x_1, \dots, x_N; t + \Delta t) = q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1 - \sigma_N \Delta t, \dots, x_N - \sigma_N \Delta t; t) \times \left[1 - \sum_{i=1}^N \frac{b(x_i) \sigma_k}{1 - B(x_i)} \Delta t \right] + o(\Delta t), \quad \sum_{i=1}^N \mathbf{r}_i \leq \mathbf{R}. \quad (8)$$

Поделив левые и правые части уравнений (6)–(8) на Δt и устремив Δt к нулю, получим:

$$\frac{\partial q_0(t)}{\partial t} + \lambda_0 q_0(t) F(\mathbf{R}) = \sigma_1 \int_{0 \leq \mathbf{r} \leq \mathbf{R}} \int_0^\infty q_1(d\mathbf{r}, x; t) \frac{b(x)}{1 - B(x)} dx; \quad (9)$$

$$\begin{aligned} & \frac{\partial q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t)}{\partial t} + \sigma_k \sum_{i=1}^k \frac{\partial q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k; t)}{\partial x_i} = \\ & = - \int_{\substack{0 \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1, 2, \dots, k}} q_k(d\mathbf{s}_1, \dots, d\mathbf{s}_k, x_1, \dots, x_k; t) \times \\ & \times \left(\lambda_k F\left(\mathbf{R} - \sum_{i=1}^k \mathbf{s}_i\right) + \sum_{i=1}^k \frac{b(x_i) \sigma_k}{1 - B(x_i)} \right) + \\ & + \sum_{j=1}^{k+1} \int_{\substack{0 \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1, 2, \dots, k, \\ \mathbf{s} \leq \mathbf{R} - \mathbf{s}_1 - \dots - \mathbf{s}_k}} \int_0^\infty q_{k+1}(d\mathbf{s}_1, \dots, d\mathbf{s}_{j-1}, d\mathbf{s}, d\mathbf{s}_j, \dots, \\ & \dots, d\mathbf{s}_k, x_1, \dots, x_{j-1}, x, x_j, x_k; t) \frac{b(x) \sigma_{k+1}}{1 - B(x)} dx, \\ & \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R}, \quad 1 \leq k \leq N - 1; \quad (10) \end{aligned}$$

$$\begin{aligned} & \frac{\partial q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1, \dots, x_N; t)}{\partial t} + \sigma_N \sum_{i=1}^N \frac{\partial q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1, \dots, x_N; t)}{\partial x_i} = \\ & = -q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1, \dots, x_N; t) \sum_{i=1}^N \frac{b(x_i) \sigma_N}{1 - B(x_i)}, \\ & \sum_{i=1}^N \mathbf{r}_i \leq \mathbf{R}. \quad (11) \end{aligned}$$

Для вывода граничных условий рассмотрим поведение системы в моменты поступления заявок:

$$\begin{aligned} & \int_0^{\sigma_k \Delta t} q_k(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}, \mathbf{r}_j, \dots, \mathbf{r}_k, x_1 + \sigma_k \Delta t, \dots, \\ & \dots, x_{i-1} + \sigma_k \Delta t, x, x_i + \sigma_k \Delta t, \dots, x_{k-1} + \sigma_k \Delta t; \\ & t + \Delta t) dx = \frac{\lambda_k \Delta t}{k} F(\mathbf{r}) \times \\ & \times q_{k-1}(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, x_1, \dots, x_{k-1}; t) + o(\Delta t), \\ & \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R} - \mathbf{r}, \quad 0 < k \leq N. \end{aligned}$$

Аналогично, поделив на Δt и устремив Δt к нулю, получим:

$$\begin{aligned} & \sigma_k q_k(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}, \mathbf{r}_i, \dots, \mathbf{r}_k, x_1, \dots, x_{i-1}, 0, x_i, \dots \\ & \quad \dots, x_{k-1}; t) = \\ & = \frac{\lambda_k}{k} F(\mathbf{r}) q_{k-1}(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, x_1, \dots, x_{k-1}; t), \\ & \quad \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R} - \mathbf{r}, \quad 0 < k \leq N. \quad (12) \end{aligned}$$

Чтобы получить уравнения для стационарного режима, приравняем к нулю производные функций (4)–(5) по времени t . В результате уравнения (9)–(11) и граничное условие (12) примут вид:

$$\begin{aligned} \lambda_0 q_0 F(\mathbf{R}) &= \sigma_1 \int_{0 \leq \mathbf{r} \leq \mathbf{R}} \int_0^\infty q_1(d\mathbf{r}, x) \frac{b(x)}{1-B(x)} dx; \quad (13) \\ \sigma_k \sum_{i=1}^k \frac{\partial q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k)}{\partial x_i} &= \\ &= - \int_{\substack{\mathbf{0} \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1,2,\dots,k}} q_k(d\mathbf{s}_1, \dots, d\mathbf{s}_k, x_1, \dots, x_k) \times \\ & \times \left(\lambda_k F\left(\mathbf{R} - \sum_{i=1}^k \mathbf{s}_i\right) + \sum_{i=1}^k \frac{b(x_i) \sigma_k}{1-B(x_i)} \right) + \\ & + \sum_{j=1}^{k+1} \int_{\substack{\mathbf{0} \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1,2,\dots,k, \\ \mathbf{s} \leq \mathbf{R} - \mathbf{s}_1 - \dots - \mathbf{s}_k}} \int_0^\infty q_{k+1}(d\mathbf{s}_1, \dots, d\mathbf{s}_{j-1}, d\mathbf{s}, d\mathbf{s}_j, \dots \\ & \dots, d\mathbf{s}_k, x_1, \dots, x_{j-1}, x, x_j, x_k) \frac{b(x) \sigma_{k+1}}{1-B(x)} dx, \\ & \quad \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R}, \quad 1 \leq k \leq N-1; \quad (14) \end{aligned}$$

$$\begin{aligned} \sigma_N \sum_{i=1}^N \frac{\partial q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1, \dots, x_N)}{\partial x_i} &= \\ &= -q_N(\mathbf{r}_1, \dots, \mathbf{r}_N, x_1, \dots, x_N) \sum_{i=1}^N \frac{b(x_i) \sigma_N}{1-B(x_i)}, \\ & \quad \sum_{i=1}^N \mathbf{r}_i \leq \mathbf{R}; \quad (15) \end{aligned}$$

$$\begin{aligned} \sigma_k q_k(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}, \mathbf{r}_i, \dots, \mathbf{r}_k, x_1, \dots, x_{i-1}, 0, x_i, \dots \\ \dots, x_{k-1}) &= \\ &= \frac{\lambda_k}{k} F(\mathbf{r}) q_{k-1}(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, x_1, \dots, x_{k-1}), \\ & \quad \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R} - \mathbf{r}, \quad 0 < k \leq N. \quad (16) \end{aligned}$$

Решение системы (13)–(16) задается выражениями:

$$\begin{aligned} q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k) &= \\ &= q_0 F(\mathbf{x}_1) \dots F(\mathbf{x}_k) \frac{1}{k!} \prod_{i=1}^k \frac{\lambda_{i-1}}{\sigma_i} (1-B(x_i)), \\ & \quad 0 \leq k \leq N, \quad \mathbf{x}_1 \geq \mathbf{0}, \dots, \mathbf{x}_k \geq \mathbf{0}, \quad \sum_{i=1}^k \mathbf{x}_i \leq \mathbf{R}. \quad (17) \end{aligned}$$

Это следует из того, что для любой непрерывной функции $B(x)$

$$\frac{d(1-B(x))}{dx} = -b(x),$$

и тогда каждое i -е слагаемое в левой части уравнений (14) и (15) равно i -му слагаемому в первой сумме правой части. Кроме того, так как

$$\int_0^\infty b(x) dx = 1,$$

то

$$\begin{aligned} \sum_{j=1}^{k+1} \int_{\substack{\mathbf{0} \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1,2,\dots,k, \\ \mathbf{s} \leq \mathbf{R} - \mathbf{s}_1 - \dots - \mathbf{s}_k}} \int_0^\infty q_{k+1}(d\mathbf{s}_1, \dots, d\mathbf{s}_{j-1}, d\mathbf{s}, d\mathbf{s}_j, \dots \\ \dots, d\mathbf{s}_k, x_1, \dots, x_{j-1}, x, x_j, x_k) \frac{b(x) \sigma_{k+1}}{1-B(x)} dx - \\ - \lambda_k F\left(\mathbf{R} - \sum_{i=1}^k \mathbf{s}_i\right) \times \\ \times \int_{\substack{\mathbf{0} \leq \mathbf{s}_i \leq \mathbf{r}_i, \\ i=1,2,\dots,k}} q_k(d\mathbf{s}_1, \dots, d\mathbf{s}_k, x_1, \dots, x_k) = 0. \end{aligned}$$

Таким образом, выражение (17) является решением системы (13)–(16). Доказательство того, что система (13)–(16) имеет единственное вероятностное решение, можно провести аналогично работе [9]. Таким образом, доказана следующая теорема.

Теорема. Если объемы работ имеют ФР $B(x)$ и конечное математическое ожидание b , то стационарное распределение (5) процесса $\tilde{X}(t)$ имеет вид:

$$\begin{aligned} Q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, x_1, \dots, x_k) &= \\ &= q_0 F(\mathbf{r}_1) \dots F(\mathbf{r}_k) \frac{b^k}{k!} \prod_{i=1}^k \frac{\lambda_{i-1}}{\sigma_i} \tilde{B}(x_i), \\ & \quad 1 \leq k \leq N, \quad \mathbf{r}_1 \geq \mathbf{0}, \dots, \mathbf{r}_k \geq \mathbf{0}, \quad \sum_{i=1}^k \mathbf{r}_i \leq \mathbf{R}, \end{aligned}$$

где $\tilde{B}(x)$ есть ФР остаточного объема работы:

$$\tilde{B}(x) = \frac{1}{b} \int_0^x (1-B(y)) dy.$$

Отсюда, в частности, следует, что формула для маргинальных вероятностей $Q_k(\mathbf{r}_1, \dots, \mathbf{r}_k, \infty, \dots, \infty)$ идентична выражению (3) для случая экспоненциального распределения объема работы. Таким образом, доказано, что стационарное распределение вероятностей СМО с ограниченными ресурсами и с интенсивностями поступления и обслуживания заявок, зависящими от состояния системы, не зависит от вида распределения объема работы заявки, а зависит только от математического ожидания.

4 Заключение

В работе был проведен анализ стационарного распределения СМО с ограниченными ресурсами и интенсивностями поступления и обслуживания, зависящими от состояния системы, в случае произвольного распределения объема работы, принятого заявкой. Была доказана теорема о мультипликативном виде стационарного распределения системы. Кроме того, было показано, что стационарное распределение числа заявок в системе и объема занимаемых ими ресурсов не зависит от вида распределения объема работы заявки, а зависит только от математического ожидания.

Литература

1. Naumov V., Samouylov K., Yarkina N., Sopin E., Andreev S., Samouylov A. LTE performance analysis using queuing systems with finite resources and random requirements // 7th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops. — IEEE, 2015. P. 100–103. doi: 10.1109/ICUMT.2015.7382412.
2. 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.
3. 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.
4. Севастьянов Б. А. Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами // Теория вероятностей и ее применения, 1957. Т. 2. Вып. 1. С. 106–116.
5. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. — М.: Наука, 1966. 432 с.
6. Taylor P. G. Insensitivity in stochastic models // Queueing networks / Eds. R. Boucherie, N. van Dijk. — International ser. in operations research & management science. — Springer, 2011. Vol. 154. P. 121–140. doi: 10.1007/978-1-4419-6472-4.
7. Samouylov K. E., Gaidamaka Y. V., Sopin E. S. Simplified analysis of queueing systems with random requirements // Statistics and simulation / Eds. J. Pilz, D. Rasch, V. Melas, K. Moder. — Springer proceedings in mathematics & statistics ser. — Springer, 2018. Vol. 231. P. 381–390. doi: 10.1007/978-3-319-76035-3_27.
8. Whitt W. Continuity of generalized semi-Markov processes // Math. Oper. Res., 1980. Vol. 5. Iss. 4. P. 494–501. doi: 10.1287/moor.5.4.494.
9. Miyazawa M., Yamazaki G. The basic equations for a supplemented GSMP and its applications to queues // J. Appl. Probab., 1988. Vol. 25. Iss. 3. P. 565–578. doi: 10.2307/3213985.

Поступила в редакцию 15.06.18

ON THE INSENSITIVITY OF THE STATIONARY DISTRIBUTION OF THE LIMITED RESOURCES QUEUEING SYSTEM WITH STATE-DEPENDENT ARRIVAL AND SERVICE RATES

E. S. Sopin^{1,2}, V. A. Naumov³, and K. E. Samouylov^{1,2}

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

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

³Service Innovation Research Institute (PIKE), 8A Annankatu, Helsinki 00120, Finland

Abstract: The authors consider further generalization of the queueing systems, in which customers require not only a server but also a certain amount of limited resources. In the considered queueing system, arrival and serving intensities depend on the state of the system. The authors assume an arbitrary distribution of the service time. The authors prove that the stationary distribution of the system has product form in the case of Poisson arrivals.

Moreover, it was shown that the steady-state probability distribution of number of customers in the system and volumes of occupied resources depends on the service time distribution only through its mean.

Keywords: queuing system; limited resources; insensitivity; service time

DOI: 10.14357/19922264180306

Acknowledgments

The publication was supported by the Ministry of Education and Science of the Russian Federation (project No. 2.882.2017/4.6).

References

1. Naumov, V., K. Samouylov, N. Yarkina, E. Sopin, S. Andreev, and A. Samuylov. 2015. LTE performance analysis using queuing systems with finite resources and random requirements. *7th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings*. IEEE. 100–103. doi: 10.1109/ICUMT.2015.7382412.
2. Tikhonenko, O. M., and K. G. Klimovich. 2001. Analysis of queuing systems for random-length arrivals with limited cumulative volume. *Probl. Inform. Transm.* 37(1):70–79.
3. 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.
4. Sevast'yanov, B. A. 1957. An ergodic theorem for Markov processes and its application to telephone systems with refusals. *Theor. Probab. Appl.* 2(1):104–112.
5. Gnedenko, B. V., and I. N. Kovalenko. 1966. *Vvedenie v teoriyu massovogo obsluzhivaniya* [Introduction to queuing theory]. Moscow: Nauka. 432 p.
6. Taylor, P. G. 2011. Insensitivity in stochastic models. *Queueing networks*. Eds. R. Boucherie and N. van Dijk. International ser. in operations research & management science. Springer. 154:121–140. doi: 10.1007/978-1-4419-6472-4.
7. Samouylov, K. E., Y. V. Gaidamaka, and E. S. Sopin. 2018. Simplified analysis of queuing systems with random requirements. *Statistics and simulation*. Eds. J. Pilz, D. Rasch, V. Melas, and K. Moder. Springer proceedings in mathematics & statistics ser. Springer. 231:381–390. doi: 10.1007/978-3-319-76035-3_27.
8. Whitt, W. 1980. Continuity of generalized semi-Markov processes. *Math. Oper. Res.* 5(4):494–501. doi: 10.1287/moor.5.4.494.
9. Miyazawa, M., and G. Yamazaki. 1988. The basic equations for a supplemented GSMP and its applications to queues. *J. Appl. Probab.* 25(3):565–578. doi: 10.2307/3213985.

Received June 15, 2018

Contributors

Sopin Eduard S. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sopin_es@rudn.university

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

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya 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, samuylov_ke@rudn.university

РЕСУРСНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ КАК МОДЕЛИ БЕСПРОВОДНЫХ СИСТЕМ СВЯЗИ*

А. В. Горбунова¹, В. А. Наумов², Ю. В. Гайдамака³, К. Е. Самуйлов⁴

Аннотация: Представлен обзор ресурсных систем массового обслуживания (СМО), используемых для моделирования широкого класса реальных систем, в которых ресурсы являются заведомо ограниченными. Несмотря на объективную важность исследования подобных систем, работ, посвященных их анализу, до последнего времени существовало совсем немного, что было связано со сложностью построения случайного процесса, описывающего их функционирование, и, соответственно, получения численных результатов. Однако за последние годы произошел существенный сдвиг в изучении ресурсных систем, были предложены новые методы их анализа, позволяющие строить рекуррентные алгоритмы, пригодные для численных расчетов. В этой связи в обзоре отражена только часть полученных результатов, а именно: рассмотрены ресурсные системы без мест для ожидания с экспоненциальным временем обслуживания. Рассмотрены модели беспроводных систем связи, основанные на ресурсных СМО (РСМО), выражения для оценки основных вероятностно-временных характеристик и алгоритмы их вычисления.

Ключевые слова: ресурсная система массового обслуживания; непрерывный ресурс; дискретный ресурс; ограниченный ресурс; рекуррентный алгоритм; гетерогенная сеть; стационарное распределение; полумарковский процесс; беспроводные системы связи

DOI: 10.14357/19922264180307

1 Введение

В классических СМО приборы и места ожидания играют роль необходимых для обслуживания ресурсов. В РСМО кроме приборов и мест ожидания заявкам могут потребоваться различные дополнительные ресурсы. Это может быть некоторый случайный объем ресурса, занимаемого на время ожидания начала обслуживания, либо на время обслуживания, либо на все время пребывания заявки в системе. Если у системы нет достаточного числа свободных ресурсов, поступившая заявка теряется. В дальнейшем будем использовать термин «ресурс» только для обозначения дополнительного ресурса, отличного от приборов или мест ожидания.

Интерес к РСМО объясняется возможностью их применения для моделирования достаточно широкого спектра технических устройств и в целом информационно-вычислительных систем. В частности, если говорить о единственном типе ресурса ограниченного объема, то таким образом может моделироваться ограниченность памяти некоторого устройства или отдельной системы. Таким образом, увеличивается реалистичность модели и, соответ-

ственно, ее практическая ценность. Если же говорить о множественных ресурсах, то стоит вспомнить услуги беспроводных сетей, таких, например, как Long Term Evolution (LTE) [1]. Рост их популярности делает необходимым создание эффективных инструментов для оценки телекоммуникационными операторами работы радиоинтерфейсов [2, 3]. В этих сетях каждая активная сессия занимает определенный объем радиоресурсов (например, ширину полосы пропускания спектра частот, мощности передачи радиочастотного усилителя и др.), которые являются заведомо ограниченными и должны быть распределены при поступлении вызова пользователя и освобождены по завершении сессии [4].

Стоит отметить, что моделированию беспроводных систем связи с помощью СМО с множественными ресурсами начиная с [5] посвящено большое число публикаций. Однако основной акцент в них делается на анализ различных схем распределения ресурсов в системах с детерминированными требованиями заявок к ресурсам. Обзор этих работ можно найти в [6–8].

Статья организована следующим образом: в разд. 2 описываются основные типы РСМО без

* 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@pfu.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

мест для ожидания, методы их исследования и полученные результаты в виде выражений для основных вероятностно-временных характеристик функционирования указанных систем. В разд. 3 представлены подходы, позволяющие провести численные расчеты с помощью полученных соотношений. В заключении кратко подведены итоги работы.

2 Ресурсные системы массового обслуживания

Более подробно остановимся на описании общей модели РСМО без мест для ожидания (рис. 1). Система может располагать ограниченным или неограниченным объемом ресурсов как одного, так и нескольких типов. Схему ее функционирования можно описать следующим образом:

- (1) для обслуживания каждой заявки требуется один прибор и некоторый объем ресурса каждого типа;
- (2) поступившая заявка теряется, если в момент поступления объем требуемого ей ресурса превышает объем свободного ресурса этого типа либо все приборы заняты;
- (3) в момент начала обслуживания заявки суммарный объем занятого ресурса каждого типа увеличивается на величину ресурса, выделенного этой заявке;
- (4) в момент окончания обслуживания заявки суммарный объем занятого ресурса каждого типа уменьшается на величину ресурса, выделенного этой заявке.

В СМО может поступать один или несколько классов заявок, для которых $A_l(t)$ — функция распределения времени между поступлениями заявок класса l , $H_l(t, \mathbf{x})$ — совместная функция распределения длительности обслуживания и вектора объема ресурсов, необходимых поступившей заявке

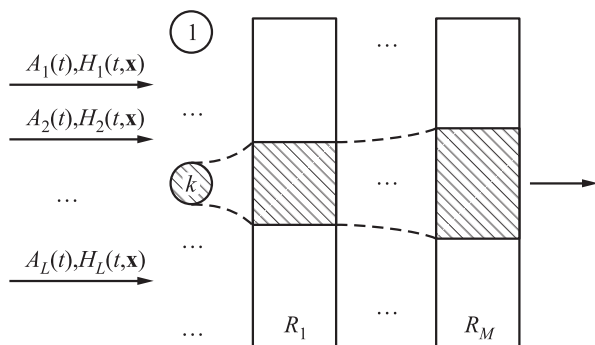


Рис. 1 Схема функционирования РСМО общего вида

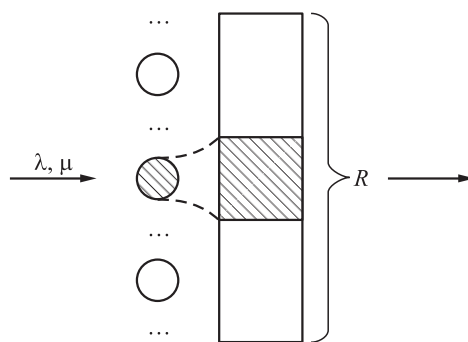


Рис. 2 Схема функционирования простейшей РСМО

класса $l, l = \overline{1, L}$. Для случая, когда случайные величины длительности обслуживания и вектора объема необходимых ресурсов независимы, имеем

$$H_l(t, \mathbf{x}) = B_l(t)F_l(\mathbf{x}),$$

где $B_l(t)$ — функция распределения времени обслуживания заявки класса l ; $F_l(\mathbf{x})$ — функция распределения вектора объема ресурсов, требуемых заявкам класса l .

Первые статьи, посвященные анализу СМО с выделением каждой поступающей заявке помимо прибора некоторого случайного объема ресурса единственного типа появились в начале 1970-х гг. [9, 10]. В частности, в работе [9] рассматривалась бесконечно линейная СМО с пуассоновским входящим потоком и экспоненциальным временем обслуживания (рис. 2). Величины требуемых ресурсов — независимые одинаково распределенные случайные величины с функцией распределения $F(x)$. В качестве емкости системы, т.е. максимально допустимого объема ресурсов, выступает величина R .

Система уравнений равновесия (СУР) для случайного процесса, описывающего систему, фактически представляет собой обобщение системы Эрланга. В результате решения СУР были получены стационарные вероятности того, что в системе находится k заявок:

$$p_k = \frac{(1/k!)(\lambda/\mu)^k F^{(k)}(R)}{\sum_{i=0}^{\infty} (1/i!)(\lambda/\mu)^i F^{(i)}(R)},$$

где $F^{(k)}(x)$ является k -кратной сверткой функции распределения $F(x)$, $k = 2, 3, \dots$, $F^{(0)}(x) = 1$, $F^{(1)}(x) = F(x)$. В условиях описанной модели потеря заявки или отказ в обслуживании происходят только тогда, когда разность между величиной объема всей системы и суммарным объемом ресурсов, занятых находящимися в системе заявками,

меньше, чем величина требуемого объема ресурсов у вновь поступившей заявки. Таким образом, вероятность потери заявки равна

$$B = 1 - \frac{\sum_{k=0}^{\infty} (1/k!) (\lambda/\mu)^k F^{(k+1)}(R)}{\sum_{k=0}^{\infty} (1/k!) (\lambda/\mu)^k F^{(k)}(R)}.$$

Для того чтобы более детально ознакомиться с особенностями построения и анализа РСМО, подробнее остановимся на статье [4]. Здесь рассматривается многолинейная СМО с $N \leq \infty$ приборами. Поступающий поток является пуассоновским с параметром λ , длительности обслуживания заявок независимы между собой и от поступающего потока и имеют экспоненциальное распределение с параметром μ . Система располагает ограниченным объемом ресурсов M типов. Обозначим через R_m общий объем ресурса типа m , $\mathbf{R} = (R_1, \dots, R_M)$, и через $\mathbf{r}_j = (r_{j1}, r_{j2}, \dots, r_{jM})$ — вектор объемов ресурсов, необходимых j -й поступившей заявке, $j = 1, 2, \dots$. Будем считать, что случайные векторы \mathbf{r}_j не зависят от процессов поступления и обслуживания заявок, независимы в совокупности и одинаково распределены с функцией распределения $F(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_M)$. Состояние такой системы в момент t можно описать полумарковским процессом $X(t) = \{\xi(t), \Gamma(t)\}$ [4]. Здесь $\xi(t)$ — число заявок в системе, а $\Gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_{\xi(t)}(t))$, где $\gamma_i(t)$ — вектор объемов всех типов ресурсов, занимаемых i -й обслуживаемой заявкой. Находящиеся на обслуживании заявки перенумеровываются в порядке убывания остаточного времени обслуживания. Рассмотрим предельное распределение процесса $X(t)$:

$$p_0 = \lim_{t \rightarrow \infty} P\{\xi(t) = 0\};$$

$$P_k(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) = \lim_{t \rightarrow \infty} P\{\xi(t) = k;$$

$$\gamma_1(t) \leq \mathbf{x}_1, \gamma_2(t) \leq \mathbf{x}_2, \dots, \gamma_k(t) \leq \mathbf{x}_k\}.$$

После решения соответствующей СУР получаем

$$p_0 = \left(1 + \sum_{i=1}^N F^{(i)}(\mathbf{R}) \frac{\rho^i}{i!}\right)^{-1};$$

$$P_k(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) = p_0 F(\mathbf{x}_1) F(\mathbf{x}_2) \cdots F(\mathbf{x}_k) \frac{\rho^k}{k!},$$

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \geq \mathbf{0}, \sum_{i=1}^k \mathbf{x}_i \leq \mathbf{R}, 1 \leq k \leq N,$$

где $\rho = \lambda/\mu$; $F^{(k)}(\mathbf{x})$ — k -кратная свертка функции $F(\mathbf{x})$; $\mathbf{x}_i = (x_{i1}, \dots, x_{iM})$, $i = \overline{1, k}$. Далее, если обозначить вектор суммарных объемов занятых ресурсов каждого типа $\delta(t) = \sum_{i=1}^{\xi(t)} \gamma_i(t)$,

$\delta(t) = (\delta_1(t), \dots, \delta_M(t))$, стационарное распределение $Q_k(\mathbf{x})$ случайного процесса $X(t) = (\xi(t); \delta(t))$ примет вид:

$$Q_k(\mathbf{x}) = \lim_{t \rightarrow \infty} P\{\xi(t) = k; \delta(t) \leq \mathbf{x}\} = p_0 F^{(k)}(\mathbf{x}) \frac{\rho^k}{k!},$$

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{R}, 1 \leq k \leq N.$$

В [11] исследуется РСМО с единственным типом ограниченного ресурса объема R , но уже с L входящими пуассоновскими потоками с интенсивностями $\lambda_1, \dots, \lambda_L$ и с $N \leq \infty$ приборами. Длительности обслуживания заявок независимы между собой, от поступающих потоков и экспоненциально распределены с параметром μ_l для заявок класса l , $l = \overline{1, L}$. Предполагается, что объем ресурса, требуемого заявкам класса l , является случайной величиной с функцией распределения $F_l(x)$, не зависящей от процессов поступления и обслуживания заявок. Обслуживающимся заявкам присваивается номер, причем так, чтобы заявка с номером i имела i -е по величине остаточное время обслуживания. Этот номер следует отличать от порядкового номера заявки. При поступлении новой заявки все находящиеся на обслуживании заявки перенумеровываются. Состояние системы в момент t описывается полумарковским процессом $X(t) = (\xi(t); \theta(t); \gamma(t))$. Здесь, как и прежде, $\xi(t)$ — число заявок в системе; $\theta(t) = (\theta_1(t), \theta_2(t), \dots, \theta_{\xi(t)}(t))$; $\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_{\xi(t)}(t))$, где $\theta_i(t)$ — класс i -й обслуживаемой заявки; $\gamma_i(t)$ — объем занимаемого ею ресурса.

Введем стационарное распределение процесса $X(t)$

$$p_0 = \lim_{t \rightarrow \infty} P\{\xi(t) = 0\};$$

$$p_{l_1, \dots, l_k}^k(x_1, \dots, x_k) = \lim_{t \rightarrow \infty} P\{\xi(t) = k;$$

$$\theta_1(t) = l_1, \dots, \theta_k(t) = l_k;$$

$$\gamma_1(t) \leq x_1, \dots, \gamma_k(t) \leq x_k\}.$$

В результате решения соответствующей СУР получены выражения для стационарных вероятностей описанной системы. Кроме того, в [11] показано, что стационарные вероятности того, что в системе находятся k_j заявок типа j и суммарный объем занимаемого ими ресурсов не превосходит x_j , $j = \overline{1, L}$, имеют мультипликативный вид:

$$P_{k_1, \dots, k_L}(x_1, \dots, x_k) = p_0 \prod_{j=1}^L F_j^{(k_j)}(x_j) \frac{\rho_j^{k_j}}{k_j!}.$$

В [12] исследуются показатели эффективности сетей ЛТЕ. Ресурсная СМО, моделирующая систему, аналогична представленной в [11], но уже

с M типами ограниченных ресурсов, и потому стационарные вероятности

$$p_0 = \lim_{t \rightarrow \infty} P\{\xi(t) = 0\};$$

$$p_{l_1, \dots, l_k}^k(\mathbf{x}_1, \dots, \mathbf{x}_k) = \lim_{t \rightarrow \infty} P\{\xi(t) = k;$$

$$\theta_1(t) = l_1, \dots, \theta_k(t) = l_k;$$

$$\gamma_1(t) \leq \mathbf{x}_1, \dots, \gamma_k(t) \leq \mathbf{x}_k\};$$

после решения соответствующей СУР примут вид:

$$p_0 = \left(1 + \sum_{r=1}^N \sum_{k_1+\dots+k_r=r} \left(F_1^{(k_1)} * F_2^{(k_2)} * \dots * F_r^{(k_r)} \right) (\mathbf{R}) \times \right.$$

$$\left. \times \frac{\rho_1^{k_1}}{k_1!} \frac{\rho_2^{k_2}}{k_2!} \dots \frac{\rho_r^{k_r}}{k_r!} \right)^{-1};$$

$$p_{l_1, \dots, l_k}^k(\mathbf{x}_1, \dots, \mathbf{x}_k) =$$

$$= p_0 F_{l_1}(\mathbf{x}_1) F_{l_2}(\mathbf{x}_2) \dots F_{l_k}(\mathbf{x}_k) \prod_{n=1}^k \frac{\lambda_{l_n}}{\sum_{i=1}^n \mu_i},$$

$$1 \leq l_1, \dots, l_k \leq L, \quad \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \geq \mathbf{0},$$

$$\sum_{i=1}^k \mathbf{x}_i \leq \mathbf{R}, \quad 1 \leq k \leq N,$$

где символ $*$ означает свертку функции распределения.

В работе [13] моделируется ситуация, когда объем ресурсов, запрашиваемых пользователями, может быть не только положительным, но и отрицательным. Запросы на отрицательный объем ресурса увеличивают объем доступного ресурса для пользователей, запрашивающих его положительные объемы. В [13] предполагается зависимость времени обслуживания и интервалов между поступлениями заявок от числа заявок в системе. В результате анализа моделей получены формулы для расчета основных вероятностно-временных характеристик.

В [14, 15] для анализа сетей LTE с динамически меняющимися требованиями к ресурсам рассматриваются РСМО с добавлением пуассоновского потока сигналов, инициирующего перераспределение ресурсов для активных пользователей. Развитием работы [15] стали статьи [16, 17]. Были исследованы два сценария перераспределения ресурсов и сопоставлены посредством численного анализа.

В серии работ [15, 18–24] исследуются РСМО, в которых объем выделяемых заявке ресурсов имеет дискретное распределение, т. е. для i -й поступившей в систему заявки с вероятностью $p_j = P(r_i = j)$

потребуется ресурс объема j . Так, в [20] анализируется РСМО с L входящими пуассоновскими потоками и M типами ресурсов. Получены выражения для стационарных вероятностей:

$$q_{k_1, \dots, k_L}^k(\mathbf{r}_1, \dots, \mathbf{r}_L) =$$

$$= q_0 \sum_{k_1+\dots+k_L=k} p_{1, \mathbf{r}_1}^{(k_1)} \dots p_{1, \mathbf{r}_L}^{(k_L)} \frac{\rho_1^{k_1}}{k_1!} \dots \frac{\rho_L^{k_L}}{k_L!};$$

$$q_0 =$$

$$= \left(1 + \sum_{k=0}^N \sum_{k_1+\dots+k_L=k} \sum_{\mathbf{r}_1+\dots+\mathbf{r}_L \leq \mathbf{R}} p_{1, \mathbf{r}_L}^{(k_L)} \frac{\rho_1^{k_1}}{k_1!} \dots \frac{\rho_L^{k_L}}{k_L!} \right)^{-1},$$

где $q_{k_1, \dots, k_L}^k(\mathbf{r}_1, \dots, \mathbf{r}_L)$ — это вероятность того, что в системе находятся k заявок, из которых k_1 — класса 1, k_2 — класса 2 и т. д., а суммарный объем ресурсов каждого типа, занятых заявками класса 1, равен \mathbf{r}_1 и т. д.

В статье [22] при исследовании той же СМО, что и в [20], было получено распределение стационарных вероятностей $q_k(\mathbf{r})$ с объединенным входящим потоком и средневзвешенным требованием

$$p_{\mathbf{r}} = \sum_{l=1}^L \frac{\rho_l}{\rho} p_{l, \mathbf{r}},$$

где $\rho = \sum_{l=1}^L \rho_l$:

$$q_k(\mathbf{r}) = q_0 \frac{\rho^k}{k!} p_{\mathbf{r}}^{(k)}, \quad q_0 = \left(\sum_{k=0}^N \sum_{\mathbf{r}=\mathbf{0}}^{\mathbf{R}} p_{\mathbf{r}}^{(k)} \right)^{-1},$$

что позволило выразить вероятность блокировки и среднего объема занятых ресурсов в аналитическом виде.

В случае рассматриваемой СМО, но с заявками одного класса, в [24] представлены выражения для стационарных вероятностей состояний числа заявок в системе и суммарного объема занятых ресурсов, а также формулы для вероятности блокировки и среднего объема занятых ресурсов.

3 Вычисление характеристик ресурсных систем массового обслуживания

В работе [25] для системы с одним типом ресурса показано, что в предположении о гамма-распределении необходимого заявкам ресурса плотность распределения высвобождаемого заявкой ресурса при заданном числе заявок в системе и заданном

векторе суммарных объемов занятых ресурсов совпадает с бета-распределением, позволяющим легко рассчитывать многократные свертки, к которым приводит необходимость учитывать объемы всех заявок в системе. В остальных случаях наличие в формулах большого числа сверток создает значительную вычислительную сложность при расчете стационарных характеристик РСМО. Так, для расчета характеристик СМО из [20] необходимо для каждого $k \in \{0, \dots, N\}$, а также всех наборов векторов $\mathbf{r} \leq \mathbf{R}$ хранить в памяти компьютера значения сверток вероятностей $p_{\mathbf{r}}$. А при больших значениях N и \mathbf{R} вычисление вероятностей блокировок системы и также объемов занятого ресурса по представленным формулам вообще нерационально. Поэтому задача получения действенных численных методов является крайне важной. В работе [21] для модели СМО из [20], чтобы сократить вычисления, был предложен рекуррентный алгоритм вычисления нормировочной константы $G(N, \mathbf{R}) = q_0^{-1}$, основанный на алгоритме Бузена [26]. Кроме того, на основе разработанного алгоритма были получены рекуррентные формулы для вычисления вероятностных характеристик системы: вероятности блокировки системы, среднего объема дисперсии занятых ресурсов. Если обозначить

$$G(n, \mathbf{r}) \sum_{k=0}^n \sum_{\mathbf{j}=0}^{\mathbf{r}} p_{\mathbf{j}}^{(k)} \frac{\rho^k}{k!}, \quad n \geq 0, \quad \mathbf{r} \geq \mathbf{0},$$

то функция $G(n, \mathbf{r})$ будет удовлетворять следующему рекуррентному соотношению:

$$G(n, \mathbf{r}) = G(n-1, \mathbf{r}) + \frac{\rho}{n!} \sum_{\mathbf{j}=0}^{\mathbf{r}} p_{\mathbf{j}} (G(n-1, \mathbf{r}-\mathbf{j}) - G(n-2, \mathbf{r}-\mathbf{j}))$$

с начальными условиями

$$G(0, \mathbf{r}) = 1, \quad \mathbf{r} \geq \mathbf{0}; \quad G(1, \mathbf{r}) = 1 + \sum_{\mathbf{j}=0}^{\mathbf{r}} p_{\mathbf{j}}.$$

При анализе РСМО, описывающих М2М (machine-to-machine) трафик в сетях LTE, аналогичный рекуррентный алгоритм для вычисления нормировочной константы был разработан в [24]. Матричные методы анализа РСМО, применимые при моделировании соты сети LTE с двумя типами трафика, М2М и Н2Н (human-to-human), предложены в работах [27, 28].

4 Заключение

В настоящем обзоре кратко представлены основные разновидности РСМО, существующие

методы их анализа, выражения для оценки основных вероятностно-временных характеристик и алгоритмы их вычисления.

Литература

1. Andrews J. G., Buzzi S., Choi W., Hanly S. V., Lozano A., Soong A. C. K., Zhang J. C. What will 5G be? // IEEE J. Sel. Area. Comm., 2014. Vol. 32. No. 6. P. 1065–1082.
2. Buturlin I. A., Gaidamaka Y. V., Samuylov A. K. Utility function maximization problems for two cross-layer optimization algorithms in OFDM wireless networks // 4th Congress (International) on Ultra Modern Telecommunications and Control Systems, 2012. P. 63–65.
3. Galinina O., Andreev S. D., Gerasimenko M., Koucheryavy Y., Himayat N., Yeh S. P., Talwar S. Capturing spatial randomness of heterogeneous cellular/WLAN deployments with dynamic traffic // IEEE J. Sel. Area. Comm., 2014. Vol. 32. No. 6. P. 1083–1099.
4. Наумов В. А., Самуйлов К. Е., Самуйлов А. К. О суммарном объеме ресурсов, занимаемых обслуживаемыми заявками // Автоматика и телемеханика, 2016. № 8. С. 125–135.
5. Gimpelson L. A. Analysis of mixtures of wide- and narrow-band traffic // IEEE T. Commun. Techn., 1968. Vol. 13. No. 3. P. 258–266.
6. Kelly F. P. Loss networks // Ann. Appl. Probab., 1991. No. 1. P. 319–378.
7. Ross K. W. Multiservice loss models for broadband telecommunication networks. — London: Springer-Verlag, 1995. 343 p.
8. Basharin G. P., Samouylov K. E., Yarkina N. V., Gudkova I. A. A new stage in mathematical teletraffic theory // Automat. Rem. Contr., 2009. Vol. 70. No. 12. P. 1954–1964.
9. Ромм Э. Л., Скитович В. В. Об одном обобщении задачи Эрланга // Автоматика и телемеханика, 1971. № 6. С. 164–168.
10. Кац Б. А. Об обслуживании сообщений случайной длины // Теория массового обслуживания: Тр. 3-й Всесоюз. школы-совещания по теории массового обслуживания, 1976. С. 157–168.
11. Наумов В. А., Самуйлов А. К. Модель выделения ресурсов беспроводной сети объемами случайной величины // Вестник РУДН. Серия: Математика, информатика, физика, 2015. № 2. С. 38–45.
12. Naumov V., Samouylov K., Yarkina N., Sopin E., Andreev S., Samuylov A. LTE performance analysis using queuing systems with finite resources and random requirements // 7th Congress on Ultra Modern Telecommunications and Control Systems. — IEEE, 2015. P. 100–103.
13. 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.

14. *Samouylov K., Sopin E., Vikhrova O.* Analysis of queueing system with resources and signals // *Comm. Com. Inf. Sc.*, 2017. Vol. 800. P. 358–369.
15. *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. — IEEE, 2017. P. 101–106.
16. *Petrov V., Solomitskii D., Samuylov A., Lema Maria A., Gapeyenko M., Moltchanov D., Andreev S., Naumov V., Samouylov K., Dohler M., Koucheryavy Ye.* Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments // *IEEE J. Sel. Area. Comm.*, 2017. Vol. 35. No. 9. P. 2038–2055.
17. *Наумов В. А., Самуйлов К. Е.* Анализ сетей ресурсных систем массового обслуживания // *Автоматика и телемеханика*, 2018. № 5. С. 59–68.
18. *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.
19. *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.
20. *Sopin E., Samouylov K., Vikhrova O., Kovalchukov R., Moltchanov D., Samuylov A.* Evaluating a case of downlink uplink decoupling using queueing system with random requirement // *Internet of Things, smart spaces, and next generation networks and systems* / Eds. O. Galinina, S. I. Balandin, Y. Koucheryavy. — Lecture notes in computer science ser. — Springer, 2016. Vol. 9870. P. 440–450.
21. *Samouylov K., Sopin E., Vikhrova O., Shorgin S.* Convolution algorithm for normalization constant evaluation in queueing system with random requirements // *AIP Conf. Proc.*, 2017. Vol. 1863. Art. No. 090004. 4 p.
22. *Вихрова О. Г.* К вычислению вероятностных характеристик СМО ограниченной емкости со случайными требованиями к ресурсам // *Вестник РУДН. Серия: Математика, информатика, физика*, 2017. № 3. С. 203–210.
23. *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*, 2018. Vol. 16. Art. No. 01008. 4 p.
24. *Sopin E., Gaidamaka Yu., Markova E., Vikhrova O.* Performance analysis of M2M traffic in LTE network using queueing systems with random resource requirements // *Autom. Control Comp. S.*, 2018 (in press).
25. *Наумов В. А., Самуйлов К. Е.* О моделировании систем массового обслуживания с множественными ресурсами // *Вестник РУДН. Серия: Математика, информатика, физика*, 2014. № 3. С. 60–64.
26. *Buzen J. P.* Computational algorithms for closed queueing networks with exponential servers // *Commun. ACM*, 1973. Vol. 16. P. 527–531.
27. *Вишневецкий В. М., Самуйлов К. Е., Наумов В. А., Яркина Н. В.* Модель соты LTE с межмашинным трафиком в виде мультисервисной системы массового обслуживания с эластичными и потоковыми заявками и марковским входящим потоком // *Вестник РУДН. Серия: Математика, информатика, физика*, 2016. № 4. С. 26–36.
28. *Vishnevsky V., Samouylov K., Naumov V., Krishnamoorthy A., Yarkina N.* Multiservice queueing system with map arrivals for modelling LTE cell with H2H and M2M communications and M2M aggregation // *Comm. Com. Inf. Sc.*, 2017. Vol. 700. P. 63–74.

Поступила в редакцию 16.06.18

RESOURCE QUEUING SYSTEMS AS MODELS OF WIRELESS COMMUNICATION SYSTEMS

A. V. Gorbunova¹, V. A. Naumov², Yu. V. Gaidamaka^{1,3}, and K. E. Samouylov^{1,3}

¹Peoples' Friendship University of Russia (RUDN University), 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 presents an overview of the resource queuing systems used for modeling of a wide class of real systems with admittedly limited resources. Despite the objective importance of studying of such systems, there have been very few works devoted to their analysis until recently, which was due to the complexity of constructing a random process to describe their functioning and, accordingly, of obtaining the numerical results. However, in recent years, there has been a significant shift in the study of the resource systems — new methods for their analysis have been proposed, which made it possible to construct recursive algorithms suitable for the numerical calculations. In this regard, the current review reflects only a part of the previously obtained results, namely, it considers

the resource systems without waiting space with exponentially distributed service time. The authors consider the models of wireless communication systems based on resource queuing systems, expressions for estimating the main probabilistic, and temporal characteristics and algorithms for their calculation.

Keywords: resource queuing systems; continuous resource; discrete resource; limited resource; recursive algorithm; heterogeneous network; stationary distribution; semi-Markov process; wireless communication systems

DOI: 10.14357/19922264180307

Acknowledgments

The work was partly supported by the Russian Ministry of Education and Science (project 2.882.2017/4.6).

References

- Andrews, J. G., S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. 2014. What will 5G be? *IEEE J. Sel. Area. Comm.* 32(6):1065–1082.
- Buturlin, I. A., Y. V. Gaidamaka, and A. K. Samuylov. 2012. Utility function maximization problems for two cross-layer optimization algorithms in OFDM wireless networks. *4th Congress (International) on Ultra Modern Telecommunications and Control Systems.* 63–65.
- Galinina, O. S., D. Andreev, M. Gerasimenko, Y. Koucheryavy, N. Himayat, S. P. Yeh, and S. Talwar. 2014. Capturing spatial randomness of heterogeneous cellular/WLAN deployments with dynamic traffic. *IEEE J. Sel. Area. Comm.* 32(6):1083–1099.
- 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.
- Gimpelson, L. A. 1968. Analysis of mixtures of wide- and narrow-band traffic. *IEEE T. Commun. Techn.* 13(3):258–266.
- 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. Skitovich. 1971. Ob odnom obobshchenii zadachi Erlanga [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. Tr. 3 Vsesoyuzn. shkoly-soveshchaniya po teorii massovogo obsluzhivaniya* [Queuing Theory: 3rd All-Union School-Seminar on Queuing Theory Proceedings]. 157–168.
- Naumov, V. A., and A. K. Samuylov. 2015. Model' vydeleniya resursov besprovodnoy seti ob"emami sluchaynoy velichiny [Queuing system with resource allocation of the random volume]. *RUDN J. Math. Information Sci. Phys.* 2:38–45.
- Naumov, V., K. Samouylov, N. Yarkina, E. Sopin, S. Andreev, and A. Samuylov. 2015. LTE performance analysis using queuing systems with finite resources and random requirements. *7th Congress on Ultra Modern Telecommunications and Control Systems.* IEEE. 100–103.
- 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.
- Samouylov, K., E. Sopin, and O. Vikhrova. 2017. Analysis of queuing 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.* 101–106.
- Petrov, V., D. Solomitckii, A. Samuylov, A. Maria Lema, M. Gapeyenko, D. Moltchanov, S. Andreev, V. Naumov, K. Samouylov, M. Dohler, and Ye. Koucheryavy. 2017. Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments. *IEEE J. Sel. Area. Comm.* 35(9):2038–2055.
- Naumov, V. A., and K. E. Samuilov. 2018. Analysis of networks of the resource queuing systems. *Automat. Rem. Contr.* 79(5):822–829.
- Samouylov, K., E. Sopin, and O. Vikhrova. 2015. Analyzing blocking probability in LTE wireless network via queuing system with finite amount of resources. *Comm. Com. Inf. Sc.* 564:393–403.
- 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.
- Sopin, E., K. Samouylov, O. Vikhrova, R. Kovalchukov, D. Moltchanov, and A. Samuylov. 2016. Evaluating a case of downlink uplink decoupling using queuing system with random requirement. *Internet of Things, smart spaces, and next generation networks and systems.* Eds. O. Galinina, S. I. Balandin, and Y. Koucheryavy. Lecture notes in computer science ser. Springer. 9870:440–450.
- Samouylov, K., E. Sopin, O. Vikhrova, and S. Shorgin. 2017. Convolution algorithm for normalization constant evaluation in queuing system with random requirements. *AIP Conf. Proc.* 1863:090004. 4 p.
- Vikhrova, O. G. 2017. K vychisleniyu veroyatnostnykh kharakteristik SMO ogranichennoy emkosti so sluchayny-

- mi trebovaniyami k resursam [About probability characteristics evaluation in queuing system with limited resources and random requirements]. *RUDN J. Math. Information Sci. Phys.* 25(3):203–210.
23. Sopin, E., and K. Samouylov. 2018. On the analysis of the limited resources queuing system under MAP arrivals. *Conference (International) on Applied Mathematics, Computational Science and Systems Engineering*. 16:01008. 4 p.
 24. Sopin, E., Yu. Gaidamaka, E. Markova, and O. Vikhrova. 2018 (in press). Performance analysis of M2M traffic in LTE network using queuing systems with random resource requirements. *Autom. Control Comp. S.*
 25. Naumov, V. A., and K. E. Samouylov. 2014. O modelirovanii sistem massovogo obsluzhivaniya s mnozhestvennyimi resursami [On the modeling of queueing systems with multiple resources]. [*RUDN J. Math. Information Sci. Phys.* 3:60–64.
 26. Buzen, J. P. 1973. Computational algorithms for closed queueing networks with exponential servers. *Commun. ACM* 16:527–531.
 27. Vishnevsky, V. M., K. E. Samouylov, V. A. Naumov, and N. V. Yarkina. 2016. Model' soty LTE s mezhmashinnyim trafikom v vide mul'tiservisnoy sistemy massovogo obsluzhivaniya s elastichnymi i potokovymi zayavkami i markovskim vkhodyashchim potokom [Multiservice queuing system with elastic and streaming flows and markovian arrival process for modelling LTE cell with M2M traffic]. *RUDN J. Math. Information Sci. Phys.* 4:26–36.
 28. Vishnevsky, V., K. E. Samouylov, V. Naumov, A. Krishnamoorthy, and N. Yarkina. 2017. Multiservice queueing system with map arrivals for modelling LTE cell with H2H and M2M communications and M2M aggregation. *Comm. Com. Inf. Sc.* 700:63–74.

Received June 16, 2018

Contributors

Gorbunova Anastasiya V. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Peoples' Friendship University of Russia (RUDN University), 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 (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gaydamaka_yuv@rudn.university

Samuylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; samuylov_ke@rudn.university

ОБУЧАЕМАЯ КЛАССИФИКАЦИЯ ДАННЫХ С УЧЕТОМ АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ

М. П. Кривенко¹

Аннотация: Рассматриваются вопросы обучаемой классификации с учетом результатов анализа главных компонент (PCA — Principal Component Analysis). Построение байесовского классификатора становится возможным после представления ковариаций через параметры вероятностной модели PCA. Выделен случай сингулярных распределений данных, для него оценивание параметров модели предлагается проводить при ограничениях на собственные значения ковариационных матриц. Исследуется качество классификации с учетом реальной размерности данных. Продемонстрировано, что при ее правильном задании классификатор обладает наименьшими вероятностями ошибки. Превышение наилучшего значения размерности обычно ухудшает качество классификации в меньшей степени, чем его занижение. Смесь вероятностных анализаторов главных компонент позволяет моделировать объемные данные с помощью относительно небольшого числа свободных параметров. Число свободных параметров можно контролировать с помощью выбора латентной размерности данных.

Ключевые слова: анализ главных компонент; смеси нормальных распределений; EM-алгоритм; обучаемая классификация

DOI: 10.14357/19922264180308

1 Введение

Один из способов снижения размерности данных заключается в применении анализа главных компонент (PCA). Популярность PCA определяется рядом свойств, важнейшим из которых является его оптимальность при сжатии множества векторов высокой размерности в множество векторов более низкой размерности, а затем их восстановления.

Использовать PCA в задаче обучаемой классификации данных можно двояко. Во-первых, безотносительно к сложной структуре результатов наблюдений, подразумевающей наличие классов данных. В этом случае данные без уточнения их статистической модели сжимаются, а затем подвергаются анализу. Более сложным оказывается второй подход, когда PCA проводится индивидуально для каждого класса в отдельности. В связи с его применением возникают два вопроса:

1. Как проводить классификацию данных, объединяя результаты PCA для каждого класса в отдельности?
2. Может ли подобное сжатие данных стать источником повышения качества классификации данных?

Задача классификации данных становится традиционной после перехода к PPCA(k) — вероят-

ностной модели PCA (PPCA — probabilistic PCA) для сниженной размерности k :

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{a} + \boldsymbol{\varepsilon},$$

где \mathbf{y} — d -мерная наблюдаемая переменная, $\mathbf{y} \sim N(\mathbf{a}, \mathbf{C}(k))$; \mathbf{W} — $(d \times k)$ -матрица преобразования; \mathbf{x} — k -мерная латентная переменная, $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$; $\boldsymbol{\varepsilon}$ — d -мерная переменная, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$; $\mathbf{C}(k) = \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}$. Здесь d — исходная размерность данных; k — сниженная размерность сжатых данных; σ^2 и \mathbf{W} суть параметры модели (данные принимаются центрированными).

Пусть задана $(d \times N)$ -матрица «признак–объект» \mathbf{Y} и найдена выборочная ковариационная матрица \mathbf{S} . Справедливо спектральное разложение вида $\mathbf{S} = \mathbf{U}\mathbf{V}\mathbf{U}^T$, где \mathbf{V} — диагональная матрица, ее элементы v_1, \dots, v_d суть собственные значения матрицы \mathbf{S} , а \mathbf{U} является ортогональной матрицей, столбцы которой — ортонормированные собственные векторы матрицы \mathbf{S} . Тогда согласно [1] могут быть найдены оценки параметров модели:

$$\hat{\sigma}^2 = \frac{1}{d-k} \sum_{i=k+1}^d v_i; \quad \hat{\mathbf{W}} = \mathbf{U}_k (\mathbf{V}_k - \hat{\sigma}^2 \mathbf{I})^{1/2},$$

где столбцы $(d \times k)$ -матрицы \mathbf{U}_k суть оси первых k главных компонент; \mathbf{V}_k — диагональная

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

$(k \times k)$ -матрица соответствующих дисперсий. После этого можно рассматривать случайную нормально распределенную величину $\mathbf{y} \sim N(\mathbf{a}, \hat{\mathbf{C}}(k))$, где

$$\hat{\mathbf{C}}(k) = \hat{\mathbf{W}}\hat{\mathbf{W}}^T + \hat{\sigma}^2\mathbf{I},$$

и становится возможным построение байесовского классификатора. Заметим, что для различных значений k ковариационная матрица $\hat{\mathbf{C}}(k) \neq \mathbf{S}$, кроме случаев $k = d - 1, d$.

2 Смесь вероятностных моделей анализа главных компонент

Анализ главных компонент определяет только линейную проекцию данных, по этой причине область его применения несколько ограничена. Это, естественно, мотивировало различные разработки нелинейного PCA.

Связь вероятностной модели со стандартным PCA открывает заманчивую перспективу моделировать сложные структуры данных с помощью комбинации локальных подмоделей PCA и реализации механизма смеси вероятностных анализаторов главных компонент.

Этот подход позволяет определять все параметры модели путем максимизации правдоподобия, в ходе которого автоматически происходит разбиение данных и определение соответствующих главных осей. Логарифм правдоподобия для такой модели смеси есть

$$L = \sum_{n=1}^N \ln \{p(\mathbf{y}_n)\} = \sum_{n=1}^N \ln \left\{ \sum_{j=1}^M \pi_j p(\mathbf{y}_n|j) \right\},$$

где $p(\mathbf{y}_n|j)$ отвечает элементарной PPCA-модели; π_j — соответствующий вес элемента смеси с $\pi_j \geq 0$ и $\sum_{j=1}^M \pi_j = 1$. Заметим, что с каждым j -м элементом смеси связаны свои параметры \mathbf{a}_j , \mathbf{W}_j и σ_j^2 .

При этом генерирующая модель для смеси требует случайного выбора элемента в соответствии с пропорциями π_j , после чего формирование наблюдений для \mathbf{x} и ϵ происходит согласно модели PPCA(k) с соответствующими параметрами. Кроме того, для некоторой точки \mathbf{y} теперь имеется апостериорное распределение, связанное с каждым латентным пространством.

Можно разработать итеративный EM (expectation-maximization) алгоритм для оценивания всех параметров модели π_j , \mathbf{a}_j , \mathbf{W}_j и σ_j^2 . Если

$q_{nj} = p(j|\mathbf{y}_n)$ — вероятность принадлежности \mathbf{y}_n к j -му элементу смеси и

$$q_{nj} = \frac{\pi_j p(\mathbf{y}_n|j)}{p(\mathbf{y}_n)},$$

то согласно приложению С [1] обновления для параметров принимают обычный вид для смеси нормальных распределений. Более того, в приложении С [1] также показано: комбинация E- и M-шагов приводит к интуитивно ясному результату, что оси \mathbf{W}_j и дисперсии шума σ_j^2 определяются из взвешенной ковариационной матрицы

$$\mathbf{S}_j = \frac{1}{\tilde{\pi}_j N} \sum_{n=1}^N q_{nj} (\mathbf{y}_n - \hat{\mathbf{a}}_j) (\mathbf{y}_n - \hat{\mathbf{a}}_j)^T$$

с помощью обычной факторизации так же, как и для элементарного PPCA. Однако, как отмечено в разд. 3.4 и приложении А.5 [1], для больших значений размерности данных d могут быть получены вычислительные преимущества, если оценки \mathbf{W}_j и σ_j^2 обновляются итеративно в соответствии со схемой EM-алгоритма.

До сих пор предполагалось, что $|\mathbf{S}_j| \neq 0$. Иная ситуация с наличием сингулярных распределений может возникать при малых выборках, когда их объем не превосходит размерности выборочного пространства, или при применении EM-алгоритма оценивания параметров, характеризующих смесь нормальных распределений. Если это так, то на помощь может прийти подход и результаты из разд. 2.2.2 [2], заключающиеся в использовании уточненной (невыврожденной) модели многомерного нормального распределения, для которой введены ограничения на множество возможных значений ковариационной матрицы.

Пусть плотность распределения смеси нормальных распределений есть

$$f(\mathbf{u}) = \sum_{j=1}^M \pi_j \varphi(\mathbf{u}, \mathbf{a}_j, \mathbf{C}_j),$$

где $\varphi(\mathbf{u}, \mathbf{a}_j, \mathbf{C}_j)$ — плотность нормального распределения со средним \mathbf{a}_j и ковариационной матрицей \mathbf{C}_j . При этом все собственные значения $v_i(\mathbf{C}_j)$ ковариационных матриц \mathbf{C}_j ограничены снизу некоторой положительной константой v_0 , т. е. $v_i(\mathbf{C}_j) \geq v_0 > 0, i = 1, \dots, d$. Доказано, что при этих условиях на t -м шаге итерации EM-алгоритма максимум функции $\sum_{j=1}^M \sum_{n=1}^N q_{nj}^{(t)} \ln \varphi(\mathbf{y}_j, \mathbf{a}_j, \mathbf{C}_j)$ достигается при значениях параметров, которые для каждого значения допустимого j последовательно находятся следующим образом (реализация M-шага):

(1) вычислить

$$\mathbf{a}_j^{(t+1)} = \frac{\sum_{n=1}^N q_{nj}^{(t)} \mathbf{y}_n}{\sum_{n=1}^N q_{nj}^{(t)}};$$

(2) вычислить

$$\tilde{\mathbf{C}}_j = \frac{\sum_{n=1}^N q_{nj}^{(t)} (\mathbf{y}_n - \mathbf{a}_j^{(t+1)}) (\mathbf{y}_n - \mathbf{a}_j^{(t+1)})^T}{\sum_{n=1}^N q_{nj}^{(t)}};$$

(3) найти матрицы $\tilde{\mathbf{U}}_j$ и $\tilde{\mathbf{V}}_j$, которые задают спектральное разложение матрицы $\tilde{\mathbf{C}}_j$;

(4) определить элементы $v_{jl}^{(t+1)}$ диагональной матрицы $\mathbf{V}_j^{(t+1)}$ через элементы \tilde{v}_{jl} матрицы $\tilde{\mathbf{V}}_j$ по формулам $v_{jl}^{(t+1)} = \max\{\tilde{v}_{jl}, v_0\}, l = 1, \dots, d$;

(5) вычислить $\mathbf{C}_j^{(t+1)} = \tilde{\mathbf{U}}_j \mathbf{V}_j^{(t+1)} (\tilde{\mathbf{U}}_j)^T$.

Заметим, что в рассматриваемом случае пространство параметров ограничено и максимум функции правдоподобия может лежать на границе. В силу этого полученные оценки не подпадают под обычные условия о сходимости EM-алгоритма (см., например, [3]). Требуемые результаты были получены в [4].

Задание ограничения снизу на собственные значение ковариационных матриц необходимо для предотвращения появления недопустимо больших (малых) значений функции правдоподобия. При этом возникает необходимость выбора этого ограничения v_0 . С одной стороны, значение v_0 должно быть достаточно большим, чтобы обеспечить корректное выполнение операций с плавающей точкой. С другой стороны, неразумное увеличение этого значения может дать снижение качества классификации данных на основе модели смеси (например, слишком большие значения v_0 могут привести к потере индивидуальности отдельных элементов смеси).

3 Последствия неправильного выбора размерности

Исследуем влияние ошибочного представления о реальной модели данных РРСА, принятой при классификации. Подобная постановка задачи актуальна в связи с вопросом, может ли снижение размерности данных на основе анализа главных компонент привести к повышению качества классификации данных.

Аналогичная ситуация рассматривалась в [1]. На примере задачи распознавания рукописных цифр исследовалась эффективность представления плотности распределения данных с помощью модели смеси РРСА. Было продемонстрировано снижение ошибочной классификации за счет рассмотрения не просто нормального распределения, а смеси нормальных распределений (увеличение числа элементов смеси с $M = 1$ до 10) и снижения размерности данных (с $d = 64$ до 10). Но данное улучшение, скорее всего, является просто результатом использования смеси, а выбор малых значений k лишь обеспечивает снижение вычислительной сложности, но не ясно, как он влияет на качество классификации. Таким образом, фактически было лишь показано, что качество классификации повышается в связи с усложнением модели данных.

Чтобы составить представление о поведении качества классификации с учетом РРСА, рассмотрим различие двух классов ω_1 и ω_2 нормально распределенных данных $N(\mathbf{a}_1, \mathbf{C}_1(k))$ и $N(\mathbf{a}_2, \mathbf{C}_2(k))$, формируемых в соответствии с моделью РРСА(k). При этом байесовский классификатор будет строиться с помощью модели РРСА(q). В этом случае он будет обучаться по выборке \mathbf{Y}_n объема n в соответствии с моделью РРСА(q), что даст для каждого класса оценки $\hat{\mathbf{a}}_i$ и $\hat{\mathbf{C}}_i(q)$, $i = 1, 2$. Поэтому решающая функция для некоторого вектора \mathbf{x} примет вид:

$$d_i(\mathbf{x}) = \ln \pi_i - \frac{1}{2} \ln |\hat{\mathbf{C}}_i(q)| - \frac{1}{2} (\mathbf{x} - \hat{\mathbf{a}}_i)^T \hat{\mathbf{C}}^{-1}(q) (\mathbf{x} - \hat{\mathbf{a}}_i),$$

где π_i — вероятности появления классов.

Для построенного классификатора теперь можно найти условную вероятность ошибки классификатора P_e или ее оценку \hat{P}_e , т.е. реализовать следующие шаги:

- генерирование выборки \mathbf{Y}_n для $\pi_1 N(\mathbf{a}_1, \mathbf{C}_1(k)) + \pi_2 N(\mathbf{a}_2, \mathbf{C}_2(k))$;
- обучение классификатора в соответствии с моделью РРСА(q), т.е. нахождение на основе сгенерированной выборки \mathbf{Y}_n оценок параметров модели РРСА(q) и с их помощью $\hat{\mathbf{C}}_i(q)$;
- нахождение P_e или \hat{P}_e .

Данные шаги могут быть многократно повторены для различных выборок u_n , что позволит найти оценки безусловных характеристик вероятности ошибки классификации.

Нахождение \hat{P}_e можно реализовать с помощью метода моделирования, т.е. реализовать следующие шаги:

- генерирование обучающей выборки \mathbf{X}_m для $f(\mathbf{u}) = \hat{\pi}_1 N(\hat{\mathbf{a}}_1, \hat{\mathbf{C}}_1(q)) + \hat{\pi}_2 N(\hat{\mathbf{a}}_2, \hat{\mathbf{C}}_2(q))$;
- классификация элементов \mathbf{X}_m на основе $f(\mathbf{u})$, затем получение значения \hat{P}_e путем сравнения смоделированной и оцененной классификаций.

С помощью некоторого упрощения постановки задачи классификации удается добиться того, что P_e можно найти аналитически, в частности, обобщая прием разд. 4.4 [5].

Действительно, рассмотрим различие двух классов и нормально распределенных данных с различными векторами средних \mathbf{a}_1 и \mathbf{a}_2 , но одинаковыми ковариационными матрицами. Пусть наблюдения формируются в соответствии с моделью РРСА(k), а для их классификации используется модель РРСА(q). Обучение байесовского классификатора по выборке \mathbf{Y}_n в соответствии с моделью РРСА(q) дает оценку $\hat{\mathbf{C}}(q)$. После этого классификация некоторого вектора \mathbf{x} осуществляется с помощью функции

$$u_{12}(\mathbf{x}) = 2\mathbf{x}^T \hat{\mathbf{C}}^{-1}(q) (\mathbf{a}_1 - \mathbf{a}_2)^T - (\mathbf{a}_1 + \mathbf{a}_2)^T \hat{\mathbf{C}}^{-1}(q) (\mathbf{a}_1 - \mathbf{a}_2).$$

При единичной функции потерь и вероятностях появления классов π_i условие, определяющее принадлежность \mathbf{x} к ω_1 , имеет вид $u_{12}(\mathbf{x}) > t$, где $t = \ln(\pi_2/\pi_1)$. Случайная величина $u_{12}(\mathbf{x})$, где $\mathbf{x} \sim N(\mathbf{a}_i, \mathbf{C}(k))$ для $i = 1, 2$, как линейная комбинация нормально распределенных случайных величин имеет также нормальное распределение. Поэтому достаточно найти первые моменты распределений $u_{12}(\mathbf{x})$ для каждого из двух классов, а именно:

$$2E_i \{u_{12}(\mathbf{x})\} = \begin{cases} \rho, & i = 1; \\ -\rho, & i = 2, \end{cases}$$

где

$$\rho = (\mathbf{a}_1 - \mathbf{a}_2)^T \hat{\mathbf{C}}^{-1}(\mathbf{a}_1 - \mathbf{a}_2);$$

$$\begin{aligned} E_i \left\{ (u_{12}(\mathbf{x}) - E_i \{u_{12}(\mathbf{x})\})^2 \right\} &= (\mathbf{a}_1 - \mathbf{a}_2)^T \hat{\mathbf{C}}^{-1}(q) \times \\ &\times E_i \left\{ (\mathbf{x} - \mathbf{a}_i) (\mathbf{x} - \mathbf{a}_i)^T \hat{\mathbf{C}}^{-1}(q) (\mathbf{a}_1 - \mathbf{a}_2) \right\} = \\ &= (\mathbf{a}_1 - \mathbf{a}_2)^T \hat{\mathbf{C}}^{-1}(q) \cdot \mathbf{C}(k) \hat{\mathbf{C}}^{-1}(q) \cdot (\mathbf{a}_1 - \mathbf{a}_2) = v^2 \\ &\text{(по определению)}. \end{aligned}$$

В результате имеем:

$$\begin{aligned} P_e &= \pi_1 \Pr \{u_{12}(\mathbf{x}) < t | \omega_1\} + \pi_2 \Pr \{u_{12}(\mathbf{x}) > t | \omega_2\} = \\ &= \pi_1 \Phi \left(\frac{t - \rho/2}{\sqrt{v^2}} \right) + \pi_2 \left(1 - \Phi \left(\frac{t + \rho/2}{\sqrt{v^2}} \right) \right), \end{aligned}$$

где $\Phi(u)$ — функция стандартного нормального распределения.

При $\pi_1 = \pi_2 = 1/2$

$$P_e = \Phi \left(-\frac{\rho}{2\sqrt{v^2}} \right).$$

Если $q = k$, то $v^2 = \rho$ и получаем ранее известную формулу:

$$P_e = \Phi \left(-\frac{\sqrt{\rho}}{2} \right).$$

Для того чтобы составить представление о реальной зависимости качества классификации от знания фактической размерности данных, рассматривался случай: $d = 50; k = 5; \mathbf{a}_1 - \mathbf{a}_2 = (0, 1; \dots; 0, 1)^T$; \mathbf{C} — некоторая случайно выбранная ковариационная матрица. Параметры выборок были таковы: $n = 300; N_{\text{exp}} = 100$. Результаты моделирования позволили получить оценки 95%-ных доверительных интервалов для оценки вероятности ошибки байесовского классификатора при различных значениях q .

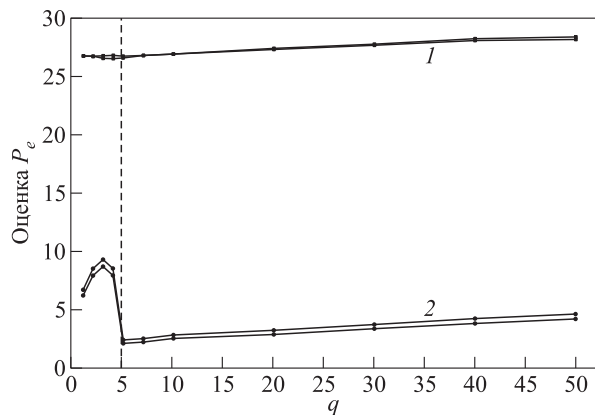
В первую очередь ковариационные матрицы для классов принимались одинаковыми:

$$\mathbf{C}_1 = \mathbf{C}_2 = (\mathbf{u}_1, \dots, \mathbf{u}_k) \mathbf{V}_k (\mathbf{u}_1, \dots, \mathbf{u}_k)^T.$$

Соответствующие результаты отражены на рисунке в виде практически слившихся линий с пометкой 1 (границы доверительных интервалов приблизительно равны).

Затем рассматривался случай ковариационных матриц:

$$\begin{aligned} \mathbf{C}_1 &= (\mathbf{u}_1, \dots, \mathbf{u}_k) \mathbf{V}_k (\mathbf{u}_1, \dots, \mathbf{u}_k)^T; \\ \mathbf{C}_2 &= (\mathbf{u}_2, \dots, \mathbf{u}_{k+1}) \mathbf{V}_k (\mathbf{u}_2, \dots, \mathbf{u}_{k+1})^T, \end{aligned}$$



Зависимость качества классификации P_e от размерности данных q , принятой при анализе данных

использующих несколько разные подпространства главных компонент и совпадающие дисперсии для них.

Соответствующие результаты представлены на рисунке уже в виде двух линий, помеченных как 2. При оценивании вероятности ошибки классификации бралась контрольная выборка объема $N_{\text{cont}} = 300$. Заметим, что только изменение взаимной ориентации главных компонент для классов приводит к существенному повышению качества классификации.

Продемонстрированный пример, а также множество дополнительных проведенных экспериментов позволяют сформулировать следующие результаты:

- оценка P_e как функция от размерности q для модели данных имеет минимум для $q = k$, т.е. при правильном задании реальной размерности данных классификатор обладает наилучшим качеством (на рисунке наилучшее значение выделено вертикальной штриховой прямой);
- превышение наилучшего значения q обычно ухудшает качество классификации в меньшей степени, чем его занижение (на рисунке это более ярко проявляется для графика 2).

4 Заключение

Внимание к модели смеси в рамках РРСА определяется в первую очередь необходимостью повышения эффективности сжатия и восстановления данных. Но эта модель позволяет также детализировать описание реальных данных и тем самым создать предпосылки для повышения качества классификации. Как показывают проведенные эксперименты, вероятность ошибок классификатора может снижаться достаточно существенно.

Модель смеси нормальных распределений является основой для популярного подхода к комбинированной оценке плотности.

Однако такая модель обладает следующим недостатком: если каждая гауссовская компонента описывается полной ковариационной матрицей, то для каждой компоненты смеси должны оцениваться $d(d + 1)/2$ отдельных ковариационных параметров. Очевидно, что по мере роста размерности пространства данных, да еще при естественном желании увеличить число элементов смеси, количество точек данных, необходимых для надежного определения этих параметров, становится непомерно высоким. Альтернативный подход заключается в уменьшении числа параметров путем введения ограничения на форму ковариационной

матрицы (другой прием состоял во введении предположений о параметрах полной ковариационной матрицы [6]). При этом обычно используются два общих ограничения: задать матрицу ковариаций изотропной или диагональной.

Изотропная модель сильно ограничена, поскольку она присваивает только один параметр для описания всей структуры ковариации для полно-размерных данных. Диагональная модель более гибкая, с d параметрами, но главные оси эллипсоидов для элементов смеси должны быть выровнены с осями данных, и, таким образом, каждый отдельный элемент смеси не способен описывать корреляции между переменными. Поэтому смесь моделей РРСА, где ковариации каждого элемента параметризуются с помощью соотношения $\mathbf{C} = \sigma^2 \mathbf{I} + \mathbf{W}\mathbf{W}^T$, может содержать существенно меньшее число параметров.

Одним из преимуществ методологии РРСА является то, что определение модели плотности позволяет вычислять апостериорные вероятности принадлежности некоторого наблюдаемого значения элементу смеси (некоторому классу) и использовать их для последующей классификации, не находя ошибку восстановления.

Возможным недостатком вероятностного подхода к объединению локальных моделей РСА является то, что, оптимизируя функцию правдоподобия, модель смеси РРСА напрямую не минимизирует квадратичную ошибку реконструкции. Для приложений, где это ключевой критерий, следует ожидать, что алгоритмы, которые явно минимизируют ошибку восстановления, будут эффективнее. Эксперименты действительно показали, что это, как правило, имеет место, но важны две оговорки, прежде чем можно будет сделать какие-либо твердые выводы относительно пригодности данной модели.

Во-первых, есть задачи, где окончательная модель смеси РРСА оказывалась фактически лучше в смысле ошибки реконструкции, даже на обучающем наборе.

Второе соображение заключается в том, что имеются также свидетельства того, что сглаживание, подразумеваемое мягкой кластеризацией, присущей модели смеси РРСА, помогает уменьшить переобучение, особенно в случае эксперимента сжатия данных, где статистика набора тестовых данных отличается от данных обучения гораздо больше, чем для других примеров (см., например, [1]).

В терминах модели гауссовой смеси смесь вероятностных анализаторов главных компонент позволяет моделировать данные больших размеров с относительно небольшим числом свободных параметров, не налагая в целом неуместного ограни-

чения на структуру ковариации. Число свободных параметров можно контролировать с помощью выбора скрытой пространственной размерности k , что позволяет проводить интерполяцию по сложности модели от изотропной до полной ковариационной структуры.

Литература

1. Tipping M. E., Bishop C. M. Mixtures of probabilistic principal component analyzers // *Neural Comput.*, 1999. Vol. 11. Iss. 2. P. 443–482.
2. Кривенко М. П. Прикладные методы оценивания распределения многомерных данных малой выборки. — М.: ИПИ РАН, 2011. 146 с.
3. Wu C. F. J. On convergence properties of the EM algorithm // *Ann. Stat.*, 1983. Vol. 11. P. 95–103.
4. Nettleton D. Convergence properties of the EM algorithm in constrained parameter spaces // *Can. J. Stat.*, 1999. Vol. 27. Iss. 3. P. 639–648.
5. Ту Дж., Гонсалес Р. Принципы распознавания образов / Пер. с англ. — М.: Мир, 1978. 414 с. (Tou J., Gonzalez R. C. Pattern recognition principles. — Reading, MA, USA: Addison-Wesley Publ. Co., 1974. 377 p.)
6. Ormoneit D., Tresp V. Improved gaussian mixture density estimates using Bayesian penalty terms and network averaging // *Advances in neural information processing systems* / Eds. D. S. Touretzky, M. C. Mozer, M. E. Hasselmo. — Cambridge, MA, USA: MIT Press, 1996. Vol. 8. P. 542–548.

Поступила в редакцию 30.05.18

SUPERVISED LEARNING CLASSIFICATION OF DATA TAKING INTO ACCOUNT PRINCIPAL COMPONENT ANALYSIS

M. P. Krivenko

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

Abstract: The article examines questions of supervised learning classification of data taking into account principal component analysis (PCA) results. Construction of a Bayesian classifier becomes possible after representation of covariances through the parameters of the probabilistic PCA model. The case of singular data distributions is singled out; for this case, it is suggested to estimate the parameters of the model under constraints on the eigenvalues of covariance matrices. The quality of classification is studied in respect to the actual data dimension. It is demonstrated that, when correctly assigned, the classifier has the least error probabilities. Exceeding the best value of the dimension usually worsens the quality of the classification to a lesser extent than its underestimation. The mixture of probabilistic principal component analyzer allows modeling big data by means of a relatively small number of free parameters. The number of free parameters can be controlled by choosing the latent dimension of the data.

Keywords: principal component analysis; mixtures of normal distributions; EM algorithm; supervised learning classification

DOI: 10.14357/19922264180308

References

1. Tipping, M. E., and C. M. Bishop. 1999. Mixtures of probabilistic principal component analyzers. *Neural Comput.* 11(2):443–482.
2. Krivenko, M. P. 2011. *Prikladnye metody otsenivaniya raspredeleniya mnogomernykh dannykh maloy vyborki* [Applied methods for estimating the distribution of small sample multidimensional data]. — Moscow: IPI RAN. 146 p.
3. Wu, C. F. J. 1983. On convergence properties of the EM algorithm. *Ann. Stat.* 11:95–103.
4. Nettleton, D. 1999. Convergence properties of the EM algorithm in constrained parameter spaces. *Can. J. Stat.* 27(3):639–648.
5. Tou, J., and R. C. Gonzalez. 1974. *Pattern recognition principles*. Reading, MA: Addison-Wesley Publ. Co. 377 p.
6. Ormoneit, D., and V. Tresp. 1996. Improved gaussian mixture density estimates using Bayesian penalty terms and network averaging. Eds. D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo. *Advances in neural information processing systems*. Cambridge, MA: MIT Press. 8:542–548.

Received May 30, 2018

Contributor

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

ПАРАМЕТРИЗАЦИЯ В ПРИКЛАДНЫХ ЗАДАЧАХ ПОИСКА ЭМПИРИЧЕСКИХ ПРИЧИН*

А. А. Грушо¹, Н. А. Грушо², М. И. Забейайло³, Д. В. Смирнов⁴, Е. Е. Тимонина⁵

Аннотация: Представление конечного класса объектов в форме множества характеристик (параметров) этих объектов назовем параметризацией рассматриваемого класса. Кроме идентификации объектов множествами характеристик существует задача выявления причины того, что некоторые объекты класса обладают свойством P . Для решения этой задачи в условиях появления новых объектов исходного множества характеристик может не хватить. В этом случае необходимо изменять параметризацию. Работа посвящена построению методов изменения начальной параметризации в задаче уточнения эмпирической причины появления свойства P при расширении исходных данных. Построенные методы продемонстрированы на практических примерах.

Ключевые слова: ДСМ-методы искусственного интеллекта; параметризация классов объектов; эмпирическая причина; аутентификация

DOI: 10.14357/19922264180309

1 Введение

Работа посвящена исследованию следующей проблемы. Если в классе объектов часть их обладает свойством P , а часть не обладает этим свойством, то возникает задача выявления причины появления свойства P у части объектов. Объекты описываются множеством характеристик этих объектов, или параметров — по Эшби [1]. Поэтому можно отождествлять объекты и соответствующие множества их характеристик.

Множество характеристик полно, если каждый объект однозначно выделяется по своему подмножеству характеристик. Однако при исследовании причины появления свойства P в некоторых объектах можно прийти к выводу, что полнота множества характеристик не гарантирует описания причин появления свойства P . Тогда необходимо изменять параметризацию класса объектов (множество параметров описания объектов), уточняя ее в такой степени, чтобы в этом описании содержались причины появления свойства P .

В работе рассматривается несколько способов расширения множества характеристик данного класса объектов при сохранении свойства полноты.

2 Эмпирические причины

Рассмотрим простейшую модель ДСМ-метода [2] интеллектуального анализа данных (ИАД), построенную на языке теории множеств. Пусть $U = \{u_1, \dots, u_m\}$ является множеством характеристик наблюдаемых объектов O_1, \dots, O_n , т.е. объекты полностью описываются наборами характеристик из U . Можно считать, что любой объект O — это подмножество U , а множество всех возможных объектов — это множество всех подмножеств множества U . Само множество U будем называть параметризацией наблюдаемых объектов. Кроме характеристик определим понятие свойства объекта. Свойство P объекта O отражает некоторую интегральную характеристику объекта O , которая может присутствовать в O или отсутствовать в O . Если объекты появляются последовательно, то вновь появившийся объект должен проверяться на наличие свойства P .

Предположим, что рассматриваемое свойство P удовлетворяет следующим условиям:

(А) если свойство P выявлено в объектах O_1 и O_2 , то оно есть в объекте $O = O_1 \cap O_2$;

*Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00274-а, 15-29-07981-офи-м).

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

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

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

⁴ПАО Сбербанк России, dvlsmirnov@sberbank.ru

⁵Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

(В) если объект O обладает свойством P , то для любого объекта O_1 , содержащего O , объект O_1 также обладает P .

Определим понятие эмпирической причины некоторого свойства P в наблюдаемых данных. Основываясь на понимании причины Д. С. Миллем [3], в рассматриваемой модели эмпирической причиной свойства P для множества объектов O_1, O_2, \dots, O_l будем называть производный объект $O = \bigcap_i O_i$ по всем индексам i , для которых O_i обладает свойством P , но ни один из остальных O_j свойством P не обладает. Причина должна быть единственной, и от нее ничего нельзя убавить или к ней добавить [4].

Пример 1. Пусть $U = \{а, б, \dots, я\}$ — русский алфавит. Свойство P означает множество букв, из которых можно собрать слово, выражающее понятие «дом». Пусть есть два объекта $O_1 = \{д, о, м\}$ и $O_2 = \{с, у, к\}$. Ясно, что O_1 является эмпирической причиной свойства P . При этом O_2 не обладает свойством P .

3 Различные параметризации

Однозначное представление объектов в виде подмножеств множества характеристик не означает, что не существует другого множества характеристик для описания того же множества наблюдаемых объектов и свойств.

Пример 2. Понятие «дом» можно выразить на английском языке словом *home*. Тогда множество характеристик есть английский алфавит, а аналогичное свойство примера 1 выражается другой причиной, выраженной другим множеством характеристик.

Представление объектов из множества характеристик можно строить с помощью упорядоченных наборов характеристик. Тогда множество описываемых объектов совпадает с множеством слов конечной длины U^* из исходного множества характеристик. Это множество бесконечно и $U^{**} = U^*$. Эмпирическая причина для U^* определяется словом, описывающим интегральное свойство P в языке U^* . При этом свойство P примера 1 выражается проще. Объект «дом» обладает свойством P , а объект «мода» свойством P уже не обладает. Однако в данном представлении объектов может возникнуть ситуация, когда эмпирическая причина отсутствует.

Пример 3. Объект «дом» соответствует понятию, описанному словом *дом*, и является эмпирической

причиной свойства P , а объект «строение» не является словом *дом*, но соответствует смыслу свойства P . Таким образом, возникает объект, не обладающий эмпирической причиной и одновременно обладает свойством P .

Чаще всего добавление новых объектов фальсифицирует эмпирическую причину, найденную по исходному набору объектов. Фальсификация происходит в следующих формах:

- (1) найденная эмпирическая причина появляется в новом объекте, который не обладает свойством P ;
- (2) найденная эмпирическая причина не появляется в новом объекте, который, как предполагается, обладает свойством P .

Рассмотрим некоторые другие примеры изменения параметризации, позволяющие тоньше идентифицировать причины их фальсификации.

4 Схема аутентификации как поиск причины в модифицированной параметризации

В данном разделе рассматривается задача изменения параметризации за счет добавления множеств характеристик из различных информационных пространств.

Пусть A — субъект, который должен подтвердить субъекту W свое имя. Эта процедура называется аутентификацией [5] и может быть проведена следующим образом.

Пусть Σ_1 — это информационное пространство, содержащее объекты $O_1^{(1)}, O_2^{(1)}, \dots, O_{s_1}^{(1)}$. Каждый из этих объектов классифицируется двумя значениями t (truth) и f (false). Свойство P состоит в знании для этих объектов правильного вектора $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_{s_1}^{(1)})$, где

$$x_i^{(1)} = \begin{cases} t; \\ f \end{cases} \quad \text{для } i = 1, \dots, s_1.$$

Пусть вектор $x^{(1)}$ известен A и W . Предположим, что существует субъект B , который назвал себя именем A и также пытается подтвердить это имя. Возможны два случая.

1. B случайно выбирает значение вектора $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_{s_1}^{(1)})$.
2. Пространство Σ_1 известно B , и он знает вектор $x^{(1)}$.

Субъект W проверяет знание вектора $x^{(1)}$, считая что он общается с некоторым субъектом Φ , который может принимать значения $\Phi = A$ или B . Субъект W предъявляет объекты $O_1^{(1)}, O_2^{(1)}, \dots, O_{s_1}^{(1)}$ субъекту Φ . Тогда, получив вектор $x^{(1)}$, субъект W может считать, что субъект Φ знает эмпирическую причину свойства P . При этом эту причину может знать как A , так и B . Причем B может случайно угадать вектор $x^{(1)}$ (ложная аутентификация).

Проверка имени A может быть продолжена, если к информационному пространству Σ_1 добавить информационные пространства $\Sigma_2, \dots, \Sigma_m$. Из характеристик каждого из добавленных пространств можно сформировать объекты $O_1^{(i)}, O_2^{(i)}, \dots, O_{s_i}^{(i)}$, $i = 2, \dots, m$, для которых W и A знают векторы $x^{(i)}$, $i = 2, \dots, m$. Таким образом, пара субъектов W и A расширяет параметризацию для определения причин свойства P до знания векторов $x^{(1)}, \dots, x^{(m)}$.

Если субъект B выбирает значение вектора $x_j^{(i)}$ случайно, то с вероятностью, как угодно близкой к 1, он на каком-то шаге ошибется. Это позволит субъекту W понять, что субъект Φ не обладает свойством P .

Если субъект B скомпрометировал пространства $\Sigma_{i_1}, \dots, \Sigma_{i_r}$, $r < m$, то субъект W знает о возможной компрометации каких-то информационных пространств, но не знает каких. Тогда субъект W предъявляет Φ объекты, созданные в информационных пространствах $\Sigma_1, \Sigma_2, \dots, \Sigma_m$, попадает на объекты нескомпрометированных пространств и определяет фальсификацию эмпирической причины P . Если субъект Φ определяет правильно все векторы $x^{(1)}, \dots, x^{(m)}$, то причина P подтверждена и аутентификация прошла успешно.

Рассмотренный пример показывает, что расширение множества характеристик за счет привлечения дополнительной информации позволяет уточнять причину исследуемого свойства P . Идея подтверждения эмпирической причины за счет расширения исходной параметризации с помощью добавления характеристик рассматривалась в работах [6, 7].

5 Изменение параметризации за счет разбиения параметров

Пусть $U = \{u_1, \dots, u_m\}$ — это хосты сети. Пусть свойство P соответствует обработке информационной технологии за время $\tau > T_0$. Пусть объект O — это множество хостов, выделяемое провайдером для реализации информационной технологии. При повторях технологии выделяется множество объектов $\{O^+\}$, на которых наблюдаются задержки

$\tau > T_0$ выполнения информационной технологии, и множество объектов $\{O^-\}$, для которых $\tau \leq T_0$, т. е. в сети возникает так называемая «мерцающая» ошибка [8, 9].

Предположим, что причина o свойства P ищется с помощью пересечения объектов из множества $\{O^+\}$. Эмпирическая причина является под-объектом o , если o не встречается в множестве $\{O^-\}$. Пусть к множеству объектов добавляется новый объект O и пусть $o \subseteq O$, но информационная технология реализуется так, что $\tau \leq T_0$. Таким образом, происходит фальсификация эмпирической причины.

При исследовании характеристик U оказалось, что u_1 — это локальная сеть из двух машин $u_1^{(1)}$ и $u_1^{(2)}$ с прокси-сервером для выхода в общую сеть. Для реализации информационной технологии любая из этих машин выбирается случайно. Оказывается, что $u_1^{(2)}$ работает всегда быстро, а $u_1^{(1)}$ — всегда медленно. Рассмотрим новое пространство характеристик $U' = \{u_1^{(1)}, u_1^{(2)}, u_2, \dots, u_m\}$. Тогда в новом множестве характеристик причина свойства $P = (\tau > T_0)$ определяется тем же методом, что и ранее, но эмпирическая причина не фальсифицируется.

6 Заключение

Выявление устойчивых при увеличении исходных данных эмпирических причин исследуемых свойств не всегда возможно в условиях исходной параметризации. Изменение параметризации требует выполнения полноты описания известных и вновь поступающих объектов.

Возможно дополнение изменения параметризации требованием сохранения описания эмпирических причин других свойств, поэтому необходимо строить новую параметризацию путем уточнения старой параметризации. В работе рассмотрены примеры построения таких уточняющих параметризаций. Возможны гибридные варианты применения изложенных методов.

В дальнейшем предполагается исследовать возможность доказательства того, что других путей уточнения параметризации нет.

Литература

1. Эшби У. Р. Конструкция мозга / Пер. с англ. — М.: Иностранная литература, 1962. 397 с. (*Ashby W. R. Design for a brain.* — New York, NY, USA: Wiley, 1954. 260 p.)
2. Анишаков О. М., Фабрикантова Е. Ф. ДСМ-метод автоматического порождения гипотез: Логические и эпи-

- стемологические основания. — М.: Либроком, 2009. 432 с.
3. *Милль Дж. С.* Система логики силлогической и индуктивной: Изложение принципов доказательства в связи с методами научного исследования / Пер. с англ. — 5-е изд., испр. и доп. — М.: ЛЕНАНД, 2011. 832 с. (*Mill J. S.* A system of logic ratiocinative and inductive, being a connected view of the principles of evidence and the methods of scientific investigation. — 1st ed. — London: John W. Parker, 1843. 622 p.)
 4. *Финн В. К.* Искусственный интеллект: Методология, применения, философия. — М.: Красанд, 2011. 448 с.
 5. *Грушо А. А., Применко Э. А., Тимонина Е. Е.* Теоретические основы компьютерной безопасности. — М.: Академия, 2009. 272 с.
 6. *Грушо А. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е.* О комплексной аутентификации // Системы и средства информатики, 2017. Т. 27. № 3. С. 3–10.
 7. *Грушо А. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е.* Модель множества информационных пространств в задаче поиска инсайдера // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 65–69.
 8. *Грушо А. А., Забежайло М. И., Зацаринный А. А., Николоаев А. В., Писковский В. О., Тимонина Е. Е.* Классификация ошибочных состояний в распределенных вычислительных системах и источники их возникновения // Системы и средства информатики, 2017. Т. 27. № 2. С. 30–41.
 9. *Грушо А. А., Забежайло М. И., Зацаринный А. А., Николоаев А. В., Писковский В. О., Сенчило В. В., Сударинов И. В., Тимонина Е. Е.* Об анализе ошибочных состояний в распределенных вычислительных системах // Системы и средства информатики, 2018. Т. 28. № 1. С. 99–109.

Поступила в редакцию 10.06.18

PARAMETRIZATION IN APPLIED PROBLEMS OF SEARCH OF EMPIRICAL REASONS

A. A. Grusho¹, N. A. Grusho¹, M. I. Zabezhailo², D. V. Smirnov³, and E. E. Timonina¹

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

²A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119133, Russian Federation

³Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation

Abstract: The authors define description of a finite class of objects in the form of a set of characteristics (parameters) of these objects as parametrization of the considered class. Besides identification of objects, the problem of the causality that some objects have property P exists. For the solution of this task, in the conditions of emergence of new objects, an initial set of characteristics cannot be enough. In this case, it is necessary to change parametrization. The paper is devoted to creation of methods of changes of the initial parametrization in the problem of specification of the empirical causality of emergence of the property P at expansion of basic data. The constructed methods are shown on practical examples.

Keywords: JSM-methods of artificial intelligence; parametrization of classes of objects; empirical reason; authentication

DOI: 10.14357/19922264180309

Acknowledgments

The paper was partially supported by the Russian Foundation for Basic Research (projects 18-07-00274-a and 15-29-07981-ofi-m).

References

1. Ashby, W. R. 1954. *Design for a brain*. New York, NY: Wiley. 260 p.
2. Anshakov, O. M., and E. F. Fabrikantova. 2009. *DSM-metod avtomaticheskogo porozhdeniya gipotez: Logicheskie i epistemologicheskie osnovaniya* [JSM-method of automatic hypothesis generation: Logical and epistemological]. Moscow: Librokom. 432 p.
3. Mill, J. S. 1843. *A system of logic ratiocinative and inductive, being a connected view of the principles of evidence and the methods of scientific investigation*. 1st ed. London: John W. Parker. 622 p.

4. Finn, V. K. 2011. *Iskusstvennyy intellekt: Metodologiya, primeneniya, filosofiya* [Artificial intelligence: Methodology, applications, philosophy]. Moscow: Krasand. 448 p.
5. Grusho, A., Ed. Primenko, and E. Timonina. 2009. *Teoreticheskie osnovy komp'yuternoy bezopasnosti* [Theoretical bases of computer security]. Moscow: Academy. 272 p.
6. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. O kompleksnoy autentifikatsii [About complex authentication]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(3): 3–10.
7. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primneneniya — Inform. Appl.* 11(4):65–69.
8. Grusho, A. A., M. I. Zabezhailo, A. A. Zatsarinnyy, A. V. Nikolaev, V. O. Piskovski, and E. E. Timonina. 2017. Klassifikatsiya oshibochnykh sostoyaniy v raspredelennykh vychislitel'nykh sistemakh i istochniki ikh vozniknoveniya [Erroneous states classification in distributed computing systems and sources of their occurrence]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(3):30–41.
9. Grusho, A. A., M. I. Zabezhailo, A. A. Zatsarinnyy, A. V. Nikolaev, V. O. Piskovski, V. V. Senchilo, I. V. Sudarikov, and E. E. Timonina. 2018. Ob analize oshibochnykh sostoyaniy v raspredelennykh vychislitel'nykh sistemakh [About the analysis of erratic statuses in the distributed computing systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(1):99–109.

Received June 10, 2018

Contributors

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

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

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, associate professor; principal scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119133, Russian Federation; m.zabezhailo@yandex.ru

Smirnov Dmitry V. (b. 1984) — business partner for IT security department of Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation; dvlsmirnov@sberbank.ru

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

МЕТОДЫ И СРЕДСТВА ОБНАРУЖЕНИЯ НЕШТАТНЫХ СИТУАЦИЙ, ВОЗНИКАЮЩИХ НА ЭЛЕМЕНТАХ ЖИЛИЩНО-КОММУНАЛЬНОЙ ИНФРАСТРУКТУРЫ*

И. А. Шанин¹, С. А. Ступников², В. Н. Захаров³

Аннотация: Работа относится к области разработки специализированных информационных систем на основе технологии Интернета вещей. Рассматривается подход к программной реализации компонентов модуля обнаружения нештатных ситуаций как составной части информационной системы, обеспечивающей поддержку технического обслуживания элементов жилищной инфраструктуры в целях оперативного контроля их состояния, предиктивного ремонта и оповещения о возникающих нештатных ситуациях и регулярных событиях. Описаны алгоритмы и особенности реализации компонентов, осуществляющих построение моделей функционирования элементов жилищной инфраструктуры и обнаружение нештатных ситуаций. Приводятся экспериментальные результаты применения подхода для обнаружения нештатных ситуаций на модельных наборах данных.

Ключевые слова: Интернет вещей; анализ данных; обнаружение нештатных ситуаций; жилищно-коммунальная инфраструктура

DOI: 10.14357/19922264180310

1 Введение

Активное развитие технологии Интернета вещей, предполагающей, в частности, оснащение физических предметов средствами взаимодействия друг с другом или с внешней средой, открывает перспективы появления и совершенствования новых классов информационных систем для решения насущных задач различного рода. Например, в области жилищно-коммунальной инфраструктуры оснащение элементов инфраструктуры (таких как трансформаторы, насосы, кондиционеры и др.) разнообразными датчиками — температуры, тока, напряжения, звука — позволяет собирать с них данные и осуществлять их анализ с целью поддержки технического обслуживания.

Данная работа проводится в рамках проекта, нацеленного на создание комплекса программно-технических решений по созданию информационной системы, обеспечивающей поддержку технического обслуживания элементов жилищной инфраструктуры в целях оперативного контроля их состояния, предиктивного ремонта и оповещения о возникающих нештатных ситуациях и ре-

гулярных событиях. К нештатным ситуациям относятся, в частности, отказ элементов жилищной инфраструктуры в эксплуатационный период, заявленный производителем; к регулярным событиям — техническое обслуживание и замена элементов по истечении эксплуатационного периода. В рамках проекта рассматриваются такие элементы жилищной инфраструктуры, как трансформаторы, насосы, вентиляционные установки и кондиционеры, газовые котлы, трубчатые электронагреватели (ТЭНы), электрические лампы.

В работе [1] авторами была предложена архитектура информационной системы, включающая модули сбора данных, хранения данных, обнаружения и предсказания нештатных ситуаций, информирования пользователя о произошедших и потенциально возможных нештатных ситуациях. Одной из основных составляющих архитектуры является модуль обнаружения и предсказания нештатных ситуаций, включающий, в частности, компоненты построения моделей функционирования элементов жилищной инфраструктуры и обнаружения нештатных ситуаций [1]. В данной работе рассматривается подход к программной реализации упо-

*Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (уникальный идентификатор проекта RFMEFI60717X0176).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, v08shinin@gmail.com

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

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vazakharov@ipiran.ru

мянутых компонентов модуля. Описаны алгоритмы функционирования и особенности реализации компонентов модуля (разд. 2). Приводятся экспериментальные результаты применения подхода для обнаружения нештатных ситуаций на модельных наборах данных (разд. 3).

2 Алгоритмы функционирования компонентов модуля обнаружения нештатных ситуаций

2.1 Построение моделей функционирования элементов жилищной инфраструктуры

Компонент реализует следующие подходы к анализу временных рядов: сезонная интегрированная модель авторегрессии — скользящего среднего (SARIMA — seasonal autoregressive integrated moving average) и скрытые марковские модели (HMM — hidden Markov models).

Модель SARIMA. Являясь обобщением классической модели авторегрессии — скользящего среднего (ARMA — autoregressive moving average) [2], данная модель находит массу применений в анализе временных рядов, в частности в задаче прогнозирования и интерполяции значений. Модель использует понятие тренд-сезонного разложения временного ряда [3], представляя наблюдаемые значения в виде суммы (или произведения) трех компонент: *тренда*, *сезонной составляющей* и *остатков*. Трендовая составляющая отражает долгосрочное изменение показателей, сезонная составляющая отражает периодические колебания временного ряда. Остатки не имеют ни сезонных, ни трендовых особенностей. Для обучения модели необходимо оценить гиперпараметры p, d, q, P, D, Q и s :

- целочисленные параметры p и q соответствуют порядкам моделей авторегрессии (AR — autoregression) и скользящего среднего (MA — moving average), их значение отображает число элементов временного ряда, участвующих в построении линейных моделей AR и MA;
- параметр d соответствует порядку дифференцирования, необходимому для удаления несезонных трендов и приведения ряда к стационарному виду;
- целочисленный параметр s соответствует периоду сезонности — периодических колебаний временного ряда;

- параметры P, D и Q имеют тот же смысл, что и p, d и q , но относятся к сезонной составляющей ряда.

Параметры p, d, q, P, D и Q принимают значения из множества $\{0, 1, 2, 3\}$. Для выбранных наборов значений гиперпараметров оцениваются коэффициенты моделей AR и MA.

Входные данные (`clean_data`) при построении модели представляют собой набор кортежей вида `(timestamp: time, value: real)` — временной ряд значений некоторого датчика, установленного на элементе жилищной инфраструктуры, соответствующий штатному режиму работы элемента. Построение модели включает следующие шаги:

- вывести на экран графики функций автокорреляции (ACF — autocorrelation function) и частичной автокорреляции (PACF — partial ACF) данных. По свойствам этих графиков можно сделать вывод о возможных значениях гиперпараметров модели, тем самым существенно ускорив время обучения;
- считать введенные пользователем множества возможных значений гиперпараметров `p_range, d_range, q_range, P_range, D_range` и `Q_range`, являющиеся подмножествами множества $\{0, 1, 2, 3\}$ и значение s^* параметра s ;
- для всех возможных наборов значений гиперпараметров p, d, q, P, D и Q и входных данных `clean_data` вычисляется значение информационного критерия Акаике (AIC — Akaike information criterion), выбираются конкретные значения гиперпараметров $p^*, d^*, q^*, P^*, D^*, Q^*$ и s^* , соответствующие минимальному значению AIC;
- обучить модель SARIMA на основании `clean_data` и значений $p^*, d^*, q^*, P^*, D^*, Q^*$ и s^* , при этом определить AR^* и MA^* — кортежи вещественных коэффициентов линейных моделей AR и MA;
- вычислить остатки построенной модели на входных данных `clean_data`;
- проверить статистическую гипотезу о нестационарности остатков временного ряда с использованием расширенного теста Дики–Фуллера (ADF — augmented Dickey–Fuller test). В случае если p -значение теста больше 0,05, вернуть предупреждение о неприменимости SARIMA, иначе вернуть набор параметров модели в виде кортежа $(p^*, d^*, q^*, P^*, D^*, Q^*, AR^*, MA^*)$.

Для реализации указанных шагов использовались функции пакета p^* , statsmodels¹ языка Python.

Модель НММ. Данная модель [4] основывается на гипотезе о том, что у анализируемой системы имеется конечное множество «скрытых» состояний. Таким образом, помимо последовательности значений анализируемого временного ряда существует скрытая последовательность состояний (система может принимать ровно одно состояние в каждый момент времени). Скрытая марковская модель характеризуется набором параметров $\langle T, N, P, M, A \rangle$, где T — множество возможных скрытых состояний; N — множество возможных наблюдаемых значений; P — множество условных вероятностных распределений $p(x|t_i)$ для каждого скрытого состояния t_i ; M — матрица размера $|T| \times |T|$, элементы M_{ij} которой являются вероятностями перехода системы из состояния t_i в состояние t_j ; A — вероятностное распределение первого состояния скрытой марковской модели (априорное распределение).

В данном подходе метки классов неисправностей в каждый момент времени будут напрямую соответствовать скрытым состояниям модели. Используются два принципиально разных способа применения скрытых марковских моделей: в первом случае активно используется обучающая выборка, в которой сигнал уже разделен на сегменты, соответствующие исправным и неисправным режимам работы; разметка используется на этапе обучения параметров модели по методу максимума правдоподобия. Второй способ не предполагает наличие размеченной выборки, в этом случае для обучения модели применяется алгоритм Баума–Уэлша (частный случай EM (expectation-maximization) алгоритма) [5]. Главным ограничением такого подхода является необходимость заранее оценить число типов различных нештатных ситуаций, а следовательно, применимость НММ заметно снижается в случае отсутствия размеченной обучающей выборки. На практике также оказывается эффективным смешанный подход: проводить инициализацию параметров модели с помощью обучающей выборки по методу максимума правдоподобия, чтобы потом улучшить параметры с помощью алгоритма Баума–Уэлша.

Входные данные (training_data) при построении модели представляют собой набор кортежей вида $\langle \text{value: real, hidden_state: int} \rangle$, где value — элемент последовательности значений анализируемого временного ряда, а hidden_state — разметка элементов последовательности по классам

аномалий. При реализации используется класс HiddenMarkovModel пакета pomegranate² языка Python; для обучения на неразмеченных данных используется вызов метода fit на входных данных с параметром algorithm = 'baum-welch'. Для обучения на размеченных данных используется вызов метода fit на входных данных с параметрами labels = training_data.hidden_state, algorithm = labeled'; а затем — вызов метода fit на входных данных с параметром algorithm = 'baum-welch'.

В результате обучения формируется модель, содержащая конкретные значения вышеупомянутых параметров $\langle T, N, P, M, A \rangle$.

2.2 Обнаружение нештатных ситуаций

Компонент получает на вход параметры построенных моделей SARIMA и НММ и анализируемые временные ряды показаний датчиков, прошедшие предобработку. Цель работы данного компонента — бинарная классификация показаний датчиков на показания, соответствующие штатной работе исправного оборудования, и показания, соответствующие различным нештатным ситуациям (аномалиям во временном ряду).

Обнаружение нештатных ситуаций при помощи модели SARIMA. Модель SARIMA позволяет прогнозировать следующее значение временного ряда, зная несколько предыдущих значений. Таким образом, для обнаружения аномалий требуется сравнить реальные значения временного ряда с границами доверительного интервала SARIMA. Входными данными при обнаружении аномалий являются кортеж $\text{model} = \langle p, d, q, P, D, Q, \text{AR}, \text{MA} \rangle$ параметров модели и набор data кортежей вида $\langle \text{timestamp: time, value: real} \rangle$. В качестве выходных данных выступает бинарный массив labels, по числу элементов совпадающий с data, элементы которого принимают значения из множества $\{0, 1\}$, где 0 означает штатность соответствующего элемента data, а 1 — его аномальность. Процедура обнаружения аномалий включает следующие шаги:

- инициализировать переменную filtered_data значением data;
- для каждого i начиная с $\max(p, q)$ до $|\text{data}| - 1$ на основании первых i элементов временного ряда filtered_data предсказать $(i + 1)$ -й элемент ряда с использованием модели model, вычислить предсказанное значение forecast_value и доверительный интервал confidence_interval;
- проверить вхождение значения $\text{data}[i].\text{value}$ в интервал confidence_interval. В случае вхож-

¹<http://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>.

²<https://pomegranate.readthedocs.io/en/latest/>.

дения присвоить 0 элементу $labels[i]$, иначе присвоить 1 элементу $labels[i]$ и присвоить значение $\langle data[i].timestamp, forecast_value \rangle$ элементу $filtered_data[i]$ для обеспечения возможности дальнейшего предсказания: фактически, при этом во временном ряду аномальное значение заменяется на штатное, предсказанное моделью.

Для реализации указанных шагов использовались функции пакета `statsmodels` языка Python.

Обнаружение нештатных ситуаций при помощи модели НММ. Здесь задача детектирования нештатных режимов работы оборудования рассматривается как задача сегментации сигнала: по известному набору параметров НММ требуется восстановить наиболее вероятную последовательность скрытых состояний системы. Данная задача решается при помощи алгоритма Витерби — классического алгоритма динамического программирования.

На входе процедура обнаружения аномалий получает обученную скрытую марковскую модель `model` и анализируемые данные `data` — набор кортежей вида $\langle timestamp: time, value: real \rangle$, соответствующие реальной (возможно, нештатной) работе оборудования — источника данных. При реализации используется вызов метода `predict` класса `HiddenMarkovModel` из пакета `pomegranate` языка Python на наборе `data` с параметром `algorithm = 'viterbi'`. Результатом работы является целочисленный массив `labels`, по числу элементов совпадающий с `data`, значения которого соответствуют скрытым состояниям (режимам работы) системы. Значение 0 соответствует штатному режиму работы, а все остальные — аномальным режимам.

3 Экспериментальные результаты

В рамках проекта RFMEFI60717X0176 для проведения экспериментов создан исследовательский стенд, развернутый на нескольких объектах жилищной инфраструктуры, включающий аппаратно-программные контроллеры и датчики (расходомеры, датчики входного/выходного тока/напряжения, температуры, звука, освещенности, давления), установленные на трансформаторах, водяных насосах, вентиляционных установках, кондиционерах, газовых котлах, ТЭНах, электрических лампах. В настоящее время продолжается сбор данных с компонентов исследовательского стенда для их последующего анализа. В рамках же данной статьи эксперименты проводятся на двух готовых наборах данных: Intel Lab Data и LUCE. Краткое описание наборов данных приведено ниже.

Данные исследовательской лаборатории Intel в Беркли. Набор данных состоит из показаний бытовых домашних метеостанций, размещенных в помещении лаборатории: измерений датчиков освещенности, влажности, температуры и напряжения [6]. Измерения снимались раз в 31 секунду на протяжении 36 дней, таким образом было собрано 2,3 млн записей. Данные представлены в виде набора кортежей $\langle date: yyyy-mm-dd, time: hh:mm:ss.xxx, temperature: real, light: real \rangle$. Для анализа использовались данные по температуре и освещенности. Одной из важных особенностей этого набора данных является ярко выраженная сезонность измерений (периодические колебания) с периодом 24 ч, связанная с суточным изменением температуры и освещенности.

Данные эксперимента LUCE проекта SensorScope. Данный набор данных также представляет собой показания метеостанций, но датчики расположены не в помещении, а на улице (использовались уличные метеостанции) [7]. Измерения считывались каждые 120 с, в набор данных вошли измерения за 44 дня. Среди измеренных значений присутствуют данные об окружающей среде, такие как температура, относительная влажность, солнечная радиация, скорость ветра и др. Для анализа была выбрана температурная составляющая измерений. Таким образом, кортеж анализируемых данных имеет вид $\langle date: yyyy-mm-dd, time: hh:mm:ss.xxx, temperature: real \rangle$. В данных также ярко выражена сезонная составляющая с периодом 24 ч.

Классификация возможных аномалий в наборах данных Intel Lab и SensorScope предложена в [8]. Позднее в [9] было сформулировано следующее определение. Пусть r_i — значение в момент времени i , предсказываемое моделью, а $r_i + eps_i$ — наблюдаемое значение. Тогда предлагается рассматривать следующие типы аномалий:

- *случайные аномалии*: одиночно встречающиеся значения-выбросы;
- *неисправности*: частое наличие в данных неверных значений таких, что $eps_i > t$ (где t — некоторое пороговое значение); в последовательности неверных значений не предполагается наличие закономерности.

В работе [10] разработаны алгоритмы моделирования данных классов аномалий, в частности построен контрольный набор данных, в который включена размеченная выборка, включающая в себя смоделированные аномалии. Результаты генерации неисправностей опубликованы в виде контрольного набора данных [11], в котором содержатся данные Intel Lab и SensorScope, включающие синтезированные неисправности. Кроме то-

го, для каждого измерения добавлено контрольное значение, соответствующее классу нештатной ситуации либо ее отсутствию в конкретный момент. Классы размечены следующим образом: «0» — нет аномалий; «1» — случайные аномалии; «2» — неисправности; «4» — смещение; «8», «16», «32» — полиномиальные искажения, сгенерированы разными алгоритмами. Эти данные были выбраны для контроля качества детектирования аномалий разрабатываемой системы.

На рассматриваемых данных были применены методы обнаружения аномалий, основанные на моделях SARIMA и HMM. Из набора данных Intel Lab Data были использованы данные по датчикам света, из набора данных LUCE — данные по датчикам температуры.

В результате настройки моделей SARIMA наиболее часто встречались следующие наборы гиперпараметров $\langle p, d, q, P, D, Q \rangle$: $\langle 0, 1, 1, 1, 1, 1 \rangle$ и $\langle 0, 1, 1, 0, 1, 1 \rangle$, при сезонном параметре $s = 24$ ч (сутки). Эксперименты показали, что модели SARIMA лучше подходят для обнаружения краткосрочных случайных аномалий, так как в случае долгосрочных аномалий доверительные интервалы ее предсказаний становятся шире, а следовательно, слабее.

Метод HMM показал себя более успешным на продолжительных аномалиях (неисправностях), при этом заметно уступая методу SARIMA в краткосрочных аномалиях (случайных выбросах). В работе [12] также для долгосрочных аномалий отдается предпочтение скрытым марковским моделям.

Так, на наборе Intel Lab Data средняя доля обнаруженных случайных аномалий составила 92,6% для SARIMA и 29,3% для HMM, а доля обнаруженных неисправностей составила 43,2% для SARIMA и 69,1% для HMM. На наборе LUCE средняя доля обнаруженных случайных аномалий составила 89,2% для SARIMA и 32,7% для HMM, а доля обнаруженных неисправностей составила 40,1% для SARIMA и 71,9% для HMM. Доля ложных срабатываний при этом не превышает 2% от общего числа измерений.

В работе [8] метод SARIMA был применен к тем же наборам данных, но моделирование аномалий проводилось иначе. Кроме того, параметры модели SARIMA использовались фиксированные: $\langle 0, 1, 1, 0, 1, 1 \rangle$. Результаты на случайных аномалиях оказались схожими — более 96% ARIMA (других сравнимых типов аномалий в этой работе не было).

Литература

1. Kovalev D., Shanin I., Stupnikov S., Zakharov V. Data mining methods and techniques for fault detection and predictive maintenance in housing and utility infrastructure // Conference (International) on Engineering Technologies and Computer Science. — IEEE, 2018. doi: 10.1109/EnT.2018.00016.
2. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. Time series analysis: Forecasting and control. — 5th ed. — Wiley, 2015. 712 p.
3. Cleveland R. B., Cleveland W. S., McRae J. E., Terpenning I. STL: A seasonal-trend decomposition // J. Off. Stat., 1990. Vol. 6. Iss. 1. P. 3–73.
4. Bengio Y., Frasconi P. An input output HMM architecture // NIPS Proceedings. — MIT Press, 1995. P. 427–434.
5. Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition // Readings in speech recognition / Eds. A. Waibel, K.-F. Lee. — San Francisco, CA, USA: Morgan Kaufmann, 1990. P. 267–296.
6. Bodik P., Hong W., Guestrin C., Madden S., Paskin M., Thibaux R. Intel Lab Data. Intel Berkeley Research lab, 2004. <http://db.csail.mit.edu/labdata/labdata.html>.
7. Sensorscope: Sensor networks for environmental monitoring. Lausanne Urban Canopy Experiment (LUCE). — EPFL, 2006. <https://lcav.epfl.ch/page-145180-en.html>.
8. Sharma A. B., Golubchik L., Govindan R. Sensor faults: Detection methods and prevalence in real-world datasets // ACM Trans. Sens. Netw., 2010. Vol. 6. Iss. 3. P. 23.
9. Baljak V., Tei K., Honiden S. Fault classification and model learning from sensory Readings — Framework for fault tolerance in wireless sensor networks // 8th Conference (International) on Intelligent Sensors, Sensor Networks and Information Processing. — IEEE, 2013. P. 408–413.
10. De Bruijn B., Nguyen T. A., Bucur D., Tei K. Benchmark datasets for fault detection and classification in sensor data // 5th Conference (International) on Sensor Networks Proceedings. — SCITEPRESS, 2016. P. 185–195.
11. De Bruijn B., Nguyen T. A., Bucur D., Tei K. Benchmark datasets for fault detection and classification in sensor data, 2015. <http://tuananh.io/datasets>.
12. Warriach E. U., Aiello M., Tei K. A machine learning approach for identifying and classifying faults in wireless sensor network // 15th Conference (International) on Computational Science and Engineering. — IEEE, 2012. P. 618–625.

Поступила в редакцию 15.07.18

METHODS AND TOOLS FOR FAULT DETECTION ON ELEMENTS OF HOUSING AND UTILITY INFRASTRUCTURE

I. A. Shanin, S. A. Stupnikov, and V. N. Zakharov

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 work belongs to the area of development of specific information systems based on the Internet of Things technology. An approach for program implementation of a module intended for detection of faults on elements of housing and utility infrastructure is proposed. The module is considered as a part of an information system aimed at technical maintenance of mentioned elements: condition monitoring, predictive maintenance, fault detection, and reporting. Operation algorithms of module components are described: building of operation models for housing and utility infrastructure elements and fault detection. The approach is applied on a couple of datasets for fault detection, experimental results are presented.

Keywords: Internet of Things; data analysis; fault detection; housing and utility infrastructure

DOI: 10.14357/19922264180310

Acknowledgments

The research is supported by Ministry of Education and Science of the Russian Federation (project’s unique identifier RFMEFI60717X0176).

References

1. Kovalev, D., I. Shanin, S. Stupnikov, and V. Zakharov. 2018. Data mining methods and techniques for fault detection and predictive maintenance in housing and utility infrastructure. *Conference (International) on Engineering Technologies and Computer Science*. IEEE. doi: 10.1109/EnT.2018.00016.
2. Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time series analysis: Forecasting and control*. 5th ed. Wiley. 712 p.
3. Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning. 1990. STL: A seasonal-trend decomposition. *J. Off. Stat.* 6(1):3–73.
4. Bengio, Y., and P. Frasconi. 1995. An input output HMM architecture. *NIPS Proceedings*. MIT Press. 427–434.
5. Rabiner, L. R. 2004. A tutorial on hidden Markov models and selected applications in speech recognition. *Readings in speech recognition*. Eds. A. Waibel and K.-F. Lee. San Francisco, CA: Morgan Kaufmann. 267–296.
6. Bodik, P., W. Hong, C. Guestrin, S. Madden, M. Paskin, and R. Thibaux. 2004. Intel Lab Data. Intel Berkeley Research lab. Available at: <http://db.csail.mit.edu/labdata/labdata.html> (accessed July 16, 2018).
7. Sensorscope. 2006. Sensorscope: Sensor networks for environmental monitoring. Lausanne Urban Canopy Experiment (LUCE). EPFL. Available at: <https://icav.epfl.ch/page-145180-en.html> (accessed July 16, 2018).
8. Sharma, A. B., L. Golubchik, and R. Govindan. 2010. Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Trans. Sens. Netw.* 6(3):23.
9. Baljak, V., K. Tei, and S. Honiden. 2013. Fault classification and model learning from sensory Readings — Framework for fault tolerance in wireless sensor networks. *8th Conference (International) on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE. 408–413.
10. De Bruijn, B., T. A. Nguyen, D. Bucur, and K. Tei. 2016. Benchmark datasets for fault detection and classification in sensor data. *5th Conference (International) on Sensor Networks Proceedings*. SCITEPRESS. 185–195.
11. De Bruijn, B., T. A. Nguyen, D. Bucur, and K. Tei. 2015. Benchmark datasets for fault detection and classification in sensor data. Available at: <http://tuananh.io/datasets/> (accessed July 16, 2018).
12. Warriach, E. U., M. Aiello, and K. Tei. 2012. A machine learning approach for identifying and classifying faults in wireless sensor network. *15th Conference (International) on Computational Science and Engineering*. IEEE. 618–625.

Received July 15, 2018

Contributors

Shanin Ivan A. (b. 1991) — junior 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; v08shanin@gmail.com

Stupnikov Sergey A. (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; sstupnikov@ipiran.ru

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

ИМПЛИЦИРОВАННЫЕ ЗНАНИЯ: ОСНОВАНИЯ И ТЕХНОЛОГИИ ИЗВЛЕЧЕНИЯ*

И. М. Зацман¹

Аннотация: Дано описание теоретических оснований разработки информационных технологий, которые обеспечивают целенаправленность формирования лингвистических типологий как форм представления нового знания о языке. Они формируются в процессе контрастивного анализа параллельных выровненных текстов, служащих источниками нового знания. В параллельных текстах встречаются случаи имплицирования субъективных знаний переводчиков, которые не представлены в системе современного знания о языке. Их экспликация возможна с помощью информационных технологий, позволяющих обрабатывать параллельные тексты и целенаправленно извлекать имплицированные знания. Цель статьи состоит в описании нового подхода к разработке технологий, обеспечивающих целенаправленность формирования лингвистических типологий, и его сопоставлении с существующими подходами и моделями процессов формирования (роста) знания. Сформулированы те условия, при которых технологически может быть обеспечен целенаправленный рост объективного знания (в терминах К. Р. Поппера). Предлагаемый подход иллюстрируется на примере задачи формирования типологии конструкций русского языка с модальным значением, возникающих в переводе немецких модальных конструкций.

Ключевые слова: параллельные тексты; корпусная лингвистика; имплицированные знания; извлечение новых знаний; эмерджентность; информационная технология; целенаправленность; формирование типологий

DOI: 10.14357/19922264180311

1 Введение

Разработка информационной технологии, которая обеспечивает формирование лингвистических типологий как форм представления нового знания о языке, является одной из задач проекта по гранту РФФИ, который в настоящее время выполняется в Институте проблем информатики ФИЦ ИУ РАН.

Проектируемая информационная технология формирования (ИТФ) лингвистических типологий² предназначена для использования в процессе решения трех взаимосвязанных задач:

- (1) обнаружение лагун в системе современного знания о языке относительно исследуемых языковых единиц;
- (2) целенаправленное извлечение новых знаний для заполнения лагун;
- (3) формирование и развитие лингвистических типологий на основе извлеченных знаний.

Цель статьи состоит в описании подхода к созданию ИТФ, обеспечивающей целенаправленность развития современного знания о языке на основе

экспликации субъективных знаний переводчиков, имплицированных в параллельных текстах, за счет формирования новых или развития существующих типологий.

Существует принципиальная разница между знаниями, представленными в параллельных текстах и имплицированными в них.

В первом случае смысловое содержание оригинального и переводного текстов непосредственно передается с помощью естественных языков оригинала или перевода соответственно.

Второй случай предполагает знание двух языков и переводных соответствий языковых единиц оригинального и переводного текстов. Для понимания текста перевода необязательно знать язык оригинала и переводные соответствия, но для выполнения перевода знание переводных соответствий является необходимым. При этом в процессе перевода возможны случаи использования концептов субъективных знаний переводчика, которые не представлены в системе современного знания о языке. Такое их использование предлагается называть *имплицированием* субъективных знаний в параллельных текстах.

* Работа выполнена в Институте проблем информатики ФИЦ ИУ РАН при поддержке РФФИ (проект 18-07-00192).

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

² В лингвистике типологии используются, как правило, для описания сходства и различий между языками. В статье этот термин используется в другом значении: для описания сходства и различий между языковыми единицами.

Фрагменты параллельных текстов, в которых имплицитны концепты субъективных знаний переводчиков, будем называть *объектами интерпретации*, которые по определению являются двуязычными. Создаваемая ИТФ предназначена для поиска таких объектов в параллельных текстах и обеспечения целенаправленной экспликации имплицитных знаний в процессе семантического анализа объектов интерпретации. Здесь важно сказать о той точке зрения, которая широко распространена в эпистемологии, что рост научного знания эмерджентен и поэтому непредсказуем в принципе. А раз его рост является спонтанным, то и нельзя говорить об обеспечении целенаправленности его роста. Поэтому первый вопрос, рассматриваемый в статье, можно было бы сформулировать так: всегда ли рост знания является эмерджентным?

Если ответ «нет», то второй вопрос, являющийся актуальным для выполняемого проекта, можно было бы сформулировать так: если современная система знания о языке не удовлетворяет новым образовательным, технологическим (например, машинный перевод) или иным потребностям общества, то можно ли обеспечить целенаправленность его роста с использованием средств информатики?

2 Что говорит эпистемология?

Согласно К. Р. Попперу, рост научного знания непредсказуем в принципе: «Мои доводы строятся на предположении, что существует такая вещь, как подлинный рост научного знания, или, выражаясь практически, что завтра или через год мы, возможно, выдвинем и подвергнем проверке важные теории, о которых до сих пор никто *всерьез не думал* (курсив мой — ИЗ). Если существует рост научного знания в этом смысле, то он не может быть предсказуем средствами науки. Ведь тот, кто смог бы сегодня средствами науки предсказать завтрашние открытия, мог бы сегодня их и сделать, это означало бы, что рост научного знания закончился» [1, с. 283].

В этой книге описывается четырехчленная итерационная схема спонтанного роста научного знания, представленного в форме теорий. Каждая итерация этой схемы имеет следующий вид: $P_1 \rightarrow TT \rightarrow EE \rightarrow P_2$, где P_1 означает «проблема в начале итерации»; TT — «пробная теория»; EE — «устранение ошибок»; P_2 — «проблема в конце итерации». С помощью этой схемы К. Р. Поппер «пытался показать, что результатом критики или устранения ошибок в применении к пробной теории, как правило, становится возникновение новой

проблемы или даже нескольких новых проблем» [1, с. 273–274].

Итерационный подход к моделированию процессов роста знания используется и в экономике. В этой научной дисциплине первая итерационная модель была создана в последнем десятилетии прошлого века: спиральная модель процессов формирования новых знаний. Она была предложена И. Нонака и впервые описана им в работах [2, 3]. В процессе ее построения рассматривались индивидуальные знания человека и коллективные знания группы людей. Каждая из этих двух категорий знаний была разделена на эксплицитные (*explicit*) и имплицитные (*tacit*) знания. Следовательно, спиральная модель включает в рассмотрение следующие четыре категории знаний: индивидуальные имплицитные, коллективные имплицитные, коллективные и индивидуальные эксплицитные знания.

С использованием этих четырех категорий И. Нонака определил понятие «спираль формирования знаний». Было показано на примерах, что эта спираль может служить качественной моделью итерационного процесса формирования новых знаний во время проведения «мозгового штурма». Развитие спиральной модели и описание примеров ее использования дано в работах [4, 5].

Обобщение и принципиально новое развитие результатов И. Нонака было предложено в работах Й. Накамори и А. Вежицкого в рамках создаваемой ими научной дисциплины, которую они называют «Наука о знаниях» [6–8].

В этих работах знания разделены на индивидуальные знания человека, коллективные и конвенциональные знания. Это деление Вежицкий и Накамори называют социальным аспектом или измерением, так как перечисленные знания отличаются тремя уровнями социализации (от первого индивидуального уровня до третьего конвенционального). С учетом деления на эксплицитные и имплицитные знания добавляются две новые категории знаний, которых нет в спиральной модели: конвенциональные имплицитные и конвенциональные эксплицитные знания. Развитие спиральной модели, включающее шесть категорий знаний, получило название «креативное пространство» [6].

Отметим, что в четырехчленной схеме, спиральной модели и креативном пространстве нет явно определенной оси времени. Это не дает возможности фиксировать моменты времени создания пробной теории или генерации каждого нового концепта как структурного элемента создаваемой теории или развиваемой системы знания. Также в рассмотренных подходах явно не указываются

Таблица 1 Три подхода к описанию роста знания

Позиции сопоставления	1. Четырехчленная схема [1]	2. Спиральная модель [2–5]	3. Креативное пространство [6–8]
Социальное измерение (включает три позиции номинативной шкалы: индивидуальное, коллективное, конвенциональное)	Индивидуальные (субъективные) знания мира 2 и конвенциональные (объективные) знания мира 3	Индивидуальные и коллективные знания	Индивидуальные, коллективные и конвенциональные знания
ИмPLICITные/эксплицитированные знания	ИмPLICITные индивидуальные знания (мир 2), эксплицитированные индивидуальные и конвенциональные знания (мир 3)	ИмPLICITные и эксплицитированные знания (и для индивидуальных, и для коллективных)	ИмPLICITные и эксплицитированные знания (для индивидуальных, коллективных и конвенциональных)
Источник роста знания	Явно не указан (подразумеваются индивидуальные и конвенциональные эксплицитированные знания)	Явно не указан (подразумеваются индивидуальные и коллективные эксплицитированные знания)	Явно не указан (подразумеваются индивидуальные, коллективные и конвенциональные эксплицитированные знания)

информационные источники роста знания, но они подразумеваются.

Сопоставим три рассмотренных подхода к описанию роста знания, которые пронумеруем согласно первой строке табл. 1. Строго говоря, в описании самой четырехчленной схемы не указаны категории знания, но они отмечены для первого подхода в позиции «ИмPLICITные/эксплицитированные знания» в этой таблице. Почему? В той же самой работе К. Р. Поппер проводит границу между субъективным и объективным знанием, строит теорию познания, в контексте которой и дано описание рассматриваемой четырехчленной схемы.

В процессе ее построения он описывает идею «трех миров» следующим образом: «Примерами объективного знания являются теории, опублико-

ванные в журналах и книгах и хранящиеся в библиотеках, обсуждения этих теорий, трудности или проблемы, на которые было указано в связи с такими теориями и т. д. Мы можем назвать физический мир «миром 1», мир наших осознанных переживаний — «миром 2», а мир логического содержания книг, библиотек, компьютерной памяти и тому подобного — «миром 3» [1, с. 78]. Следовательно, если рассматривать четырехчленную схему в контексте описания идеи «трех миров», то это является основанием указать для нее три категории знания в табл. 2: имPLICITные индивидуальные, эксплицитированные индивидуальные и конвенциональные знания.

Вернемся к первому вопросу статьи: возможно ли именно целенаправленное выявление лаку-

Таблица 2 Семь значений слова *face* по НТЕ [9, 10]

Номер концепта в НТЕ	Концепт	Слово и год появления концепта
01.02.03.08.01.04 n.	The body :: Face	Matching word(s): face (1290–)
01.02.03.08.01.04 04 n.	The body :: Face :: with reference to beauty	Matching word(s): face (1591–)
01.02.03.08.01.04.01 n.	The body :: Face with expression/expression	Matching word(s): face (1330–)
01.02.03.08.01.04.01 01 n.	The body :: Face with expression/expression :: grimace/distortion	Matching word(s): face (1602–)
01.12.05.03.01 13 n.	Relative position :: Surface :: one of several surfaces of a thing	Matching word(s): face (1340–)
01.12.05.03.01 19 n.	Relative position :: Surface :: front surface	Matching word(s): face (1611 + 1820–)
01.12.05.03.01 19.01 n.	Relative position :: Surface :: front surface :: specifically of a coin/medal/seal/die, etc.	Matching word(s): face (1515–)

Примечание: если в третьем столбце после «–» не указан год, то это говорит о том, что это значение встречается в текстах до настоящего времени.

и их заполнение, обеспечивающее рост знания? Отметим, что во всех трех рассмотренных подходах категория *имплицитных* знаний отсутствует, т. е. для них первый вопрос остается открытым. Во всех трех подходах можно найти неявные указания на источники нового знания. В табл. 1 указаны по два источника для первого и второго подходов, три — для третьего. Источники не указаны в явном виде, но они подразумеваются в описании связей между знаниями разных категорий. Например, зависимость имплицитного индивидуального знания от конвенционального в первом подходе описана так: «В мире 3 мы можем открыть новые проблемы, которые были там до того, как их открыли, и до того, как они были осознаны, то есть до того, как что-либо *соответствующее им появилось в мире 2* (...). Основной тезис: наше *осознанное субъективное знание (знание в мире 2) зависит от мира 3, то есть от теорий, сформулированных (хотя бы виртуально) на определенном языке* (курсив мой — ИЗ)» [1, с. 78].

Рассмотренные три теоретических подхода сопоставим с практикой конкурсного финансирования фундаментальных исследований, ориентированных на рост научного знания.

3 Практика формирования нового знания

Для примера возьмем проект, поддержанный Национальным научным фондом (ННФ) США, который посвящен моделированию процесса возникновения новых значений (= концептов) слов. *Целью* проекта ННФ является разработка ретроспективной модели, описывающей временную траекторию возникновения новых концептов. Исполнители проекта исходили из того, что в естественном языке используется конечное число слов, обозначающих концепты, но при этом слова должны быть способны выражать значительно большее число концептов, чем число слов в языке. Из этого они сделали вывод, что в процессе развития языка существующие слова приобретают все новые и новые концепты согласно некоторой модели, что и было ими показано в результате выполнения проекта.

Для достижения цели проекта его исполнители поставили *задачу* разработать ретроспективную модель (*ожидаемый результат проекта*), описывающую процесс появления в прошлом концептов, исследуя эволюцию английского языка [11]. В этой модели есть ось времени, на которой указаны моменты появления новых значений у исследуемого слова. Эти моменты определялись по Историческому тезаурусу английского языка (The Historical

Thesaurus of English — HTE), в котором содержится информация о годах появления новых концептов слов, начиная с 1000 г. [9].

Ретроспективная модель описывает связи между концептами каждого исследуемого слова и степень их семантической близости, которая определяется на основе иерархического номера, присвоенного каждому концепту в HTE [10]. В качестве примера приведем семь концептов слова *face* по HTE с указанием их номеров и годов появления в текстах Британского национального корпуса (British National Corpus — BNC) [12] (см. табл. 2). Отметим, что для всех семи концептов не указаны годы их исчезновения из текстов, т. е. они встречаются в современных текстах BNC.

В ретроспективной модели источниками нового знания о временных траекториях возникновения новых концептов являются HTE и BNC. Для рассмотренного проекта три позиции сравнения, приведенных в табл. 1, будут иметь следующий вид:

- (1) конвенциональные знания, отраженные в BNC, а также индивидуальные и коллективные знания исполнителей проекта;
- (2) эксплицитные конвенциональные знания, а также эксплицитные и имплицитные индивидуальные и коллективные знания исполнителей проекта;
- (3) источником нового знания о временных траекториях являются HTE и BNC.

Возникает закономерный вопрос, в какой степени новое знание в форме моделей и временных траекторий возникновения новых концептов является спонтанным? Авторы проекта сформулировали его цель и задачу, описали ожидаемый результат, доступные информационные ресурсы HTE и BNC, запланировали выполнение проекта, включая проведение вычислительных экспериментов с использованием методов корпусной лингвистики, получили ожидаемый результат и опубликовали статью [11]. При этом они сформировали новое знание в форме моделей и их действия в целях его получения были *целенаправленными*. Таким образом, этот пример говорит о возможности целенаправленного роста научного знания.

Естественно, что возможны ситуации, когда не удается получить ожидаемые результаты проектов фундаментальных исследований и/или спонтанно возникают незапланированные новые результаты. Для положительного ответа на первый вопрос статьи достаточно было бы и одного приведенного примера, но практика научных фондов позволяет утверждать, что это далеко не единичный случай.

Таблица 3 Примеры объектов интерпретации с модальными конструкциями в оригинале

Оригинал	Перевод
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)]
”Das Kind kann nichts dafür”, hörte sie die Stimme sagen, “aber du, Kassiopiea — warum hast du das nur getan?” [Michael Ende. Momo (1973)]	— Ребенок не виноват, — услышала она говорившего. — Но ты, Кассиопея, почему ты это сделала? [Михаэль Энде. Момо (Ю. И. Коринец, 1982)]
“Mag sein,” sagte der Advokat, “wir wollen aber trotzdem nichts übereilen.” [Franz Kafka. Der Prozess (1914)]	— Возможно, — сказал адвокат, — и все же не будем торопиться. [Франц Кафка. Процесс (Р. Райт-Ковалева, 1965)]

Таким образом, ответ на первый вопрос предлагается такой: многолетняя практика работы отечественных и зарубежных фондов по финансированию проектов фундаментальных исследований говорит о возможности целенаправленного формирования нового научного знания.

Рассмотрим второй вопрос — о возможности обеспечить целенаправленность роста научного знания с использованием средств информатики. При его рассмотрении ограничимся только системой знания о языке.

4 Теоретические основания создания информационной технологии формирования

Разрабатываемая в рамках проекта РФФИ концепция создания ИТФ предназначена в первую очередь для решения трех задач, которые перечислены в начале статьи. Первая основная идея предлагаемой концепции и ее отличие от трех подходов, приведенных в табл. 1, заключается в явном описании источника нового знания о языке и фрагментации его информации на объекты интерпретации (пять их примеров приведены в табл. 3).

В разрабатываемой концепции фрагментация информации источника знания на объекты интерпретации позиционируется как необходимое условие создания ИТФ. Проиллюстрируем это условие на примере задачи построения типологии конструкций русского языка с модальным значением, возникающих в переводе конструкций немецкого языка с модальными глаголами, которая является сегодня актуальной в лингвистике [13]. Для решения этой задачи естественно использовать

параллельные тексты с предложениями на немецком языке и их переводами на русский язык (см. табл. 3).

В настоящее время наиболее представительный массив немецко-русских текстов в электронной форме объемом 2,6 млн словоупотреблений хранится в Параллельном немецком корпусе (ПНК), который находится в открытом доступе [14]. Так как эти тексты являются выровненными (оригинальным предложениям поставлены в соответствие их переводы), то необходимая фрагментация в ПНК уже есть, при этом объект интерпретации — это пара немецкого и русского предложений (см. табл. 3). Когда используются невыровненные тексты оригинала и перевода, тогда операция фрагментации является необходимым компонентом ИТФ [15].

Для решения задачи построения типологии необходимо из корпуса выровненных текстов отобрать с помощью поиска по леммам предложения с исследуемыми немецкими модальными глаголами и одновременно переводы этих предложений на русский язык. Таблица 4 содержит данные прове-

Таблица 4 Число объектов интерпретации в немецко-русских текстах ПНК

Модальный глагол	Число объектов интерпретации
Dürfen	758
Können	5782
Mögen	937
Müssen	3209
Sollen	2041
Wollen	3541
<i>Всего объектов интерпретации</i>	16268

Таблица 5 Пример завершенной аннотации

<i>Sie sollten diese ungemütliche Sache jetzt lieber sein lassen.</i>	sollen <+Inf I> <Sie> <Praet> <Konj II> <sollen-3>	– <i>Давай оставим</i> это дело <i>в покое.</i>	<1pl> <Imperat> <2sg> <давай>
---	--	---	--

Примечания: <+Inf I> — управляет глаголом в форме инфинитива I; <Sie> — вежливая форма обращения ко 2-му лицу; <Praet> — Präteritum; <Konj II> — Konjunktiv II; <sollen-3> — соответствует 3-му значению в словарной статье модального глагола *sollen*; <1pl> — 1-е лицо, множественное число (глагола, управляемого формой *давай*); <Imperat> — императив (характеристика формы *давай*); <2sg> — характеристика формы *давай*.

Таблица 6 Пример незавершенной аннотации

Gut, sagte ich, soll er dich <i>verehren</i> , aber so viel kostbare Blumen, das ist aufdringlich.	Sollen <+Inf I> <3sg> <Praes> <sollen-x> <Verb-Initial>	– Очень мило, — сказал я, — <i>поклонник поклонником</i> , но дарить такой большой букет дорогих цветов — значит <i>навязываться</i> .	<xN-Nomin+xN-Instr>
---	---	--	---------------------

Примечания: <sollen-x> — означает, что значение модального глагола в этом предложении не найдено в немецко-русском словаре в процессе лингвистического аннотирования; <xN-Nomin+xN-Instr> — конструкция «существительное в именительном падеже + то же существительное в творительном».

денного поиска: для рассматриваемой задачи исследователям доступны в ПНК более 16 тыс. объектов интерпретации (в таблице они распределены по шести немецким модальным глаголам).

Вторая идея предлагаемой концепции заключается в обеспечении с помощью ИТФ лингвистического аннотирования отобранных объектов интерпретации, позволяющего фиксировать лакуны в системе современного знания о языке. До начала аннотирования для рассматриваемой задачи был выбран немецко-русский словарь [16], который отражает современный уровень, и он используется в процессе аннотирования. Результатом аннотирования объектов интерпретации могут быть как завершенные (табл. 5), так и незавершенные (табл. 6) двуязычные аннотации¹.

Массив отобранных завершенных аннотаций содержит структурированное описание конструкций немецкого языка с модальными глаголами и конструкций русского языка с модальным значением. Исследование соотношения между количеством завершенных и незавершенных аннотаций, которое меняется в процессе аннотирования, является самостоятельной задачей. Есть первые результаты ее решения для переводов французских глаголов на русский язык [17]. Был проведен эксперимент по аннотированию 2 500 объектов

интерпретации, т. е. пар предложений, включающих переводы французских глаголов, которые были выровнены так же, как и немецкие предложения в табл. 3. В результате их аннотирования было сформировано 97,7% завершенных аннотаций. При этом семантический анализ 2,3% незавершенных аннотаций позволил лингвистам извлечь четыре низкочастотных модели перевода французских глаголов прошедшего времени несовершенного вида (*imparfait*) на русский язык, отсутствующих в современных контрастивных грамматиках [18].

Завершенность процесса аннотирования объекта интерпретации говорит о том, что значение немецкого модального глагола, которое встретилось в объекте интерпретации, присутствует в словаре. Если этого значения в словаре нет, то в аннотации ставятся специальные теги, которые фиксируют ее незавершенность, причину незавершенности и тем самым описывают потенциальные лакуны в системе современного знания о языке. Таким образом, массив незавершенных аннотаций фиксирует некоторый спектр потенциальных лакун. Одной из целей создания ИТФ является обеспечение процесса аннотирования и разметки незавершенных аннотаций с помощью тегов, которые передаются на этап их семантического анализа, выполняемого лингвистами-экспертами. Результатом анали-

¹ Аннотации сформированы В. И. Карповым (см. табл. 5) и А. А. Гончаровым (см. табл. 6). Примечания к табл. 5 и 6 подготовлены Д. О. Добровольским и Анной А. Зализняк.

за является описание новых значений немецких модальных глаголов и пополнение ими немецко-русского словаря с использованием переводов тех предложений, в которых эти новые значения были обнаружены в процессе аннотирования.

Третья идея предлагаемой концепции заключается в обеспечении с помощью ИТФ многоаспектной кластеризации завершенных аннотаций. Выбор аспектов кластеризации зависит от решаемой лингвистической задачи и исследуемых языковых единиц. Например, для формирования типологии конструкций русского языка с модальным значением, возникающих в переводе конструкций немецкого языка с модальными глаголами, завершённые аннотации группируются по немецким модальным глаголам (см. табл. 4), но это является только одним из аспектов кластеризации.

5 Заключение

Применение ИТФ в процессе лингвистического аннотирования дает возможность лингвистам сформировать массив незавершённых аннотаций, отражающих потенциальные лакуны в системе современного знания о языке. Последующий семантический анализ таких аннотаций позволяет выявить новые значения модальных глаголов и сформировать искомую типологию [13]. Сопоставим предлагаемые теоретические основания создания ИТФ с ретроспективной моделью возникновения новых концептов [9–11]. Основное отличие заключается в том, что в первом случае ИТФ помогает лингвистам целенаправленно выявлять и описывать новые концепты, имплицированные в параллельных текстах, а во втором случае информационные ресурсы BNC используются для описания ретроспективной временной траектории появления концептов, которые эксплицированы, так как они уже были описаны в НТЕ до начала проекта ННФ.

В первом случае ИТФ обеспечивает извлечение имплицированных субъективных знаний переводчиков (мир 2) и формирование на их основе новых эксплицированных знаний лингвистов в мире 3 как объективного знания в терминах К. Р. Поппера. Во втором случае были построены временные траектории и новые модели появления концептов с использованием информационных ресурсов BNC и НТЕ без извлечения из них имплицированных знаний. Из НТЕ копировалась информация о концептах и годах их первого появления в текстах BNC, на основе которой строились траектории и модели появления концептов исследуемых слов.

Необходимыми условиями создания ИТФ, обеспечивающих рост научного знания в лингвистике за счет экспликации имплицированных знаний, являются следующие положения:

- наличие источника нового знания для обозначенной предметной области и возможность фрагментации его информации на объекты интерпретации, в которых имплицированы индивидуальные (субъективные) знания;
- возможность отражения потенциальных лагун в системе современного знания с помощью разметки тегами объектов интерпретации;
- возможность заполнения лагун результатами семантического анализа соответствующих объектов интерпретации в процессе экспликации имплицированных знаний.

Для ряда задач контрастивной лингвистики эти три условия выполняются [17, 18], и они являются необходимыми теоретическими основаниями разработки ИТФ. В настоящее время разработан ее первый компонент, ориентированный на обнаружение потенциальных лагун в системе современного знания о переводах конструкций с немецкими модальными глаголами на русский язык. Первые результаты экспериментальной эксплуатации позволяют говорить о возможности обеспечения целенаправленности экспликации имплицированных знаний с использованием средств информатики. Что касается других предметных областей, то вопрос применимости в них аналогичных средств информатики, обеспечивающих целенаправленность роста знания, в настоящее время является открытым, так как требуется отдельная проверка выполнимости в этих областях трех перечисленных условий.

Литература

1. *Popper K. R.* Объективное знание. Эволюционный подход / Пер. с англ. — М.: Эдиториал УРСС, 2002. 384 с. (*Popper K. R.* Objective knowledge. An evolutionary approach. — Oxford: Clarendon Press, 1979. 395 p.)
2. *Nonaka I.* The knowledge-creating company // *Harvard Bus. Rev.*, 1991. Vol. 69. No. 6. P. 96–104.
3. *Nonaka I.* A dynamic theory of organizational knowledge creation // *Organ. Sci.*, 1994. Vol. 5. No. 1. P. 14–37.
4. *Nonaka I., Takeuchi H.* Компания — создатель знания. — М.: Олимп-бизнес, 2003. 384 с. (*Nonaka I., Takeuchi H.* The knowledge-creating company. — Oxford, NY, USA: Oxford University Press, 1995. 284 p.)
5. *Nonaka I., Toyama R.* The knowledge-creating theory revisited: Knowledge creation as a synthesizing process // *Knowl. Man. Res. Pract.*, 2003. Vol. 1. No. 1. P. 2–10.

6. *Wierzbicki A. P., Nakamori Y.* Basic dimensions of creative space // *Creative space: Models of creative processes for knowledge civilization age* / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin: Springer Verlag, 2006. P. 59–90.
7. *Wierzbicki A. P., Nakamori Y.* Knowledge sciences: Some new developments // *Z. Betriebswirt.*, 2007. Vol. 77. No. 3. P. 271–295.
8. *Nakamori Y.* Knowledge and systems science — enabling systemic knowledge synthesis. — London–New York: CRC Press, Taylor & Francis Group, 2013. 234 p.
9. *Kay C., Roberts J., Samuels M., Wotherspoon I., Alexander M.* The historical thesaurus of English. Version 4.2. — Glasgow, U.K.: University of Glasgow, 2015. <https://historicalthesaurus.arts.gla.ac.uk>.
10. *Kay C., Roberts J., Samuels M., Wotherspoon I., Alexander M.* The historical thesaurus of English: Face. — Glasgow, U.K.: University of Glasgow, 2015. <http://historicalthesaurus.arts.gla.ac.uk/category-selection/?qsearch=face>.
11. *Ramiro C., Srinivasan M., Malt B. C., Xu Y.* Algorithms in the historical emergence of word senses // *P. Natl. Acad. Sci. USA*, 2018. Vol. 115. No. 10. P. 2323–2328.
12. The British National Corpus (Oxford University Computing Services). www.natcorp.ox.ac.uk.
13. *Добровольский Д. О., Зализняк Анна А.* Немецкие конструкции с модальными глаголами и их русские со-ответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. С. 172–184.
14. Параллельный немецкий корпус. <http://www.ruscorpora.ru/search-para-de.html>.
15. *Loiseau S., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M.* Information technologies for creating the database of equivalent verbal forms in the Russian–French multivariant parallel corpus // *Информатика и её применения*, 2013. Т. 7. Вып. 2. С. 100–109.
16. *Немецко-русский словарь: актуальная лексика* / Под ред. Д. О. Добровольского. — М.: Лексрус, 2018 (в печати).
17. *Zatsman I., Buntman N.* Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery // *16th European Conference on Knowledge Management Proceedings*. — Reading, MA, USA: Academic Publishing International Ltd., 2015. P. 851–860.
18. *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. P. 976–983.

Поступила в редакцию 10.06.18

IMPLIED KNOWLEDGE: FOUNDATIONS AND TECHNOLOGIES OF EXPLICATION

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 theoretical foundations of the development of information technologies that provide the goal-oriented creation of new linguistic typologies are described. They are formed in the process of contrastive analysis of parallel aligned texts, which are sources of new knowledge of the language. In parallel texts, there are the implications of subjective knowledge of translators that are not represented in the system of modern knowledge of language. Their explication is possible with the help of information technologies, which allow processing parallel texts and goal-oriented extraction of the implied knowledge. The aim of the paper is to describe a new approach to the development of technologies that provide purposefulness and compare it with existing approaches and models. The conditions under which the growth of objective knowledge (in terms of K. R. Popper) can be technologically ensured are formulated. The proposed approach is illustrated by the example of the task of forming a typology of Russian language constructions with a modal value that arise in the translation of German modal constructions.

Keywords: parallel texts; corpus linguistics; implied knowledge; extraction of new knowledge; emergence; information technology; purposefulness; formation of typologies

DOI: 10.14357/19922264180311

Acknowledgments

The work was fulfilled at the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences and was supported by the Russian Foundation for Basic Research (project 18-07-00192).

References

1. Popper, K. R. 1979. *Objective knowledge. An evolutionary approach*. Oxford: Clarendon Press. 395 p.
2. Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
3. Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.
4. Nonaka, I., and H. Takeuchi. 1995. *The knowledge-creating company*. Oxford, NY: Oxford University Press. 284 p.
5. Nonaka, I., and R. Toyama. 2003. The knowledge-creating theory revisited: Knowledge creation as a synthesizing process. *Knowl. Man. Res. Pract.* 1(1):2–10.
6. Wierzbicki, A. P., and Y. Nakamori. 2006. Basic dimensions of creative space. *Creative space: Models of creative processes for knowledge civilization age*. Eds. A. P. Wierzbicki and Y. Nakamori. Berlin: Springer Verlag. 59–90.
7. Wierzbicki, A. P., and Y. Nakamori. 2007. Knowledge sciences: Some new developments. *Z. Betriebswirt.* 77(3):271–295.
8. Nakamori, Y. 2013. *Knowledge and systems science — enabling systemic knowledge synthesis*. London — New York: CRC Press, Taylor & Francis Group. 234 p.
9. Kay, C., J. Roberts, M. Samuels, I. Wotherspoon, and M. Alexander. 2015. The historical thesaurus of English. Version 4.2. Glasgow, U.K.: University of Glasgow. Available at: <https://historicalthesaurus.arts.gla.ac.uk> (accessed May 16, 2018).
10. Kay, C., J. Roberts, M. Samuels, I. Wotherspoon, and M. Alexander. 2015. The historical thesaurus of English: Face. Glasgow, U.K.: University of Glasgow. Available at: <https://historicalthesaurus.arts.gla.ac.uk/category-selection/?qsearch=face> (accessed May 5, 2018).
11. Ramiro, C., M. Srinivasan, B. C. Malt, and Y. Xu. 2018. Algorithms in the historical emergence of word senses. *P. Natl. Acad. Sci. USA* 115(10):2323–2328.
12. The British National Corpus (Oxford University Computing Services). Available at: <http://www.natcorp.ox.ac.uk> (accessed May 16, 2018).
13. Dobrovol'skij, D. O., and Anna A. Zalizniak. 2018. Nemetskie konstruksii s modal'nymi glagolami i ikh russkie sootvetstviya: proyekt 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.
14. Parallel'nyy nemetskiy korpus [Parallel German corpus]. Available at: <http://www.ruscorpora.ru/search-para-de.html> (accessed May 19, 2018).
15. Loiseau, S., D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2013. Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):100–109.
16. Dobrovol'skij, D. O., ed. 2018 (in press). *Nemetsko-russkiy slovar': aktual'naya leksika* [German–Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
17. 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, MA: Academic Publishing International Ltd. 851–860.
18. Zatsman, I., N. Buntman, A. Coldefy-Faucard, and V. Nuriev. 2016. WEB knowledge base for asynchronous brainstorming. *17th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 976–983.

Received June 10, 2018

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

STATISTICAL ANALYSIS OF LANGUAGE SPECIFICITY OF CONNECTIVES BASED ON PARALLEL TEXTS

O. Yu. Inkova¹ and M. G. Kruzhkov¹

Abstract: In recent decades, problems of language specificity in the Russian language attract considerable attention of researchers, although until recently, they have not been thoroughly examined using corpus-based methods. This paper presents a new method of investigating language specificity of Russian connectives based on statistical analysis of annotated parallel texts. Russian–French and French–Russian parallel texts are processed with the help of the Supracorpora Database (SCDB) of Connectives designed specifically for annotation of translation correspondences (TCs) found in parallel texts. Each TC includes annotations of a Russian connective and its translation equivalent (TE), which enables one to obtain statistical data on various translation models (TMs) based on several proposed parameters of language specificity of connectives. As an example, in this work, language specificity of two Russian connectives will be examined: *unu* and *a mo*. Based on the proposed statistical parameters, it will be demonstrated that *unu* has a very low degree of language specificity in the context of the Russian–French language pair, while *a mo* is a highly language-specific connective. The results of this research are applicable to informatics (machine translation and statistical analysis of textual data) and comparative study of languages, such as lexical typology, lexicography, and theory and practice of translation.

Keywords: supracorpora databases; statistical analysis; contrastive corpus analysis; language specificity; parallel corpora; linguistic information resources; connectives; discourse relations; semantics

DOI: 10.14357/19922264180312

1 Introduction

Language-specific phenomena (the term introduced in [1]) are examined in Russian linguistics since 1990s (see, e. g., [2, 3]). According to [4], a language-specific lexical unit “includes a **unique conceptual configuration**, meaning that all of its existing translation equivalents either lack a certain semantic component, or include an irremovable superfluous component.” Until recently, language-specific phenomena had been studied based almost exclusively on comparative semantic analysis. By now, semantic analysis has been supplemented by quantitative statistical analysis of annotated parallel texts.

The goal of this paper is to describe new approaches to identifying the degree of language specificity of connectives based on data from parallel corpora. The language specificity of Russian connectives is examined in comparison with French TEs that appear in parallel texts.

The SCDB of Connectives is a new information resource that contains aligned texts of parallel corpora from the Russian National Corpus (RNC) and makes it possible to create and annotate Russian connectives together with their TEs in parallel texts. Such annotated pairs are called translation correspondences. Annotations of connectives and their TEs include relevant context properties of Russian connectives and their

parallels, such as type of relation, position, structural features, etc. The concept of the SCDBs and their use are described in [5, 6], and the annotation system and statistical features — in [7].

The quantitative analysis requires adequate interpretation and should be accompanied by semantic and functional analysis. This analysis makes it possible to determine the type and degree of language specificity of various Russian connectives, as will be demonstrated by the example of 210 annotations of the Russian connective *unu* with rather obvious semantics and 210 annotations of the polysemic Russian connective *a mo*.

2 Data and Information Resources

The primary source of data for this research are parallel texts of Russian–French corpus of the RNC. These texts include morphological annotation for individual words (such as part of speech, tense, number, etc.) and they are aligned to mark correspondences between large text segments of Russian and French texts ranging from one to several sentences. An example fragment of an aligned Russian–French parallel text is presented in Table 1.

For many tasks of contrastive analysis, such as analysis of language specificity of Russian connectives in

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

Table 1 Example fragment of an aligned Russian–French parallel text*

Original Russian Text	[English translation]	French Translation Text
— Ну что ты, Алеша... лежи, молчи...	[— Come, come, Alyosha... lie down, don't talk...]	— Voyons, Aliocha... Allons, reste couché et tais-toi...
Пальто-то этой дамы у нас пока будет?	[Should we keep the coat of this lady for now?]	Le manteau de cette dame, on le garde ici, pour l'instant?
— Да, да. Чтобы Николка не вздумал тащить его.	[— Yes, yes. Don't let Nikolka even dream of bringing it back.]	— Oui, oui. Et que Nikolka ne s'avise pas de le porter.
А то на улице... Слышишь?	[Because on the street... Do you hear?]	Sinon, dans la rue... Tu comprends?
Вообще, ради бога, не пускай его никуда.	[For god's sake, don't let him go anywhere at all.]	En général, pour l'amour du ciel, ne le laisse aller nulle part.

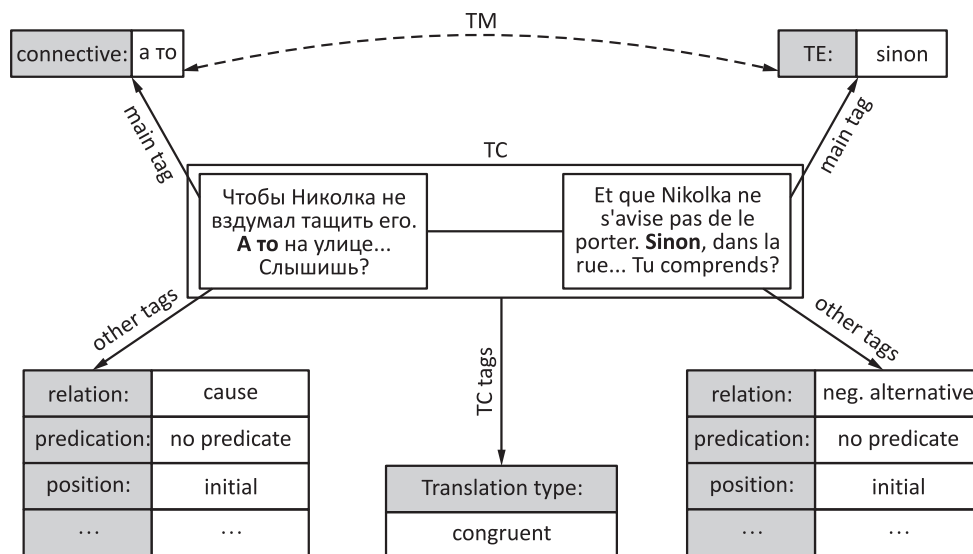
* Mikhail Boulgakov. The White Guard. 1926. French translation by Claude Ligny, 1970. Literal English translation of Russian texts here and hereafter is provided by the authors of the paper.

contrast with French, an additional layer of annotation is needed. The goal of this additional layer is to enable annotation of TCs between specific language units (LUs), in this case — Russian connectives, and their TEs in French. Note that Russian connectives may include more than one word while in contemporary corpora, only individual words are usually annotated. To support this type of annotation, a new information resource was created — the SCDB of Connectives.

The SCDB of Connectives includes information about TCs that experts record and annotate for Russian connectives using customizable sets of features (tags). All LUs in the SCDB of Connectives (both Russian connectives and their corresponding TEs) are annotated with the following information:

- word composition and relevant context of the LU;
- one main feature (tag) that specifies the type of the LU. A combination of main features of a Russian connective and its French TE defines one of the possible TMs for that connective; and
- several additional features (tags) that describe context properties of the LU. The most notable among these are the tags that specify the meaning of each connective; in other words, discourse relation between the corresponding text fragments that this connective marks.

In addition, some features apply to the TC as a whole and not to any of the LUs composing it. For example, translation type of a TC is tagged *congruent* if a Russian



The framework of the SCDB of Connectives and an example of an annotated TC (the TC presented in this figure is based on the fragment of the parallel text presented in Table 1)

connective corresponds to a French LU of the same category (i. e., a French connective), and *divergent* if it corresponds to a French LU of a different category (see [8, p. 23–25]). The figure presents an example of an annotated TC in the framework of the SCDB of Connectives.

3 Method of Calculation of Language Specificity Parameters

In [9], the following features are proposed as parameters that may signal language specificity of a Russian connective.

In Russian–French parallel texts:

1. A large number of different TMs.
2. A high proportion of divergent TEs (i. e., units other than connectives in translation).
3. A high proportion of zero-equivalents (connectives are not translated).

In French–Russian parallel texts:

4. A large number of different translation stimuli (TSs).
5. A high proportion of divergent TSs.
6. A high proportion of zero-stimuli (no stimuli for the connective is found in the original text).

Table 2 Normalization of the language specificity parameter 1 per 100 TCs based on the assumption of linear growth

Connective	Number of TCs	Number of TMs	Estimated number of TMs per 100 TCs
K_1	300	100	33.3
K_2	50	20	40

Parameters 2–3 and 5–6 are calculated as percentages from the total number of TCs for each connective. Evaluation of parameters 1 and 4 is less obvious. It is obvious that as the number of annotated TCs for a certain connective grows in the SCDB, the number of identified TMs for that connective will also grow. For instance, let us suppose that in the Russian–French parallel corpus of the SCDB 50 TCs were annotated for connective K_1 and 300 TCs — for connective K_2 . In addition, 20 TMs were identified for connective K_1 and 100 TMs — for connective K_2 . Then, based on the assumption of linear growth, one would arrive at the estimate shown in Table 2.

Based on this estimate, it could be concluded that connective K_2 generates new translation models slightly more often than connective K_1 . This is, however, not

true, because increase of the number of TMs relative to the number of TCs is not linear. It is intuitively clear that as more and more annotated TCs for a certain connective are registered in the database, growth of the number of new, previously unidentified TMs will be gradually slowing down. This assumption is supported by the data.

This means that in order to compare language specificity of these two connectives based on the parameter 1, it is necessary to approximate the ratio of increase in the number of TMs relative to the total number of TCs for a given connective (the same logic applies to parameter 4, which is calculated based on reverse French–Russian translations).

Analysis of available data showed that the growth of the number of TMs relative to number of TCs can be approximately described by the following formula¹:

$$n_i = R_i m_i^{0.65}$$

where m_i is the number of TCs created for connective K_i in the SCDB; n_i is the number of TMs identified for connective K_i in these TCs; and R_i is the *TM-factor* of language specificity of the connective K_i . The higher is the TM-factor, the higher is the likelihood for a certain connective to generate new TMs with each new TC created in the SCDB. Based on this formula, the TM-factor is calculated as follows:

$$R_i = \frac{n_i}{m_i^{0.65}}.$$

The proposed method for generating more accurate language specificity parameters is based on this formula. The updated estimate for the above speculative example (Table 3) clearly shows that Russian connective K_1 has a significantly higher TM-factor of language specificity and, thus, generates new TMs more often compared to connective K_2 .

In order to calculate language specificity parameters based on the number of different TSs (parameter 4), more experimental data should be gathered and processed to test and adopt the above formula to correctly reflect the growth of the number of TSs relative to number of TCs.

4 Examination of Language Specificity Parameters for *a mo* and *uuu*

Table 4 shows the language specificity parameters calculated based on 210 annotations of *a mo* and 210 annotations of *uuu* in the SCDB’s Russian–French corpus.

All the parameters differ significantly for the two connectives, their values for *a mo* are at least twice as

¹The formula was proposed and tested by M. Kruzchkov.

Table 3 Language specificity parameters based on the formula $n_i = R_i m_i^{0.65}$

Connective	Number of TCs, m_i	Number of TMs, n_i	TM-factor, R_i	Estimated number of TMs per 100 TCs
K_1	300	100	2.45	49.0
K_2	50	20	1.57	31.4

Table 4 Language specificity parameters of *a to* and *или* based on the SCDB data

Russian connectives	Total TCs	Zero ¹	Zero, %	Dvrg ²	Dvrg, %	TMs	TM-factor	TMs per 100
а то	210	54	25,71	6	2,86	33	1.02	20.37
или	210	20	9,52	1	0,48	18	0.56	11.11

¹Number of zero-equivalents (see language specificity parameter 3).

²Number of divergent TEs (see language specificity parameter 2).

high as for *или*. Of course, further semantic analysis is required to clarify the nature of these differences.

4.1 Connective ‘или’

Connective *или* expresses the relation of alternative (1) and in 70.48% of TCs in the SCDB, its French functional equivalent is connective *ou*, which has similar semantics.

- (1) Закон, устанавливающий *или* отягчающий ответственность, обратной силы не имеет.¹

La loi établissant *ou* aggravant la responsabilité d'une personne n'a pas d'effet rétroactif. [Constitution de la Fédération de Russie (M.: Lesage, 2000)]

Nevertheless, in different contexts the opposition of disjuncts can be more or less strong. For example, it is minimal in (2), where alternative is perceived almost as a connective relation, and maximal in (3), where alternative is interpreted as one that excludes one of the disjuncts.

- (2) Здесь чужой акцент только намечается, но он уже порождает оговорку *или* заминку в речи.²

Ici l'accent “étranger” est seulement esquissé, mais cela suffit à faire naître dans le discours des réserves et des hésitations. [Mikhaïl Bakhtine. La poésie de Dostoïevski (Isabelle Kolitcheff, 1970)]

- (3) Но правило для всех одно: «Прогибайся *или* прогибай».³

Mais la règle est la même pour tout le monde: *ou* tu te fais écraser, *ou* tu écrases les autres. [Svetlana Alexievitch. La fin de l'homme rouge ou le temps du désenchantement (Sophie Benech, 2013)]

These subtleties of meaning can be made explicit by translators. When opposition between disjuncts weakens, translators choose copulative *et* (2), *ainsi que*, and in the presence of negation of disjuncts — *ni*. Translation models with copulative semantics account for 7.62% of TCs with *или*. Some of the zero-equivalents (9.52%) can also be attributed to weakening the opposition: in such cases, disjuncts in translation are separated by a comma. When the opposition is strong, translators choose duplicate conjunctions *ou||ou* (3), *soit||soit*, single or duplicate *ou bien*. They account for 6.68% of the total TCs (5.24% of which account for single *ou bien*).

The remaining TMs (mostly, metalinguistic alternatives, such as *ou plutôt*, *autrement dit*, *c'est-à-dire*) account for 8.58%. These factors explain the number of translation models (18) and the TM-factor (0.56). However, they are all rather homogeneous in their semantics and could be replaced by the most frequent TM *ou*. Finally, the relation of alternative is extremely rarely expressed by other means than connectives: a divergent translation (a temporal alteration adverbial *de temps en temps*) was encountered only once (0.48%). This usage can be viewed as a context-driven translator's choice.

4.2 Connective ‘а то’

As opposed to *или*, *а то* is a polysemic connective; researchers distinguish up to 6 meanings for it (i. e., discourse relations that can be expressed by it, see [10]). It is important to note that language specificity parameters of *а то* vary considerably for different meanings of this connective (Table 5).

The most frequent meaning of *а то* is the negative alternative relation — it accounts for 56% of all its uses (4).

¹The law that establishes or aggravates responsibility does not have a retroactive effect.

²Here, the foreign accent is barely hinted at, but it already entails a fault or a halt in discourse.

³But the rule is the same for everybody: “Yield, or force others to yield.”

Table 5 Language specificity parameters of *a mo* combined with relations expressed by it

Relation	TCs	TCs, %	Zero	Zero, %	Dvrg	Dvrg, %	TMs	TM-factor	TMs per 100
negative alternative	116	56	11	9.48	2	1.72	19	0.86	17.25
cause	52	25	33	63.46	0	0.00	9	0.69	13.77
discrepancy	15	7	3	20.00	1	6.67	7	1.20	24.04
alternative	13	6	3	23.08	1	7.69	8	1.51	30.11
complementarity	9	4	3	33.33	0	0.00	6	1.44(!)	28.71(!)*
exception	2	1	0	0.00	2	100.00	2	—	—

*The estimate is approximate since there is not enough data.

(4) — Говори, *a mo* я заплачу.¹

— Parle, *sinon* je pleure. [Sergueï Dovlatov. L'Étranger (Jacques Michaut-Paterno, 2001)]

When combined with this relation, the language specificity parameters of *a mo* are relatively low. Zero-equivalents are less than 10%, divergent translations are also rare (1,72%), and TM-factor is also relatively low (0.86). The most frequent TM is *sinon* (63.79%), which can be combined with *mais* and *et* (1.72%); there are also a few synonyms of *sinon*: *sans cela*, *sans quoi*, *sans ça*, *autrement*, and *ou* (12.07%). Translation models from this semantic group account for 77.58% of all TCs for negative alternative *a mo* in the SCDB, and if we include zero-equivalents, the total reaches 87.06%.

The second most frequent meaning of *a mo* is the one that expresses cause (25%):

(5) По крайней мере позвольте объяснить господину Вольдемару, в чем дело, (...) *a mo* он совсем растерялся.²

Permettez au moins que nous expliquions le jeu à M. Voldémar, (...). *Car* il a complètement perdu le nord... [Ivan Tourgueniev. Premier Amour (Michel-Rostislav Hofmann, 1974)]

The most frequent equivalent here is the zero-equivalent, which is not unusual for causal relations (see [11, 12]), but the actual percentage (63.46%) surely attracts attention. In cases where a TE can be identified, equivalents are usually markers of casual relations: *car* (17.31%), *parce que* (3.85%), *à cause de* (1.92%). One should also note that in some cases, *a mo* with negative alternative relation is translated into French by causal conjunctions (1.74%) and at the same time, casual *a mo* in some cases (7.72%, even more often) is translated into French by units that express the negative alternative relation (*sinon*, *sans quoi*). Compare (6), where in the original text connective *a mo* expresses cause, while in the translation, the present tense is substituted by the future and, as a result, the causal motivation is transformed

into motivation through possible negative consequences of failure to implement the initial proposition.

(6) Нужно физическое движение, *a mo* мой характер решительно портится.³

J'ai besoin d'un exercice violent, *sinon* mon caractère deviendra intraitable [Léon Tolstoï. Anna Karénine (Henri Mongault, 1952)]

Nevertheless, the TM-factor for causal relation is lower than for negative alternative relation: 0.69 which is evidently caused by extremely high percentage of zero-equivalents.

Next in terms of frequency are relations of discrepancy (7%) and alternative (6%). The TM-factors for these relations are quite high: 1.2 and 1.51, respectively. This may have to do with the fact that *a mo* belongs to an informal style of discourse abundant with fixed expressions for which it is often difficult to find an exact equivalent; so, the choice of a TMs is often dependent on the context. However, as in the above cases, all TMs belong to the same semantic class.

The most frequent TM for *a mo* with discrepancy meaning is *mais* (26.67%, see (7)); next in frequency are *tandis que* and zero-equivalent (20% each).

(7) Ну, захотел помочь — дай пятнадцать, дай двадцать, ну да хоть три целковых себе оставь, *a mo* все двадцать пять так и отвалил!⁴

Ma foi, si tu voulais l'aider, tu n'avais qu'à donner quinze, vingt roubles, et en garder pour toi ne serait-ce que trois, *mais* il les a lâchés d'un coup tous les vingt-cinq! [Fédor Dostoïevski. Crime et châtiment (Élisabeth Guertik, 1947)]

Another possible equivalent is conjunction *et* (13.33%), which is also used here in an oppositional context (8) and, finally, there is *au lieu que* (6.67%), so that the total amounts to 86.66%.

¹Speak, or else I will cry.

²At least, let me explain to Mr. Voldemar what is the matter, (...) *since* he is completely at a loss.

³Physical movement is necessary, *since* my character clearly deteriorates.

⁴Well, if you want to help, then give 15, give 20, but keep at least three rubles for yourself, *but* he just likes that gave away all 25.

(8) — Он женится! Хочешь об заклад, что не женится? — возразил он. — Да ему Захар и спать-то помогает, *a mo* жениться!¹

— Lui, se marier! je parie qu'il ne se mariera pas! répliqua-t-il. Il a besoin de Zakhar même pour dormir, *et tu veux qu'il se marie!* [Ivan Gontcharov. Oblomov (Luba Jurgenson, 1988)]

High TM-factor of *a mo* coupled with relation of alternative (1.51) has to do with the fact that different TMs express different shades of this relation: temporal alternation (*tantôt, de temps en temps*), gradation (*même*), or intensification of alternative (*ou encore, ou bien*). Nevertheless, in more than 50% of cases, we encounter either TMs belonging to different semantic classes (30.76%), or zero-equivalents (23.8%).

Unfortunately, there are insufficient data at this point to obtain representative statistics and make reliable conclusions about the two remaining relations — complementarity (4.21%) and exception (0.53%).

5 Concluding Remarks

The new information resource, the SCDB of Connectives, and the associated new method of analyzing corpus data offer researchers valuable insights that can produce tangible evidence that *или* is not a language specific connector in the context of Russian–French language pair. In 70.48% of cases, its TE is conjunction *ou* that expresses the same relation; in most other cases, the TMs selected by translators may be replaced by *ou* without distortion of meaning.

On the other hand, analysis of the SCDB data suggests that Russian connective *a mo* has a relatively high degree of language specificity, especially for some relations that can be expressed by it. This is due to the fact that no connective with similar semantic configuration exists in French. Its most frequent TE is *sinon*, but it is found in only 38% of TCs, and what's more, it is associated with a different range of meanings (*sinon* is translated into Russian by *a mo* in 44.75% of cases, mainly expressing the relation of negative alternative, which is common for both connectives and has one of the lowest TM-factors). These findings allow us to attribute *a mo* to language specificity Type 4, where “the connective K_A of language A has in language B a systemic equivalent K_B which is the most frequent in translation, but does not reproduce all the values and uses of K_A ” [9]. In addition, statistical analysis of the corpus data allowed us to arrange 6 meanings of the Russian connective *a mo* according to their frequencies.

Thus, the proposed method of statistical analysis can lend greater objectivity to linguistic research by supple-

menting semantic analysis with quantitative data that can either support or question it. The results of this research are applicable to many areas related to informatics (machine translation) and comparative study of languages, such as lexical typology, lexicography, and theory and practice of translation.

Acknowledgments

The work was carried out at the Institute of Informatics Problems (FRC CSC RAS) and funded by Russian Science Foundation according to the research project No. 16-18-10004.

References

1. Wierzbicka, A. 1992. *Semantics, culture, and cognition. Universal human concepts in culture-specific configurations*. Oxford: Oxford University Press. 496 p.
2. Zaliznyak, Anna A., and I. B. Levontina. 1996. Otrazhenie natsional'nogo kharaktera v leksike russkogo yazyka [The reflection of the national character in the vocabulary of the Russian language]. *Russ. Linguist.* 20:237–264.
3. Zaliznyak, Anna A., I. B. Levontina, and A. D. Shmelev. 2012. *Konstanty i peremennye russkoy yazykovoy kartiny mira* [Constants and variables of the Russian language picture of the world]. Moscow: Yazyki Slavyanskikh Kul'tur [Languages of Slavic Cultures]. 696 p.
4. Zaliznyak, Anna A. 2015. Lingvospetsifichnye edinitny russkogo yazyka v svete kontrastivnogo lingvisticheskogo analiza [Russian language-specific words as an object of contrastive corpus analysis]. *Computational Linguistics and Intellectual Technologies: Conference (International) “Dialogue 2015” Proceedings*. Moscow: RGGU. 14(21):683–695.
5. Kruzchkov, M. 2016. Supracorpora Databases as corpus-based superstructure for manual annotation of parallel corpora. *8th Conference (International) on Corpus Linguistics*. EPiC ser. in language and linguistics. 1:236–248. Available at: <https://easychair.org/publications/paper/270289> (accessed May 29, 2017).
6. Inkova, O., and M. Kruzchkov. 2016. Nadkorpornye russko-frantsuzskie bazy dannykh glagol'nykh form i konnektorov [Supracorpora databases of Russian and French verbal forms and connectors]. *Lingue slave a confronto*. Bergamo: Bergamo University Press. 365–392.
7. Inkova, O., and N. Popkova. 2017. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):123–131.
8. Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam: John Benjamins B.V. 355 p.
9. Inkova, O. 2017. Printsipy opredeleniya stepeni lingvospetsifichnosti konnektorov [Principles of how to determine the degree of language-specificity of connec-

¹“He’s getting married! Want a bet that he will not marry?” he retorted. “Zakhar even helps him to sleep, *and* [you talk of] marriage!”

- tives]. *Computational Linguistics and Intellectual Technologies: Conference (International) "Dialogue 2017" Proceedings*. Moscow: RGGU. 16(23):139–149.
10. Inkova-Manzotti, O. 2005. Encore sur la conjonction russe a to. *Revue des études slaves* 76(4):485–497.
11. Sanders, T. J. M. 2005. Coherence, causality and cognitive complexity in discourse. *1st Symposium (International) on the Exploration and Modelling of Meaning Proceedings*. Toulouse: University of Toulouse-le-Mirail. 105–114.
12. Hoek, J., J. Evers-Vermeul, and T. Sanders. 2015. The role of expectedness in the implicature and explicitation of discourse relations. *Discourse in Machine Translation (DiscoMT 2015): 2nd Workshop Proceedings*. Lisbon, Portugal: Association for Computational Linguistics. 41–46.

Received July 6, 2018

Contributors

Inkova Olga Yu. (b. 1965) — Doctor of Science (PhD) in philology, 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; olyainkova@yandex.ru

Kruzhkov Mikhail G. (b. 1975) — 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; magnit75@yandex.ru

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЛИНГВОСПЕЦИФИЧНОСТИ КОННЕКТОРОВ (НА МАТЕРИАЛЕ ПАРАЛЛЕЛЬНЫХ КОРПУСОВ)*

О. Ю. Инькова, М. Г. Кружков

Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Аннотация: В последние десятилетия проблемы лингвоспецифичности в русском языке привлекают пристальное внимание исследователей, хотя до последнего времени они рассматривались без привлечения корпусных методов. В этой работе описываются новые методы исследования лингвоспецифичности русских коннекторов с привлечением статистического анализа аннотированных параллельных корпусов. Русско-французские и франко-русские параллельные тексты обрабатываются с помощью Надкорпусной базы данных коннекторов, разработанной специально для аннотирования переводных соответствий, выявляемых в параллельных корпусах. Каждое переводное соответствие включает в себя аннотацию русского коннектора и его переводного эквивалента, что позволяет генерировать статистические данные по различным типам моделей перевода — на основе предлагаемых для коннекторов параметров лингвоспецифичности. В качестве примера в данной работе рассматриваются два русских коннектора: *или* и *a to*. На основе предлагаемых статистических параметров будет показано, что *или* имеет очень низкий уровень лингвоспецифичности в сопоставлении с французским языком, в то время как коннектор *a to* обладает высокой лингвоспецифичностью. Результаты данного исследования могут быть востребованы в информатике (машинный перевод, статистический анализ текстовых данных), а также в различных областях, связанных с контрастивными лингвистическими исследованиями, таких как лексическая типология, лексикография и переводоведение.

Ключевые слова: надкорпусные базы данных; статистический анализ; контрастивный корпусный анализ; лингвоспецифичность; параллельные корпуса; лингвистические информационные ресурсы; коннекторы; дискурсивные отношения; семантика

DOI: 10.14357/19922264180312

Литература

1. *Wierzbicka A.* Semantics, culture, and cognition. Universal human concepts in culture-specific configurations. — Oxford: Oxford University Press, 1992. 496 p.
2. *Зализняк Анна А., Левонтина И. Б.* Отражение национального характера в лексике русского языка // *Russ. Linguist.*, 1996. Вып. 20. С. 237–264.
3. *Зализняк Анна А., Левонтина И. Б., Шмелев А. Д.* Константы и переменные русской языковой картины мира. — М.: Языки славянских культур, 2012. 696 с.
4. *Зализняк Анна А.* Лингвоспецифичные единицы русского языка в свете контрастивного корпусного анализа // *Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Между-*

* Работа выполнена при финансовой поддержке РФФИ (проект 16-18-10004).

- нар. конф. «Диалог». — М.: РГГУ, 2015. Вып. 14(21). С. 683–695.
5. *Kruzikov M.* Supracorpora Databases as corpus-based superstructure for manual annotation of parallel corpora // 8th Conference (International) on Corpus Linguistics. — EPIC ser. in language and linguistics. — 2016. Vol. 1. P. 236–248. <https://easychair.org/publications/paper/270289>.
 6. *Инькова О., Кружков М.* Надкорпусные русско-французские базы данных глагольных форм и коннекторов // *Lingue slave a confronto*. — Bergamo: Bergamo University Press, 2016. С. 365–392.
 7. *Inkova O., Popkova N.* Statistical data as information source for linguistic analysis of Russian connectors // *Информатика и её применения*, 2017. Т. 11. Вып. 3. С. 123–131.
 8. *Johansson S.* Seeing through Multilingual Corpora. — Amsterdam: John Benjamins B.V., 2007. 355 p.
 9. *Инькова О.* Принципы определения степени лингвоспецифичности коннекторов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2017. Вып. 16(23). С. 139–149.
 10. *Inkova-Manzotti O.* Encore sur la conjonction russe a to // *Revue des études slaves*, 2005. Vol. 76. No. 4. P. 485–497.
 11. *Sanders T. J. M.* Coherence, causality and cognitive complexity in discourse // 1st Symposium (International) on the Exploration and Modelling of Meaning Proceedings. — Toulouse: University of Toulouse-le-Mirail, 2005. P. 105–114.
 12. *Hoek J., Evers-Vermeul J., Sanders T.* The role of expect-edness in the implicitation and explicitation of discourse relations // *Discourse in Machine Translation (DiscoMT 2015): 2nd Workshop Proceedings*. — Lisbon, Portugal: Association for Computational Linguistics, 2015. P. 41–46.

Поступила в редакцию 06.07.2018

СЕМАНТИЧЕСКАЯ ОБРАБОТКА НЕСТРУКТУРИРОВАННЫХ ТЕКСТОВЫХ ДАННЫХ НА ОСНОВЕ ЛИНГВИСТИЧЕСКОГО ПРОЦЕССОРА PullEnti

Е. Б. Козеренко¹, К. И. Кузнецов², Д. А. Романов³

Аннотация: Представлена методика создания систем извлечения знаний, основанная на подходе, главным инструментом которого является программный пакет PullEnti, включающий алгоритмы морфологического и семантико-синтаксического анализа для выделения сущностей определенных типов из текстов естественного языка (персоны, организации, локации и другие целевые семантические объекты). В системе PullEnti используются динамически подключаемые компоненты (плагины), что позволяет без перекомпилирования активировать различные функциональные возможности. Именно таким образом запускается блок семантического анализа. В процессе анализа выделяются семантические единицы (токены), которые представляют собой типизированные фразы: текстовые, числовые и др. Приводятся примеры реализованных проектов для различных предметных областей.

Ключевые слова: семантическое моделирование; извлечение именованных сущностей; области с интенсивным использованием данных; автоматизированные системы извлечения знаний; семантический поиск; интеллектуальные интернет-технологии

DOI: 10.14357/19922264180313

1 Введение

Задача автоматического анализа текстовой информации, представленной в Интернете, является актуальной во всем мире. В данной статье представлены результаты исследований и разработок, направленных на решение научной проблемы создания оптимальной методики логико-статистического моделирования механизмов целевого семантического анализа в информационных системах с интенсивным использованием знаний, выполняющих функции извлечения знаний, поддержки аналитических решений, в том числе в среде нескольких естественных языков.

Для решения полного спектра задач обработки естественного языка создан семантически-ориентированный лингвистический процессор (СОЛП). Центральным компонентом СОЛП является инструментальный пакет (SDK-модуль) PullEnti. Этот процессор в рамках проводимых соревнований конференции «Диалог-2016» занял несколько первых мест при анализе текстов в рамках решения задач извлечения именованных сущностей. Разработчик PullEnti — Кузнецов Константин Игоревич. В системе PullEnti используются динамически подключаемые компоненты (плагины), что

позволяет без перекомпилирования запускать различные функциональные возможности. Именно таким образом активируется блок семантического анализа.

В процессе анализа выделяются семантические единицы (токены), которые представляют собой типизированные фразы, такие как текстовые, числовые и др. Например, в результате анализа фразы «В 2017 году» будут выделены три токена: «В» — текстовый; «году» — текстовый; «2017» — числовой. Такие токены можно назвать простыми. Кроме того, выделяются *метатокены* — сложные токены, которые объединяют несколько простых токенов, например существительные с определителями, скобки, кавычки и т. п.

В системе существует пополняемый статический словарь терминов. В него можно добавлять термины и затем проверять их наличие в тексте. Кроме того, в системе можно формировать динамически подобные словари на основе анализа текста.

При анализе текста создается аналитический контейнер, в который помещаются выделяемые сущности, токены в определенной последовательности, статистические данные и др.

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

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, k.smith@mail.ru

³Национальный исследовательский университет «Высшая школа экономики», DRomanov@it.ru

2 Лингвистическое моделирование в системах обработки знаний в многоязычной среде

Способы представления информации, знаний многообразны. Огромный объем данных представлен в виде текстов естественного языка, что делает задачу извлечения и структурирования информации из текстов весьма важной. Это относится к различным предметным областям. Для оперирования данными на компьютере необходимо выделить из текста объекты, их атрибуты, связи между объектами, процессы, в которых эти объекты задействованы, другую важную информацию, которая бы позволяла не только описать ситуацию, но и строить выводы, характерные для конкретной предметной области, прогнозировать развитие ситуации.

Для решения поставленных задач проведены эксперименты с различными грамматическими формализмами, в том числе с грамматикой категориального типа [1]. Проведено сравнительное исследование методов классификации применительно к лингвистическим задачам; выработан эффективный метод отображения вектора естественно-языковых структур в расширенное пространство признаков для классификации новых языковых объектов и структур; сформирована фокусная выборка параллельных текстов деловых и научных документов на русском и английском языках по различным отраслям науки и техники; сформирована расширенная система новых категорий для повышения изобразительных возможностей двуязычной грамматики; выработаны пути расширения базовых представлений на основе аппарата расширенных семантических сетей [2] и результатов применения метода векторных пространств, направленного на разрешение неоднозначности языковых структур для синтаксического разбора при распознавании текста в процессе извлечения знаний из текстов на разных естественных языках. Разработаны алгоритмы автоматического выравнивания параллельных текстов для развития грамматических компонент систем обработки знаний в многоязычном режиме.

Основной результат исследований — модель лингвистической составляющей интеллектуальных информационных систем, работающих в многоязычном пространстве для поиска информации, обеспечения оптимальных аналитических и управленческих решений в сферах деятельности с интенсивным использованием данных. Результаты исследований применяются в логико-семантиче-

ских и статистических процедурах обработки слабоструктурированной текстовой информации, при разработке технологии и инструментальных средств построения лингвистических компонент интеллектуальных систем и систем машинного перевода.

3 Представление лингвистических знаний на основе векторных пространств

Процедуры анализа и синтеза естественно-языковых высказываний отражают динамический характер языка как деятельности; соответственно, в модели, которая кладется в основу проекта системы обработки естественно-языковых высказываний, должен быть заложен механизм, позволяющий строить представления движения.

Методы машинного обучения на основе векторных моделей развиваются и используются в различных областях знаний, применительно к лингвистическим задачам эти методы вполне эффективны для разрешения лексической многозначности [3–8].

Более сложной задачей и новым направлением исследований возможности применения векторных моделей для представления и обработки лингвистических данных является моделирование грамматических преобразований на основе векторных пространств и тензоров. Тензор (от лат. *tensus*, напряженный) — объект линейной алгебры, преобразующий элементы одного линейного пространства в элементы другого. Часто тензор представляют как многомерную таблицу, заполненную числами — компонентами тензора $d \cdot d \cdot \dots \cdot d$, где d — размерность, над которой задан тензор, а число сомножителей совпадает с так называемой валентностью, или рангом тензора. Важно, что такое представление (кроме скаляров, т.е. тензоров валентности ноль) возможно только после выбора базиса (или системы координат): при смене базиса компоненты тензора меняются определенным образом. Сам тензор как «геометрическая сущность» от выбора базиса не зависит, компоненты вектора меняются при смене координатных осей, но сам вектор, образ которого может быть просто нарисованная стрелка, от этого не изменяется. Тензор обычно обозначают некоторой буквой с совокупностью верхних (контрвариантных) и нижних (ковариантных) индексов: $X_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r}$. При смене базиса ковариантные компоненты меняются так же, как и базис (с помощью того же преобразования), а контрвариантные — обратно изменению бази-

са (обратным преобразованием). Тензор является сущностью любой системы реального мира и сохраняется, несмотря на происходящие изменения в этой системе [9]. Эта особенность тензора чрезвычайно актуальна для моделирования языковых преобразований в лингвистических процессорах, когда необходимо выявлять сходные значения, выраженные многочисленными способами, системой разнородных языковых средств.

В работе, представленной в данной статье, используются два основных подхода к представлению смысла в вычислительной лингвистике: символьный подход [10, 11] и подход на основе дистрибутивной семантики [4–6, 9]; решение заключается в сочетании методов компьютерной лингвистики и когнитивной науки, в котором символьное и «коннекционистское» (от англ. *connectionist*, т.е. основанное на нейронных сетях как модели машинного обучения) представления объединяются с помощью тензорных произведений. Исследованы возможные применения данного метода для обработки синтаксических структур и контекстов в русском и английском языках и проведены межязыковые сопоставления.

4 Семантико-ориентированный лингвистический процессор

Методы, описанные выше, используются в процедурах семантической обработки текстовых знаний в СОЛП, который решает задачу извлечения структурированной информации из текстов на русском и английском языках. Ядром СОЛП является программный пакет PullEnti, включающий алгоритмы морфологического и синтаксического анализа, который позволяет выделять сущности определенных типов из текстов естественного языка (персоны, организации, локации и другие семантические объекты). «Именованная сущность» — это объект, содержащий набор значений атрибутов, отличающий его от других объектов этого же типа. В тексте находятся именованные сущности и устанавливаются семантические связи между ними, при этом учитывается возможность обозначения одной сущности несколькими способами (синонимия). Все множество сущностей, выделенных из текста или нескольких текстов, представляет собой ориентированный граф.

Предварительный этап обработки текстов включает в себя морфологический и синтаксический анализ. При морфологическом анализе текст разбивается на словоформы, также называемые токенами (от англ. *token* — пример использования

лингвистической единицы в тексте). Основными наследными классами базового класса Token являются TextToken и MetaToken. TextToken — это исходный фрагмент текста, содержащий результат морфологического анализа. TextToken ссылается на MorphToken, содержащий все морфологические варианты разбора. MetaToken — это группа токенов, соответствующих одной синтаксической или семантической группе. К классу метатокенов относятся NumberToken, представляющий число, и ReferentToken, представляющий сущность.

Приведем пример морфологического анализа предложения.

Исходный текст:

«По словам директора департамента экономического развития автономного округа Павла Сидорова, на эти цели планируется привлечь 200 млн рублей из федерального бюджета и еще 450 млн рублей из внебюджетных источников».

В результате морфологического разбора текст был разбит на словоформы, для каждой словоформы указана начальная форма, часть речи и морфологические характеристики. В случае многозначности или омонимии указываются все варианты морфологического разбора.

Также были выделены следующие метатокены:

Павел Сидоров — текстовый фрагмент «директора департамента экономического развития автономного округа Павла Сидорова»

200.000.000 RUB — текстовый фрагмент «200 миллионов рублей»

450.000.000 RUB — текстовый фрагмент «450 миллионов рублей»

Вслед за морфологическим анализом проводится выделение именованных сущностей различных типов. Семантико-ориентированный лингвистический процессор извлекает из текстов объекты следующих типов: дата, временной период, территориальное образование, денежная сумма, телефон, URL, адрес, организация, транспорт, свойство персоны, персона, декрет, часть декрета. Каждому типу соответствуют свои свойства и связи с объектами других типов.

Базовым классом для сущностей является класс Referent. Тип сущностей задается классом Referent Class, наследным от Referent, который содержит набор атрибутов. Значение может быть как простым (строка, число), так и ссылкой на другую сущность. Помимо значений атрибутов сущность содержит список ссылок на участки исходного текста, в которых эта сущность располагается. Для задач, в которых требуется обрабатывать множество текстов и хранить получаемые сущности, в СОЛП используется базовый класс Repository Base, облегчающий реализацию хранилища сущностей. Место хране-

ния сериализованных данных от сущностей определяется в наследном классе (например, это может быть реляционная система управления базами данных или файловая система). Repository Base берет на себя функции отождествления новых данных со старыми данными и поддержки непротиворечивости семантической сети.

Извлекаемая из текста информация должна быть адресной, поэтому из одного и того же текста можно извлекать совершенно различные виды информации, характерные для конкретной предметной области. В результате анализа текстовой информации выделяются типизированные объекты предметной области. Программа PullEnti стала основой для построения множества систем, таких как программа «Доктор Ватсон», система поиска экспертов, процессор BRef и др. Программа «Доктор Ватсон» предназначена для исследования массивов текстовой информации с целью выявления сущностей и связей между ними. При этом пользователь может добавить недостающие сущности и связи (которые не были выделены программой), настроить выдаваемую информацию, сформировать отчет о результатах работы программы. Данная программа может использоваться в таких сферах деятельности, как криминалистика, конкурентная разведка, маркетинг, реклама, безопасность. Результат работы программы — отчет об исследуемом объекте, диаграммы сущностей и связей — представлен на рис. 1. Из текущего текста выделены организации, персоны, их связи.

На рис. 2 представлены выделенные объекты, их связи. Для каждой связи выделяется тип связи и название (например, тип связи — «родственные», заголовок связи — «отец»; тип связи — «владение», заголовок связи — «особняк в центре Вашингтона» и т. д.), определяются попарно объекты-участники связи. Для более полного определения ситуации выделяется не только время, характеризующее текущую ситуацию, но и интервалы времени. Дополнительные параметры позволяют выяснить, является ли связь симметричной для данной пары выделенных объектов (субъектов). Также для каждого выделенного объекта (субъекта) выделяются атрибуты. Например, для типа объекта «персона» выделяются имя, фамилия, отчество, дата рождения и др.

Результаты работы программы могут быть представлены в виде графа (вкладка «Диаграммы») (см. рис. 3). В отчете выводятся обнаруженные объекты (персоны, организации, локации, атрибуты), их связи в удобном для анализа виде.

Программа «Логика ЕСМ. Правовая экспертиза» предназначена для автоматизации процесса проведения экспертизы проектов нормативно-правовых актов, организационно-распорядительных документов, договоров и других документов. Система значительно упрощает процесс проведения правовой экспертизы и сокращает его сроки, выполняя рутинные операции и кардинально снижая затраты рабочего времени квалифицированных юристов. Система «Логика ЕСМ. Правовая экспер-



Рис. 1 Результаты работы программы «Доктор Ватсон». Выделение сущностей

Проект • Типы объектов и связей

Объекты и тексты | Связи | Диаграммы | Отчеты

Одинаковые атрибуты | Новая связь

Все связи

Тип	Заголовок связи	Первый объект	Второй объект	Параметр	Симметричная	Комментарий
Коллеги		Нуланд Виктория	Чейни Ричард	с 2003 год по 2005 год	<input type="checkbox"/>	
Родство...	отец	Нуланд Виктория	Шерин Нуланд		<input type="checkbox"/>	
Родство...	дед	Нуланд Виктория	Нудельман Мейер		<input type="checkbox"/>	
Учеба	училась	Нуланд Виктория	Шоат Розмари Холл	с 1967 год по 1978 год	<input type="checkbox"/>	
Учеба	училась	Нуланд Виктория	Брауновский Университет	с 1979 год по 1984 год	<input type="checkbox"/>	
Родство...	муж	Нуланд Виктория	Каган Роберт		<input checked="" type="checkbox"/>	
Владение	особняк на второй линии в Майами	Нуланд Виктория	Дом	1986 год по настоящий момент	<input type="checkbox"/>	
Владение	пентхаус в центре Вашингтона	Нуланд Виктория	Кварира	1995 год по настоящий момент	<input type="checkbox"/>	
Владение	владелец	Нуланд Виктория	ИнформАгентство ФорвардН...	1997 год по настоящий момент	<input type="checkbox"/>	
Работа	представитель	Псака Дженифер	Государственный Департам...	1987 год по настоящий момент	<input type="checkbox"/>	
Родство...	отец	Каган Роберт	Каган Дональд		<input type="checkbox"/>	
Родство...	братья	Каган Роберт	Каган Фредерик		<input type="checkbox"/>	


Первый объект связи
Краткое описание
Нуланд Виктория
Комментарий

Второй объект связи
Краткое описание
Каган Роберт
Комментарий

Атрибуты

Тип	Значение	Комментарий
Фамилия	НУЛАНД	
Имя	ВИКТОРИЯ	
ФИО	Нуланд Виктория	
ФИО	Nuland Victoria	
Фамилия	Nuland	

Картинка



Атрибуты

Тип	Значение	Комментарий
Фамилия	КАГАН	
Имя	РОБЕРТ	
ФИО	Каган Роберт	
ФИО	Каган Robert	
Дата рождения	26 сентября 1958 года	

Картинка




Рис. 2 Выделение связей, периодов

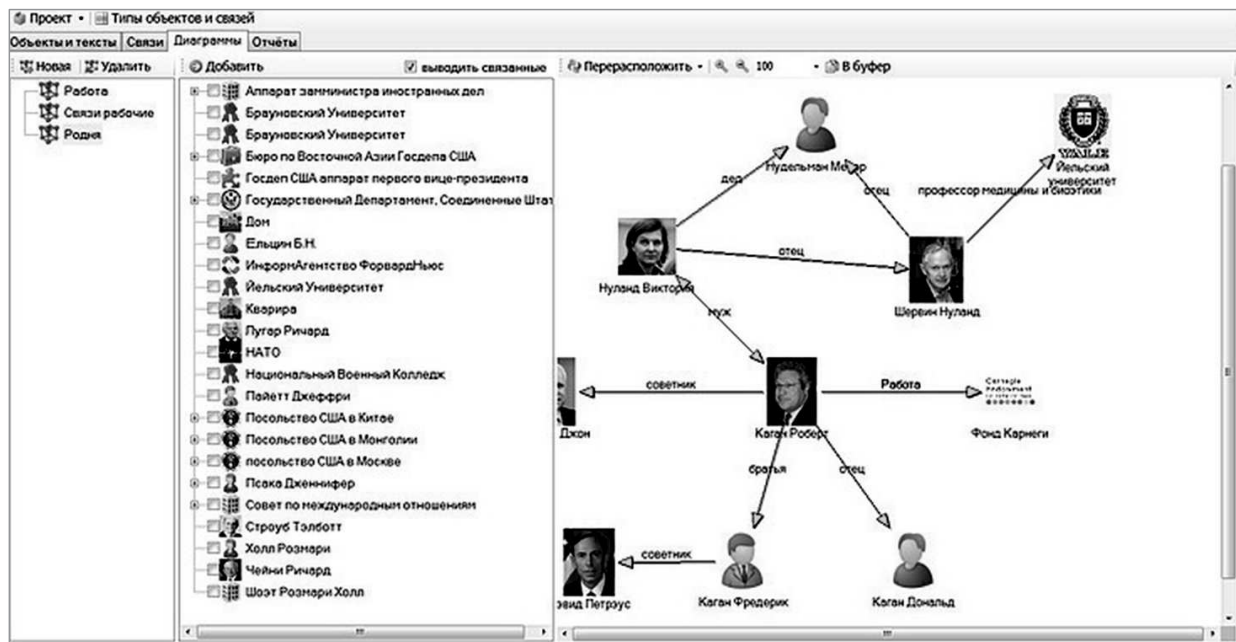


Рис. 3 Графическое представление результатов работы программы «Доктор Ватсон»

тиза» автоматически за несколько секунд поможет, например, установить:

- не содержатся ли в проверяемом документе ссылки на нормативные правовые акты, которые утратили силу;
- нет ли в проверяемом документе фрагментов других документов, не возникает ли избыточное дублирование нормативной документации;

- соответствует ли оформление и структура документа установленным в организации правилам;
- нет ли ошибок в оформлении цифровой информации в договоре, соответствуют ли друг другу суммы, указанные цифрами и прописью, правильно ли рассчитан налог на добавленную стоимость и т. п. На основе лингвистического процессора PullEnti был реализован процессор обработки ссылок и списка литературы BREF,

который позволяет по выделенной информации построить *Граф Цитирования* и *Граф Соавторов*, отражающие формальные связи в коллекции документов.

Лингвистический процессор PullEnti под псевдонимом Pink на соревновании FactRuEval конференции «Диалог-2016» занял первые места на большинстве дорожек [12].

Соревнование проводилось на следующих дорожках:

- определение в тексте границ именованных сущностей, таких как персона, организация, локация;
- выделение именованных сущностей с определением атрибутов в нормализованном виде. Для персон это фамилия, имя и отчество. Для организаций и локаций — нормализованное название;
- извлечение фактов (например: «встреча», «покупка», ...) и наборов строковых полей (например: «участник встречи 1», «участник встречи 2», «место встречи», «дата/время начала встречи», ...).

5 Заключение

Лингвистические процессоры на основе программы PullEnti могут быть использованы в различных областях, в которых информация представлена в текстовом виде. Особенно это важно в тех случаях, когда необходимо выделять важную информацию из большого потока документов на естественном языке. Очень хорошо данная технология работает в задачах кластеризации текстов по определенным признакам. При этом существует возможность автоматической настройки программы на требования пользователя.

Описанные выше системы, созданные на основе технологии PullEnti, доказывают ее эффективность в самых различных областях. Следующими шагами исследований станут: методы уточнения границ моделируемых предметных областей за счет построения семантического ядра каждой области (в том числе с использованием методов вероятностного тематического моделирования); выделение массива неявных ссылок (упоминаний персон и идей, выраженных ключевыми фразами/значимыми словосочетаниями); расчет корреляции между явными и неявными ссылками в рамках созданной коллекции, формирование функционально-грамматических модулей естественных языков, включаемых в лингвистический процессор.

Литература

1. *Shaumyan S.* Categorical grammar and semiotic universal grammar // Conference (International) on Artificial Intelligence Proceedings. — Las Vegas, NV, USA: CSREA Press, 2003. P. 623–629.
2. *Kuznetsov I. P., Kozerenko E. B., Matskevich A. G.* Intelligent extraction of knowledge structures from natural language texts // IEEE/WIC/ACM Joint Conferences (International) on Web Intelligence and Intelligent Agent Technology Proceedings. — Washington, DC, USA: IEEE Computer Society, 2011. Vol. 3. P. 269–272.
3. *Dempster A. P., Laird N. M., Rubin D. B.* Maximum likelihood from incomplete data via the EM algorithm // J. Roy. Stat. Soc. B, 1977. Vol. 39. Iss. 1. P. 1–22.
4. *Lund K., Burgess C.* Producing high-dimensional semantic spaces from lexical co-occurrence // Behav. Res. Meth. Ins. C., 1996. Vol. 28. Iss. 2. P. 203–208.
5. *Curran J. R.* From distributional to semantic similarity. — Edinburgh: University of Edinburgh, 2004. PhD Thesis. 177 p. <https://www.inf.ed.ac.uk/publications/thesis/online/IP030023.pdf>
6. *McCarthy D., Koeling R., Weeds J., Carroll J.* Finding predominant senses in untagged text // 42nd Annual Meeting of the Association for Computational Linguistics Proceedings. — Stroudsburg, PA, USA: Association for Computational Linguistics, 2004. P. 280–287. doi: 10.3115/1218955.1218991.
7. *Clark S., Pulman S.* Combining symbolic and distributional models of meaning // AAAI Spring Symposium on Quantum Interaction Proceedings. — Palo Alto, CA, USA: AAAI Press, 2007. 4 p. <http://www.cl.cam.ac.uk/~sc609/pubs/aaai07.pdf>.
8. *Kozerenko E. B.* Parallel texts alignment strategies // Conference (International) on Artificial Intelligence Proceedings. — Las Vegas, NV, USA: CSREA Press, 2012. Vol. 2. P. 945–951.
9. *Danielson D. A.* Vectors and tensors in engineering and physics. — 2nd ed. — Boulder, CO, USA: Westview (Perseus), 2003. 287 p.
10. *Montague R.* Universal grammar // Theoria, 1970. Vol. 36. P. 373–398. (Reprinted in: Formal philosophy: Selected papers of Richard Montague / Ed. R. H. Thomason. — New Haven, CT, USA: Yale University Press, 1974. P. 7–27.)
11. *Pang B., Knight K., Marcu D.* Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences // Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings. — Stroudsburg, PA, USA: Association for Computational Linguistics. 2003. Vol. 1. P. 102–109. doi: 10.3115/1073445.1073469.
12. FACRUEVAL. Evaluation of named entity recognition and fact extraction systems for Russian, 2016. <http://www.dialog-21.ru/media/3430/starostinaetal.pdf>.

Поступила в редакцию 13.07.18

SEMANTIC PROCESSING OF UNSTRUCTURED TEXTUAL DATA BASED ON THE LINGUISTIC PROCESSOR PullEnti

E. B. Kozerenko¹, K. I. Kuznetsov¹, and D. A. Romanov²

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

²National Research University “Higher School of Economics,” 20 Myasnikskaya Str., Moscow 101000, Russian Federation

Abstract: The paper presents the method for creation of knowledge extraction systems based on the approach employing the software tool system PullEnti comprising the algorithms for morphological and semantic-syntactical analysis which makes it possible to extract entities of certain types from natural language texts (persons, organizations, locations, and other target semantic objects). The PullEnti system uses dynamically connected components (plugins) which makes it possible to activate various functions without recompiling. This is how the semantic analysis unit is incorporated. During the analysis, the semantic units (tokens) are established, which are typed phrases: text, numerical data, etc. Examples of implemented projects for different subject areas are given.

Keywords: semantic modeling; named entities recognition, data intensive domains; automated systems of knowledge extraction; semantic search; intelligent Internet technologies

DOI: 10.14357/19922264180313

References

1. Shaumyan, S. 2003. Categorical grammar and semiotic universal grammar. *Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, NV: CSREA Press. 623–629.
2. Kuznetsov, I. P., E. B. Kozerenko, and A. G. Matskevich. 2011. Intelligent extraction of knowledge structures from natural language texts. *IEEE/WIC/ACM Conferences (International) on Web Intelligence and Intelligent Agent Technology Proceeding*. Washington, DC: IEEE Computer Society. 3:269–272. doi: 10.1109/WI-IAT.2011.235.
3. Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B* 39(1):1–22.
4. Lund, K., and C. Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Meth. Ins. C*. 28(2):203–208.
5. Curran, J. R. 2004. From distributional to semantic similarity. Edinburgh: University of Edinburgh. PhD Thesis. 177 p. Available at: <https://www.inf.ed.ac.uk/publications/thesis/online/IP030023.pdf> (accessed July 19, 2018).
6. McCarthy, D., R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant senses in untagged text. *42nd Annual Meeting of Association for Computational Linguistics Proceedings*. Stroudsburg, PA: Association for Computational Linguistics. 280–287. doi: 10.3115/1218955.1218991.
7. Clark, S., and S. Pulman. 2007. Combining symbolic and distributional models of meaning. *AAAI Spring Symposium on Quantum Interaction Proceedings*. Palo Alto, CA: AAAI Press. 4 p. Available at: <http://www.cl.cam.ac.uk/~sc609/pubs/aaai07.pdf> (accessed July 19, 2018).
8. Kozerenko, E. B. 2012. Parallel texts alignment strategies. *Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, NV: CSREA Press. 2:945–951.
9. Danielson, D. A. 2003. *Vectors and tensors in engineering and physics*. 2nd ed. Boulder, CO: Westview Press. 287 p.
10. Montague, R. 1970. Universal grammar. *Theoria* 36:373–398. (Reprinted in: 1974. *Formal philosophy: Selected papers of Richard Montague*. Ed. R. H. Thomason. New Haven, CT: Yale University Press. 7–27.)
11. Pang, B., K. Knight, and D. Marcu. 2003. Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences. *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings*. Stroudsburg, PA: Association for Computational Linguistics. 1:102–109. doi: 10.3115/1073445.1073469.
12. FACRUEVAL. 2016. Evaluation of named entity recognition and fact extraction systems for Russian. Available at: <http://www.dialog-21.ru/media/3430/starostinaetal.pdf> (accessed July 19, 2018).

Received July 13, 2018

Contributors

Kozerenko Elena B. (b. 1959) — Candidate of Science (PhD) in linguistics, 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; kozerenko@mail.ru

Kuznetsov Konstantin I. (b. 1968) — leading engineer, 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; k.smith@mail.ru

Romanov Dmitri A. (b. 1967) — Candidate of Science (PhD) in technology, associate professor, National Research University “Higher School of Economics,” 20 Myasnitskaya Str., Moscow 101000, Russian Federation; DRomanov@it.ru

УПРАВЛЕНИЕ ВЫХОДОМ СТОХАСТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ СИСТЕМЫ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ. I. ОПТИМАЛЬНОЕ РЕШЕНИЕ МЕТОДОМ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ*

А. В. Босов¹, А. И. Стефанович²

Аннотация: Решается задача оптимального управления для диффузионного процесса Ито и линейного управляемого выхода. Рассматриваемая постановка близка к классической линейно-квадратичной гауссовской задаче управления (linear-quadratic Gaussian (LQG) control). Отличия состоят в том, что состояние описывается нелинейным дифференциальным уравнением Ито $dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t$ и не зависит от управления u_t , оптимизации подлежит управляемый линейный выход $dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t$. Дополнительные обобщения внесены в квадратичный критерий качества с целью возможности постановки таких задач, как отслеживание выходом состояния или управлением — линейной комбинации состояния и выхода. Для решения используется метод динамического программирования. Функцию Беллмана позволяет найти предположение о ее структуре вида $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$. Решение дают три дифференциальных уравнения для коэффициентов α_t , $\beta_t(y)$ и $\gamma_t(y)$. Эти уравнения составляют оптимальное решение рассматриваемой задачи.

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

DOI: 10.14357/19922264180314

1 Введение

Ключевые результаты в области оптимизации стохастических динамических систем, составляющие классическую теорию управления, получены более 40 лет назад (такова работа [1] в отношении задачи управления линейно-гауссовскими стохастическими системами по квадратичному критерию). К классической теории следует относить линейные модели стохастических систем и квадратичный критерий качества. Это исходный базис, на котором основано множество успешно исследованных и решенных задач стохастического управления и оптимизации.

Дальнейшее развитие — это новые модели и критерии, но прежде всего это новые методы: от теории линейных регуляторов, метода динамического программирования и принципа максимума к адаптивному и минимаксному подходу, импульсному управлению и т. д. Множество инноваций как в части моделей, так и в части математического аппарата, имевших место в последующие годы, существенно обогатили теорию управления. Но и до

настоящего времени линейные модели и квадратичный критерий, несмотря на всю справедливую критику в отношении их адекватности и гибкости, сохраняют исследовательский интерес и находят современные области приложения.

Не претендуя на сколько-нибудь полное обоснование последнего тезиса, приведем несколько примеров, показавшихся наиболее интересными.

Так, в [2] решается линейно-квадратичная задача в игровой постановке с запаздыванием. В близкой по модели работе [3] задача управления ставится в терминах H_∞ -робастности. Точнее называть эту тематику H_2/H_∞ -управлением, и работ по этой теме очень много. Аккуратности ради следует уточнить, что под линейными понимаются модели с мультипликативными по состоянию возмущениями.

Совсем другой класс моделей, особо популярных в последние годы, составляют скачкообразные процессы. Например, линейные уравнения в сочетании с пуассоновскими скачками в [4] исполь-

* 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

зуются в моделях, описывающих различные показатели функционирования сетевых протоколов передачи данных транспортного уровня. Телекоммуникации представляют в последние годы самый популярный прикладной материал для исследований, работ по этой проблематике множество, математические техники привлекаются самые разные и самые современные, но и линейным моделям место находится. Еще один любопытный пример исследования скачкообразного процесса и оптимизации на основе квадратичного критерия можно найти в [5] применительно к задаче инвестирования на финансовом рынке. Наконец, упомянем еще работу [6], подводящую итог исследований в отношении классической детерминированной линейно-квадратичной задачи с использованием техники матричных неравенств.

В данной работе также эксплуатируются привлекательные свойства линейных моделей и квадратичного критерия, причем в стохастической постановке. Направлением для обобщения выбрана модель динамики системы: основные усилия направлены на то, чтобы сделать ее нелинейной. Кроме того, представленная постановка может рассматриваться и как обобщение ранее решенной задачи в дискретном времени [7, 8] на время непрерывное. В упомянутых работах помимо собственно модельной постановки важна еще и привлекаемая прикладная область — функционирование сложных программных систем. Результатов, ориентированных непосредственно на такие приложения, к настоящему времени пренебрежимо мало, поэтому [7, 8] — это еще и прикладное обоснование рассматриваемой далее задачи.

Оптимизируемая динамическая система описывается двумя уравнениями. Состояние задается нелинейным стохастическим дифференциальным уравнением Ито, не содержащим управляемой переменной. Возмущение здесь описывается стандартным винеровским процессом, накладываются простые условия существования и единственности решения. Поскольку состояние не управляется, то уместно его интерпретировать как сложное внешнее возмущение. Вторая переменная — управляемый выход — задается линейным стохастическим дифференциальным уравнением. Цель оптимизации выхода формируется квадратичным критерием общего вида. Формальная постановка задачи приведена в следующем разделе.

Для решения задачи используется метод динамического программирования, решается уравнение Беллмана [9]. Соответственно, в результате получаются аналитические выражения и для оптимального управления, и для значения функционала качества. Технически традиционный, стандартный

подход к задаче обременен, пожалуй, единственной проблемой — поиском верного представления структуры функции Беллмана. Справиться с этой проблемой в большей степени удастся за счет результата, полученного при решении дискретного по времени аналога рассматриваемой постановки [8]. Конечные соотношения для оптимального решения, как и во всех подобных задачах, включая классическую линейно-квадратичную, содержат решения определенных дифференциальных уравнений (обыкновенных и в частных производных). Вывод этих уравнений и составляет содержание первой части данной работы. Во второй части будет обсуждаться их приближенное численное решение и компьютерные эксперименты.

Кратко обозначим основные положения, привлекаемые далее к решению задачи, следуя в основном обозначениям и терминологии [9], а именно: будем рассматривать задачу оптимального управления в стохастической динамической системе по полной информации, применяя метод динамического программирования. В качестве целевого функционала, определяющего качество управления $U_0^T = \{u_t, 0 \leq t \leq T\}$, выступает

$$J(U_0^T) = \mathbb{E} \left\{ \int_0^T L_t(x_t, u_t) dt + l(x_T) \right\}. \quad (1)$$

Здесь $\mathbb{E}\{\cdot\}$ — оператор математического ожидания; x_t — случайный процесс, описываемый стохастическим дифференциальным уравнением Ито

$$dx_t = m_t(x_t, u_t) dt + \sigma_t(x_t) dW_t, \quad x_0 = X, \quad (2)$$

где W_t — стандартный винеровский процесс подходящей размерности; X — случайный вектор.

U_0^T будем выбирать из класса допустимых неупреждающих (по отношению к W_t) управлений [9]. Соответственно, относительно функций сноса и диффузии m_t и σ_t в (2) будем предполагать выполненными какие-либо условия существования сильного решения для заданного допустимого управления. Например, для управления с обратной связью $u_t = u_t(x_t)$ будем считать, что $m_t(x, u_t(x))$ и $\sigma_t(x)$ удовлетворяют условию линейного роста и локальному условию Липшица по x равномерно по t (т. е. условиям Ито).

Для поиска оптимального управления, минимизирующего $J(U_0^T)$, рассматривается функция Беллмана

$$V_t(x) = \inf_{U_t^T} \mathbb{E} \left\{ \int_t^T L_t(x_t, u_t) dt + l(x_T) \middle| \mathcal{F}_t^x \right\}, \quad (3)$$

где \mathcal{F}_t^x — σ -алгебра, порожденная x_τ , $0 \leq \tau \leq t$, $E\{\cdot|\mathcal{F}\}$ — оператор условного математического ожидания относительно \mathcal{F} . Соответственно, в качестве достаточного условия оптимальности воспользуемся уравнением динамического программирования

$$\frac{\partial V_t(x)}{\partial t} + \frac{1}{2} \sum_{i,j=1}^n \sigma_{t_{ij}}^2 \frac{\partial^2 V_t(x)}{\partial x_i \partial x_j} + \min_u \left[\sum_{i=1}^n m_{t_i} \frac{\partial V_t(x)}{\partial x_i} + L_t(x, u) \right] = 0, \quad V_T(x) = l(x), \quad (4)$$

где m_{t_i} — i -й элемент вектор-функции $m_t(x, u)$; $\sigma_{t_{ij}}^2 = \sum_{k=1}^m \sigma_{t_{ik}} \sigma_{t_{kj}}$, $\sigma_{t_{ij}}$ — i -й по строке, j -й по столбцу элемент матричной функции $\sigma_t(x)$; n и m — размерности x_t и W_t соответственно.

Традиционно в рамках применения метода динамического программирования будем предполагать, что функции L_t , l , m_t и σ_t обеспечивают существование хотя бы одного решения уравнения (4), а следовательно, и оптимального управления u_t^* , $0 \leq t \leq T$, доставляющего минимум целевому функционалу (1). Задача оптимизации далее получается путем указания конкретных выражений для L_t , l , m_t и σ_t .

2 Постановка задачи управления выходом

Рассматриваемые далее случайные функции будут предполагаться скалярными. Такое упрощение позволит разгрузить выкладки и итоговые выражения от не самых существенных деталей.

Рассмотрим стохастическую дифференциальную систему, состояние которой представляет диффузионный процесс y_t , описываемый нелинейным стохастическим дифференциальным уравнением Ито

$$dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t, \quad y_0 = Y, \quad (5)$$

где v_t — стандартный (одномерный) винеровский процесс; Y — случайная величина с конечным вторым моментом; функции A_t и Σ_t удовлетворяют условиям Ито:

$$|A_t(y)| + |\Sigma_t(y)| \leq C(1 + |y|) \text{ для всех } 0 \leq t \leq T;$$

$$|A_t(y_1) - A_t(y_2)| + |\Sigma_t(y_1) - \Sigma_t(y_2)| \leq C|y_1 - y_2| \text{ для всех } 0 \leq t \leq T \text{ и } y_1, y_2 \in \mathbb{R}^1,$$

обеспечивающим существование единственного сильного (потраекторного) решения уравнения.

Будем считать, что y_t описывает состояние некоторой динамической системы. Соответственно, поведение этой системы опишем выходом, линейно связанным с состоянием:

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z. \quad (6)$$

Здесь w_t — не зависящий от v_t , Y и Z стандартный (одномерный) винеровский процесс; Z — случайная величина с конечным вторым моментом; u_t — допустимое неупреждающее управление, качество которого определяется целевым функционалом следующего вида:

$$J(U_0^T) = E \left\{ \int_0^T \left(S_t (s_t y_t - g_t z_t - h_t u_t)^2 + G_t z_t^2 + H_t u_t^2 \right) dt + S_T (s_T y_T - g_T z_T)^2 + G_T z_T^2 \right\}, \quad (7)$$

где S_t , G_t и H_t — неотрицательные функции $0 \leq t \leq T$. Такой критерий отражает физический смысл задачи распределения ресурсов согласно аналогичной (5)–(7) задаче для дискретного времени, рассмотренной в [7]. В частности, функционал (7) позволяет ставить задачи отслеживания выходом состояния системы, используя слагаемое $(y_t - z_t)^2$, или управлением — линейной комбинации состояния и выхода, слагаемое типа $(y_t + z_t - u_t)^2$. Поскольку задача формулируется в предположении наличия полной информации о состоянии y_t и выходе z_t (соответствующую σ -алгебру обозначим $\mathcal{F}_t^{y,z}$), то допустимое управление ищется в классе $\mathcal{F}_t^{y,z}$ -измеримых неупреждающих функций (и, как будет показано далее, оказывается управлением с обратной связью).

Функции a_t , b_t , c_t и σ_t будем предполагать ограниченными: $|a_t| + |b_t| + |c_t| + |\sigma_t| \leq C$ для всех $0 \leq t \leq T$, процесс управления — допустимым неупреждающим [9], обеспечивая, таким образом, существование сильного решения уравнения (6) для любого допустимого управления.

Задачу составляет поиск u_t^* — допустимого управления, доставляющего минимум квадратичному функционалу $J(U_0^T)$.

Поставленная задача очевидным образом формулируется в терминах введенных выше в (1)–(3) обозначений, а именно: требуется обозначить

$$x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}; \quad m_t(x_t, u_t) = \begin{pmatrix} A_t(y_t) \\ a_t y_t + b_t z_t + c_t u_t \end{pmatrix};$$

$$\sigma_t(x_t) = \begin{pmatrix} \Sigma_t(y_t) & 0 \\ 0 & \sigma_t \end{pmatrix}; \quad W_t = \begin{pmatrix} v_t \\ w_t \end{pmatrix}$$

для записи уравнения состояния типа (2) и

$$\begin{aligned} L_t(x, u) &= L_t(y, z, u) = \\ &= S_t (s_t y - g_t z - h_t u)^2 + G_t z^2 + H_t u^2; \\ l(x) &= l(y, z) = S_T (S_T y - g_T z)^2 + G_T z^2 \end{aligned}$$

для записи целевого функционала в виде (1).

Функция Беллмана (3) принимает вид $V_t(x) = V_t(y, z)$. Для записи соответствующего (4) уравнения Беллмана для $V_t(y, z)$ заметим, что

$$\left(\sigma_{t_{ij}}^2 \right)_{i,j=1,2} = \begin{pmatrix} \Sigma_t^2(y) & 0 \\ 0 & \sigma_t^2 \end{pmatrix}.$$

С учетом перечисленных обозначений уравнение динамического программирования (4) принимает вид:

$$\begin{aligned} &\frac{\partial V_t(y, z)}{\partial t} + \frac{1}{2} \left(\Sigma_t^2(y) \frac{\partial^2 V_t(y, z)}{\partial y^2} + \sigma_t^2 \frac{\partial^2 V_t(y, z)}{\partial z^2} \right) + \\ &+ \min_u \left[A_t(y) \frac{\partial V_t(y, z)}{\partial y} + (a_t y + b_t z + c_t u) \frac{\partial V_t(y, z)}{\partial z} + \right. \\ &\left. + S_t (s_t y - g_t z - h_t u)^2 + G_t z^2 + H_t u^2 \right] = 0, \\ &V_T(y, z) = S_T (S_T y - g_T z)^2 + G_T z^2. \quad (8) \end{aligned}$$

Это и есть то самое уравнение, которое требуется решить: существование решения данного уравнения суть достаточное условие оптимальности; оптимальное управление при этом — точка минимума соответствующего слагаемого.

3 Динамическое программирование и оптимальное управление

В рассматриваемой постановке линейность выхода и квадратичность критерия дают те же преимущества, что и в классической линейно-квадратичной задаче управления [1], а именно: позволяют сразу определить вид оптимального управления и фактические условия его существования. Действительно, сохраняя в (8) под знаком \min_u только члены, зависящие от u , получаем

$$\begin{aligned} &\frac{\partial V_t(y, z)}{\partial t} + \frac{1}{2} \left(\Sigma_t^2(y) \frac{\partial^2 V_t(y, z)}{\partial y^2} + \sigma_t^2 \frac{\partial^2 V_t(y, z)}{\partial z^2} \right) + \\ &+ A_t(y) \frac{\partial V_t(y, z)}{\partial y} + (a_t y + b_t z) \frac{\partial V_t(y, z)}{\partial z} + \\ &+ S_t (s_t y - g_t z)^2 + G_t z^2 + \end{aligned}$$

$$\begin{aligned} &+ \min_u \left[\left(c_t \frac{\partial V_t(y, z)}{\partial z} - 2S_t (s_t y - g_t z) h_t \right) u + \right. \\ &\left. + (S_t h_t^2 + H_t) u^2 \right] = 0, \end{aligned}$$

откуда в предположении $S_t h_t^2 + H_t > 0$ следует, что существует оптимальное управление, которое определяется равенством

$$u_t^* = u_t^*(y, z) = -\frac{1}{2} (S_t h_t^2 + H_t)^{-1} \left(c_t \frac{\partial V_t(y, z)}{\partial z} - 2S_t (s_t y - g_t z) h_t \right) \quad (9)$$

и доставляет минимум соответствующему слагаемому в уравнении Беллмана, равный $-(S_t h_t^2 + H_t)^{-1} (c_t \partial V_t(y, z) / \partial z - 2S_t (s_t y - g_t z) h_t)^2 / 4$.

Отметим, что, как и в классической линейно-квадратичной задаче, управление из класса допустимых неупреждающих получилось управлением с обратной связью.

Таким образом, функция Беллмана описывается следующим дифференциальным уравнением:

$$\begin{aligned} &\frac{\partial V_t(y, z)}{\partial t} + \frac{1}{2} \left(\Sigma_t^2(y) \frac{\partial^2 V_t(y, z)}{\partial y^2} + \sigma_t^2 \frac{\partial^2 V_t(y, z)}{\partial z^2} \right) + \\ &+ A_t(y) \frac{\partial V_t(y, z)}{\partial y} + (a_t y + b_t z) \frac{\partial V_t(y, z)}{\partial z} + \\ &+ S_t (s_t y - g_t z)^2 + G_t z^2 - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} \times \\ &\times \left(c_t \frac{\partial V_t(y, z)}{\partial z} - 2S_t (s_t y - g_t z) h_t \right)^2 = 0. \quad (10) \end{aligned}$$

Возводя в квадрат последнее слагаемое в (10), перепишем его в виде:

$$\begin{aligned} &\frac{\partial V_t(y, z)}{\partial t} + \frac{1}{2} \left(\Sigma_t^2(y) \frac{\partial^2 V_t(y, z)}{\partial y^2} + \sigma_t^2 \frac{\partial^2 V_t(y, z)}{\partial z^2} \right) + \\ &+ A_t(y) \frac{\partial V_t(y, z)}{\partial y} + (a_t y + b_t z + \\ &+ (S_t h_t^2 + H_t)^{-1} c_t S_t (s_t y - g_t z) h_t) \frac{\partial V_t(y, z)}{\partial z} + \\ &+ \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) (s_t y - g_t z)^2 + \\ &+ G_t z^2 - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \left(\frac{\partial V_t(y, z)}{\partial z} \right)^2 = 0. \quad (11) \end{aligned}$$

Рассматривая полученное уравнение, заметим, что его решение может быть представлено в виде:

$$V_t(y, z) = \alpha_t z^2 + \beta_t(y) z + \gamma_t(y), \quad (12)$$

т.е. будем искать решение при дополнительном предположении о квадратичности функции Беллмана по переменной z , и сведем, таким образом, поиск оптимального решения к уравнениям относительно функций α_t , $\beta_t(y)$ и $\gamma_t(y)$. Отметим сразу, что явный вид функции $\gamma_t(y)$ для реализации оптимального управления не требуется, однако далее будет предложен вариант вычисления и этой функции, что представляется небесполезным, поскольку позволит выполнять расчет минимума целевого функционала. Источником для предложения (12) является уже упоминавшаяся аналогичная задача для случая дискретного времени [7, 8]. В той задаче выражение для функции Беллмана получается формально без дополнительных усилий. При этом форма (12) обнаруживается как свойство оптимального решения. В рассматриваемом случае непрерывного времени (12) постулируется, а правильность постулата подтверждается далее результирующими уравнениями для α_t , $\beta_t(y)$ и $\gamma_t(y)$. Кроме того, данное предположение представляется вытекающим из линейной структуры задачи в отношении переменной z , в частности, тем фактом, что такой вид функции Беллмана обеспечивает линейность оптимального управления (9) по z .

Граничное условие при выбранном предположении (12) принимает вид:

$$\begin{aligned} V_T(y, z) &= S_T (s_T y - g_T z)^2 + G_T z^2 = \\ &= \alpha_T z^2 + \beta_T(y) z + \gamma_T(y), \end{aligned}$$

т.е.

$$\begin{aligned} \alpha_T &= S_T g_T^2 + G_T; \\ \beta_T(y) &= -2S_T s_T g_T y; \\ \gamma_T(y) &= S_T s_T^2 y^2. \end{aligned}$$

При этом само оптимальное управление, определенное выражением (9), оказывается управлением с обратной связью по y_t и z_t :

$$\begin{aligned} u_t^* &= u_t^*(y, z) = \\ &= -\frac{1}{2} (S_t h_t^2 + H_t)^{-1} (c_t (2\alpha_t z + \beta_t(y)) + \\ &\quad + 2S_t (s_t y - g_t z) h_t). \end{aligned} \quad (13)$$

Подставляем $V_t(y, z) = \alpha_t z^2 + \beta_t(y) z + \gamma_t(y)$ в (11):

$$\begin{aligned} &\frac{\partial \alpha_t}{\partial t} z^2 + \frac{\partial \beta_t(y)}{\partial t} z + \frac{\partial \gamma_t(y)}{\partial t} + \\ &+ \frac{1}{2} \left(\Sigma_t^2(y) \left(\frac{\partial^2 \beta_t(y)}{\partial y^2} z + \frac{\partial^2 \gamma_t(y)}{\partial y^2} \right) + 2\sigma_t^2 \alpha_t \right) + \\ &+ A_t(y) \left(\frac{\partial \beta_t(y)}{\partial y} z + \frac{\partial \gamma_t(y)}{\partial y} \right) + \\ &+ \left(a_t y + b_t z + (S_t h_t^2 + H_t)^{-1} c_t S_t (s_t y - g_t z) h_t \right) \times \end{aligned}$$

$$\begin{aligned} &\times (2\alpha_t z + \beta_t(y)) + \\ &+ \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) (s_t y - g_t z)^2 + \\ &+ G_t z^2 - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 (2\alpha_t z + \beta_t(y))^2 = 0. \end{aligned}$$

Далее выделяем слагаемые при z^2 , z и z^0

$$\begin{aligned} &\frac{\partial \alpha_t}{\partial t} z^2 + \frac{\partial \beta_t(y)}{\partial t} z + \frac{\partial \gamma_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t(y)}{\partial y^2} z + \\ &+ \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \gamma_t(y)}{\partial y^2} + \sigma_t^2 \alpha_t + A_t(y) \frac{\partial \beta_t(y)}{\partial y} z + \\ &+ A_t(y) \frac{\partial \gamma_t(y)}{\partial y} + \\ &+ 2\alpha_t \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) z^2 + \\ &+ \left(2\alpha_t \left(\alpha_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \right. \\ &+ \beta_t(y) \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) \left. \right) z + \\ &+ \beta_t(y) \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ &+ \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) g_t^2 z^2 - \\ &- 2 \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t y z + \\ &+ \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 + \\ &+ G_t z^2 - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t^2 z^2 - \\ &- (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t \beta_t(y) z - \\ &- \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2(y) = 0, \end{aligned}$$

группируем их и получаем следующие уравнения:

– для α_t :

$$\begin{aligned} &\frac{\partial \alpha_t}{\partial t} + 2\alpha_t \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) + \\ &+ \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) g_t^2 + G_t - \\ &- (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t^2 = 0, \quad \alpha_T = S_T g_T^2 + G_T; \end{aligned} \quad (14)$$

– для β_t :

$$\begin{aligned} &\frac{\partial \beta_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t(y)}{\partial y^2} + A_t(y) \frac{\partial \beta_t(y)}{\partial y} + \\ &+ 2\alpha_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ &+ \beta_t(y) \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) - \\ &- 2 \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t y - \\ &- (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t \beta_t(y) = 0, \\ &\beta_T(y) = -2S_T s_T g_T y; \end{aligned} \quad (15)$$

– для γ_t :

$$\begin{aligned} & \frac{\partial \gamma_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \gamma_t(y)}{\partial y^2} + \sigma_t^2 \alpha_t + A_t(y) \frac{\partial \gamma_t(y)}{\partial y} + \\ & + \beta_t(y) \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ & + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 - \\ & - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2(y) = 0, \\ & \gamma_T(y) = S_T s_T^2 y^2. \end{aligned} \quad (16)$$

Уравнение (14), легко заметить, является уравнением Риккати, которое в силу сформулированного выше условия имеет единственное неотрицательное решение для всех $0 \leq t \leq T$. Этот факт требует дополнительного комментария. Для получения уравнения (14) рассмотрим исходную задачу при дополнительных условиях $a_t = 0$ и $s_t = 0$ для всех $0 \leq t \leq T$. Нетрудно видеть, что эти условия рассматриваемую постановку сводят фактически к классической линейно-квадратичной задаче. Имеющаяся в рассматриваемой формулировке чуть более общую форму целевой функции (принципиального значения это обобщение, конечно, не имеет) сведем к классической еще одним предположением: $S_t = 0$ для всех $0 \leq t \leq T$. Теперь уравнение для α_t принимает хорошо известный вид:

$$\frac{\partial \alpha_t}{\partial t} + 2\alpha_t b_t + G_t - H_t^{-1} c_t^2 \alpha_t^2 = 0, \quad \alpha_T = G_T. \quad (17)$$

В таком случае, как известно [10], существует единственное оптимальное управление — линейное с обратной связью по выходу z_t , с коэффициентом усиления, описываемым уравнением Риккати (17). Именно этот результат дают уравнения (14)–(16) и описываемая ими функция Беллмана (12), так как из $a_t = 0$ и $s_t = 0$ немедленно следует, что $\beta_t(y) = 0$, откуда, в свою очередь, с учетом независимости решения от y_t следует, что $\gamma_t(y) = \gamma_t$, т. е. не зависит от y и задается уравнением:

$$\frac{\partial \gamma_t(y)}{\partial t} + \sigma_t^2 \alpha_t = 0, \quad \gamma_T = 0.$$

Оптимальное управление при этом

$$u_t^* = -H_t^{-1} c_t \alpha_t z_t,$$

т. е. все полностью совпадает с известным классическим решением.

С уравнениями (15) и (16) ситуация, естественно, обстоит сложнее. Это линейные уравнения второго порядка параболического типа, поскольку $\Sigma_t^2(y) > 0$. Фактически отсутствуют конструктивные условия, гарантирующие существование их

решений (требовать, чтобы все фигурирующие в уравнениях коэффициенты были представлены аналитическими функциями на всем пространстве значений, вряд ли целесообразно), поэтому далее будем предполагать, что данные уравнения имеют на рассматриваемом интервале $0 \leq t \leq T$ хотя бы одно ограниченное решение и именно эти условия будем рассматривать как достаточные условия существования оптимального решения рассматриваемой задачи.

Таким образом, доказана следующая теорема.

Теорема. Пусть для диффузионного процесса (5) выполнены условия Ито, для процесса (6) — ограничены коэффициенты, уравнения (14)–(16) имеют ограниченные решения для $0 \leq t \leq T$. Тогда минимум функционалу (7) доставляет оптимальное управление (13), где $y = y_t$; $z = z_t$.

4 Заключение

Рассмотренная задача оптимизации в целом близка и по модели, и по критерию к классической линейно-квадратичной постановке. Принципиальным отличием является нелинейная модель для описания состояния динамической системы, в которой отсутствует управляющее воздействие. Такую модель наряду с традиционной интерпретацией «состояние–выход» можно понимать как модель неконтролируемого сложного внешнего воздействия. Небольшое дополнительное отличие дает предложенная форма квадратичного критерия, позволяющая, в частности, ставить такие задачи, как отслеживание выходом или управлением состояния системы или ее выхода.

Поскольку обсуждать возможности точного решения уравнений, определяющих оптимальное управление, не имеет смысла, наиболее актуальной далее является задача их приближенного численного решения и анализа возможности практической реализации. Этому посвящена вторая часть данной работы, планируемая к выходу в ближайшее время.

Литература

1. Athans M. Editorial on the LQG problem // IEEE T. Automat. Contr., 1971. Vol. 16. No. 6. P. 528–552. doi: 10.1109/TAC.1971.1099845.
2. Wu Z. Forward-backward stochastic differential equations, linear quadratic stochastic optimal control and nonzero sum differential games // J. Syst. Sci. Complex., 2005. Vol. 18. No. 2. P. 179–192.

3. Chen B. S., Zhang W. Stochastic H2/H1 control with state-dependent noise // *IEEE T. Automat. Contr.*, 2004. Vol. 49. No. 1. P. 45–56. doi: 10.1109/TAC.2003.821400.
4. Bohacek S. A stochastic model of TCP and fair video transmission // *IEEE INFOCOM*, 2003. Vol. 2. P. 1134–1144. doi: 10.1109/INFCOM.2003.1208950.
5. Домбровский В. В., Обьедко Т. Ю. Управление с прогнозированием системами с марковскими скачками при ограничениях и применение к оптимизации инвестиционного портфеля // *Автомат. телемех.*, 2011. № 5. С. 96–112. doi: 10.1134/S0005117911050079.
6. Баландин Д. В., Коган М. М. Оптимальное линейно-квадратичное управление: от матричных уравнений к линейным матричным неравенствам // *Автомат. телемех.*, 2011. № 11. С. 60–69. doi: 10.1134/S0005117911110038.
7. Босов А. В. Обобщенная задача распределения ресурсов программной системы // *Информатика и её применения*, 2014. Т. 8. Вып. 2. С. 39–47. doi: 10.14357/19922264140204.
8. Босов А. В. Управление линейным выходом дискретной стохастической системы по квадратичному критерию // *Изв. РАН. Теория и системы управления*, 2016. № 3. С. 19–35. doi: 10.1134/S1064230716030060.
9. Флеминг У., Рашел Р. Оптимальное управление детерминированными и стохастическими системами / Пер. с англ. — М.: Мир, 1978. 316 с. (*Fleming W. H., Rishel R. W. Deterministic and stochastic optimal control.* — New York, NY, USA: Springer-Verlag, 1975. 222 p.)
10. Девис М. Х. А. Линейное оценивание и стохастическое управление / Пер. с англ. — М.: Наука, 1984. 206 с. (*Davis M. H. A. Linear estimation and stochastic control.* — London: Chapman and Hall, 1977. 224 p.)

Поступила в редакцию 30.03.18

STOCHASTIC DIFFERENTIAL SYSTEM OUTPUT CONTROL BY THE QUADRATIC CRITERION. I. DYNAMIC PROGRAMMING OPTIMAL 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 problem of optimal control for the Ito diffusion process and a controlled linear output is solved. The considered statement is close to the classical linear-quadratic Gaussian control (LQG control) problem. Differences consist in the fact that the state is described by the nonlinear differential Ito equation $dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t$ and does not depend on the control u_t , optimization subject is controlled linear output $dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t$. Additional generalizations are included in the quadratic quality criterion for the purpose of statement such problems as state tracking by output or a linear combination of state and output tracking by control. The method of dynamic programming is used for the solution. The assumption about Bellman function in the form $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$ allows one to find it. Three differential equations for the coefficients α_t , $\beta_t(y)$, and $\gamma_t(y)$ give the solution. These equations constitute the optimal solution of the problem under consideration.

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

DOI: 10.14357/19922264180314

Acknowledgments

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

References

1. Athans, M. 1971. Editorial on the LQG problem. *IEEE T. Automat. Contr.* 16(6):528–552. doi: 10.1109/TAC.1971.1099845.
2. Wu, Z. 2005. Forward-backward stochastic differential equations, linear quadratic stochastic optimal control and nonzero sum differential games. *J. Syst. Sci. Complex.* 18(2):179–192.
3. Chen, B. S. and W. Zhang. 2004. Stochastic H2/H1 control with state-dependent noise. *IEEE T. Automat. Contr.* 49(1):45–56. doi: 10.1109/TAC.2003.821400.
4. Bohacek, S. 2003. A stochastic model of TCP and fair video transmission. *IEEE INFOCOM*. 2:1134–1144. doi: 10.1109/INFCOM.2003.1208950.

5. Dombrovskii, V. V., and T. Yu. Ob"edko. 2011. Predictive control of systems with Markovian jumps under constraints and its application to the investment portfolio optimization. *Automat. Rem. Contr.* 72(5):989–1003.
6. Balandin, D. V., and M. M. Kogan. 2011. Optimal linear-quadratic control: From matrix equations to linear matrix inequalities. *Automat. Rem. Contr.* 72(11):2276–2284.
7. Bosov, A. V. 2014. Obobshchennaya zadacha raspredeleniya resursov programmnoy sistemy [The generalized problem of software system resources distribution]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):39–47. doi: 10.14357/19922264140204.
8. Bosov, A. V. 2016. Discrete stochastic system linear output control with respect to a quadratic criterion. *J. Comput. Syst. Sc. Int.* 55(3):349–364.
9. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p.
10. Davis, M. H. A. 1977. *Linear estimation and stochastic control*. London: Chapman and Hall. 224 p.

Received March 30, 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

МОДЕЛЬ ПЕРЕДВИЖЕНИЯ ПОЕЗДОВ И МАНЕВРОВЫХ ЛОКОМОТИВОВ НА ЖЕЛЕЗНОДОРОЖНОЙ СТАНЦИИ В ПРИЛОЖЕНИИ К ОЦЕНКЕ И АНАЛИЗУ ВЕРОЯТНОСТИ БОКОВОГО СТОЛКНОВЕНИЯ*

А. В. Босов¹, А. Н. Игнатов², А. В. Наумов³

Аннотация: Предложена математическая модель для решения задачи управления движением маневровых локомотивов на железнодорожной станции при заданном расписании движения пассажирских/грузовых поездов через станцию и фиксированном графике маневровых работ, под которыми понимается отцепка и прицепка вагонов, выпуск и расформирование поездов. Модель используется для постановки и решения задачи минимизации времени передвижения по станции маневрового локомотива для осуществления очередной маневровой работы с учетом занятости некоторых путей для движения вследствие наличия на них пассажирских/грузовых поездов, а также с учетом ограничений на время исполнения маневровых работ. Исходная постановка сводится к задаче смешанного целочисленного линейного программирования. Представленная модель использована для оценки вероятности бокового столкновения на станции с учетом возможных случайных задержек в движении пассажирских поездов. Приведены результаты численных экспериментов.

Ключевые слова: имитационная модель; расписание; интенсивность; смешанное целочисленное линейное программирование

DOI: 10.14357/19922264180315

1 Введение

Основной задачей управления рисками на железнодорожном транспорте является достижение и поддержание допустимого уровня риска различных неблагоприятных событий [1]. К таким событиям относятся, например, столкновения на железнодорожных переездах поездов с автотранспортом, сход вагонов при поездной работе, пожары на локомотивах, излом рельса под поездом и др.

Изучение риска таких событий привлекало внимание как российских [2–5], так и западных [6] исследователей. Работа [6] в этой связи выделяется тем, что в ней исследовались вопросы оптимального положения вагонов с опасными грузами в поезде. Отдельный пласт задач связан с оценкой и анализом рисков происшествий, происходящих на железнодорожных станциях.

На крупных станциях, где маневровые работы осуществляются при интенсивном движении поездов, в качестве возможных неблагоприятных событий рассматриваются столкновения между маневровыми составами и пассажирскими/грузовыми

поездами, взрез стрелки (случайный перевод стрелки колесами подвижного состава) или сход с рельсов маневрового состава. Для описания соответствующих рисков в работе [7] была предложена оценка вероятности хотя бы одного столкновения на станции за произвольный промежуток времени, а в [8] оценено число взрезов и сходов с рельсов. Вычисление этих величин основано на интенсивностях пересечения маневровыми составами стрелочных переводов, используемых для перевода подвижного состава с одного пути на другой.

Для уточнения приведенной в [7] оценки интенсивностей можно провести длительные натурные наблюдения за работой конкретной станции, которые могут оказаться весьма дорогостоящими. При этом результаты (оценки интенсивности) будут получены только для одной этой станции и не дадут каких-либо оценок для других.

Другим способом уточнения интенсивностей является построение имитационной модели передвижения поездов на железнодорожной станции, которую можно было бы использовать на различных станциях, задавая небольшой набор входных

* Работа выполнена при поддержке Российского научного фонда (проект № 16-11-00062).

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

²Московский авиационный институт (национальный исследовательский университет), alexei.ignatov1@gmail.com

³Московский авиационный институт (национальный исследовательский университет), naumovav@mail.ru

данных, таких как схема станции, график движения поездов и маневровых работ, под которыми понимаются отцепка и прицепка вагонов, выпуск и расформирование поездов. Набор этих работ будет далее предполагаться фиксированным, поскольку он может быть получен из суточного плана работы станции. Кроме того, важно отметить, что сами маневровые работы могут быть отсортированы по времени начала исполнения. С течением суток вследствие исполнения маневровых работ их общее число уменьшается. Для предотвращения нарушения суточного плана работы станции необходимо оптимальным образом прокладывать траекторию движения маневрового состава от момента исполнения последней работы к моменту начала новой работы.

Применение имитационных моделей для задач на железнодорожном транспорте является обычной практикой. Чаще всего моделируют назначение локомотивов для составов с целью уменьшения общего числа используемых локомотивов [9–12], передвижение пассажиров по станции с целью увеличения пропускной способности станции [13]. Схожими по смыслу являются Комплексная автоматизированная система ведения технологических процессов работы железнодорожных станций (АС ВТП) и Автоматизированная система разработки и мониторинга выполнения Единых технологических процессов работы железнодорожных путей необщего пользования и станций примыкания (АС ЕТП). В них заложены описания, модели всех технологических процессов, происходящих на станции. Разрабатываемая в данной работе модель интенсивностей пересечения стрелочных переводов маневровыми локомотивами строится с учетом компромисса между учетом всех технологических процессов, происходящих на станции, и наименьшим объемом входных данных с целью быстрой, но близкой к точной оценки искомого интенсивностей. Построенная система моделирования также позволит учитывать случайные задержки в прибытии и отправлении поездов на станции, позволяя проанализировать, что произойдет на станции в случае неисполнения расписания.

Далее в статье сформулированы принципы и описана структура имитационной модели передвижения поездов на железнодорожной станции. Станция представляется в виде неориентированного нагруженного графа, вершинами которого являются стрелочные переводы, а также места перехода кривых участков пути в прямые и места перехода прямых участков пути в кривые, а также точки входа/выхода со станции. Ребрами являются железнодорожные пути, связывающие вершины. Каждому ребру сопоставляется число, равное расстоянию от одной вершины до другой на плоскости, т. е. дли-

на пути. На станции предполагается возможным движение транзитных пассажирских поездов, пассажирских поездов местного формирования, грузовых поездов, а также маневровых составов. Пассажирские и грузовые поезда следуют по расписанию, которое в силу различных происшествий на железной дороге может исполняться не полностью, т. е. предусматривается возможность задержек поездов. В заключительной части статьи приводятся примеры расчета интенсивности пересечения стрелочных переводов маневровыми составами, выполненные на основе построенной имитационной модели.

2 Постановка задачи

Пусть имеется определяющий модель станции неориентированный граф $G = \langle V, E \rangle$, где V — множество вершин (стрелочных переводов, стыков между рельсами и точек входа и выхода со станции (границ станции)); E — множество ребер (железнодорожных путей), соединяющих данные вершины. Также задана функция $D : E \rightarrow \mathbb{R}_+$, характеризующая длину ребра. Пусть $|E| = m$. Пронумеровав ребра графа G от 1 до m , составим новый граф $G' = \langle V', E' \rangle$, множеством вершин V' которого являются номера ребер графа G , т. е. $V' = \{1, 2, \dots, m\}$. Множество ребер E' включает в себя ребра между вершинами из V' , если эти вершины являются смежными ребрами в графе G . На элементах множества V' введем функцию $D' : V' \rightarrow \mathbb{R}_+$, характеризующую «вес» вершин в графе G' , т. е. длину соответствующих ребер в графе G .

Предположим, что расписание движения пассажирских поездов и набор маневровых работ являются корректными в том смысле, что некоторым назначением траекторий движения маневровых локомотивов можно исполнить суточный план работы станции без его нарушения. Пусть максимальная скорость передвижения по станции маневрового локомотива равна v_{\max} ; номер ребра графа G , на котором заканчивается предыдущая маневровая работа, равен $j_0 \in V'$; номер ребра графа G , на котором начинается следующая маневровая работа, равен $j_T \in V'$, где T — время окончания передвижения по станции маневрового локомотива, назначенного для осуществления маневровой работы. С учетом данных параметров поставим задачу по отысканию маршрута передвижения по станции маневрового локомотива для выполнения очередной еще не исполненной маневровой работы в плане на сутки с учетом ряда физических ограничений, описываемых ниже, с целью минимизации времени T . Будем отсчитывать время от момента окончания последней маневровой работы. Пусть $u(t)$ — номер ребра, которое проходит маневровый локомотив

в момент времени t от окончания последней маневровой работы ($u(t) \in V'$).

Поскольку на станции осуществляется движение пассажирских поездов и других маневровых составов, часть ребер закрыта для проезда. В связи с этим введем функцию $F : V' \times \mathbb{R}_+^1 \rightarrow \{0, 1\}$ вида

$$F(j, t) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{ребро (графа } G) \text{ с номером } j \\ & \text{свободно в момент времени } t; \\ 1 & \text{иначе,} \end{cases}$$

которая характеризует занятость ребра для движения маневрового локомотива в момент времени t от начала выполнения маневровой работы.

Перебор всех возможных путей (называемых в дальнейшем маршрутами) в графе G' из вершины j_0 в вершину j_T невозможен в силу их бесконечного количества, связанного с наличием циклов в графе G . Поэтому выберем несколько таких маршрутов, общим числом равных L . Составим из этих последовательностей множество J . Для минимизации времени T необходимо найти минимальное время, за которое можно пройти каждый маршрут (каждую последовательность вершин) из множества J , а затем выбрать среди найденных времен минимальное.

Сформулируем задачу по минимизации времени прохождения маневрового состава через станцию по маршруту, задаваемому произвольной последовательностью из множества J . Для этого отметим, что произвольный элемент \bar{J}_l множества J имеет вид:

$$\bar{J}_l = \{j_{0,l}, j_{1,l}, \dots, j_{k,l}, \dots, j_{K_l,l}\},$$

где $j_{k,l} \in V'$, причем $j_{0,l} = j_0$ и $j_{K_l,l} = j_T$, а $k = \overline{1, K_l}, l = \overline{1, L}$.

Пусть $t_{k,l}$ — момент перехода с вершины с номером $j_{k-1,l}$ на вершину с номером $j_{k,l}$ графа G' , а $t_{0,l} = 0, t_{K_l+1,l} = T$. Тогда для каждого элемента \bar{J}_l множества J условие физической реализуемости можно записать в виде:

$$t_{k+1,l} - t_{k,l} \geq \frac{D'(j_{k,l})}{v_{\max}}, k = \overline{0, K_l},$$

т.е. исключить нереализуемую возможность проехать любое ребро графа G за бесконечно малое время.

Условие на движение только по свободным ребрам графа G записывается в виде:

$$\forall j_{k,l} \forall t \in [t_{k,l}, t_{k+1,l}] \quad F(j_{k,l}, t) = 0, k = \overline{0, K_l},$$

которое эквивалентно ограничениям:

$$\int_{t_{k,l}}^{t_{k+1,l}} F(j_{k,l}, t) dt = 0, k = \overline{0, K_l}.$$

Также имеет место ограничение вида:

$$t_{\min} \leq t_{K_l+1,l} \leq t_{\max},$$

которое гарантирует, что очередная маневровая работа не начнется позже момента времени t_{\max} и/или раньше t_{\min} . Величины t_{\min} и t_{\max} определяются исходя из графика маневровых работ и связаны с тем, что маневровый локомотив должен подъехать к месту осуществления работ в определенный промежуток времени $[t_{\min}, t_{\max}]$, так как в иное время очередная работа может быть пропущена или еще не начаться согласно графику работ.

Таким образом, задача по минимизации времени прохождения маневрового состава через станцию по маршруту, задаваемому произвольной последовательностью \bar{J}_l из множества J , имеет вид:

$$t_{K_l+1,i} \rightarrow \min_{t_{k,l}, k=\overline{1, K_l+1}} \quad (1)$$

при ограничениях

$$\left. \begin{aligned} t_{k+1,l} - t_{k,l} &\geq \frac{D'(j_{k,l})}{v_{\max}}, k = \overline{0, K_l}; \\ \int_{t_{k,l}}^{t_{k+1,l}} F(j_{k,l}, t) dt &= 0, k = \overline{0, K_l}; \\ t_{\min} &\leq t_{K_l+1,j} \leq t_{\max}, t_{0,l} = 0. \end{aligned} \right\} \quad (2)$$

3 Сведение задачи о выборе маршрута движения к задаче смешанного целочисленного линейного программирования

Заметим, что используемые выше функции

$$H(j_{k,l}, t) \stackrel{\text{def}}{=} \int_0^t F(j_{k,l}, y) dy, k = \overline{1, K_l},$$

являются кусочно-линейными, поэтому ограничения (2) в задаче (1) являются нелинейными, что делает поиск решения в задаче (1) при ограничениях (2) весьма затруднительным. Но указанное свойство функций $H(j_{k,l}, t), k = \overline{0, K_l}$, позволяет путем введения целочисленных переменных свести исходную задачу нелинейного программирования к задаче смешанного целочисленного линейного программирования. Для этого сформируем множество $T_{k,l}$, состоящее из левой и правой границ интервалов времени, когда ребро с номером $j_{k,l}$ оказывалось свободным для движения маневрового состава, $k = \overline{0, K_l}$.

С помощью множества $\Gamma_{k,l}$ можно определить «окна» (т. е. интервалы времени, в которые ребро с номером $j_{k,l}$ свободно). Упорядочив элементы множества $\Gamma_{k,l}$ по возрастанию, составим из них вектор $\tau_{k,l}$. Пусть $\dim \tau_{k,l} = 2I_{k,l}$, где $I_{k,l}$ — число «окон». Введем новые переменные $\delta_{k,l}^i$, равные единице, если движение по ребру с номером $j_{k,l}$ осуществляется в промежуток времени между $\tau_{k,l}^{2i-1}$ и $\tau_{k,l}^{2i}$, и равные нулю, если движение по ребру с номером $j_{k,l}$ в промежуток времени между $\tau_{k,l}^{2i-1}$ и $\tau_{k,l}^{2i}$ не осуществляется, $k = \overline{0, K_l}$, $i = \overline{1, I_{k,l}}$. С учетом новых переменных $\delta_{k,l}^i$ задача (1) при ограничениях (2) эквивалентным образом сводится к следующей задаче:

$$t_{K_l+1,l} \rightarrow \min_{t_{k,l}, t_{K_l+1,l}, \delta_{0,l}^i, \delta_{k,l}^i, k=\overline{1, K_l}, i=\overline{1, I_{k,l}}} \quad (3)$$

при ограничениях

$$\left. \begin{aligned} t_{k+1,l} - t_{k,l} &\geq \frac{D'(j_{k,l})}{v_{\max}}, \quad k = \overline{0, K_l}; \\ \sum_{i=1}^{I_{k,l}} \delta_{k,l}^i &= 1, \quad k = \overline{0, K_l}; \\ t_{k+1,l} &\leq \delta_{k,l}^i \tau_{k,l}^{2i} + (1 - \delta_{k,l}^i) t_{\max}, \\ &\quad k = \overline{0, K_l}, i = \overline{1, I_{k,l}}; \\ t_{k,l} &\geq \delta_{k,l}^i \tau_{k,l}^{2i-1}, \quad k = \overline{0, K_l}, i = \overline{1, I_{k,l}}; \\ t_{\min} &\leq t_{K_l+1,l} \leq t_{\max}, \quad \delta_{k,l}^i \in \{0, 1\}, \\ &\quad i = \overline{1, I_{k,l}}, k = \overline{0, K_l}, t_{0,l} = 0. \end{aligned} \right\} \quad (4)$$

Поскольку задача (1) при ограничениях (2) может не иметь решения, то введем новую величину:

$$T_l \stackrel{\text{def}}{=} \begin{cases} T_l^*, & \text{решение задачи (1)} \\ & \text{при ограничениях (2) существует;} \\ +\infty & \text{иначе,} \end{cases}$$

$l = \overline{1, L}$. Здесь T_l^* — оптимальное значение критерия в задаче (3)–(4). С использованием величин T_l^* заключаем, что можно оценить величину T^* сверху как $T^* \leq \min_{l=\overline{1, L}} T_l^*$, где T^* — минимальное время, за которое можно добраться из вершины j_0 в вершину j_T в графе G' , не нарушая и график маневровых работ, и график движения пассажирских поездов.

Таким образом, решение задачи, сформулированной выше, дает возможность оценить сверху время передвижения маневрового локомотива по станции до исполнения очередной маневровой работы.

После передвижения по станции к месту выполнения маневровой работы и ее исполнения маневровому локомотиву назначается любая маневровая

работа из перечня оставшихся маневровых работ, такая что она может быть выполнена вовремя указанным маневровым локомотивом.

4 Применение полученных результатов для оценки интенсивности пересечения стрелочного перевода маневровым составом

Наряду с максимальным числом пассажирских/грузовых поездов, пересекающих станцию, которое характеризует доход от функционирования станции, важной характеристикой является вероятность хотя бы одного бокового столкновения между пассажирскими и грузовыми поездами и маневровыми составами, которая характеризует степень безопасности движения. Для увеличения максимального числа поездов, пересекающих станцию, вначале необходимо рассчитать вероятность хотя бы одного бокового столкновения, например за сутки. Для этого необходимо рассчитать вероятность столкновения произвольного пассажирского/грузового поезда с маневровым составом на произвольной стрелке при пересечении пассажирским/грузовым поездом станции.

В [7] была предложена следующая оценка вероятности бокового столкновения, получаемая на основе предположения о том, что поток пересеченный маневровыми составами стрелок — пуассоновский:

$$P(A) = \left(\lambda_M (P_M + P_M P_{\Pi/\Gamma} + P_{\Pi/\Gamma}) \left(\frac{l_{\Pi/\Gamma}}{v_{\Pi/\Gamma}} + \frac{l_M}{v_M} \right) + \lambda_C P_{\Pi/\Gamma} \tau_C + \lambda_M P_M P_{\Pi/\Gamma} \tau_{\Pi/\Gamma} \right) \kappa_C, \quad (5)$$

$$\lambda_M = \frac{1}{d} \sum_{i=1}^Q \frac{N_i}{N}.$$

Здесь λ_M — интенсивность пересечения стрелки маневровыми составами в направлении, при котором возможно боковое столкновение между пассажирским/грузовым поездом и маневровым составом (рис. 1) [1/ч]; Q — число маневровых локомотивов на станции; N — число стрелок на станции; N_i — число стрелок, пересекаемых в час i -м маневровым составом; d — общее число направлений движения по стрелке маневрового состава (на рис. 1 $d = 4$); $l_{\Pi/\Gamma}$ — средняя длина пассажирского/грузового поезда [км]; l_M — средняя длина маневрового состава [км]; $v_{\Pi/\Gamma}$ — средняя скорость пассажирского/грузового поезда

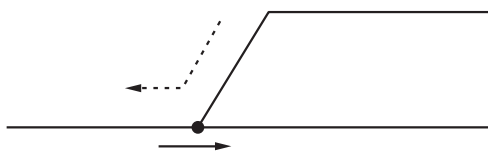


Рис. 1 Направление следования (отмечено пунктиром) маневрового состава через стрелочный перевод (отмечен кружком), при котором возможно столкновение с пассажирским составом

[км/ч]; v_m — средняя скорость маневрового состава [км/ч]; τ_c — среднее время нахождения маневрового состава на стрелочном переводе при условии остановки на нем [ч]; $\tau_{пс/гс}$ — среднее время стоянки пассажирского/грузового поезда на стрелочном переводе [ч]; $P_{пс/гс}$ — вероятность остановки пассажирского/грузового поезда на стрелочном переводе; $P_{п/г}$ — вероятность проезда на запрещающий сигнал светофора пассажирского/грузового поезда; P_m — вероятность проезда на запрещающий сигнал светофора маневрового состава; κ_c — коэффициент, характеризующий неизолированность стрелочного перевода (возможность бокового столкновения), принимающий значение 1, если стрелочный перевод неизолированный, и 0, если изолированный.

Вероятности $P_{пс/гс}$, $P_{п/г}$ и P_m задаются согласно [7, 14]. Как видно из формулы (5), интенсивность λ_m (при одинаковых d) получается одной и той же для стрелок и с интенсивным, и с неинтенсивным движением по ним маневровых составов. При этом сама величина λ_m существенно влияет на вероятность $P(A)$. Поэтому на основе составленного расписания движения маневрового локомотива по станции предложим иной способ вычисления интенсивностей λ_m .

Наличие предложенной имитационной модели позволяет использовать для оценки интенсивности пересечения стрелочных переводов маневровыми локомотивами классическую оценку в виде частоты.

Пусть w — число пересечений маневровыми локомотивами стрелочного перевода в направлении, при котором возможно боковое столкновение между пассажирским/грузовым поездом и маневровым составом, в течение суток. Тогда $\lambda_m = w/24$.

5 Результаты численного эксперимента

Рассмотрим некоторую станцию, схема части которой отображена на рис. 2.

Общее число стрелок на данной станции равно 149. Пусть данную станцию в некоторые сутки пересекают 84 пассажирских поезда (включая поездные локомотивы), общее число маневровых работ равно 76, а максимальная скорость v_m передвижения маневровых локомотивов по станции равна 10,8 км/ч. Очевидно, что чем больше маршрутов передвижения маневрового локомотива по станции, найденных по графу G' , к месту выполнения очередной маневровой работы исследуется, тем большая доля маневровых работ может быть выполнен, а время передвижения по станции маневрового локомотива уменьшится. Однако в этом случае также увеличится суммарное время, необходимое для нахождения решения задач (3) при ограничениях (4).

Проанализируем время счета и доля выполненных маневровых работ в зависимости от различного числа маршрутов передвижения маневрового локомотива по станции к месту выполнения очередной маневровой работы.

Расчеты проводились на персональном компьютере со следующими характеристиками: процессор Intel Core i7 2,8 ГГц, память 8 Гб 1 600 МГц DDR3.

Как следует из табл. 1 и 2, время счета ожидается растет при увеличении числа L , однако растет и доля выполненных маневровых работ. При этом для выполнения 100% маневровых работ достаточно рассмотреть небольшое число маршрутов

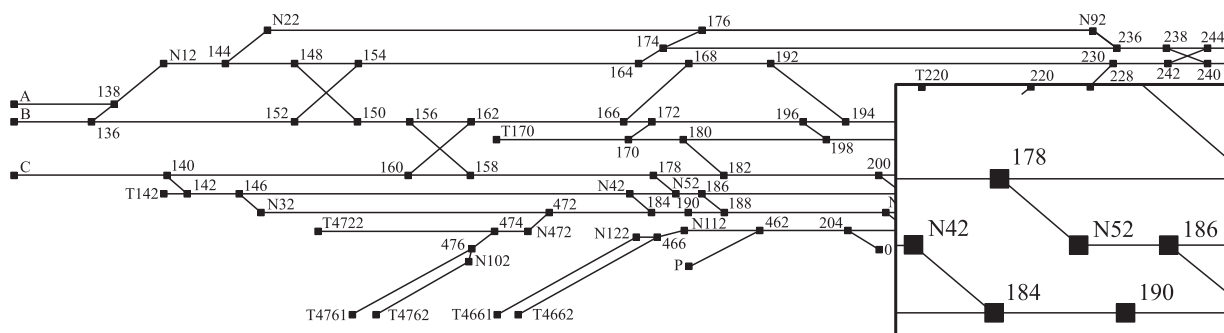


Рис. 2 Схема станции

Таблица 1 Время передвижения по станции для осуществления i -й маневровой работы [мин]/ время поиска решения в задачах (3) при ограничениях (4) [мин]

L	i							
	...	14	15	16	17	18	19	...
1	...	32/0,003	25/0,01	4/0,002	3/0,005	6/0,009	6/0,01	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
7	...	30/0,02	23/0,1	2/0,03	2/0,04	4/0,17	5/0,11	...

Таблица 2 Общее время счета и доля выполненных маневровых работ

L	Общее время счета, ч	Доля выполненных маневровых работ, %
1	0,26	73
3	2,71	83,8
5	2,95	97,3
7	3,03	100

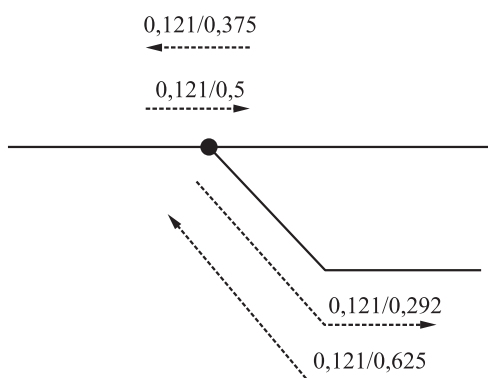


Рис. 3 Интенсивности пересечения стрелочного перевода № 178 по формуле (5) по данным из [7]/ по предлагаемой имитационной модели

пересечения маневровым локомотивом станции до места исполнения очередной маневровой работы.

Теперь отдельно рассмотрим стрелочный перевод № 178 и приведем число пересечений маневровыми локомотивами данного перевода в различных направлениях.

Как следует из рис. 3, интенсивности пересечения стрелочного перевода № 178, полученные по предлагаемой имитационной модели и по формуле (5), отличаются. Это связано с тем, что интенсивности, полученные по формуле (5), назначаются одинаковыми каждому стрелочному переводу, а реальная загруженность того или иного стрелочного перевода (удаленность или близость конкретного стрелочного перевода от места проведения большинства маневровых работ) в расчет не принимается.

6 Заключение

В работе рассмотрена задача по составлению расписания движения по станции маневровых локомотивов с учетом различных технических ограничений, суточного расписания движения поездов, а также набора маневровых работ.

Сформулирована задача нелинейного программирования, которая сведена к задаче смешанного целочисленного линейного программирования.

На основе построенной имитационной модели получена модифицированная оценка интенсивности пересечения стрелочных переводов в различных направлениях маневровыми локомотивами, которая в дальнейшем используется для уточнения оценки вероятности бокового столкновения между маневровым локомотивом и пассажирским/грузовым поездом на стрелочном переводе.

Литература

- ГОСТ 33433-2015. Безопасность функциональная. Управление рисками на железнодорожном транспорте. — М.: Стандартиформ, 2016. 34 с.
- Шубинский И. Б., Проневич О. Б., Данилова А. Д. Особенности оценки вероятности возникновения пожаров на тепловозах различных серий // Надежность, 2016. Т. 16. № 4. С. 24–29. doi: 10.21683/1729-2646-2016-16-4-24-29.
- Крутиков А. М. Оценка надежности рельсов Р65 по ресурсу: экспериментальные исследования. — М.: Финансы и статистика, 2016. 151 с.
- Замышляев А. М., Игнатов А. Н., Кибзун А. И., Новожилов Е. О. Функциональная зависимость между количеством вагонов в сходе из-за неисправностей вагонов или пути и факторами движения // Надежность, 2018. Т. 18. № 1. С. 53–60. doi: 10.21683/1729-2646-2018-18-1-53-60.
- Кибзун А. И., Игнатов А. Н. О задаче распределения инвестиций в установку средств, предотвращающих несанкционированный проезд автотранспортом железнодорожных переездов, для различных статистических критериев // Надежность, 2018. Т. 18. № 2. С. 31–37. doi: 10.21683/1729-2646-2018-18-2-31-37.

6. *Bagheri M., Saccomanno F., Chenouri S., Fu L.* Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist // *Accident Anal. Prev.*, 2011. Vol. 43. Iss. 4. P. 613–620. doi: 10.1016/j.aap.2010.09.008.
7. *Игнатов А. Н., Кибзун А. И., Платонов Е. Н.* Оценка вероятности столкновения железнодорожных составов на железнодорожных станциях на основе пуассоновской модели // *Автоматика и телемеханика*, 2016. № 11. С. 43–59. doi: 10.1134/S0005117916110035.
8. *Шубинский И. Б., Замышляев А. М., Игнатов А. Н., Кан Ю. С., Кибзун А. И., Платонов Е. Н.* Оценка рисков, связанных с проездом запрещающего сигнала светофора маневровым составом или пассажирским поездом // *Надежность*, 2016. Т. 16. № 3. С. 39–46. doi: 10.21683/1729-2646-2016-16-3-39-46.
9. *Иванов С. В., Кибзун А. И., Осокин А. В.* Оптимизационная стохастическая модель назначения локомотивов для перевозки грузовых составов // *Автоматика и телемеханика*, 2016. № 11. С. 80–95. doi: 10.1134/S0005117916110059.
10. *Лазарев А. А., Мусатова Е. Г., Тарасов И. А.* Решение задачи планирования двухстороннего движения на однопутном участке железной дороги с разъездом // *Автоматика и телемеханика*, 2016. № 11. С. 158–174. doi: 10.1134/S0005117916110047.
11. *Гайнанов Д. Н., Рассказова В. А.* Математическое моделирование в задаче оптимального назначения и перемещения локомотивов методами теории графов и комбинаторной оптимизации // *Труды МАИ*, 2017. № 92. 24 с.
12. *Буянов М. В., Иванов С. В., Кибзун А. И., Наумов А. В.* Развитие математической модели управления грузоперевозками на участке железнодорожной сети с учетом случайных факторов // *Информатика и её применения*, 2017. Т. 11. Вып. 4. С. 85–93. doi: 10.14357/19922264170411.
13. *Искаков Т. А.* Имитационное моделирование функционирования транспортного узла // *Интеллектуальные системы управления на железнодорожном транспорте. Компьютерное и математическое моделирование: Труды V научно-технич. конф. с международным участием.* — М.: НИИАС, 2016. С. 221–225.
14. *Шубинский И. Б.* Функциональная надежность информационных систем. Методы анализа. — Ульяновск: Печатный двор, 2012. 296 с.

Поступила в редакцию 01.02.18

MODEL OF TRANSPORTATION OF TRAINS AND SHUNTING LOCOMOTIVES AT A RAILWAY STATION FOR EVALUATION AND ANALYSIS OF SIDE-COLLISION PROBABILITY

A. V. Bosov¹, A. N. Ignatov², and A. V. Naumov²

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

²Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125993, Russian Federation

Abstract: A mathematical model for solution of the shunting locomotives traffic control problem is proposed for a fixed schedule of passenger/freight trains traffic across a station and a fixed time-table of shunting operations: set off and attaching of cars, output operation and deconsolidation of trains. The model is used for formulation and solution of the problem to minimize time of shunting locomotive transportation across the station to perform next scheduled operation with respect to busy condition of some tracks for transportation owing to presence of passenger/freight trains on them and with respect to restriction on shunting operation execution time. The original statement is reduced to mixed integer linear programming. The presented model was used for evaluation of side-collision probability at a railway station with respect to possible random drag in passenger trains traffic. The results of numerical experiments are presented.

Keywords: simulation model; schedule; intensity; mixed integer linear programming

DOI: 10.14357/19922264180315

Acknowledgments

This work was supported by the Russian Science Foundation (grant No. 16-11-00062).

References

1. GOST 33433-2015. 2016. *Bezopasnost' funktsional'naya. Upravlenie riskami na zheleznodorozhnom transporte* [Functional safety. Risk control in railroad transport]. Moscow: Standartinform. 34 p.
2. Shubinsky, I. B., O. B. Pronevich, and A. D. Danilova. 2016. Osobennosti otsenki veroyatnosti vozniknoveniya pozharov na teplovozhakh razlichnykh seriy [Special aspects of estimating the probability of fire occurrence on diesel locomotives of various types]. *Dependability* 16(4):24–29. doi: 10.21683/1729-2646-2016-16-4-24-29.
3. Krutikov, A. M. 2016. *Otsenka nadezhnosti rel'sov R65 po resursu: eksperimental'nye issledovaniya* [Evaluation of the dependability of rails R65 on the resource: Experimental researches]. Moscow: Finansy i statistika. 151 p.
4. Zamyshlyayev, A. M., A. N. Ignatov, A. I. Kibzun, and E. O. Novozhilov. 2018. Funktsional'naya zavisimost' mezhdru kolichestvom vagonov v skhode iz-za neispravnostey vagonov ili puti i faktorami dvizheniya [Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors]. *Dependability* 18(1):53–60. doi: 10.21683/1729-2646-2018-18-1-53-60.
5. Kibzun, A. I., and A. N. Ignatov. 2018. O zadache raspredeleniya investitsiy v ustanovku sredstv, predotvrashchayushchikh nesanksionirovanny proezd avtotransportom zheleznodorozhnykh perezhdov, dlya razlichnykh statisticheskikh kriteriev [On the task of allocating investment to facilities preventing unauthorized movement of road vehicles across level crossings for various statistical criteria]. *Dependability* 18(2):31–37. doi: 10.21683/1729-2646-2018-18-2-31-37.
6. Bagheri, M., F. Saccomanno, S. Chenouri, and L. Fu. 2011. Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. *Accident Anal. Prevent.* 43(3):613–620. doi: 10.1016/j.aap.2010.09.008.
7. Ignatov, A. N., A. I. Kibzun, and E. N. Platonov. 2016. Estimating collision probabilities for trains on railroad stations based on a Poisson model. *Automat. Rem. Contr.* 77(11):1914–1927.
8. Shubinsky, I. B., A. M. Zamyshlyayev, A. N. Ignatov, Yu. S. Kan, A. I. Kibzun, and E. N. Platonov. 2016. Otsenka riskov, svyazannykh s proездом zapreshchayushchego signala svetofora, manevrovym sostavom ili passazhirskim proездом [Estimation of risks related to stop signal passed by shunting loco or passenger train]. *Dependability* 16(3):39–46. doi: 10.21683/1729-2646-2016-16-3-39-46.
9. Ivanov, S. V., A. I. Kibzun, and A. V. Osokin. 2016. Stochastic optimization model of locomotive assignment to freight trains. *Automat. Rem. Contr.* 77(11):1944–1956.
10. Lazarev, A. A., E. G. Musatova, and I. A. Tarasov. 2016. Two-directional traffic scheduling problem solution for a single-track railway with siding. *Automat. Rem. Contr.* 77(12):2118–2131.
11. Gainanov, D. N., and V. A. Rasskazova. 2017. Matematicheskoe modelirovanie v zadache optimal'nogo naznacheniya i peremeshcheniya lokomotivov metodami teorii grafov i kombinatornoy optimizatsii [Mathematical modelling of locomotives' traffic problem by graph theory and combinatorial optimization methods]. *Trudy MAI* 92. 24 p.
12. Buyanov, M. V., S. V. Ivanov, A. I. Kibzun, and A. V. Naumov. 2017. Razvitie matematicheskoy modeli upravleniya gruzoperevozkami na uchastke zheleznodorozhnoy seti s uchedom sluchaynykh faktorov [Development of the mathematical model of cargo transportation control on a railway network segment taking into account random factor]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):85–93. doi: 10.14357/19922264170411.
13. Isakov, T. A. 2016. Imitatsionnoe modelirovanie funktsionirovaniya transportnogo uzla [Simulation modelling of functioning transport hub]. *Trudy V nauchno-tekhnich. konf. s mezhdunarodnym uchastiem "Intellectual'nye sistemy upravleniya na zheleznodorozhnom transporte. Komp'yuternoe i matematicheskoe modelirovanie"* [5th Science and Technological Conference (with International Participation) "Intellectual Control Systems in Railroad Transport. Computer and Mathematical Modeling" Proceedings]. Moscow. 221–225.
14. Shubinskiy, I. B. 2012. *Funktsional'naya nadezhnost' informatsionnykh sistem. Metody analiza* [International functional dependability of information systems. Methods of analysis]. Ul'yanovsk: Pechatnyy dvor. 296 p.

Received February 1, 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

Ignatov Alexey N. (b. 1991) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125993, Russian Federation; alexei.ignatov1@gmail.com

Naumov Andrey V. (b. 1966) — Doctor of Science in physics and mathematics, associate professor, Department “Probability Theory and Computational Modeling,” Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125993, Russian Federation; naumovav@mail.ru

ФИЛЬТРАЦИЯ СОСТОЯНИЙ МАРКОВСКИХ СКАЧКООБРАЗНЫХ ПРОЦЕССОВ ПО ДИСКРЕТИЗОВАННЫМ НАБЛЮДЕНИЯМ*

А. В. Борисов¹

Аннотация: Статья посвящена решению задачи оптимальной фильтрации состояний однородного марковского скачкообразного процесса (МСП). Наблюдения представляют собой приращения случайных процессов — интегральных преобразований состояний, зашумленные винеровскими процессами, интенсивность которых также зависит от оцениваемого состояния. Оптимальная оценка в моменты получения нового наблюдения вычисляется как функция предыдущей оценки и новых наблюдений, а между моментами наблюдений — простейшим прогнозом в силу системы уравнений Колмогорова. Рекуррентная формула пересчета ресурсозатратна, так как содержит интегралы — масштабно-сдвиговые смеси многомерных гауссиан, где в качестве смешивающих выступают распределения времени пребывания состояния в каждом из возможных значений. Предложены более простые аппроксимации, основанные на предположении об ограниченности числа скачков состояния за время между наблюдениями. Получены универсальные локальная и глобальная характеристики точности аппроксимаций, зависящие от параметров оцениваемого процесса, величины временного шага между наблюдениями и максимального числа учитываемых скачков.

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

DOI: 10.14357/19922264180316

1 Введение

Фильтр Вонэма [1] — один из редких удачных случаев, когда оценка оптимальной фильтрации состояния стохастической системы наблюдения выражается в виде решения некоторой замкнутой конечномерной системы стохастических дифференциальных уравнений.

Алгоритм данного фильтра позволяет вычислить оценку фильтрации состояния *марковского скачкообразного процесса* с конечным множеством состояний по наблюдениям в присутствии аддитивных винеровских шумов. Теоретически оптимальная оценка состояния — его условное распределение в текущий момент времени — обладает очевидными свойствами неотрицательности и нормировки. При численной реализации данного фильтра классическим методом Эйлера—Маруямы [2] данные свойства могут не сохраняться и процедура вычисления становится неустойчивой. В связи с этим обстоятельством разрабатывались другие алгоритмы численного решения уравнения фильтра Вонэма, обладающие требуемыми свойствами устойчивости (см. [3, 4] и библиографию в них). В части этих работ доказана лишь слабая сходимость предлагаемых аппроксимационных схем к оценке фильтра

Вонэма, в то время как какая-либо характеристика точности этих приближений отсутствует.

В [5] было представлено распространение фильтра Вонэма на случай наблюдений с мультипликативными шумами. При этом уравнение обобщенного фильтра содержит в правой части квадратичную характеристику шумов в наблюдениях. Данный процесс на практике никогда не наблюдается непосредственно, а является лишь некоторым нелинейным интегральным преобразованием наблюдений. Очевидно, что имеющиеся в настоящий момент времени алгоритмы приближенного вычисления оценки фильтрации Вонэма для данной системы не подходят.

Целью предлагаемой работы является использование результатов оптимальной фильтрации состояний систем с дискретным временем для аппроксимации решения аналогичной задачи для стохастических дифференциальных систем.

Статья организована следующим образом. Раздел 2 содержит формальную постановку задачи фильтрации состояний однородного МСП с конечным множеством состояний по наблюдениям, полученным путем временной дискретизации процессов с непрерывным временем — интегральных преобразований состояния системы в при-

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

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

сутствии мультипликативных винеровских шумов. В разд. 3 представлено решение поставленной задачи фильтрации: пересчет оценок состояний в момент получения новых дискретизованных наблюдений выполняется в соответствии с некоторыми рекуррентными интегральными соотношениями, в то время как между моментами наблюдений оценка корректируется в соответствии с прогнозом в силу системы уравнений Колмогорова. Вычислительная сложность упомянутых выше интегральных соотношений связана с тем, что в расчет принимается возможность того, что между моментами наблюдений оцениваемое состояние может совершить произвольное число скачков. В разд. 4 представлен более простой алгоритм приближенного вычисления оценки фильтрации, основанный на ограничении возможного числа учитываемых скачков МСП. Доказана теорема, определяющая как локальную (одношаговую), так и глобальную (многшаговую) характеристики точности предложенного приближения — ℓ_1 -нормы ошибки аппроксимации. Полученные характеристики являются универсальными, т. е. не асимптотическими по шагу дискретизации, и зависят от характеристик самого МСП, шага временной дискретизации и числа скачков состояния, учитываемых на шаге. Обсуждение результатов и заключительные комментарии представлены в разд. 5.

2 Постановка задачи фильтрации

На полном вероятностном пространстве с фильтрацией $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \geq 0})$ рассматривается система наблюдений

$$\left. \begin{aligned} X_t &= X_0 + \int_0^t \Lambda^\top X_s ds + \mu_t; \\ Y_k &= \int_{t_{k-1}}^{t_k} f X_s ds + \int_{t_{k-1}}^{t_k} \sum_{n=1}^N X_s^n g_n dW_s, \\ \{t_k\}_{k \geq 0} &: 0 = t_0 < t_1 < t_2 \dots, \end{aligned} \right\} \quad (1)$$

где

- $X_t \triangleq \text{col}(X_t^1, \dots, X_t^N) \in \mathbb{S}^N$ — ненаблюдаемое состояние системы, являющееся однородным МСП с конечным множеством состояний $\mathbb{S}^N \triangleq \{e_1, \dots, e_N\}$ (\mathbb{S}^N — множество единичных векторов евклидова пространства \mathbb{R}^N), матрицей интенсивностей переходов Λ и начальным распределением π ;
- $\mu_t \triangleq \text{col}(\mu_t^1, \dots, \mu_t^N) \in \mathbb{R}^N$ — \mathcal{F}_t -согласованный мартингал;

- $\{Y_k\}_{k \in \mathbb{N}} : Y_k \triangleq \text{col}(Y_k^1, \dots, Y_k^M) \in \mathbb{R}^M$ — последовательность дискретизованных наблюдений, доступных в известные неслучайные моменты времени $\{t_k\}_{k \in \mathbb{N}}$, в которых $W_t \triangleq \text{col}(W_t^1, \dots, W_t^M) \in \mathbb{R}^M$ является \mathcal{F}_t -согласованным стандартным винеровским процессом, определяющим шум в наблюдениях, f — $(M \times N)$ -мерная матрица плана наблюдений, а набор матриц $\{g_n\}_{n=1, \dots, N}$ характеризует интенсивности шумов в зависимости от текущего состояния X_t .

Введем также в рассмотрение неубывающие семейства σ -алгебр $\mathcal{O}_k \triangleq \sigma\{Y_\ell : 1 \leq \ell \leq k\}$ и $\mathcal{O}_t \triangleq \mathcal{O}_{k(t)}$, где $k(t) \triangleq \sum_{j \in \mathbb{N}} \mathbf{I}(t - t_j)$; $\mathcal{O}_0 \triangleq \{\emptyset, \Omega\}$.

Задача оптимальной фильтрации состояния X по наблюдениям Y заключается в нахождении условного математического ожидания (УМО)

$$\hat{X}_t \triangleq \mathbb{E}\{X_t | \mathcal{O}_t\}.$$

Относительно системы (1) сделаны следующие предположения:

- (а) $\mathcal{F}_t \equiv \mathcal{F}_t^X \vee \mathcal{F}_t^W$ для любого $t \geq 0$;
- (б) шум в наблюдениях равномерно невырожденный, т. е. $g_n g_n^\top \geq \alpha I > 0$ для всех $n = 1, \dots, N$ и некоторого $\alpha > 0$.

3 Уравнения оптимального фильтра

Для получения уравнений оптимального фильтра воспользуемся подходом, применяемым для решения аналогичной задачи в стохастических системах наблюдения с дискретным временем [6]. Воспользуемся методом математической индукции.

При $r = 0$

$$\hat{X}_{t_0} = \mathbb{E}\{X_0 | \mathcal{O}_0\} = \mathbb{E}\{X_0\} = \pi. \quad (2)$$

Пусть для некоторого $r \geq 0$ известна оценка оптимальной фильтрации $\hat{X}_{t_r} = \mathbb{E}X_{t_r} | \mathcal{O}_r$. Определим оценку оптимальной фильтрации \hat{X}_t для $t \in (t_r, t_{r+1}]$.

Для произвольного момента $t \in (t_r, t_{r+1})$ в силу мартингального разложения МСП X_t и свойств УМО верна следующая цепочка равенств:

$$\begin{aligned} \hat{X}_t &= \mathbb{E}\{X_t | \mathcal{O}_r\} = \\ &= \mathbb{E}\left\{ \mathcal{P}^\top(t_r, t) X_{t_r} + \int_{t_r}^t \mathcal{P}^\top(t_r, s) dM_s | \mathcal{O}_r \right\} = \end{aligned}$$

$$\begin{aligned}
 &= \mathcal{P}^\top(t_r, t) \widehat{X}_{t_r} + \mathbb{E} \left\{ \mathbb{E} \left\{ \int_{t_r}^t \mathcal{P}^\top(t_r, s) dM_s | \mathcal{F}_{t_r} \right\} | \mathcal{O}_r \right\} = \\
 &= \mathcal{P}^\top(t_r, t) \widehat{X}_{t_r}, \quad (3)
 \end{aligned}
 \qquad
 \begin{aligned}
 &= \sum_{\ell=1}^N e_\ell \int_{\mathcal{A}} \left[\sum_{k=1}^N \int_{\mathcal{D}_{r+1}} \mathcal{N} \left(y, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \times \right. \\
 &\qquad \left. \times \rho_{r+1}^{k,\ell}(du) \widehat{X}_{t_r}^k \right] dy,
 \end{aligned}$$

где $\mathcal{P}(s, t)$ ($s \leq t$) — матрица переходной вероятности МСП на промежутке $[s, t]$, являющаяся решением системы дифференциальных уравнений Колмогорова

$$\mathcal{P}'_t(s, t) = \mathcal{P}(s, t)\Lambda, \quad t > s, \quad \mathcal{P}(s, s) = I.$$

В случае однородного МСП $\mathcal{P}(s, t) = e^{(t-s)\Lambda}$.

Далее необходимо определить совместное распределение $(X_{t_{r+1}}, Y_{r+1})$ относительно \mathcal{O}_r . Из модели наблюдений следует, что распределение Y_{r+1} относительно σ -алгебры $\mathcal{F}_{t_{r+1}}^X \vee \mathcal{O}_r$ — гауссовское с параметрами

$$\begin{aligned}
 \mathbb{E} \{ Y_{r+1} | \mathcal{F}_{t_{r+1}}^X \} &= f\tau_{r+1}; \\
 \text{cov} \left(Y_{r+1}, Y_{r+1} | \mathcal{F}_{t_{r+1}}^X \right) &= \sum_{n=1}^N \tau_{r+1}^n g_n g_n^\top,
 \end{aligned}$$

где $\tau_{r+1} = \tau_{r+1}(X(\omega)) = \text{col}(\tau_{r+1}^1, \dots, \tau_{r+1}^N) \triangleq \int_{t_r}^{t_{r+1}} X_s ds$ — случайный вектор, n -я компонента которого равна времени пребывания процесса X в состоянии e_n на интервале времени $[t_r, t_{r+1}]$. Обозначим через $\mathcal{D}_{r+1} \triangleq \{u = \text{col}(u^1, \dots, u^N) : u_m \geq 0, \sum_{m=1}^M u_m = t_{r+1} - t_r\}$ ($M-1$)-мерный симплекс в пространстве \mathbb{R}^M , являющийся носителем распределения вектора τ_{r+1} . Пусть $\rho_{r+1}^{k,\ell}(du)$ — распределение вектора $\tau_{r+1} X_{t_{r+1}}^\ell$ при условии $X_{t_r} = e_k$, т. е. для любого $\mathcal{A} \in \mathcal{B}(\mathbb{R}^M)$ верно тождество:

$$\begin{aligned}
 \mathbf{P} \{ \omega : X_{t_{r+1}}(\omega) = e_\ell, \\
 \tau_{r+1}(X(\omega)) \in \mathcal{A} | X_{t_r} = e_k \} &\equiv \rho_{r+1}^{k,\ell}(\mathcal{A}).
 \end{aligned}$$

Обозначим через

$$\begin{aligned}
 \mathcal{N}(y, m, K) &\triangleq (2\pi)^{-M/2} \det^{-1/2} K \times \\
 &\times \exp \left\{ -\frac{1}{2} (y - m)^\top K^{-1} (y - m) \right\}
 \end{aligned}$$

M -мерную плотность гауссовского распределения с математическим ожиданием m и ковариационной матрицей K .

Из марковского свойства $\{X_{t_r}, Y_r\}_{r \geq 0}$ относительно \mathcal{F}_{t_r} [7] и теоремы Фубини следует, что для любого множества $\mathcal{A} \in \mathcal{B}(\mathbb{R}^M)$ верна следующая цепочка равенств:

$$\begin{aligned}
 \mathbb{E} \{ X_{t_{r+1}} \mathbf{I}_{\mathcal{A}}(Y_{r+1}) | \mathcal{O}_r \} &= \\
 = \mathbb{E} \left\{ \mathbb{E} \left\{ X_{t_{r+1}} \mathbf{I}_{\mathcal{A}}(Y_{r+1}) | \mathcal{F}_{t_{r+1}}^X \vee \mathcal{O}_r \right\} | \mathcal{O}_r \right\} &=
 \end{aligned}$$

из чего следует, что интегранд в квадратных скобках в последнем выражении определяет искомое совместное распределение $(X_{t_{r+1}}, Y_{r+1})$ относительно \mathcal{O}_r . Оценка $\widehat{X}_{t_{r+1}}$ покомпонентно определяется [6] с помощью обобщенного варианта формулы Байеса:

$$\begin{aligned}
 \widehat{X}_{t_{r+1}}^j &= \\
 &= \frac{\int_{\mathcal{D}_{r+1}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \sum_{k=1}^N \widehat{X}_{t_r}^k \rho_{r+1}^{k,j}(du)}{\int_{\mathcal{D}_{r+1}} \mathcal{N} \left(Y_{r+1}, fv, \sum_{q=1}^N v^q g_q g_q^\top \right) \sum_{i,\ell=1}^N \widehat{X}_{t_r}^i \rho_{r+1}^{i,\ell}(dv)}, \\
 &\qquad j = \overline{1, N}. \quad (4)
 \end{aligned}$$

Таким образом, доказана следующая

Лемма 1. Если для системы наблюдения (1) верны условия (а) и (б), то оценка \widehat{X}_t оптимальной фильтрации определяется формулой (2) при $t = 0$, рекуррентным соотношением (4) — в моменты t_{r+1} получения наблюдений Y_{r+1} и формулой (3) — в промежутках времени между моментами получения наблюдений.

Несмотря на компактную запись (4), их прямая численная реализация ресурсозатратна. Во-первых, в (4) требуется вычислять распределения масштабно-сдвиговых смесей многомерных нормальных распределений, что является трудоемкой процедурой. Во-вторых, распределения $\rho_{r+1}^{k,j}$ времени пребывания представляют собой сумму бесконечного ряда, слагаемые которого вычисляются с помощью некоторой рекуррентной процедуры [8]. В-третьих, распределения $\rho_{r+1}^{k,j}$ не являются абсолютно непрерывными относительно меры Лебега.

Следующий раздел посвящен численной аппроксимации (4) и исследованию ее точностных характеристик.

4 Приближенное вычисление оценки фильтрации

Без ограничения общности будем считать, что сетка $\{t_r\}_{r \geq 0}$ является равномерной с шагом Δ , т. е. $t_r = r\Delta$ и $\mathcal{D}_r \equiv \mathcal{D}$. Обозначим через N_{r+1} об-

шее число скачков процесса X_t , имевших место на промежутке $(t_r, t_{r+1}]$. Тогда из формулы полной вероятности следует, что (4) представима в виде:

$$\begin{aligned} \widehat{X}_{t_{r+1}}^j &= \left(\int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \times \right. \\ &\quad \times \left. \sum_{h=0}^{\infty} \sum_{k=1}^N \widehat{X}_{t_r, \rho_{r+1}}^{k,j,h}(du) \right) / \\ &\quad \left(\int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fv, \sum_{q=1}^N v^q g_q g_q^\top \right) \times \right. \\ &\quad \times \left. \sum_{m=0}^{\infty} \sum_{i,\ell=1}^N \widehat{X}_{t_r, \rho_{r+1}}^{i,\ell,m}(dv) \right), \quad j = \overline{1, N}, \quad (5) \end{aligned}$$

где $\rho_{r+1}^{k,j,h}(du)$ — распределение вектора $\tau_{r+1} X_{t_{r+1}}^j \mathbf{I}_{\{h\}}(N_{r+1})$ при условии $X_{t_r} = e_k$, т. е. для любого $\mathcal{A} \in \mathcal{B}(\mathbb{R}^M)$ верно тождество

$$\begin{aligned} \mathbf{P} \{ \omega : X_{t_{r+1}}(\omega) = e_j, N_{r+1} = h, \\ \tau_{r+1}(X(\omega)) \in \mathcal{A} \mid X_{t_r} = e_k \} \equiv \rho_{r+1}^{k,j,h}(\mathcal{A}). \end{aligned}$$

В качестве аппроксимации оценок можно использовать $\overline{X}_{t_{r+1}}^n \triangleq \text{col}(\overline{X}_{t_{r+1}}^{n,1}, \dots, \overline{X}_{t_{r+1}}^{n,N})$, полученные из (5) путем урезания сумм ряда в числителе и знаменателе:

$$\begin{aligned} \overline{X}_{t_{r+1}}^{n,j} &= \left(\int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \times \right. \\ &\quad \times \left. \sum_{h=0}^n \sum_{k=1}^N \overline{X}_{t_r, \rho_{r+1}}^{k,j,h}(du) \right) / \\ &\quad \left(\int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fv, \sum_{q=1}^N v^q g_q g_q^\top \right) \times \right. \\ &\quad \times \left. \sum_{m=0}^n \sum_{i,\ell=1}^N \overline{X}_{t_r, \rho_{r+1}}^{i,\ell,m}(dv) \right), \quad j = \overline{1, N}. \quad (6) \end{aligned}$$

Ниже по формуле полной вероятности получены интегралы из (6) для $h = 0, 1, 2$:

$$\begin{aligned} \int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \rho_{r+1}^{k,j,0}(du) &= \\ &= \delta_{kj} \mathcal{N} \left(Y_{r+1}, \Delta f^j, \Delta g_j g_j^\top \right) e^{\lambda_{jj} \Delta}; \\ \int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \rho_{r+1}^{k,j,1}(du) &= \end{aligned}$$

$$\begin{aligned} &= (1 - \delta_{kj}) \lambda_{kj} e^{\lambda_{jj} \Delta} \int_0^\Delta e^{(\lambda_{kk} - \lambda_{jj})u^k} \mathcal{N} \left(Y_{r+1}, u^k f^k + \right. \\ &\quad \left. + (\Delta - u^k) f^j, u^k g_k g_k^\top + (\Delta - u^k) g_j g_j^\top \right) du^k; \quad (7) \end{aligned}$$

$$\begin{aligned} \int_{\mathcal{D}} \mathcal{N} \left(Y_{r+1}, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) du &= \\ &= \sum_{\substack{\ell: \ell \neq k, \\ \ell \neq j}} \lambda_{k\ell} \lambda_{\ell j} e^{\lambda_{jj} \Delta} \times \\ &\quad \times \int_0^\Delta \int_0^{\Delta - u^k} e^{(\lambda_{kk} - \lambda_{\ell\ell})u^k + (\lambda_{\ell\ell} - \lambda_{jj})u^\ell} \times \\ &\quad \times \mathcal{N} \left(Y_{r+1}, u^k f^k + u^\ell f^\ell + (\Delta - u^k - u^\ell) f^j, \right. \\ &\quad \left. u^k g_k g_k^\top + u^\ell g_\ell g_\ell^\top + (\Delta - u^k - u^\ell) g_j g_j^\top \right) du^\ell du^k, \quad (8) \end{aligned}$$

где δ_{ij} — символ Кронекера. Интегралы для $h > 2$ также могут быть получены в явном виде, однако их сложность резко возрастает.

Так как система (1) является автономной, то в качестве локальной характеристики близости $\{\overline{X}_{t_r}\}$ к $\{\widehat{X}_{t_r}\}$ может быть выбрана величина

$$\begin{aligned} \overline{\sigma}(\pi) &\triangleq \mathbf{E} \left\{ \left\| \widehat{X}_{t_1}(\pi, Y_1) - \overline{X}_{t_1}(\pi, Y_1) \right\| \right\} = \\ &= \sum_{j=1}^N \mathbf{E} \left\{ \left| \widehat{X}_{t_1}^j(\pi, Y_1) - \overline{X}_{t_1}^{n,j}(\pi, Y_1) \right| \right\}. \end{aligned}$$

При этом начальное распределение $\pi \in \mathcal{D}_1 \triangleq \{ \text{col}(\pi^1, \dots, \pi^N) : \pi^j > 0, \sum_{j=1}^N \pi^j = 1 \}$ является начальным условием применения одного шага рекурсии (4) или (6) для вычисления оценки \widehat{X}_{t_1} или \overline{X}_{t_1} соответственно. Фактически, характеристика $\overline{\sigma}(\pi)$ определяет, насколько сильно рекурсивные схемы (4) и (6) разойдутся за один шаг, стартуя из общей точки π .

Рекуррентные схемы (4) и (6), примененные r раз, позволяют вычислить оценки \widehat{X}_{t_r} и \overline{X}_{t_r} в точке t_r . В качестве характеристики точности глобальной аппроксимации в этом случае естественно рассмотреть величину

$$\overline{\Sigma}_{t_r}(\pi) \triangleq \mathbf{E} \left\{ \left\| \widehat{X}_{t_r} - \overline{X}_{t_r} \right\| \right\} = \sum_{j=1}^N \mathbf{E} \left\{ \left| \widehat{X}_{t_r}^j - \overline{X}_{t_r}^{n,j} \right| \right\}.$$

Следующее утверждение определяет оценки локальной и глобальной точности схемы аппроксимации (6).

Теорема 1. *Выполняются неравенства*

$$\sup_{\pi \in \mathcal{D}_1} \overline{\sigma}(\pi) \leq 2 \frac{(\overline{\lambda} \Delta)^{n+1}}{(n+1)!}; \quad (9)$$

$$\sup_{\pi \in \mathcal{D}_1} \bar{\Sigma}_{t_r}(\pi) \leq 2r \frac{(\bar{\lambda}\Delta)^{n+1}}{(n+1)!} + r(r-1) \left(\frac{(\bar{\lambda}\Delta)^{n+1}}{(n+1)!} \right)^2 \left(1 - \frac{(\bar{\lambda}\Delta)^{n+1}}{(n+1)!} \right)^{r-2}, \quad (10)$$

где $\bar{\lambda} \triangleq \max_{1 \leq j \leq N} |\lambda_{jj}|$.

Доказательство теоремы 1 приведено в приложении.

Данное утверждение представляет полезные оценки точности. Во-первых, они являются равномерными по начальному распределению $\pi \in \mathcal{D}_1$. Во-вторых, оценки носят универсальный, а не асимптотический характер. Это существенно в практических задачах оценивания по дискретизованным наблюдениям с физическими или алгоритмическими ограничениями на шаг по времени. Например, в случае наблюдаемого процесса восстановления в силу центральной предельной теоремы для процессов восстановления [9] его приращения можно рассматривать как гауссовские случайные величины. Однако данная аппроксимация обладает удовлетворительной точностью только в случае, когда шаг дискретизации по времени достаточно большой. В-третьих, неравенство (10) позволяет получить порядок аппроксимации при $\Delta \rightarrow 0$. Зафиксируем момент времени $t = T$ и рассмотрим характеристику $\sup_{\pi \in \mathcal{D}_1} \bar{\Sigma}_T(\pi)$ при $r = T/\Delta$ и $\Delta \rightarrow 0$. Как только Δ становится настолько мало, что $\max((\bar{\lambda}\Delta)^{n+1}/(n+1)!, \Delta(T\bar{\lambda}^{n+1}/(n+1)!)) < 1$, из (10) следует неравенство $\sup_{\pi \in \mathcal{D}_1} \bar{\Sigma}_T(\pi) \leq (3\bar{\lambda}^{n+1}/(n+1)!T\Delta)^n$. Это значит, что с ростом времени T ошибка аппроксимации копится пропорционально T и при этом порядок точности по Δ равен n .

5 Заключение

В работе решена задача оценивания состояния однородного МСП по дискретизованным наблюдениям. Получено аналитическое решение и его численные аппроксимации. Локальные и глобальные показатели точности этих приближений в статье также представлены. Примечательно, что частный случай аппроксимаций (6) при $n = 0$ и $\Lambda = 0$ был ранее представлен в [10, 11] для решения задачи байесовской классификации случайного вектора по непрерывным наблюдениям с мультипликативными шумами. Алгоритм оптимальной фильтрации и его субоптимальные версии могут рассматриваться в качестве основы численной реализации обобщения фильтра Вонэма для систем с мультипликативными шумами в наблюдениях. Однако для их непосредственного использования необходимо решить следующие проблемы. Во-первых,

в (7) и (8) присутствуют многомерные интегралы. Следует выяснить, какую результирующую погрешность будут вносить ошибки их вычисления. Во-вторых, представляется интересным определить характеристики точности оптимальной фильтрации по дискретизованным наблюдениям по отношению к оптимальной фильтрации по непрерывным наблюдениям: каков порядок точности по шагу временной дискретизации Δ ? Для случая вычисления классического фильтра Вонэма с помощью алгоритма Эйлера–Маруямы подобный результат известен: порядок глобальной ошибки равен $1/2$. Перечисленные задачи являются предметом дальнейших исследований.

Приложение

Доказательство теоремы 1. Введем следующие обозначения для случайных величин и матриц, составленных из них:

$$\begin{aligned} \xi^{ji}(\ell) &\triangleq \sum_{h=0}^n \int_{\mathcal{D}} \mathcal{N} \left(Y_\ell, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \rho_1^{j,i,h}(du); \\ \theta^{ji}(\ell) &\triangleq \sum_{h=n+1}^{\infty} \int_{\mathcal{D}} \mathcal{N} \left(Y_\ell, fu, \sum_{p=1}^N u^p g_p g_p^\top \right) \rho_1^{j,i,h}(du); \\ \xi(\ell) &\triangleq \|\xi^{ji}(\ell)\|_{j,i=1,\overline{N}}, \quad \Xi(r) \triangleq \xi(r)\xi(r-1)\cdots\xi(1); \\ \theta(\ell) &\triangleq \|\theta^{ji}(\ell)\|_{j,i=1,\overline{N}}, \quad \Theta(r) \triangleq \theta(r)\theta(r-1)\cdots\theta(1). \end{aligned}$$

Рекуррентные формулы (4) и (6) можно записать в явной форме

$$\begin{aligned} \hat{X}_{t_r} &= (\mathbf{1}(\Xi(r) + \Theta(r))\pi)^{-1}(\Xi(r) + \Theta(r))\pi; \\ \bar{X}_{t_r} &= (\mathbf{1}\Xi(r)\pi)^{-1}\Xi(r)\pi, \end{aligned}$$

где $\mathbf{1} \triangleq (1, \dots, 1)$ — вектор-строка подходящей размерности, составленная из единиц.

Верна следующая цепочка неравенств:

$$\begin{aligned} \bar{\Sigma}_{t_r}(\pi) &= \mathbb{E} \left\{ \left\| \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r))\pi} (\Xi(r) + \Theta(r))\pi - \frac{1}{\mathbf{1}\Xi(r)\pi} \Xi(r)\pi \right\|_1 \right\} = \\ &= \mathbb{E} \left\{ \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r))\pi \mathbf{1}\Xi(r)\pi} \|\mathbf{1}\Xi(r)\pi\Theta(r)\pi - \mathbf{1}\Theta(r)\pi\Xi(r)\pi\|_1 \right\} \leq \\ &\leq \mathbb{E} \left\{ \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r))\pi \mathbf{1}\Xi(r)\pi} (\mathbf{1}\Xi(r)\pi\|\Theta(r)\pi\|_1 + \mathbf{1}\Theta(r)\pi\|\Xi(r)\pi\|_1) \right\} = \\ &= 2 \mathbb{E} \left\{ \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r))\pi} \mathbf{1}\Theta(r)\pi \right\}. \quad (11) \end{aligned}$$

Рассмотрим случайные события $a_\ell \triangleq \{\omega \in \Omega : N_\ell(\omega) \leq n\}$, $\ell = \overline{1, r}$, и $A_r \triangleq \{\omega \in \Omega : \max_{1 \leq \ell \leq r} N_\ell(\omega) \leq n\} = \prod_{\ell=1}^r a_\ell$ и оценку $\tilde{X}_{t_r}(\pi, Y_1, \dots, Y_r) \triangleq$

$\triangleq \mathbf{E} \{X_{t_r}(\omega) \mathbf{I}_{A_r}(\omega) | \mathcal{O}_r\}$. Используя введенные выше обозначения и абстрактный вариант формулы Байеса, получаем, что

$$\begin{aligned} \tilde{X}_{t_r} &= \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r)) \pi} \Xi(r) \pi; \\ \hat{X}_{t_r} - \tilde{X}_{t_r} &= \mathbf{E} \{X_{t_r}(\omega) \mathbf{I}_{A_r}(\omega) | \mathcal{O}_r\} = \\ &= \frac{1}{\mathbf{1}(\Xi(r) + \Theta(r)) \pi} \Theta(r) \pi. \end{aligned} \quad (12)$$

Из (11) и (12) для $r = 1$ следует, что

$$\begin{aligned} \bar{\sigma}(\pi) &\leq 2 \mathbf{E} \{ \|\mathbf{E} \{X_{t_1}(\omega) \mathbf{I}_{\bar{a}_1}(\omega) | \mathcal{O}_1\}\|_1 \} = \\ &= 2 \mathbf{E} \left\{ \sum_{n=1}^N \mathbf{E} \{X_{t_1}^n(\omega) \mathbf{I}_{\bar{a}_1}(\omega) | \mathcal{O}_1\} \right\} = \\ &= 2 \mathbf{E} \{ \mathbf{E} \{ \mathbf{I}_{\bar{a}_1}(\omega) | \mathcal{O}_1 \} \} = 2 \mathbf{P} \{ \bar{a}_1(\omega) \}. \end{aligned} \quad (13)$$

Процесс N_t^X общего числа скачков состояния X_t является считающим, и его квадратическая характеристика равна

$$\langle N^X, N^X \rangle_t = - \int_0^t \sum_{n=1}^N \lambda_{nn} X_s^n ds,$$

поэтому искомая вероятность ограничена сверху:

$$\mathbf{P} \{ \bar{a}_1(\omega) \} \leq e^{-\bar{\lambda} \Delta} \sum_{k=n+1}^{\infty} \frac{(\bar{\lambda} \Delta)^k}{k!} < \frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!}.$$

Из последнего неравенства и (13) следует, что для любого начального распределения π выполняется неравенство $\bar{\sigma}(\pi) < 2((\bar{\lambda} \Delta)^{n+1}/(n+1)!)$, т.е. локальная оценка (9) верна.

С помощью марковского свойства пары (X_t, N_t^X) и последнего неравенства можно оценить сверху вероятность $\mathbf{P} \{ \bar{A}_r(\omega) \}$:

$$\begin{aligned} \mathbf{P} \{ \bar{A}_r(\omega) \} &\leq 1 - \left(1 - \frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} \right)^r \leq r \frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} + \\ &+ \left| \sum_{k=2}^r C_r^k \left(-\frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} \right)^k \right| \leq r \frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} + \\ &+ \frac{r(r-1)}{2} \left(\frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} \right)^2 \left(1 - \frac{(\bar{\lambda} \Delta)^{n+1}}{(n+1)!} \right)^{r-2}, \end{aligned}$$

из чего следует истинность глобальной оценки (10). Теорема 1 доказана.

Литература

1. Wonham W. Some applications of stochastic differential equations to optimal nonlinear filtering // SIAM J. Control, 1965. Vol. 2. P. 347–369.
2. Kloeden P., Platen E. Numerical solution of stochastic differential equations. — Berlin: Springer, 1992. 636 p.
3. Yin G., Zhang Q., Liu Y. Discrete-time approximation of Wonham filters // J. Control Theory Applications, 2004. Iss. 2. P. 1–10.
4. Platen E., Rendek R. Quasi-exact approximation of hidden Markov chain filters // Communicat. Stoch. Analys., 2010. Vol. 4. Iss. 1. P. 129–142.
5. Борисов А. Фильтрация Вонэма по наблюдениям с мультипликативными шумами // Автоматика и телемеханика, 2018. № 1. С. 52–65.
6. Бертсекас Д., Шрив С. Стохастическое оптимальное управление. Случай дискретного времени / Пер. с англ. — М.: Наука, 1985. 280 с. (Bertsekas D. P., Shreve S. E. Stochastic optimal control: The discrete-time case. — Orlando, FL, USA: Academic Press Inc., 1978. 323 p.)
7. Жакод Ж., Ширяев А. Предельные теоремы для случайных процессов, I. / Пер. с англ. — М.: Физматлит, 1995. 544 с. (Jacod J., Shiryaev A. Limit theorems for stochastic processes. — Berlin: Springer, 2003. 664 p.)
8. Sericola B. Occupation times in Markov processes // Commun. Stat. Stochastic Models, 2000. Vol. 16. Iss. 5. P. 479–510.
9. Боровков А. Асимптотические методы в теории массового обслуживания. — М.: Физматлит, 1995. 384 с.
10. Борисов А. Классификация по непрерывным наблюдениям с мультипликативными шумами. I. Формулы байесовской оценки // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 11–19. doi: 10.14357/19922264170102.
11. Борисов А. Классификация по непрерывным наблюдениям с мультипликативными шумами. II. Алгоритм численной реализации оценки // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 33–41. doi: 10.14357/19922264170204.

Поступила в редакцию 10.07.18

FILTERING OF MARKOV JUMP PROCESSES BY DISCRETIZED OBSERVATIONS

A. V. Borisov

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

Abstract: The article is devoted to a solution of the optimal filtering problem of a homogenous Markov jump process state. The available observations represent time increments of the integral transformations of the Markov

state corrupted by Wiener processes. The noise intensity is also state-dependent. At the instant of the consecutive observation obtaining, the optimal estimate is calculated recursively as a function of previous estimate and the new observation, meanwhile between observations the filtering estimate is a simple forecast by virtue of the Kolmogorov differential system. The recursion is rather expensive because of need to calculate the integrals, which are the location-scale mixtures of Gaussians. The mixing distributions represent the occupation of the state in each of possible values during the mid-observation intervals. The paper contains numerically cheaper approximations, based on the restriction of the state transitions number between the observations. Both the local and global characteristics of approximation accuracy are obtained as functions of the dynamics parameters, mid-observation interval length, and upper bound of transitions number.

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

DOI: 10.14357/19922264180316

Acknowledgments

The work was supported in part by the Russian Foundation for Basic Research (Project No. 16-07-00677).

References

1. Wonham, W. 1965. Some applications of stochastic differential equations to optimal nonlinear filtering. *SIAM J. Control* 2:347–369.
2. Kloeden, P., and E. Platen. 1992. *Numerical solution of stochastic differential equations*. Berlin: Springer. 636 p.
3. Yin, G., Q. Zhang, and Y. Liu. 2004. Discrete-time approximation of Wonham filters. *J. Control Theory Applications* 2:1–10.
4. Platen, E., and R. Rendek. 2010. Quasi-exact approximation of hidden Markov chain filters. *Communicat. Stoch. Analys.* 4(1):129–142.
5. Borisov, A. 2018. Wonham filtering by observations with multiplicative noises. *Automat. Rem. Contr.* 79(1):39–50. doi: 10.1134/S0005117918010046.
6. Bertsekas, D., and S. Shreve. 1996. *Stochastic optimal control: The discrete-time case*. Nashua, NH: Athena Scientific. 330 p.
7. Jacod, J., and A. Shiryaev. 2003. *Limit theorems for stochastic processes*. Berlin: Springer. 664 p.
8. Sericola, B. 2000. Occupation times in Markov processes. *Commun. Stat.* 16(5):479–510.
9. Borovkov, A. 1984. *Asymptotic methods in queueing theory*. Hoboken, NJ: Wiley-Blackwell. 304 p.
10. Borisov, A. 2017. Klassifikatsiya po nepreryvny nablyudemyam s mul'tiplikativnymi shumami. I. Formuly bayesovskoy otsenki [Classification by continuous-time observations in multiplicative noise. I. Formulae for Bayesian estimate]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):11–19. doi: 10.14357/19922264170102.
11. Borisov, A. 2017. Klassifikatsiya po nepreryvny nablyudemyam s mul'tiplikativnymi summami. II. Formuly bayesovskoy otsenki [Classification by continuous-time observations in multiplicative noise. II. Numerical algorithm]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):33–41. doi: 10.14357/19922264170204.

Received July 10, 2018

Contributor

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

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

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

Горбунова Анастасия Владимировна (р. 1986) — кандидат физико-математических наук, ассистент 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

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забжайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследователь-

ского центра «Информатика и управление» Российской академии наук

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Игнатов Алексей Николаевич (р. 1991) — кандидат физико-математических наук, старший преподаватель Московского авиационного института (национального исследовательского университета)

Инькова Ольга Юрьевна (р. 1965) — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Козеренко Елена Борисовна (р. 1959) — кандидат филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кружков Михаил Григорьевич (р. 1975) — старший научный сотрудник 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

Кузнецов Константин Игоревич (р. 1968) — ведущий инженер 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

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 Андрей Викторович (р. 1966) — доктор физико-математических наук, доцент Московского авиационного института (национального исследовательского университета)

Наумов Валерий Арсентьевич (р. 1950) — кандидат физико-математических наук, научный руководитель Исследовательского института инноваций, г. Хельсинки, Финляндия

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

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Российского университета дружбы народов

Романов Дмитрий Александрович (р. 1967) — кандидат технических наук, доцент Национального

исследовательского университета «Высшая школа экономики»

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

Смирнов Дмитрий Владимирович (р. 1984) — бизнес-партнер по информационным технологиям департамента безопасности ПАО «Сбербанк России»

Сопин Эдуард Сергеевич (р. 1986) — кандидат физико-математических наук, доцент Российского университета дружбы народов; старший 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

Ступников Сергей Александрович (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шанин Иван Андреевич (р. 1991) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, доцент 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

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 Mezhunarodnogo 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 slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author’s name in the “Subject” field;
- an article and additional materials have to be attached using the “attach” function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)

<http://www.ipiran.ru/english/journal.asp>