

Информатика и её применения

Том 8 Выпуск 3 Год 2014

СОДЕРЖАНИЕ

Об условиях сходимости распределений экстремальных порядковых статистик к распределению Вейбулла В. Ю. Королев, И. А. Соколов	3
Аналитическое моделирование нормальных процессов в стохастических системах со сложными нелинейностями И. Н. Сеницын, В. И. Сеницын	12
Об оценках скорости сходимости и устойчивости для некоторых моделей массового обслуживания А. И. Зейфман, А. В. Коротышева, К. М. Киселева, В. Ю. Королев, С. Я. Шоргин	19
Стационарные вероятности состояний в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом Л. А. Мейханаджян, Т. А. Милованова, А. В. Печинкин, Р. В. Разумчик	28
Анализ системы обслуживания с входящим потоком авторегрессионного типа Н. Д. Леонтьев, В. Г. Ушаков	39
Модели для сравнительного анализа методов классификации в некоторых распределенных системах распознавания образов Я. М. Агаларов	45
Применение алгоритмов оптимальной фильтрации для решения задачи мониторинга доступности удаленного сервера А. В. Борисов	53
Сравнительный анализ процедур регрессионного анализа М. П. Кривенко	70
Методы математической статистики как инструмент двухпараметрического анализа магнитно-резонансного изображения Т. В. Яковлева, Н. С. Кульберг	79
Применение полутоновых представлений при анализе изменений цветных изображений О. П. Архипов, З. П. Зыкова	90
Текстурирование воксельных моделей на основе цветовой информации об опорных точках О. П. Архипов, Ю. А. Маньяков	100

Информатика и её применения

Том 8 Выпуск 3 Год 2014

СОДЕРЖАНИЕ

Проектирование самосинхронных схем: структурные методы в иерархическом анализе Л. П. Плеханов	105
Индикаторы тематических взаимосвязей науки и технологий: от текста к числам В. А. Минин, И. М. Зацман, В. А. Хавансков, С. К. Шубников	114
Аналитические аспекты оценки эффективности в технологии поддержки деятельности организационной системы А. А. Зацаринный, А. П. Шабанов	126
Об авторах	134

Технический редактор *Л. Кокушкина* Художественный редактор *М. Седакова*
Сдано в набор 01.07.14. Подписано в печать 12.09.14. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 16,5. Уч.-изд. л. 14,5. Тираж 100 экз.
Заказ № 1631

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43
Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов
Москва 121099, Шубинский пер., д. 6.

ОБ УСЛОВИЯХ СХОДИМОСТИ РАСПРЕДЕЛЕНИЙ ЭКСТРЕМАЛЬНЫХ ПОРЯДКОВЫХ СТАТИСТИК К РАСПРЕДЕЛЕНИЮ ВЕЙБУЛЛА*

В. Ю. Королев¹, И. А. Соколов²

Аннотация: Получены факторизационные представления для случайных величин, имеющих распределение Вейбулла, через случайные величины с устойчивым распределением. Эти результаты использованы для описания условий сходимости распределений линейно преобразованных минимальных порядковых статистик в выборках случайного объема к распределению Вейбулла. Приведенные результаты расширяют традиционные представления об условиях сходимости распределений экстремальных порядковых статистик к распределению Вейбулла и дают дополнительное теоретическое обоснование высокой адекватности распределения Вейбулла при анализе данных типа времени жизни, в частности в теории надежности.

Ключевые слова: распределение Вейбулла; показательное распределение; распределение Рэлея; строго устойчивое распределение; выборка случайного объема

DOI: 10.14375/19922264140301

1 Введение. Распределение Вейбулла

В теории вероятностей и математической статистике распределением Вейбулла принято называть специальное абсолютно непрерывное распределение, сосредоточенное на неотрицательной полуоси, хвост которого убывает экспоненциально-степенным образом. Оно названо в честь шведского ученого Валодди Вейбулла (Waloddi Weibull, 1887–1979), который в 1939 г. предложил использовать это распределение при анализе прочности материалов [1, 2] и детально описал и исследовал его в 1951 г. [3], продемонстрировав широкие возможности этого распределения при описании многих статистических закономерностей.

Пусть $\gamma > 0$. Распределением Вейбулла с параметром формы γ называется распределение случайной величины W_γ :

$$P(W_\gamma < x) = [1 - e^{-x^\gamma}] \mathbf{1}(x \geq 0), \quad x \in \mathbb{R}. \quad (1)$$

Здесь и далее $\mathbf{1}(C)$ обозначает индикатор множества C .

Однако Вейбулл не был первым, кто предложил это распределение, названное впоследствии его именем. Впервые это распределение было описано в 1927 г. в работе Мориса Фреше (Maurice Fréchet) [4], посвященной изучению предельного

поведения крайних членов вариационного ряда. Иногда распределение Вейбулла называют распределением Розина–Раммлера в честь Пауля Розина (Paul Rosin) и Эриха Раммлера (Erich Rammeler), немецких ученых, впервые применивших это распределение для описания статистических закономерностей в размерах частиц в 1933 г. [5]. Однако и это название не в полной мере соответствует исторической истине. В своей работе [6] Дитрих Штойян прямо пишет, что это распределение было открыто Розином, Раммлером, Шперлингом (Karl Sperling) [5, 7] и Беннеттом (John Godolphin Bennett) [8] в контексте моделирования размеров частиц.

Хорошо известно, что распределение Вейбулла обладает свойством замкнутости относительно операции взятия минимума независимых случайных величин. Как было показано в работах [4, 9], благодаря этому свойству оно является одним из возможных предельных распределений для экстремальных порядковых статистик.

Б. В. Гнеденко нашел необходимые и достаточные условия сходимости распределений линейно нормированных экстремальных порядковых статистик к распределению Вейбулла [10]. Поэтому это распределение иногда также называют (особенно в русскоязычной литературе) распределением Вейбулла–Гнеденко [11].

* Работа частично поддержана Российским фондом фундаментальных исследований (проекты 12-07-00115а, 12-07-00109а, 14-07-00041а).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, victoryukorolev@yandex.ru

²Институт проблем информатики Российской академии наук, ISokolov@ipiran.ru

В уже упоминавшейся работе [6] Д. Штойян замечает, что для этого распределения больше бы подошло нейтрально-техническое название *экспоненциально-степенное распределение*. Однако последний термин традиционно используется для другого абсолютно непрерывного распределения со сходным поведением хвостов [12, 13], у которого, в отличие от распределения (1), вид экспоненциально-степенной функции имеет *плотность* $l_\gamma(x) = ((\gamma/2)e^{-|x|^\gamma})/\Gamma(1/\gamma)$, $x \in \mathbb{R}$, где $\gamma > 0$, тогда как у распределения (1) экспоненциально-степенной функцией является *функция распределения*, т. е. интеграл плотности. В данной работе для распределения (1) будет использоваться традиционный термин *распределение Вейбулла*.

Распределение Вейбулла широко используется в медицине как модель распределения вероятностей выживания (survival analysis) [14], в страховании жизни как модель распределения вероятностей дожития, в рисковом страховании как модель распределения размера страховых требований [15], в экономике и финансовой математике как модель распределения доходностей ценных бумаг [16–18] и доходов фирм и отдельных лиц [19, 20], в теории надежности как модель распределения времени безотказной работы [21, 22], в промышленной технологии как модель распределения времени этапов производства (например, изготовления или поставки каких-либо деталей) или как модель распределения времени между появлениями новых технологий (так называемая модель Шарифа–Ислама [23]), в угольной промышленности для описания распределения размеров частиц угля при дроблении (здесь распределение Вейбулла используется под именем распределения Розина–Рамлера [5]), в радиотехнике и радиолокации, в метеорологии, гидрологии и многих других областях (см., например, [11, 21, 22, 24, 25]).

Огромная роль, которую играет распределение Вейбулла в теории надежности, и его популярность в инженерных приложениях не в последнюю очередь обусловлены его простотой. Еще одним обстоятельством, объясняющим популярность распределения Вейбулла среди инженеров, по-видимому, является то, что основополагающая статья [3] была опубликована в инженерном журнале *The ASME Journal of Applied Mechanics — Transactions of the American Society of Mechanical Engineers* после того, как она была отвергнута редколлегией статистического журнала *The Journal of the American Statistical Association* как не представляющая интереса (по другим данным, статья Вейбулла была изначально отвергнута одним из известных английских научных журналов [26]). Классическая статистика того времени базировалась на стереотипе нор-

мальности распределения анализируемых данных, и не-нормальность данных воспринималась как нетипичный бесполезный курьез.

В прикладной теории вероятностей принято считать, что та или иная модель может быть в достаточной мере обоснованной (адекватной) только тогда, когда она является *асимптотической аппроксимацией*, т. е. когда существует довольно простая предельная схема (например, схема максимума или схема суммирования) и соответствующая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения. Наличие такой формальной асимптотической схемы может дать дополнительную информацию о реальных механизмах, формирующих те или иные наблюдаемые статистические закономерности.

В классических работах [4, 9, 10] было показано, что распределение Вейбулла является одним из возможных предельных распределений в схеме минимума независимых случайных величин (также см., например, [27, 28]). В данной статье будет показано, что на самом деле теоретических предпосылок высокой адекватности распределения Вейбулла при анализе данных типа времени жизни намного больше, поскольку оно также может выступать в качестве предельного и для линейно преобразованных минимальных порядковых статистик в выборках случайного объема при нетривиальных асимптотически невырожденных распределениях случайного объема выборки. Приводимые результаты расширяют традиционные представления об условиях сходимости распределений экстремальных порядковых статистик к распределению Вейбулла.

Необходимо также отметить, что в работе [29] была рассмотрена версия закона больших чисел для случайных сумм и показано, что распределение Вейбулла с произвольным параметром формы может быть предельным для геометрических случайных сумм независимых неодинаково распределенных случайных величин. В работе [30] указанный результат был применен к исследованию рисков, связанных с наводнениями в Санкт-Петербурге.

2 Факторизационные представления для случайных величин с распределением Вейбулла через случайные величины с устойчивым распределением

В этом разделе будет показано, что распределение Вейбулла с произвольным параметром фор-

мы $\gamma > 0$ может быть представлено в виде «масштабной» смеси распределения Вейбулла с любым параметром формы $\delta > \gamma$, в которой в качестве смешивающего распределения выступает односторонний строго устойчивый закон. По своей сути приводимые далее результаты касаются специальных представлений вероятностных распределений. Однако без какого бы то ни было ограничения общности и для большей наглядности и компактности результаты будут формулироваться в терминах соответствующих случайных величин в предположении, что все случайные величины, появляющиеся в дальнейшем изложении, заданы на одном вероятностном пространстве $(\Omega, \mathfrak{A}, P)$.

Очевидно, что W_1 — это случайная величина со стандартной показательной функцией распределения: $P(W_1 < x) = [1 - e^{-x}] \mathbf{1}(x \geq 0)$. При $\gamma = 2$ распределение Вейбулла называется распределением Рэлея $P(W_2 < x) = [1 - e^{-x^2}] \mathbf{1}(x \geq 0)$ в честь Джона Уильяма Стрэтта лорда Рэлея (John William Strutt, Lord Rayleigh), который ввел это распределение в 1880 г. в связи с задачей сложения гармонических колебаний со случайными фазами [31].

Пусть $\Phi(x)$ — стандартная нормальная функция распределения:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad x \in \mathbb{R}.$$

Функцию распределения и плотность строго устойчивого распределения с характеристическим показателем α и параметром θ , задаваемого характеристической функцией

$$f_{\alpha,\theta}(t) = \exp \left\{ -|t|^\alpha \exp \left\{ -\frac{1}{2} i\pi\theta\alpha \operatorname{sign} t \right\} \right\},$$

$$t \in \mathbb{R}, \quad 0 < \alpha \leq 2, \quad |\theta| \leq \min \left\{ 1, \frac{2}{\alpha} - 1 \right\}, \quad (2)$$

будем обозначать соответственно $G_{\alpha,\theta}(x)$ и $g_{\alpha,\theta}(x)$ (см., например, [32]). Любую случайную величину с функцией распределения $G_{\alpha,\theta}(x)$ будем обозначать $Z_{\alpha,\theta}$.

Из (2) вытекает, что характеристическая функция симметричного ($\theta = 0$) строго устойчивого распределения имеет вид:

$$f_{\alpha,0}(t) = e^{-|t|^\alpha}, \quad t \in \mathbb{R}. \quad (3)$$

Лемма 1. *Симметричное строго устойчивое распределение с характеристическим показателем α является масштабной смесью нормальных законов, в которой смешивающим распределением является*

односторонний устойчивый закон ($\theta = 1$) с характеристическим показателем $\alpha/2$:

$$G_{\alpha,0}(x) = \int_0^\infty \Phi \left(\frac{x}{\sqrt{z}} \right) dG_{\alpha/2,1}(z), \quad x \in \mathbb{R}. \quad (4)$$

Доказательство. См., например, теорему 3.3.1 в [32].

Чтобы доказать, что каждое распределение Вейбулла с параметром $\gamma > 0$ является масштабной смесью распределения Вейбулла с параметром $\delta > \gamma$, прежде всего убедимся, что каждое распределение Вейбулла с параметром $\gamma \in (0, 2]$ является масштабной смесью распределений Рэлея.

Лемма 2. *Для любого $\gamma \in (0, 2]$ справедлива факторизация*

$$W_\gamma \stackrel{d}{=} W_2 \sqrt{V_{\gamma/2}}, \quad (5)$$

где $V_{\gamma/2} = 2Z_{\gamma/2,1}^{-1}$, а $Z_{\gamma/2,1}$ — случайная величина с односторонней строго устойчивой плотностью $g_{\gamma/2,1}(x)$, причем случайные величины в правой части (5) независимы.

Доказательство. Запишем соотношение (4) в терминах характеристических функций с учетом (3):

$$e^{-|t|^\alpha} = \int_0^\infty \exp \left\{ -\frac{1}{2} t^2 z \right\} g_{\alpha/2,1}(z) dz, \quad t \in \mathbb{R}. \quad (6)$$

Формально полагая в (6) $|t| = x$, где $x \geq 0$ — произвольное неотрицательное число, имеем:

$$P(W_\gamma > x) =$$

$$= e^{-x^\gamma} = \int_0^\infty \exp \left\{ -\frac{1}{2} x^2 z \right\} g_{\gamma/2,1}(z) dz. \quad (7)$$

В то же время, очевидно, что если W_2 и $Z_{\gamma/2,1}$ независимы, то

$$P \left(W_2 \sqrt{V_{\gamma/2}} > x \right) = P \left(W_2 > x \sqrt{\frac{1}{2} Z_{\gamma/2,1}} \right) =$$

$$= \int_0^\infty \exp \left\{ -\frac{1}{2} x^2 z \right\} g_{\gamma/2,1}(z) dz. \quad (8)$$

Поскольку правые части (7) и (8) тождественно (по $x \geq 0$) совпадают, тождественно совпадают и левые части этих соотношений. Лемма доказана.

Лемма 3. *Для любого $\gamma \in (0, 1]$ распределение Вейбулла с параметром γ является смешанным показательным распределением:*

$$W_\gamma \stackrel{d}{=} W_1 V_\gamma, \quad (9)$$

где $V_\gamma = 2Z_{\gamma,1}^{-1}$, а $Z_{\gamma,1}$ — случайная величина с односторонней строго устойчивой плотностью $g_{\gamma,1}(x)$, причем случайные величины в правой части (9) независимы.

Доказательство. Несложно видеть, что $P(W_1^{1/\gamma} \geq x) = P(W_1 \geq x^\gamma) = e^{-x^\gamma} = P(W_\gamma \geq x)$, $x \geq 0$, т. е.

$$W_\gamma \stackrel{d}{=} W_1^{1/\gamma}. \quad (10)$$

Из (10) вытекает, что $W_2 \stackrel{d}{=} \sqrt{W_1}$. Поэтому из леммы 2 вытекает, что для $\gamma \in (0, 2]$

$$W_\gamma \stackrel{d}{=} W_2 \sqrt{V_{\gamma/2}} \stackrel{d}{=} \sqrt{W_1 V_{\gamma/2}}$$

или, с учетом (10),

$$W_{\gamma/2} \stackrel{d}{=} W_\gamma^2 \stackrel{d}{=} W_1 V_{\gamma/2}.$$

Переобозначив $\gamma/2 \mapsto \gamma \in (0, 1]$, получим требуемое утверждение.

Замечание 1. Случай малых значений параметра $\gamma \in (0, 1]$ представляет особый интерес, поскольку распределения Вейбулла с такими параметрами занимают промежуточное место между распределениями с экспоненциальным убыванием хвостов (показательное распределение, гамма-распределение) и «тяжелохвостыми» распределениями со степенным убыванием хвостов типа Ципфа–Парето.

Из соотношения (9) вытекает основной результат данного раздела, обобщающий леммы 2 и 3 и устанавливающий, что распределение Вейбулла с произвольным положительным параметром формы γ является масштабной смесью распределений Вейбулла с произвольным положительным параметром формы $\delta > \gamma$.

Теорема 1. Пусть $\delta > \gamma > 0$ — произвольные числа. Тогда

$$W_\gamma \stackrel{d}{=} W_\delta V_\alpha^{1/\delta},$$

где $\alpha = \gamma/\delta \in (0, 1)$, а случайные величины в правой части независимы.

Доказательство. При доказательстве теоремы 2 было установлено, что распределение Вейбулла с параметром $\alpha \in (0, 1]$ является смешанным показательным распределением. Действительно, из (9) вытекает, что

$$\begin{aligned} e^{-x^\alpha} &= P(W_\alpha > x) = P\left(W_1 > \frac{1}{2} Z_{\alpha,1} x\right) = \\ &= \int_0^\infty e^{-(1/2)zx} g_{\alpha,1}(z) dz, \quad x \geq 0. \end{aligned}$$

Поэтому для любых $\delta > \gamma > 0$, обозначив $\alpha = \gamma/\delta$ (при этом, очевидно, $\alpha \in (0, 1)$), для любого $x \in \mathbb{R}$ получим:

$$\begin{aligned} P(W_\gamma > x) &= e^{-x^\gamma} = e^{-x^{\delta\alpha}} = \\ &= P(W_\alpha > x^\delta) = P\left(W_1 > \frac{1}{2} Z_{\alpha,1} x^\delta\right) = \\ &= \int_0^\infty e^{-(1/2)zx^\delta} g_{\alpha,1}(z) dz = \\ &= \int_0^\infty P\left(W_\delta > x \left(\frac{1}{2} z\right)^{1/\delta}\right) g_{\alpha,1}(z) dz = \\ &= P(W_\delta V_\alpha^{1/\delta} > x), \end{aligned}$$

что и требовалось доказать. Теорема доказана.

Замечание 2. Если $0 < \gamma < \delta < 2$, то результат теоремы 1 непосредственно вытекает из теоремы 3.3.1 в [32] в силу формального совпадения характеристической функции строго устойчивого закона и дополнительной функции распределения Вейбулла (см. доказательство леммы 2).

Теорема 1 будет использована ниже с целью расширения представлений об условиях сходимости распределений экстремальных порядковых статистик к распределению Вейбулла.

3 Об условиях сходимости распределений экстремальных порядковых статистик в выборках случайного объема к распределению Вейбулла

В классических задачах математической статистики объем выборки, доступной исследователю, традиционно считается детерминированным и в асимптотических постановках играет роль (неограниченно возрастающего) известного параметра. В то же время на практике часто возникают ситуации, когда размер выборки не является заранее определенным и может рассматриваться как случайный. Эти ситуации, как правило, связаны с тем, что статистические данные накапливаются в течение фиксированного времени. Это имеет место, в частности, в страховании, когда в течение разных отчетных периодов одинаковой длины (скажем, месяцев) происходит разное число страховых событий (страховых выплат и/или заключений страховых контрактов); в медицине, когда число пациентов с тем или иным заболеванием варьируется от года к году; в технике, когда при испытании

на надежность (скажем, при определении наработки на отказ) разных партий приборов (изделий), число отказавших приборов в разных партиях будет разным; в информатике при разработке методов оценки «своевременности» завершения программ, включая методы решения задач предсказания времени безотказного функционирования или времени выполнения прикладных программ в случайных вычислительных средах. В таких ситуациях заранее не известное число наблюдений, которые будут доступны исследователю, разумно считать случайной величиной. Другими словами, в таких ситуациях объем выборки не является (известным) параметром, а сам становится *наблюдением*, т. е. статистикой. В силу указанных обстоятельств вполне естественным становится изучение асимптотического поведения распределений статистик достаточно общего вида, основанных на выборках случайного объема.

На естественность такого подхода, в частности, обратил внимание Б. В. Гнеденко в работе [33], в которой рассматривались асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема, и было продемонстрировано, что при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут радикально измениться. К примеру, если объем выборки является геометрически распределенной случайной величиной, то вместо ожидаемого в соответствии с классической теорией нормального закона в качестве асимптотического распределения выборочной медианы возникает распределение Стьюдента с двумя степенями свободы, хвосты которого столь тяжелы, что у него отсутствуют моменты порядков, больших второго.

Как уже говорилось, еще с работ М. Фреше хорошо известно, что распределение Вейбулла может быть предельным для линейно преобразованных минимальных порядковых статистик в выборках неслучайного объема. Несложно убедиться, что при тех же условиях на распределение генеральной совокупности распределение Вейбулла может быть предельным для линейно преобразованных минимальных порядковых статистик в выборках *случайного* объема, если случайный объем выборки имеет асимптотически вырожденное распределение (см., например, [28] или [34]). Однако эта тривиальная ситуация отнюдь не исчерпывает все возможные условия, при которых распределения линейно преобразованных минимальных порядковых статистик в выборках случайного объема сходятся к распределению Вейбулла.

Убедимся в этом на примере, в котором объем выборки формируется в соответствии с дважды

стохастическим пуассоновским процессом, иначе называемым процессом Кокса. В книгах [34–38] показано, что такие процессы являются удобными, реалистичными и адекватными моделями неоднородных во времени хаотических потоков событий. В соответствии с описанным в указанных книгах подходом поток информативных событий, в результате каждого из которых появляется очередное «наблюденное» значение рассматриваемой характеристики, описывается с помощью точечного случайного процесса вида $M(\Lambda(t))$, где $M(t)$, $t \geq 0$, — однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t)$, $t \geq 0$, — независимый от $M(t)$ случайный процесс, обладающий следующими свойствами: $\Lambda(0) = 0$, $P(\Lambda(t) < \infty) = 1$ для любого $t > 0$, траектории $\Lambda(t)$ не убывают и непрерывны справа. Процесс $M(\Lambda(t))$, $t \geq 0$, называется дважды стохастическим пуассоновским процессом (процессом Кокса).

В соответствии с такой моделью в каждый момент времени t распределение случайной величины $M(\Lambda(t))$ является смешанным пуассоновским. Для большей наглядности рассмотрим случай, когда в рассматриваемой модели «бесконечно большой» параметр t дискретен: $\Lambda(t) = \Lambda_k$, $k \in \mathbb{N}$, а $\{\Lambda_k\}_{k \geq 1}$ — неограниченно возрастающая последовательность случайных величин, $\Lambda_{k+1}(\omega) \geq \Lambda_k(\omega)$ для каждого $\omega \in \Omega$, $k \geq 1$. Соответственно, положим $N_k = M(\Lambda_k)$, $k \geq 1$. При этом асимптотика $k \rightarrow \infty$ может интерпретироваться как то, что (случайная) интенсивность потока информативных событий считается очень большой.

Пусть Y_1, Y_2, \dots — независимые случайные величины с общей функцией распределения $F(x) = P(Y_i < x)$, $x \in \mathbb{R}$, $i \geq 1$. Обозначим $\text{lext}(F) = \inf\{x : F(x) > 0\}$. Предположим, что при каждом $k \in \mathbb{N}$ случайная величина N_k независима от последовательности Y_1, Y_2, \dots . В книге [34] доказан следующий результат, который в данном контексте играет вспомогательную роль и потому оформлен в виде леммы.

Лемма 4. *Предположим, что существуют неограниченно возрастающая последовательность положительных чисел $\{d_k\}_{k \geq 1}$ и неотрицательная случайная величина Λ такие, что имеет место сходимость*

$$d_k^{-1} \Lambda_k \implies \Lambda \quad (k \rightarrow \infty).$$

Предположим также, что $\text{lext}(F) > -\infty$ и функция распределения $W_F(x) = F(\text{lext}(F) - x^{-1})$ удовлетворяет условию: существует положительное число δ такое, что для любого $x > 0$

$$\lim_{y \rightarrow \infty} \frac{W_F(yx)}{W_F(y)} = x^{-\delta}. \quad (11)$$

Тогда существуют числовые последовательности a_k и b_k такие, что

$$P\left(\min_{1 \leq j \leq N_k} Y_j - a_k < b_k x\right) \Rightarrow \left[1 - \int_0^\infty e^{-\lambda x^\delta} dP(\Lambda < \lambda)\right] \mathbf{1}(x \geq 0), \quad k \rightarrow \infty.$$

При этом числа a_k и b_k могут быть определены как

$$\left. \begin{aligned} a_k &= \text{lext}(F); \\ b_k &= \sup\{x : F(x) \leq d_k^{-1}\} - \text{lext}(F), \quad k \geq 1. \end{aligned} \right\} (12)$$

Из леммы 4 и теоремы 1 вытекает следующий результат.

Теорема 2. *Предположим, что функция распределения F принадлежит к области \min -притяжения распределения Вейбулла с некоторым параметром формы $\delta > 0$, т. е. $\text{lext}(F) > -\infty$ и выполнено условие (11).*

I. *Предположим, что существуют неограниченно возрастающая последовательность положительных чисел $\{d_k\}_{k \geq 1}$ и число $\alpha \in (0, 1)$ такие, что имеет место сходимость*

$$d_k^{-1} \Lambda_k \Rightarrow 1 \quad (k \rightarrow \infty). \quad (13)$$

Тогда существуют числовые последовательности a_k и b_k такие, что

$$\frac{1}{b_k} \left(\min_{1 \leq j \leq N_k} Y_j - a_k\right) \Rightarrow W_\delta \quad (k \rightarrow \infty),$$

где числа a_k и b_k могут быть определены в соответствии с (12).

II. *Предположим, что существуют неограниченно возрастающая последовательность положительных чисел $\{d_k\}_{k \geq 1}$ и число $\alpha \in (0, 1)$ такие, что имеет место сходимость*

$$d_k^{-1} \Lambda_k \Rightarrow \frac{1}{2} Z_{\alpha,1} \quad (k \rightarrow \infty). \quad (14)$$

Тогда существуют числовые последовательности a_k и b_k такие, что

$$\frac{1}{b_k} \left(\min_{1 \leq j \leq N_k} Y_j - a_k\right) \Rightarrow W_\gamma \quad (k \rightarrow \infty),$$

где $\gamma = \delta\alpha$, а числа a_k и b_k могут быть определены в соответствии с (12).

Таким образом, существенная случайность объема выборки может заметно «утяжелить» хвост предельного вейбулловского закона по сравнению с тривиальной ситуацией (13) с асимптотически вырожденным (асимптотически неслучайным) индексом. Например, если $\delta = 1$ (т. е. «изначально»

предельным распределением минимальной порядковой статистики должно быть экспоненциальное), но объем выборки имеет вид $N_k = M(\Lambda_k)$, причем при некоторых d_k выполнено условие (14) с $\alpha = 1/2$, то реальное предельное распределение минимальной порядковой статистики — это распределение Вейбулла с параметром 1/2: $[1 - e^{-\sqrt{x}}] \mathbf{1}(x \geq 0)$, $x \in \mathbb{R}$. При этом, в частности, квантиль порядка 0,99 этого распределения примерно равна 21,208, что почти в 5 раз больше соответствующей квантили показательного распределения, примерно равной 4,605.

Литература

1. Weibull W. A statistical theory of the strength of materials // Ingeniörs Vetenskaps Akademien Handlingar, 1939. Nr. 151. P. 1–45.
2. Weibull W. The phenomenon of rupture in solids // Ingeniörs Vetenskaps Akademien Handlingar, 1939. Nr. 153. P. 16–53.
3. Weibull W. A statistical distribution function of wide applicability // J. Appl. Mech. Trans. ASME, 1951. Vol. 18. No. 3. P. 293–297.
4. Fréchet M. Sur la loi de probabilité de l'écart maximum // Annales de la Société polonaise de Mathématique (Cra-covie), 1927. Vol. 6. P. 93–116.
5. Rosin P., Rammler E. The laws governing the fineness of powdered coal // J. Inst. Fuel, 1933. Vol. 7. P. 29–36.
6. Stoyan D. Weibull, RRSB or extreme-value theorists? // Metrika, 2013. Vol. 76. P. 153–159. doi: 10.1007/s00184-011-0380-6.
7. Rosin P., Rammler E., Sperling K. Korngrößenprobleme des Kohlenstaubes und ihre Bedeutung für die Vermahlung. Bericht C 52 des Reichskohlenrates. — Berlin: VDI-Verlag, 1933. 25 p.
8. Bennett J. G. Broken coal // J. Inst. Fuel, 1936. Vol. 10. P. 22–39.
9. Fisher R. A., Tippett L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample // Proc. Camb. Philos. Soc., 1928. Vol. 24. P. 180–190.
10. Gnedenko B. V. Sur la distribution limite du terme maximum d'une serie aléatoire // Ann. Math., 1943. Vol. 44. No. 3. P. 423–453.
11. Johnson N. L., Kotz S., Balakrishnan N. Continuous univariate distributions. — 2nd ed. — New York: John Wiley & Sons, 1994. 756 p.
12. Vox G., Tiao G. Bayesian inference in statistical analysis. — Reading: Addison–Wesley, 1973. 578 p.
13. Григорьева М. Е., Королев В. Ю. О сходимости распределений случайных сумм к скошенным экспоненциально-степенным законам // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 66–74.
14. Elandt-Johnson R., Johnson N. Survival models and data analysis. — New York: John Wiley & Sons, 1999. 457 p.

15. *Hogg R. V., Klugman S. A.* Loss distributions. — New York: John Wiley & Sons, 1984. 235 p.
16. *D'Addario R.* Intorno ad una funzione di distribuzione // *Giorn. Econ. Ann. Econ.*, 1974. Vol. 33. P. 205–214.
17. *Mittnik S., Rachev S. T.* Stable distributions for asset returns // *Appl. Math. Lett.*, 1989. Vol. 2. No. 3. P. 301–304.
18. *Mittnik S., Rachev S. T.* Modeling asset returns with alternative stable distributions // *Economet. Rev.*, 1993. Vol. 12. P. 261–330.
19. *Bartels C. P. A.* Economic aspects of regional welfare. — Leiden: Martinus Nijhoff, 1977. 261 p.
20. *Bordley R. F., McDonald J. B., Mantrala A.* Something new, something old: Parametric models for the size distribution of income // *J. Income Distribution*, 1996. Vol. 6. P. 91–103.
21. *Lawless J. F.* Statistical models and methods for lifetime data. — New York: John Wiley & Sons, 1982. 580 p.
22. *Abernethy R. B.* The new Weibull handbook. Reliability an statistical analysis for predicting life, safety, survivability, risk, cost and warranty claims. — 5th ed. — 536 Oyster Road, North Palm Beach, FL: Robert B. Abernethy, 2004. 310 p.
23. *Nawaz Sharif M., Nazrul Islam M.* The Weibull distribution as a general model for forecasting technological change // *Technol. Forecast. Soc. Change*, 1980. Vol. 18. No. 3. P. 247–256.
24. *Johnson N. L., Kotz S.* Continuous univariate distributions. — Boston: Houghton Mifflin Company, 1970. 761 p.
25. *Kotz S., Nadarajah S.* Extreme value distributions. Theory and applications. — London: Imperial College Press, 2000. 186 p.
26. *Saunders S. C.* Birnbaum's contributions to reliability // Reliability and fault tree analysis, theoretical and applied aspects of system reliability and safety assessment / Eds. R. E. Barlow, J. B. Fussell, N. D. Singpurwalla. — Philadelphia: Society for Industrial and Applied Mathematics, 1975. XV–XXXIX.
27. *Гумбель Э.* Статистика экстремальных значений / Пер. с англ. — М.: Мир, 1965. 452 с. (*Gumbel E. J.* Statistics of extremes. — New York: Columbia University Press, 1958.)
28. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик / Пер. с англ. В. А. Егорова, В. Б. Невзорова. — М.: Наука, 1984. 303 с. (*Galambos J.* The asymptotic theory of extreme order statistics. — New York: John Wiley, 1978. 366 p.)
29. *Григорьева М. Е., Королев В. Ю., Соколов И. А.* Предельная теорема для геометрических сумм независимых неодинаково распределенных случайных величин и ее применение к прогнозированию вероятности катастроф в неоднородных потоках экстремальных событий // *Информатика и её применения*, 2013. Т. 7. Вып. 4. С. 11–19.
30. *Королев В. Ю., Григорьева М. Е., Нефедова Ю. С., Лазовский Р.* Метод оценивания вероятностей катастроф в неоднородных потоках экстремальных событий и его применение к прогнозированию наводнений в Санкт-Петербурге // *Актуарий*, 2014. Т. 5. Вып. 1. С. 53–58.
31. *Rayleigh J. W. S.* On the resultant of a large number of vibrations of the same pitch and of arbitrary phase // *Philos. Mag.* 5th ser., 1880. Vol. 10. P. 73–78.
32. *Золотарев В. М.* Одномерные устойчивые распределения. — М.: Наука, 1983. 304 с.
33. *Гнеденко Б. В.* Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // *Труды Тбилисского математического института*, 1989. Т. 92. С. 146–150.
34. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008. 192 с.
35. *Gnedenko B. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996. 267 p.
36. *Bening V., Korolev V.* Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002. 434 p.
37. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и дополн. — М.: Физматлит, 2011. 620 с.
38. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 510 с.

Поступила в редакцию 31.07.14

ON CONDITIONS OF CONVERGENCE OF THE DISTRIBUTIONS OF EXTREMAL ORDER STATISTICS TO THE WEIBULL DISTRIBUTION

V. Yu. Korolev^{1,2} and I. A. Sokolov²

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

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Some product representations are obtained for random variables with the Weibull distribution by stable random variables. These results are used to describe the conditions providing convergence of the distributions of linearly transformed minimum order statistics in samples with random sizes to the Weibull distribution. The presented results broaden traditional conceptions concerning conditions of convergence of extremal order statistics to the Weibull distribution and give additional theoretical explanation for high adequacy of the Weibull distribution in lifetime data analysis, in particular, in reliability theory.

Keywords: Weibull distribution; exponential distribution; Rayleigh distribution; strictly stable distribution; sample with random size

DOI: 10.14375/19922264140301

Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (projects Nos. 12-07-00115a, 12-07-00109a, and 14-07-00041a).

References

1. Weibull, W. 1939. A statistical theory of the strength of materials. *Ingeniörs Vetenskaps Akademien Handlingar* Nr. 151. Stockholm: Generalstabens Litografiska Anstalts Forlag. 45 p.
2. Weibull, W. 1939. The phenomenon of rupture in solids. *Ingeniörs Vetenskaps Akademien Handlingar* Nr. 153. Stockholm: Generalstabens Litografiska Anstalts Forlag. 55 p.
3. Weibull, W. 1951. A statistical distribution function of wide applicability. *J. Appl. Mech. Trans. ASME* 18(3):293–297.
4. Fréchet, M. 1927. Sur la loi de probabilité de l'écart maximum. *Annales de la Société polonaise de Mathématique*. Cracovie. 6:93–116.
5. Rosin, P., and E. Rammler. 1933. The laws governing the fineness of powdered coal. *J. Inst. Fuel* 7:29–36.
6. Stoyan, D. 2013. Weibull, RRSB or extreme-value theorists? *Metrika* 76:153–159. doi: 10.1007/s00184-011-0380-6.
7. Rosin, P., E. Rammler, and K. Sperling. 1933. Korngrößenprobleme des Kohlenstaubes und ihre Bedeutung für die Vermahlung. Bericht C 52 des Reichskohlenrates. Berlin: VDI-Verlag. 25 p.
8. Bennett, J. G. 1936. Broken coal. *J. Inst. Fuel* 10:22–39.
9. Fisher, R. A., and L. H. C. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Philos. Soc.* 24:180–190.
10. Gnedenko, B. V. 1943. Sur la distribution limite du terme maximum d'une serie aléatoire. *Ann. Math.* 44(3):423–453.
11. Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*. 2nd ed. New York: John Wiley & Sons. 756 p.
12. Box, G., and G. Tiao. 1973. *Bayesian inference in statistical analysis*. Reading: Addison-Wesley. 578 p.
13. Grigor'eva, M. E., and V. Yu. Korolev. 2013. O skhodimosti raspredeleniy sluchaynykh summ k skoshennym eksponentsial'no-stepennym zakonam [On convergence of the distributions of random sums to skew exponential power laws]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):66–74. doi: 10.14357/19922264130407.
14. Elandt-Johnson, R., and N. Johnson. 1999. *Survival models and data analysis*. New York: John Wiley & Sons. 457 p.
15. Hogg, R. V., and S. A. Klugman. 1984. *Loss distributions*. New York: John Wiley & Sons. 235 p.

16. D'Addario, R. 1974. Intorno ad una funzione di distribuzione. *Giorn. Econ. Ann. Econ.* 33:205–214.
17. Mittnik, S., and S. T. Rachev. 1989. Stable distributions for asset returns. *Appl. Math. Lett.* 2(3):301–304.
18. Mittnik, S., and S. T. Rachev. 1993. Modeling asset returns with alternative stable distributions. *Economet. Rev.* 12:261–330.
19. Bartels, C. P. A. 1977. *Economic aspects of regional welfare*. Leiden: Martinus Nijhoff. 261 p.
20. Bordley, R. F., J. B. McDonald, and A. Mantrala. 1996. Something new, something old: Parametric models for the size distribution of income. *J. Income Distribution* 6:91–103.
21. Lawless, J. F. 1982. *Statistical models and methods for lifetime data*. New York: John Wiley & Sons. 580 p.
22. Abernethy, R. B. 2004. *The new Weibull handbook. Reliability and statistical analysis for predicting life, safety, survivability, risk, cost and warranty claims*. 5th ed. North Palm Beach, FL: Robert B. Abernethy. 310 p.
23. Nawaz Sharif, M., and M. Nazrul Islam. 1980. The Weibull distribution as a general model for forecasting technological change. *Technol. Forecast. Soc. Change* 18(3):247–256.
24. Johnson, N. L., and S. Kotz. 1970. *Continuous univariate distributions*. Boston: Houghton Mifflin Company. 761 p.
25. Kotz, S., and S. Nadarajah. 2000. *Extreme value distributions. Theory and applications*. London: Imperial College Press. 186 p.
26. Saunders, S. C. 1975. Birnbaum's contributions to reliability. *Reliability and fault tree analysis, theoretical and applied aspects of system reliability and safety assessment*. Eds. R. E. Barlow, J. B. Fussell, and N. D. Singpurwalla. Philadelphia: Society for Industrial and Applied Mathematics. XV–XXXIX.
27. Gumbel, E. J. 2004. *Statistics of extremes*. Mineola, New York: Courier Publications. 375 p.
28. Galambos, J. 1987. *Asymptotic theory of extreme order statistics*. 2nd ed. Malabar: Krieger Publ. Co. 430 p.
29. Grigor'eva, M. E., V. Yu. Korolev, and I. A. Sokolov. 2013. Predel'naya teorema dlya geometricheskikh sum nezavisimyykh neodnorodnykh raspredelennykh sluchaynykh velichin i ee primeneniye k prognozirovaniyu veroyatnosti katastrof v neodnorodnykh potokakh ekstremal'nykh sobyitiy [A limit theorem for geometric sums of independent nonidentically distributed random variables and its application to the prediction of the probabilities of catastrophes in nonhomogeneous flows of extremal events]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):11–19. doi: 10.14357/19922264130402.
30. Korolev, V. Yu., M. E. Grigor'eva, Yu. S. Nefedova, and R. Lazovskiy. 2014. Metod otsenivaniya veroyatnostey katastrof v neodnorodnykh potokakh ekstremal'nykh sobyitiy i ego primeneniye k prognozirovaniyu navodneniy v Sankt-Peterburge [A method of estimation of probabilities of catastrophes in nonstationary flows of extremal events and its application to prediction of floods in Saint-Petersburg]. *Aktuariy [Actuary]* 5(1):53–58.
31. Rayleigh, J. W. S. 1880. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philos. Mag.* 5th ser. 10:73–78.
32. Zolotarev, V. M. 1986. One-dimensional stable distributions. *Translations of mathematical monographs*. Vol. 65. Providence: American Mathematical Society. 284 p.
33. Gnedenko, B. V. 1989. Ob otsenke neizvestnykh parametrov raspredeleniya pri sluchaynom chisle nezavisimyykh nablyudeniy [On estimation of unknown parameters of distributions from a random number of independent observations]. *Trudy Tbilisskogo Matematicheskogo Instituta [Proceedings of Tbilisi Mathematical Institute]*. 92:146–150.
34. Korolev, V. Yu., and I. A. Sokolov. 2008. *Matematicheskie modeli neodnorodnykh potokov ekstremal'nykh sobyitiy [Mathematical models of nonstationary flows of extremal events]*. Moscow: TORUS PRESS. 192 p.
35. Gnedenko, B. V., and V. Yu. Korolev. 1996. *Random summation: Limit theorems and applications*. Boca Raton: CRC Press. 267 p.
36. Bening, V., and V. Korolev. 2002. *Generalized Poisson models and their applications in insurance and finance*. Utrecht: VSP. 434 p.
37. Kopolev, V. Yu., V. E. Bening, and S. Ya. Shorgin. 2011. *Matematicheskie osnovy teorii riska [Mathematical foundations of risk theory]*. 2nd ed. Moscow: Fizmatlit. 620 p.
38. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov [Probabilistic and statistical methods for the decomposition of volatility of chaotic processes]*. Moscow: Moscow University Press. 510 p.

Received July 31, 2014

Contributors

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; victoryukorolev@yandex.ru

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ISokolov@ipiran.ru

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НОРМАЛЬНЫХ ПРОЦЕССОВ В СТОХАСТИЧЕСКИХ СИСТЕМАХ СО СЛОЖНЫМИ НЕЛИНЕЙНОСТЯМИ

И. Н. Сеницын¹, В. И. Сеницын²

Аннотация: Рассматриваются конечномерные дифференциальные стохастические системы (ДСтС) и эрмитарные (интегродифференциальные) стохастические системы (ЭСтС) с винеровскими и пуассоновскими шумами, приводимые к ДСтС со сложными конечными, дифференциальными и интегральными нелинейностями. Такие модели функционирования описывают поведение многих современных нано- и квантово-оптических технических средств информатики. Приводятся уравнения методов нормальной аппроксимации (МНА) и статистической линеаризации (МСЛ) для аналитического моделирования нестационарных и стационарных нормальных (гауссовских) процессов в нелинейных ДСтС и нелинейных ЭСтС путем аппроксимации эрмитарных ядер линейными операторными уравнениями для дифференцируемых нелинейностей и сингулярными ядрами для недифференцируемых нелинейностей. Рассматриваются методы вычисления типовых интегралов МНА (МСЛ) для сложных (многомерных и векторного аргумента) конечных и дифференциальных нелинейностей. Особое внимание уделяется иррациональным и дробно-рациональным нелинейностям скалярного аргумента. Приводятся примеры вычисления интегралов. Подробно рассматриваются вопросы вычисления типовых интегралов МНА (МСЛ) для сложных интегральных нелинейностей.

Ключевые слова: аналитическое моделирование; дифференциальные стохастические системы с винеровскими и пуассоновскими шумами (ДСтС); метод нормальной аппроксимации (МНА); метод статистической линеаризации (МСЛ); сложные иррациональные нелинейности; сложные конечные, дифференциальные и интегральные нелинейности; эрмитарные стохастические системы (ЭСтС), приводимые к дифференциальным

DOI: 10.14375/19922264140302

1 Введение

Моделями функционирования многих современных технических систем информатики служат стохастические системы (СтС), описываемые дифференциальными, интегральными и интегродифференциальными уравнениями со сложными дробно-рациональными, иррациональными и интегральными нелинейностями. В [1] дано систематическое изложение МНА и МСЛ для ДСтС и ЭСтС, приводимых к дифференциальным.

Обобщая [2–7], рассмотрим развитие МНА и МСЛ для аналитического моделирования нормальных стохастических процессов (СтП) на случай СтС со сложными конечными, дифференциальными и интегральными нелинейностями.

Как показано в [6], альтернативным подходом к аналитическому моделированию СтП в ДСтС и ЭСтС служит подход, основанный на дискретизации стохастических дифференциальных уравнений

на основе использования обобщенной формы Ито и кратных стохастических интегралов от винеровских и пуассоновских СтП с последующим применением дискретных версий МНА (МСЛ).

Статья состоит из введения, пяти разделов и заключения.

В разд. 2 и 3 приводятся уравнения МНА и МСЛ для аналитического моделирования одно- и двумерных распределений стационарных и нестационарных СтП в ДСтС и ЭСтС, приводимых к ДСтС.

Типовые интегралы МНА и МСЛ рассматриваются в разд. 4.

Особенности аналитического моделирования в ДСтС со сложными конечными и дифференциальными нелинейностями обсуждаются в разд. 5.

Раздел 6 посвящен аналитическому моделированию СтП в ДСтС со сложными интегральными нелинейностями.

Приводятся примеры.

¹Институт проблем информатики Российской академии наук, sinitsin@dol.ru

²Институт проблем информатики Российской академии наук, vsinitsin@ipiran.ru

2 Уравнения методов нормальной аппроксимации и статистической линеаризации для дифференциальных стохастических систем

Как известно [3, 5], уравнения конечномерных непрерывных нелинейных систем со стохастическими возмущениями путем расширения вектора состояния ДСтС могут быть записаны в виде следующего векторного стохастического дифференциального уравнения Ито:

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

Здесь $a = a(Y_t, t)$ и $b = b(y_t, t)$ — известные $(p \times 1)$ -мерная и $(p \times m)$ -мерная функции Y_t и t ; $W_0 = W_0(t)$ — r -мерный винеровский СтП интенсивности $\nu_0 = \nu_0(t)$; $c(Y_t, t, v)$ — $(p \times 1)$ -мерная функция Y_t, t и вспомогательного $(q \times 1)$ -мерного параметра v ; $\int_{\Delta} dP^0(t, A)$ — центрированная пуассоновская мера, определяемая

$$\int_{\Delta} dP^0(t, A) = \int_{\Delta} dP(t, A) = \int_{\Delta} \nu_P(t, A) dt.$$

В (1) принято: \int_{Δ} — число скачков пуассоновского СтП в интервале времени $\Delta = (t_1, t_2]$; $\nu_P(t, A)$ — интенсивность пуассоновского СтП $P(t, A)$; A — некоторое борелевское множество пространства R_0^q с выколотым началом. Начальное значение Y_0 представляет собой случайную величину, не зависящую от приращений СтП $W_0(t)$ и $P(t, A)$ на интервалах времени, следующих за t_0 , $t_0 \leq t_1 \leq t_2$ для любого множества A .

В случае аддитивных нормальных (гауссовских) и обобщенных пуассоновских возмущений уравнение (1) имеет вид:

$$\dot{Y}_t = a(Y_t, t) + b_0(t)V, \quad V = \dot{W}, \quad Y(t_0) = Y_0. \quad (2)$$

Здесь W — СтП с независимыми приращениями, представляющий собой смесь нормального и обобщенного пуассоновского СтП.

Если предположить существование конечных вероятностных моментов второго порядка для моментов времени t_1 и t_2 , то уравнения МНА примут следующий вид [3, 5]:

— для характеристических функций

$$g_1^N(\lambda; t) = \exp \left[i\lambda^T m_t - \frac{1}{2} \lambda^T K_t \lambda \right]; \quad (3)$$

$$g_{t_1, t_2}^N(\lambda_1, \lambda_2; t_1, t_2) = \exp \left[i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \bar{\lambda} \right], \quad (4)$$

где

$$\bar{\lambda} = [\lambda_1^T \lambda_2^T]^T; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T; \\ \bar{K}_2 = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix};$$

— для математических ожиданий m_t , ковариационной матрицы K_t и матрицы ковариационных функций $K(t_1, t_2)$:

$$\dot{m}_t = a_1(m_t, K_t, t), \quad m_0 = m(t_0); \quad (5)$$

$$\dot{K}_t = a_2(m_t, K_t, t), \quad K_0 = K(t_0); \quad (6)$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = K(t_1, t_2) a_{21}(m_{t_2}, K_{t_2}, t_2)^T; \\ K(t_1, t_1) = K_{t_1}. \quad (7)$$

Здесь приняты следующие обозначения:

$$a_1 = a_1(m_t, K_t, t) = M_N a(Y_t, t); \quad (8)$$

$$a_2 = a_2(m_t, K_t, t) = a_{21}(m_t, K_t, t) + \\ + a_{21}(m_t, K_t, t)^T + a_{22}(m_t, K_t, t); \quad (9)$$

$$a_{21} = a_{21}(m_t, K_t, t) = M_N a(Y_t, t) Y_t^{0T}; \quad (10)$$

$$a_{22} = a_{22}(m_t, K_t, t) = M_N \sigma(Y_t, t);$$

$$\sigma(Y_t, t) = b(Y_t, t) \nu_0(t) b(Y_t, t)^T + \\ + \int_{R_0^q} c(Y_t, t, v) c(Y_t, t, v)^T \nu_P(t, dv);$$

$$m_t = M Y_t, \quad Y_t^0 = Y_t - m_t,$$

$$K_t = M_N Y_0^0 Y_t^{0T}, \quad K(t_1, t_2) = M_N Y_{t_1}^0 Y_{t_2}^{0T};$$

M_N — символ вычисления математического ожидания для нормальных распределений (3) и (4).

Для стационарных ДСтС нормальные стационарные СтП — если они существуют, то $m_t = \bar{m}$, $K_t = \bar{K}$, $K(t_1, t_2) = k(\tau)$ ($\tau = t_1 - t_2$), — определяются уравнениями [3, 5]:

$$a_1(\bar{m}, \bar{K}) = 0; \quad a_2(\bar{m}, \bar{K}) = 0; \quad (11)$$

$$\left. \begin{aligned} \dot{k}_\tau(\tau) &= a_{21}(\bar{m}, \bar{K}) \bar{K}^{-1} k(\tau); \\ k(0) &= \bar{K} \quad (\forall \tau > 0), \quad k(\tau) = k(-\tau)^T \quad (\forall \tau < 0). \end{aligned} \right\} \quad (12)$$

При этом необходимо, чтобы матрица $a_{21}(\bar{m}, \bar{K}) = \bar{a}_{21}$ была бы асимптотически устойчивой.

Для ДСтС (2) уравнения МНА переходят в уравнения МСЛ Казакова [3, 5], если принять

$$a(Y_t, t) = a_1(m_t, K_t) + k_1^a(m_t, K_t)Y_t^0; \quad (13)$$

$$\left. \begin{aligned} b(Y_t, t) &= b_0(t); \\ \sigma(Y_t, t) &= b_0(t)\nu(t)b_0(t)^T = \sigma_0(t), \end{aligned} \right\} \quad (14)$$

$$k_1^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) a_0(m_t, K_t, t)^T \right]^T; \quad (15)$$

$$\dot{m}_t = a_1(m_t, K_t, t), \quad m_0 = m(t_0), \quad (16)$$

$$\left. \begin{aligned} \dot{K}_t &= k_1^a(m_t, K_t, t)K_t + K_t k_1^a(m_t, K_t, t)^T + \sigma_0(t); \\ K_0 &= K(t_0); \end{aligned} \right\} \quad (17)$$

$$\left. \begin{aligned} \frac{\partial K(t_1, t_2)}{\partial t_2} &= K(t_1, t_2)k_{t_2} k_1^a(m_{t_2}, K_{t_2}, t_2)^T; \\ K(t_1, t_2) &= K_{t_1}. \end{aligned} \right\} \quad (18)$$

Для стационарных ДСтС (2) при условии асимптотической устойчивости матрицы $k_1^a(\bar{m}, \bar{K})$ в основе МСЛ лежат уравнения (11), записанные в виде:

$$a_1(\bar{m}, \bar{K}) = 0; \quad (19)$$

$$k_1^a(\bar{m}, \bar{K})\bar{K} + \bar{K}k_1^a(\bar{m}, \bar{K})^T + \bar{\sigma}_0 = 0; \quad (20)$$

$$\left. \begin{aligned} k_\tau(\tau) &= k_1^a(\bar{m}, \bar{K})k(\tau), \quad k(0) = \bar{K} \quad (\forall \tau > 0), \\ k(\tau) &= k(-\tau)^T \quad (\forall \tau < 0). \end{aligned} \right\} \quad (21)$$

Уравнения (3)–(7) лежат в основе МНА для ДСтС (1), а уравнения (13)–(18) – в основе МСЛ для ДСтС (2). Для определения стационарных СтП согласно МНА служат соотношения (11) и (12), а МСЛ – (13)–(21).

3 Уравнения методов нормальной аппроксимации и статистической линеаризации для эредитарных стохастических систем, приводимых к дифференциальным

Рассмотрим ЭСтС, описываемую интегродифференциальным уравнением Ито следующего вида [7]:

$$\begin{aligned} dX_t &= \left[a(X_t, t) + \int_{t_0}^t a_1(X(\tau), \tau, t) d\tau \right] dt + \\ &+ \left[b(X_t, t) + \int_{t_0}^t b_1(X(\tau), \tau, t) d\tau \right] dW_0 + \\ &+ \int_{R_0^q} \left[c(X_t, t, v) + \int_{t_0}^t c_1(X(\tau), \tau, t, v) d\tau \right] dP^0(t, dv) \quad (22) \end{aligned}$$

с начальным условием $X(t_0) = X_0$. В (22) сохранены обозначения разд. 2.

Функции $a = a(X_t, t)$, $a_1 = a_1(X(\tau), \tau, t)$, $b = b(X_t, t)$, $b_1 = b_1(X(\tau), \tau, t)$, $c = c(X_t, t, v)$ и $c_1 = c_1(X(\tau), \tau, t, v)$ имеют соответственно размерности $p \times 1$, $p \times 1$, $p \times r$, $p \times r$, $p \times 1$ и $p \times 1$ и допускают представления следующего вида:

$$\left. \begin{aligned} a_1 &= A(t, \tau)\varphi(X(\tau), \tau); \\ b_1 &= B(t, \tau)\psi(X(\tau), \tau); \\ c_1 &= C(t, \tau)\chi(X(\tau), \tau, v). \end{aligned} \right\} \quad (23)$$

Здесь эредитарные ядра $A = A(t, \tau) = [A_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$), $B = B(t, \tau) = [B_{il}(t, \tau)]$ ($i = \overline{1, p}$; $l = \overline{1, r}$) и $C = C(t, \tau) = [C_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$) имеют соответственно размерности $p \times p$, $p \times r$ и $p \times p$. Они удовлетворяют следующим условиям физической реализуемости и асимптотического затухания:

$$\left. \begin{aligned} A_{ij}(t, \tau) &= 0; \quad B_{il}(t, \tau) = 0; \\ C_{ij}(t, \tau) &= 0 \quad \forall \tau > t; \end{aligned} \right\} \quad (24)$$

$$\left. \begin{aligned} \int_{-\infty}^{\infty} |A_{ij}(t, \tau)| d\tau < \infty; \quad \int_{-\infty}^{\infty} |B_{il}(t, \tau)| d\tau < \infty; \\ \int_{-\infty}^{\infty} |C_{ij}(t, \tau)| d\tau < \infty. \end{aligned} \right\} \quad (25)$$

В дальнейшем ограничимся случаем, когда эредитарные ядра удовлетворяют линейным операторным уравнениям [2, 4, 7].

Нелинейные в общем случае функции $\varphi = \varphi(X(\tau), \tau)$, $\psi = \psi(X(\tau), \tau)$, $\chi = \chi(X(\tau), \tau, v)$ отражают нелинейные свойства ЭСтС, зависят от $X(\tau)$ и имеют размерности $p \times 1$, $p \times p$, $p \times 1$ соответственно.

Важный класс эредитарных ядер представляют собой сингулярные (вырожденные) ядра, когда имеют место представления:

$$\left. \begin{aligned} A_{ij}(t, \tau) &= A_{ij}^+(t)A_{ij}^-(\tau); \\ B_{il}(t, \tau) &= B_{il}^+(t)B_{il}^-(\tau); \\ C_{ij}(t, \tau) &= C_{ij}^+(t)C_{ij}^-(\tau) \quad (i, l = \overline{1, p}, j = \overline{1, r}). \end{aligned} \right\} \quad (26)$$

В [2, 4, 7] показано, что для дифференцируемых нелинейных функций φ, ψ, χ путем расширения вектора состояния за счет инструментальных переменных, аппроксимируемых линейными операторными уравнениями, определяющими эрмитарные ядра в ЭСтС, (22)–(25) приводятся к ДСтС вида (1) или (2). В случае недифференцируемых нелинейных функций φ, ψ, χ ЭСтС (22)–(25) приводятся к (1) или (2) на основе аппроксимации вырожденными (сингулярными) ядрами [2, 4, 7].

Таким образом, после приведения ЭСтС (22) к ДСтС (1) или (2) можно воспользоваться уравнениями МНА и МСЛ разд. 2.

4 Типовые интегралы методов нормальной аппроксимации и статистической линеаризации

Как следует из уравнений (8)–(10), для МНА необходимо уметь вычислять следующие интегралы:

$$I_0^a = I_0^a(m_t, K_t, t) = a_{11}(m_t, K_t, t) = M_N a(Y_t, t); \quad (27)$$

$$I_1^a = I_1^a(m_t, K_t, t) = a_{21}(m_t, K_t, t) = M_N a(Y_t, t) Y_t^{0T}; \quad (28)$$

$$I_0^\sigma = I_0^\sigma(m_t, K_t, t) = a_{22}(m_t, K_t, t) = M_N \bar{\sigma}(Y_t, t). \quad (29)$$

Для МСЛ достаточно вычислить интеграл (27), причем интеграл I_1^a вычисляется по формуле [3, 5, 6]:

$$k_1^a = k_1^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) I_0^a(m_t, K_t, t) \right]^T.$$

Пример 1. В [1] для типовых степенных, тригонометрических, показательных и кусочно-постоянных нелинейностей $Z_t = \varphi(Y_t, t)$ скалярного и векторного аргумента приведены формулы для интегралов $I_0^\varphi = I_0^\varphi(m_t^y, K_t^y, t)$, а также $k_1^\varphi = k_1^\varphi(m_t^y, K_t^y, t)$.

Замечание. Важно иметь в виду, что уравнения МНА (МСЛ) содержат интегралы I_0^a, I_1^a, I_0^σ в виде соответствующих коэффициентов. Поэтому процедура вычисления интегралов должна быть согласована с методом численного решения обыкновенных дифференциальных уравнений для m_t, K_t и $K(t_1, t_2)$. Эти коэффициенты допускают дифференцирование по m_t и K_t , так как под интегралом стоит сглаживающая нормальная плотность.

5 Сложные конечные и дифференциальные нелинейности

Важный класс сложных конечных нелинейностей (многомерных и векторного аргумента) представляют собой сложные функции вида:

$$\xi = \varphi(X_t, Y_t, t), \quad X_t = \psi(Y_t, t).$$

В этом случае вычисление интегралов (см. разд. 4) проводится по совокупности переменных $[X_t^T Y_t^T]^T$. К таким нелинейностям, например, относятся дробно-рациональные, иррациональные нелинейности, выражаемые специальными функциями, многозначные нелинейности, зависящие от СтП X_t и его производных \dot{X}_t, \ddot{X}_t и др.

Пример 2. Рассмотрим вычисление интегралов (27) и (28) для сложных одномерных иррациональных нелинейностей скалярного аргумента

$$\varphi(Y_t, t) = |Y_t|^{\alpha-1} \operatorname{sgn} Y_t \quad (30)$$

(α — нецелый показатель).

Пользуясь (12) и (15), представим (30) в виде

$$\varphi(Y_t, t) = \varphi_0(m_t, D_t, t) + k_1^\varphi(m_t, D_t, t) Y_t^0.$$

Здесь введены следующие обозначения:

$$\varphi_0(m_t, D_t, t) = \Gamma(\alpha) D_t^{1/2} e^{-\xi^2/4} D_{-\alpha}(\xi);$$

$$k_1^a(m_t, D_t, t) = \frac{\partial \varphi_0(m_t, D_t, t)}{\partial m_t},$$

где $\Gamma(\alpha)$ — гамма-функция, $\xi = m_t/\sqrt{D_t}$ — отношение «сигнал–шум»; $D_{-\alpha}(\xi)$ — функция параболического цилиндра [8]. При вычислении были учтены следующие соотношения [8, 9]:

$$\int_0^\infty x^{\alpha-1} e^{-\beta x^2 - \gamma x} dx = (2\beta)^{-\alpha/2} \Gamma(\alpha) \exp\left(\frac{\gamma^2}{8\beta}\right) D_{-\alpha}\left(\frac{\gamma}{\sqrt{2\beta}}\right); \quad (31)$$

$$\frac{dD_\rho(\xi)}{d\xi} = -\frac{\xi}{2} D_\rho(\xi) - \rho D_{\rho-1}(\xi) = \frac{\xi}{2} D_\rho(\xi) - D_{\rho+1}(\xi) \quad (\operatorname{Re} \beta > 0, \operatorname{Re} \alpha > 0, \rho = -\alpha). \quad (32)$$

Соотношения (31) и (32) могут быть использованы также для вычисления интегралов (29).

Замечание. Для вычисления интегралов I_0^a, I_1^a и I_0^σ применительно к типовым иррациональным

нелинейностям вида $|Y_t|^{\alpha-1} e^{\delta Y_t}$, $|Y_t|^{\alpha-1} \cos \omega Y_t$, $|Y_t|^{\alpha-1} \sin \omega Y_t$ и более общим нелинейностям вида

$$\varphi(Y_t, t) = \Phi^\varphi \left(|Y_t|^{\alpha-1}, t \right)$$

можно рекомендовать известные численные методы вычисления функций на ЭВМ [9].

Пример 3. Для нелинейной дробно-рациональной функции

$$\varphi(Y_t, t) = \frac{a}{(b + Y_t)^2}$$

имеем

$$\varphi_0(m_t, D_t, t) = ab^{-2} [1 + \chi(m_t, D_t, t)];$$

$$k_1^\varphi(m_t, D_t, t) = ab^{-2} \frac{\partial \chi(m_t, D_t, t)}{\partial m_t}.$$

Здесь

$$\begin{aligned} \chi(m_t, D_t, t) &= \\ &= \sum_{n=1}^{\infty} \sum_{l=0}^{E(n/2)} \frac{(-1)^n (n+1)n!}{(n-2l)!(2l)!} b^{-n} m_t^n \left(\frac{D_t}{2m_t^2} \right)^l, \end{aligned}$$

где $E(n/2)$ — целая часть $n/2$; $a = a(t)$; $b = b(t)$.

6 Сложные интегральные нелинейности

Пусть сначала векторно-матричная нелинейность имеет эрeditary характер, т. е.

$$\underline{\varphi}(Y_t, t) = \int_{t_0}^t A(t, \tau) \varphi(Y(\tau), \tau) d\tau. \quad (33)$$

Тогда, как показано в [2, 4, 7], следует соответствующие интегродифференциальные соотношения путем введения инструментальных переменных привести к дифференциальным соотношениям. Для дифференцируемых функций φ и асимптотически устойчивых ядер $A(t, \tau)$ зависимость (26) имеет следующий дифференциальный вид:

$$F^A(t, D) \underline{\varphi}(Y_t, t) = H^A(t, D) \varphi(Y_t, t).$$

Здесь $F^A(t, D)$ и $H^A(t, D)$ — линейные дифференциальные операторы ($D = d/dt$).

Для недифференцируемых функций φ и асимптотически устойчивых сингулярных ядер (26) используются соотношения:

$$\underline{\varphi}(Y_t, t) = A^+ Z, \quad \dot{Z} = A^- \varphi, \quad Z(t_0) = 0.$$

Многочисленные примеры аналитического моделирования ЭСтС можно найти в [1–3, 5, 7, 10, 11].

Как отмечалось в [5], часто наряду с интегральными нелинейностями (33) рассматривают нелинейности вида:

$$Z_s = \sum_{\rho=1}^R \mathcal{A}_\rho \varphi_\rho(Y_{t_1}, \dots, Y_{t_r}),$$

где $\mathcal{A}_1, \dots, \mathcal{A}_R$ — произвольные линейные операторы, действующие над функциями r переменных t_1, \dots, t_r ; $\varphi_\rho = \varphi_\rho(Y_{t_1}, \dots, Y_{t_r})$ — линейные функции отмеченных переменных. Такие нелинейности называются приводимыми к линейным. Важным частным случаем (33) являются интегральные нелинейности вида:

$$Z_s = \int_T \varphi(Y_t, t, s) dt;$$

$$Z_s = \int_T \dots \int_T \varphi(Y_{t_1}, \dots, Y_{t_r}; t_1, \dots, t_r, s) dt_1 \dots dt_r,$$

В этом случае используется МСЛ по совокупности переменных Y_{t_1}, \dots, Y_{t_r} .

7 Заключение

Разработаны методы и алгоритмы МНА и МСЛ для ДСтС и ЭСтС, приводимых к ДСтС со сложными конечными, дробно-рациональными, иррациональными, а также дифференциальными и интегральными нелинейностями. Приведены примеры.

Результаты допускают обобщение на случай ДСтС и ЭСтС со стохастическими нелинейностями, заданными каноническими разложениями и интегральными каноническими представлениями [1, 5, 10].

Литература

1. Сеницын И. Н., Сеницын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
2. Сеницын И. Н. Stochastic hereditary control systems // Проблемы управления и теории информации, 1986. Т. 15. № 4. С. 287–298.
3. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. [Англ. пер. Stochastic differential systems. Analysis and filtering. — Chichester, New York: John Wiley, 1987. 549 p.]
4. Сеницын И. Н. Конечномерные распределения процессов в стохастических интегральных и интегродифференциальных системах // Preprints of the 2nd IFAC Symposium on Stochastic Control. — Vilnius: Pergamon Press, 1987. Vol. 1. P. 144–153.
5. Пугачев В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с. [Англ. пер.]

- Stochastic systems. Theory and applications. — Singapore: World Scientific, 2001. 908 p.].
6. Синицын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
 7. Синицын И. Н. Анализ и моделирование распределений в эредитарных стохастических системах // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 2–11.
 8. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. — М.: ГИФМЛ, 1963. 1100 с.
 9. Попов Б. А., Теслер Г. С. Вычисление функций на ЭВМ: Справочник. — Киев: Наукова Думка, 1984. 599 с.
 10. Синицын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009. 768 с.
 11. Синицын И. Н., Синицын В. И., Корепанов Э. Р., Белосусов В. В., Сергеев И. В., Басилашвили Д. А. Опыт моделирования эредитарных стохастических систем // Кибернетика и высокие технологии XXI века: Сб. докл. XIII Междунар. науч.-технич. конф. — Воронеж: Саквее, 2012. Т. 2. С. 346–357.

Поступила в редакцию 05.05.14

ANALYTICAL MODELING OF NORMAL PROCESSES IN STOCHASTIC SYSTEMS WITH COMPLEX NONLINEARITIES

I. N. Sinitsyn and V. I. Sinitsyn

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Differential stochastic systems (DStS) with Wiener and Poisson noises and complex finite, differential, and integral nonlinearities and hereditary StS reducible to DStS are considered. Equations and algorithms of analytical modeling based on the normal approximation method (NAM) and the statistical linearization method (SLM) are given. The case of complex continuous and discontinuous nonlinearities of scalar and vector arguments is considered. Special attention is paid to NAM (SLM) typical integrals for finite rational and irrational nonlinear and integral scalar and vector nonlinear functions. The general case of integral nonlinearities reducible to linear is considered. Test examples are given.

Keywords: analytical modeling; complex finite differential and integral nonlinearities; complex irrational nonlinearities differential stochastic system with Wiener and Poisson noises; method of normal approximation; method of statistical linearization; hereditary stochastic systems reducible to differential

DOI: 10.14375/19922264140302

References

1. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Leksii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 488 p.
2. Sinitsyn, I. N. 1986. Stochastic hereditary control systems. *Problems Control Inform. Theory* 15(4):287–298.
3. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York: John Wiley. 549 p.
4. Sinitsyn, I. N. 1987. Konechnomernye raspredeleniya protsessov v stokhasticheskikh integral'nykh i integrodifferentsial'nykh sistemakh [Finite dimensional distributions of processes in stochastic integral and integrodifferential systems]. *2nd Symposium (International) IFAC on Stochastic Control Preprints*. Vilnius: Pergamon Press. 1:144–153.
5. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
6. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametric statistical and analytical modeling of distributions in stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.
7. Sinitsyn, I. N. 2014. Analiz i modelirovanie raspredeleniy v ereditarnykh stokhasticheskikh sistemakh [Analysis and modeling of distributions in hereditary stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):2–11.
8. Gradshteyn, I. S., and I. M. Ryzhik. 1963. *Tablitsy integralov, summ, ryadov i proizvedeniy* [Tables of integrals, sums, series, and products]. Moscow: GIFML. 1100 p.

9. Popov, B. A., and G. S. Tesler. 1984. *Vychislenie funktsiy na EVM*. Spravochnik [Computing of functions]. Kiev: Naukova Dumka. 599 p.
10. Sinitsyn, I. N. 2009. *Kanonicheskie predstavleniya sluchaynykh funktsiy i ikh primenenie v zadachakh komp'yuternoy podderzhki nauchnykh issledovaniy* [Canonical expansions of random functions and its application to scientific computer-aided support]. Moscow: TORUS PRESS. 768 p.
11. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, V. V. Belousov, I. V. Sergeev, and D. A. Basilashvili. 2012. Opyt modelirovaniya ereditarnykh stokhasticheskikh sistem [Experience of modeling in hereditary stochastic systems]. *Kibernetika i Vysokie Tekhnologii XXI Veka: Sbornik dokladov XIII Mezhdunar. nauch.-tekhnich. konf.* [Cybernetics and High Technologies of the XXI Century: Materials of XIII Scientific and Technological Conference (International)]. Voronezh: Sakvoee. 2:346–357.

Received May 05, 2014

Contributors

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

Sinitsyn Vladimir I. (b. 1968) — Doctor of Science in physics and mathematics, associate professor, Head of Department, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VSinitsin@ipiran.ru

ОБ ОЦЕНКАХ СКОРОСТИ СХОДИМОСТИ И УСТОЙЧИВОСТИ ДЛЯ НЕКОТОРЫХ МОДЕЛЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ*

А. И. Зейфман¹, А. В. Коротышева², К. М. Киселева³, В. Ю. Королев⁴, С. Я. Шоргин⁵

Аннотация: Рассматривается некоторое обобщение известной модели Эрланга с потерями, а именно: изучается класс марковских моделей систем обслуживания, в которых допускается одновременное поступление группы требований (ограниченное максимальным общим количеством требований) и предусмотрено групповое обслуживание. Установлен критерий слабой эргодичности для процесса, описывающего число требований в такой системе, получены оценки скорости сходимости и устойчивости. Исследование опирается на общий подход, разработанный в предыдущих работах авторов для неоднородных марковских систем с групповым поступлением и обслуживанием требований. Рассмотрены также конкретные модели с периодическими интенсивностями при разном максимальном размере группы поступающих требований, строятся основные предельные характеристики этих моделей и выясняется влияние максимально допустимого размера группы одновременно поступающих требований на предельное среднее для числа требований в системе и предельную вероятность отсутствия требований в системе.

Ключевые слова: нестационарная марковская система обслуживания; модель Эрланга; групповое поступление и обслуживание требований; эргодичность; устойчивость; оценки

DOI: 10.14375/19922264140303

1 Введение

Стационарная и нестационарная модели Эрланга для системы с потерями изучались во многих работах (см., например, [1–10]). По-видимому, в большой степени это связано со сравнительной простотой исследования и удобством применения этой модели. В [11] была рассмотрена аналогичная модель, в которой допускается групповое обслуживание требований, получен критерий слабой эргодичности для такой модели и оценки скорости сходимости, а в [12] получены первые оценки устойчивости этой модели по отношению к малым возмущениям интенсивностей поступления и обслуживания требований.

В настоящей заметке рассматривается некоторый более общий класс моделей систем обслуживания, в которых допускается и групповое поступление требований, установлен критерий слабой эргодичности для них, получены оценки скорости сходимости и устойчивости. Исследование опирается на общий подход, разработанный в предыдущих работах авторов для некоторых клас-

сов неоднородных марковских систем с групповым поступлением и обслуживанием требований [12–14].

Рассмотрим систему обслуживания «с потерями», в которой общее количество требований не превосходит S , максимальный размер группы поступающих требований равен $N \leq S$, интенсивность поступления группы $k \leq N$ требований есть $\lambda_k(t) = \lambda(t)/k$. Интенсивность обслуживания группы $k \leq S$ имеющихся в системе требований есть $\mu_k(t) = \mu(t)/k$. Пусть $X = X(t)$, $t \geq 0$, – число требований в системе обслуживания, это неоднородная марковская цепь с непрерывным временем и пространством состояний $E = \{0, 1, \dots, S\}$. «Базовые» интенсивности поступления и обслуживания $\lambda(t)$ и $\mu(t)$ предполагаются локально интегрируемыми на $[0, \infty)$ функциями времени t .

Тогда для описания вероятностной динамики процесса получаем прямую систему Колмогорова в виде:

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}(t),$$

* Исследование выполнено за счет гранта Российского научного фонда (проект № 14-11-00397).

¹ Институт проблем информатики Российской академии наук; Вологодский государственный университет; Институт социально-экономического развития территорий Российской академии наук, a_zeifman@mail.ru

² Вологодский государственный университет, a_korotysheva@mail.ru

³ Вологодский государственный университет, a_zeifman@mail.ru

⁴ Институт проблем информатики Российской академии наук; Московский государственный университет им. М. В. Ломоносова, vkorolev@cs.msu.su

⁵ Институт проблем информатики Российской академии наук, SShorgin@ipiran.ru

где

$$A(t) = \begin{pmatrix} a_{00}(t) & \mu_1(t) & \mu_2(t) & \mu_3(t) & \cdots & \mu_S(t) \\ \lambda_1(t) & a_{11}(t) & \mu_1(t) & \mu_2(t) & \cdots & \mu_{S-1}(t) \\ \lambda_2(t) & \lambda_1(t) & a_{22}(t) & \mu_1(t) & \cdots & \mu_{S-2}(t) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_S(t) & \lambda_{S-1}(t) & \lambda_{S-2}(t) & \lambda_{S-3}(t) & \cdots & a_{SS}(t) \end{pmatrix},$$

причем

$$a_{ii}(t) = -\sum_{k=1}^i \mu_k(t) - \sum_{k=1}^{S-i} \lambda_k(t); \quad \lambda_k(t) = 0$$

при $k > N$.

Обозначим через

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

переходные вероятности процесса $X = X(t)$, а через

$$p_i(t) = \Pr \{X(t) = i\}$$

его вероятности состояний.

Будем обозначать через $\|\bullet\|$ l_1 -норму вектора и матрицы, т. е. $\|\mathbf{x}\| = \sum |x_i|$ и $\|B\| = \max_j \sum_i |b_{ij}|$

при $B = (b_{ij})_{i,j=0}^S$, а через Ω — множество всех стохастических векторов, т. е. множество векторов с неотрицательными координатами и единичной l_1 -нормой.

Через

$$E(t, k) = E \{X(t) | X(0) = k\}$$

будем далее обозначать математическое ожидание процесса (среднее число требований) в момент t при условии, что в нулевой момент времени он находится в состоянии k .

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

2 Слабая эргодичность и скорость сходимости

Теорема 1. *Процесс $X(t)$, описывающий число требований в рассматриваемой системе, слабо эргодичен*

при любом N ($1 \leq N \leq S$) тогда и только тогда, когда выполнено условие:

$$\int_0^\infty (\lambda(t) + \mu(t)) dt = +\infty. \quad (1)$$

Доказательство. Отметим, прежде всего, что если (1) не выполнено, то

$$\|A(t)\| = 2 \max |a_{ii}(t)| \leq \\ \leq 2 (\lambda(t) + \mu(t)) \sum_{k=1}^S \frac{1}{k} \leq 2 (1 + \ln S) (\lambda(t) + \mu(t))$$

и, значит,

$$\int_0^\infty \|A(t)\| dt < +\infty.$$

А тогда в соответствии с теоремой 3.3 из [15] $X(t)$ слабо эргодичным быть не может.

Пусть (1) выполняется. Пользуясь стандартным приемом и полагая $p_0 = 1 - \sum_{1 \leq i \leq S} p_i$, получаем:

$$\frac{dz}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t),$$

где $\mathbf{f}(t) = (\lambda_1, \lambda_2, \dots, \lambda_S)^T$;

$$B(t) = \begin{pmatrix} a_{11} - \lambda_1 & \mu_1 - \lambda_1 & \mu_2 - \lambda_1 & \mu_3 - \lambda_1 & \cdots \\ \lambda_1 - \lambda_2 & a_{22} - \lambda_2 & \mu_1 - \lambda_2 & \mu_2 - \lambda_2 & \cdots \\ \lambda_2 - \lambda_3 & \lambda_1 - \lambda_3 & a_{33} - \lambda_3 & \mu_1 - \lambda_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_{S-1} - \lambda_S & \lambda_{S-2} - \lambda_S & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \mu_{S-1} - \lambda_1 \\ \cdots & \cdots & \cdots & \mu_{S-2} - \lambda_2 \\ \cdots & \cdots & \cdots & \mu_{S-3} - \lambda_3 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \lambda_2 - \lambda_S & \lambda_1 - \lambda_S & a_{SS} - \lambda_S \end{pmatrix}.$$

Рассмотрим треугольную матрицу

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots & d_1 \\ 0 & d_2 & d_2 & \cdots & d_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & d_S \end{pmatrix}$$

с положительными элементами d_i и соответствующую норму $\|\mathbf{z}\|_D = \|D\mathbf{z}\|_1$. Тогда получим равенство:

$$DB(t)D^{-1} = \begin{pmatrix} a_{11} - \lambda_S & (\mu_1 - \mu_2) \frac{d_1}{d_2} & (\mu_2 - \mu_3) \frac{d_1}{d_3} & \cdots & (\mu_{S-1} - \mu_S) \frac{d_1}{d_S} \\ (\lambda_1 - \lambda_S) \frac{d_2}{d_1} & a_{22} - \lambda_{S-1} & (\mu_1 - \mu_3) \frac{d_2}{d_3} & \cdots & (\mu_{S-2} - \mu_S) \frac{d_2}{d_S} \\ (\lambda_2 - \lambda_S) \frac{d_3}{d_1} & (\lambda_1 - \lambda_{S-1}) \frac{d_3}{d_2} & a_{33} - \lambda_{S-2} & \cdots & (\mu_{S-3} - \mu_S) \frac{d_3}{d_S} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (\lambda_{S-1} - \lambda_S) \frac{d_S}{d_1} & (\lambda_{S-2} - \lambda_{S-1}) \frac{d_S}{d_2} & (\lambda_{S-3} - \lambda_{S-2}) \frac{d_S}{d_3} & \cdots & a_{SS} - \lambda_1 \end{pmatrix},$$

дающее соответствующее выражение для логарифмической нормы $B(t)$ (см. подробное обсуждение в [8, 16, 17]):

$$\begin{aligned} \gamma(B(t))_{1D} &= \gamma(DB(t)D^{-1}) = \\ &= \max_{1 \leq i \leq S} (a_{ii}(t) - \lambda_{S+1-i}(t) + \\ &+ \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \mu_i(t)) \frac{d_k}{d_i} + \\ &+ \sum_{k=1}^{S-i} (\lambda_k(t) - \lambda_{S+1-i}) \frac{d_{k+i}}{d_i}). \end{aligned} \quad (2)$$

Пусть вначале

$$\int_0^{+\infty} \mu(t) dt = +\infty. \quad (3)$$

Положим все $d_i = 1$. Тогда справедлива следующая оценка:

$$\begin{aligned} \gamma(B(t))_{1D} &\leq \\ &\leq \max_{1 \leq i \leq S} \left(-\sum_{k=1}^i \mu_i(t) + \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \mu_i(t)) \right) = \\ &= -\min_{1 \leq i \leq S} (k\mu_k(t)) = -\mu(t). \end{aligned}$$

Имеем при этом

$$\|D\| = \sum_{i=1}^S d_i = S; \quad \|D^{-1}\| = 2 \max_{1 \leq k \leq S} \left(\frac{1}{d_k} \right) = 2,$$

а значит,

$$\begin{aligned} \|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| &\leq 2\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\| = \\ &= 2\|D(\mathbf{z}^*(t) - \mathbf{z}^{**}(t))D^{-1}\| \leq \\ &\leq 4\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\|_{1D} \leq \\ &\leq 4e^{-\int_s^t \mu(u) du} \|\mathbf{z}^*(s) - \mathbf{z}^{**}(s)\|_{1D} \leq \\ &\leq 4Se^{-\int_s^t \mu(u) du} \|\mathbf{z}^*(s) - \mathbf{z}^{**}(s)\| \leq \\ &\leq 4Se^{-\int_s^t \mu(u) du} \|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\| \leq 8Se^{-\int_s^t \mu(u) du} \end{aligned}$$

для любых начальных условий $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$ и любых $s, t, 0 \leq s \leq t$. Отсюда вытекает слабая эргодичность $X(t)$.

Пусть теперь

$$\int_0^{\infty} \lambda(t) dt = +\infty. \quad (4)$$

Положим $d_k = 1/k$. Тогда из (2) получаем такую оценку логарифмической нормы $B(t)$:

$$\gamma(B(t))_{1D} \leq -\min_{1 \leq i \leq S} \left(\mu_i(t) + \frac{1}{i} \lambda_1(t) \right) \leq -\frac{\lambda(t)}{S}.$$

Теперь

$$\begin{aligned} \|D\| &= \sum_{i=1}^S d_i \leq 1 + \ln S; \\ \|D^{-1}\| &= 2 \max \left(\frac{1}{d_k} \right) = 2S. \end{aligned}$$

Значит, получаем слабую эргодичность процесса и следующую оценку:

$$\begin{aligned} \|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| &\leq 2\|D^{-1}D(\mathbf{z}^*(t) - \mathbf{z}^{**}(t))\| \leq \\ &\leq 8S(1 + \ln S) e^{-\int_s^t \lambda(\tau) d\tau} \end{aligned}$$

при любых начальных условиях $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$ и любых $s, t, 0 \leq s \leq t$.

Следствие 1. Если (1) справедливо, то процесс $X(t)$ слабо эргодичен, имеет предельное среднее $\phi(t)$ и справедливы следующие оценки скорости сходимости:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 8Se^{-\int_0^t \mu(u) du}$$

при любых начальных условиях,

$$|E(t, k) - \phi(t)| \leq 8S^2 e^{-\int_0^t \mu(u) du}$$

при любом k , если выполнено (3), а при выполнении (4) соответственно:

$$\|p^*(t) - p^{**}(t)\| \leq 8S(1 + \ln S) e^{-(1/S) \int_0^t \lambda(\tau) d\tau};$$

$$|E(t, k) - \phi(t)| \leq 8S^2(1 + \ln S) e^{-(1/S) \int_0^t \lambda(\tau) d\tau}.$$

Используя результаты теоремы 1 и следствия 1, а также общий подход, описанный в работах [13, 14, 18], можно получить соответствующие оценки устойчивости процесса, описывающего число требований в системе. Ограничимся здесь одной из возможных формулировок. Пусть $\bar{X} = \bar{X}(t)$ — число требований для «возмущенной» системы обслуживания. Соответствующие его характеристики будем обозначать теми же буквами с чертой сверху. Предположим, что его инфинитезимальная матрица имеет произвольную структуру (т. е. интенсивности поступления и обслуживания требований произвольны) и при этом для всех $t \geq 0$ выполнено условие малости $\|A(t) - \bar{A}(t)\| \leq \varepsilon$. Следующее утверждение вытекает из теоремы 1 и оценок (32), (33) из [13].

Следствие 2. Пусть выполнено (3) и вдобавок при некоторых положительных M, α и всех $0 \leq s \leq t$ справедливо неравенство:

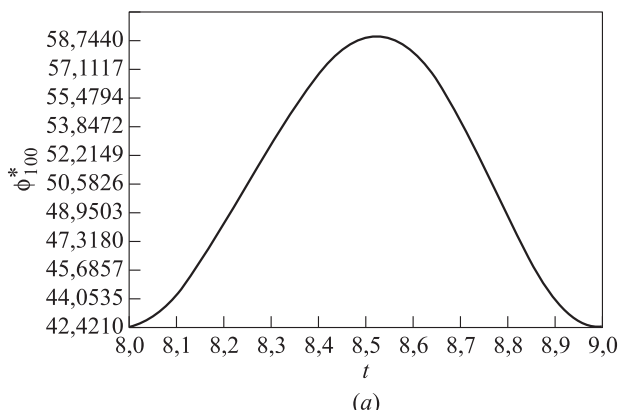
$$e^{-\int_s^t \mu(u) du} \leq M e^{-\alpha(t-s)}. \quad (5)$$

Тогда при любых начальных условиях $p(0)$ и $\bar{p}(0)$ соответственно справедливы неравенства:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| \leq \frac{\varepsilon(1 + \ln 4SM)}{\alpha};$$

$$\limsup_{t \rightarrow \infty} |E_p(t) - \bar{E}_{\bar{p}(t)}| \leq \frac{\varepsilon S(1 + \ln 4SM)}{\alpha}.$$

Отметим, что условие (5) заведомо выполнено, если интенсивность обслуживания 1-периодична, при этом



$$\alpha = \int_0^1 \mu(t) dt,$$

$$M = e^{\max_{|t-s| \leq 1} \int_s^t \mu(u) du}.$$

3 Примеры

Рассмотрим теперь несколько моделей описываемой системы обслуживания с периодически-инфинитезимальными характеристиками и для простоты вычислений, чтобы не было необходимости рассматривать усеченные процессы, как это делается обычно (см. [13, 17]), будем полагать $S = 10^2$.

1. Пусть $N = S$, а интенсивности поступления и обслуживания требований определяются функциями $\lambda^*(t) = 3 + \sin 2\pi t$ и $\mu(t) = 3 + \cos 2\pi t$ соответственно. Тогда выполнено условие (3), существует предельный 1-периодический режим, скажем $p^*(t)$, и соответствующее предельное среднее $\phi^*(t)$ и справедливы оценки скорости сходимости к ним:

$$\|p^*(t) - p^{**}(t)\| \leq 10^3 e^{-3t}, \quad |E(t, s) - \phi(t)| \leq 10^5 e^{-3t},$$

откуда вытекает, что при $t \geq 8$ предельные характеристики получаются с точностью выше 10^{-4} . Их можно найти, решая прямую систему Колмогорова с начальным условием e_0 на отрезке $[0; 9]$ и взяв затем полученные функции на отрезке $[8; 9]$. Отметим, что как в этом, так и в следующих примерах имеем $\alpha = 3, M \leq 2$ и соответствующие оценки устойчивости, основанные на следствии 2, выглядят следующим образом:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| \leq \frac{\varepsilon(1 + \ln 800)}{3} \leq 3\varepsilon;$$

$$\limsup_{t \rightarrow \infty} |E_p(t) - \bar{E}_{\bar{p}(t)}| \leq \frac{\varepsilon 10^2(1 + \ln 800)}{3} \leq 300\varepsilon.$$

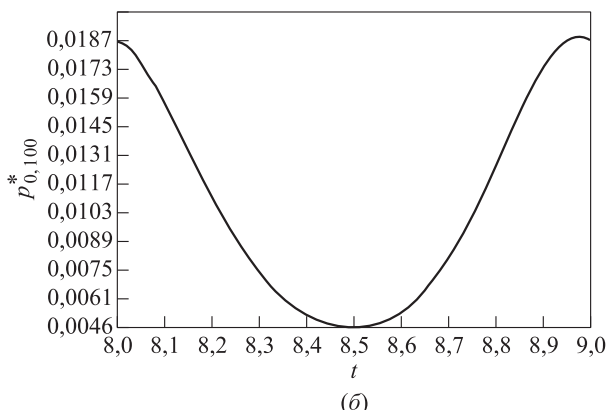


Рис. 1 Приближенные значения предельного среднего $\phi_{100}^*(t)$ (а) и предельной «пустой очереди» $p_{0,100}^*(t)$ (б)

На рис. 1 приведены приближенные графики двух предельных характеристик: предельного среднего (математического ожидания) $\phi_{100}^*(t)$ и предельной вероятности отсутствия требований в системе $p_{0,100}^*(t) = \Pr(X(t) = 0)$ при $N = S = 100$.

2. Пусть теперь $N = S/2 = 50$, интенсивности обслуживания требований предполагаются такими же, а интенсивности поступления требований подберем так, чтобы неизменной осталась «средняя интенсивность поступления требований», которая получается как величина $\lambda_1(t) + 2\lambda_2(t) + \dots + N\lambda_N(t) = N\lambda(t)$, т.е. должно выполняться равенство $N\lambda(t) = S(3 + \sin 2\pi t)$, откуда в этой ситуации $\lambda(t) = 2(3 + \sin 2\pi t)$. Поскольку полученные оценки не зависят от функции, описывающей интенсивность поступления требований, справедливы те же оценки. На рис. 2 приведены приближенные графики двух предельных характеристик: предельного среднего (математического ожидания) $\phi_{50}^*(t)$ и предельной вероят-

ности отсутствия требований в системе $p_{0,50}^*(t) = \Pr(X(t) = 0)$ при $N = S/2 = 50$.

3. Возьмем теперь убывающие значения $N = 40, 20, 10$ и 1 с интенсивностями поступления требований $\lambda(t) = 2, 5(3 + \sin 2\pi t), \lambda(t) = 5(3 + \sin 2\pi t), \lambda(t) = 10(3 + \sin 2\pi t)$ и $\lambda(t) = 10^2(3 + \sin 2\pi t)$ соответственно. Поскольку полученные оценки не зависят от функции, описывающей интенсивность поступления требований, справедливы те же оценки. На рис. 3–6 для каждого из этих случаев приведены приближенные графики двух предельных характеристик: предельного среднего (математического ожидания) $\phi_N^*(t)$ и предельной вероятности отсутствия требований в системе $p_{0,N}^*(t) = \Pr(X(t) = 0)$ при $N = 40, 20, 10$ и 1 соответственно.

Замечание. Полученные результаты показывают, что увеличение одноmomentно допустимого поступающего количества требований при той же «средней интенсивности поступления требований» при-

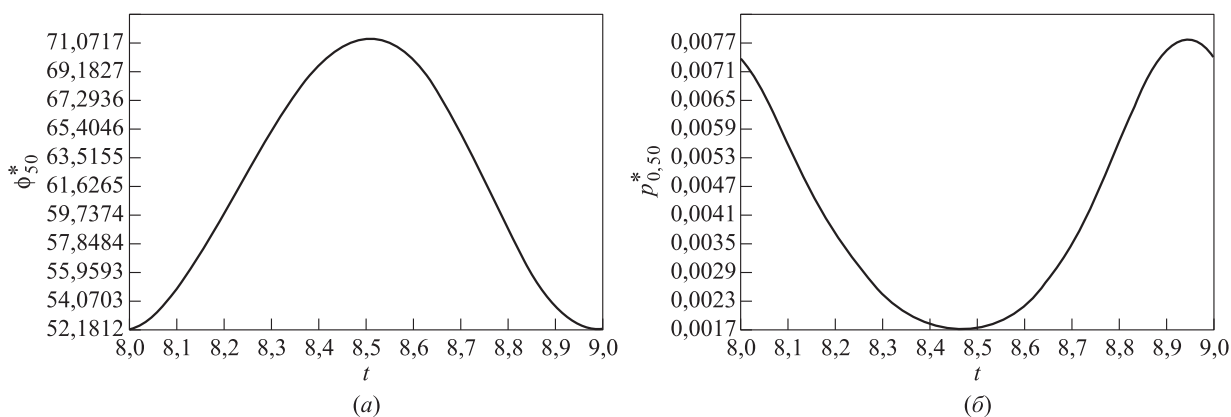


Рис. 2 Приближенные значения предельного среднего $\phi_{50}^*(t)$ (а) и предельной «пустой очереди» $p_{0,50}^*(t)$ (б)

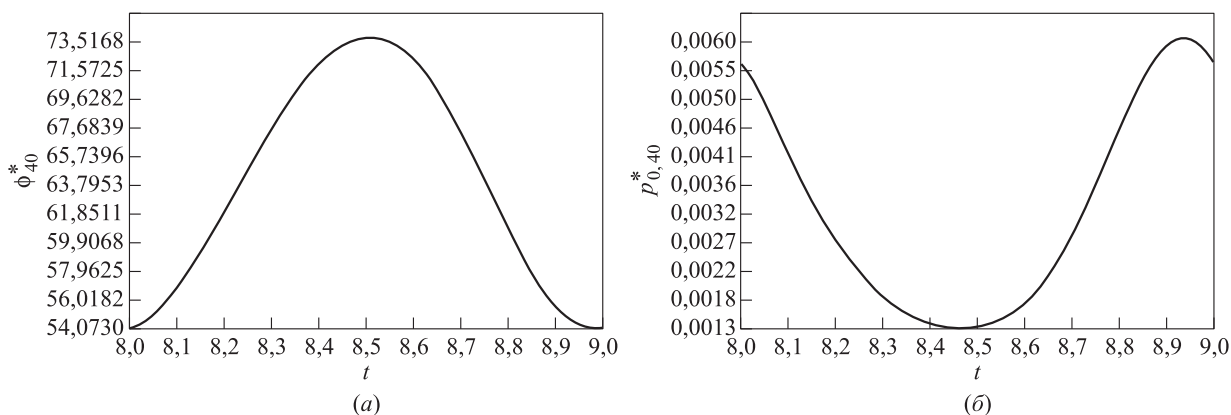


Рис. 3 Приближенные значения предельного среднего $\phi_{40}^*(t)$ (а) и предельной «пустой очереди» $p_{0,40}^*(t)$ (б)

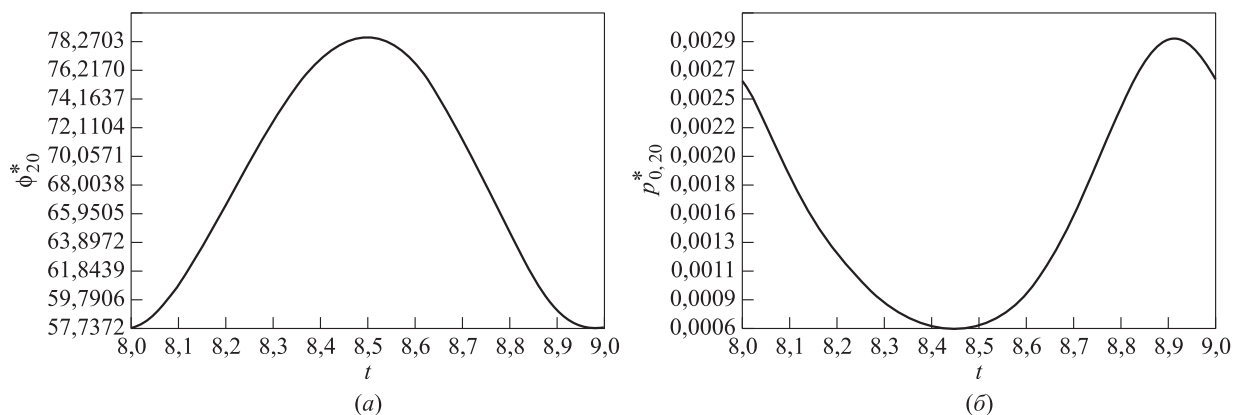


Рис. 4 Приближенные значения предельного среднего $\phi_{20}^*(t)$ (а) и предельной «пустой очереди» $p_{0,20}^*(t)$ (б)

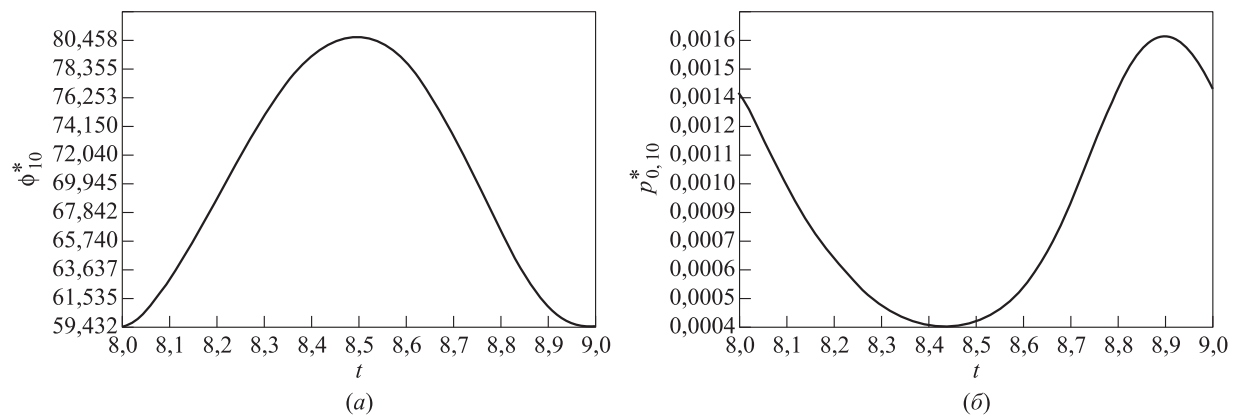


Рис. 5 Приближенные значения предельного среднего $\phi_{10}^*(t)$ (а) и предельной «пустой очереди» $p_{0,10}^*(t)$ (б)

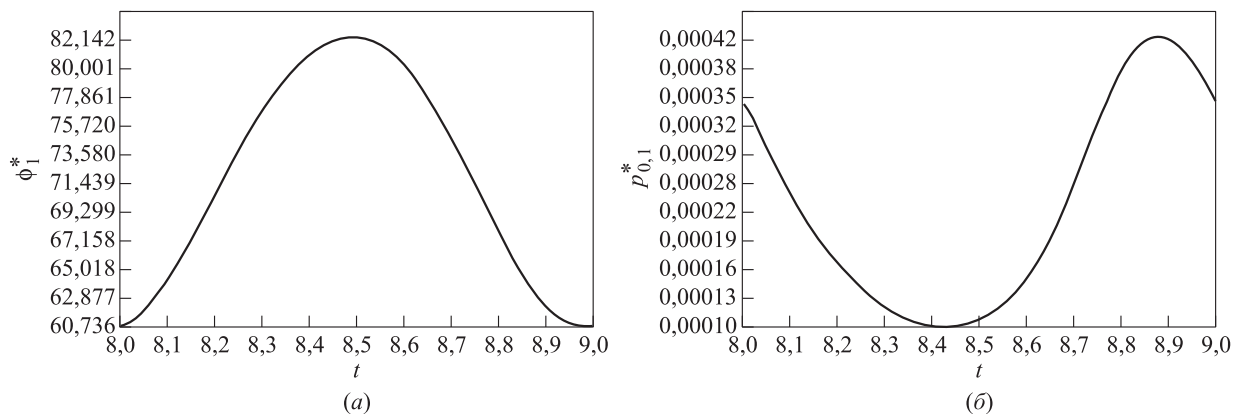


Рис. 6 Приближенные значения предельного среднего $\phi_1^*(t)$ (а) и предельной «пустой очереди» $p_{0,1}^*(t)$ (б)

водит к сокращению предельной средней длины очереди и увеличению вероятности отсутствия требований в системе.

Литература

1. Erlang A. K. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler // Elektroteknikerens, 1917. Vol. 13. P. 5–13.
2. Гнеденко Б. В., Макаров И. П. Свойства решений задачи с потерями в случае периодических интенсивностей // Дифф. уравнения, 1971. Т. 7. С. 1696–1698.
3. Зейфман А. И. Некоторые свойства системы с потерями в случае переменных интенсивностей // Автоматика и телемеханика, 1989. № 1. С. 107–113.
4. Kijima M. On the largest negative eigenvalue of the infinitesimal generator associated with $M/M/n/n$ queues // Oper. Res. Lett., 1990. Vol. 9. P. 59–64.
5. Massey W. A., Whitt W. On analysis of the modified offered-load approximation for the nonstationary Erlang loss model // Ann. Appl. Probab., 1994. Vol. 4. P. 1145–1160.
6. Fricker C., Robert P., Tibi D. On the rate of convergence of Erlang's model // J. Appl. Probab., 1999. Vol. 36. P. 1167–1184.
7. Voit M. A note of the rate of convergence to equilibrium for Erlang's model in the subcritical case // J. Appl. Probab., 2000. Vol. 37. P. 918–923.
8. Granovsky B., Zeifman A. I. Nonstationary queues: Estimation of the rate of convergence // Queueing Syst., 2004. Vol. 46. P. 363–388.
9. Van Doorn E. A., Zeifman A. I. On the speed of convergence to stationarity of the Erlang loss system // Queueing Syst., 2009. Vol. 63. P. 241–252.
10. Зейфман А. И. О нестационарной модели Эрланга // Автоматика и телемеханика, 2009. № 12. С. 71–80.
11. Zeifman A. I., Korotysheva A., Satin Ya., Shilova G., Panfilova T. On a queueing model with group services // Modern probabilistic methods for analysis of telecommunication network / Eds. A. Dudin, V. Klimenok, G. Tsarenkov, and S. Dudin. Communications in computer and information science ser. Vol. 356. P. 198–205.
12. Zeifman A. I., Korotysheva A., Shilova G., Korolev V. On perturbation bounds for a queueing model with group services. Preprint, 2014.
13. Zeifman A. I., Korotysheva A., Satin Ya., Korolev V., Benning V. Perturbation bounds and truncations for a class of Markovian queues // Queueing Syst., 2014. Vol. 76. P. 205–221.
14. Zeifman A. I., Korolev V. Y. On perturbation bounds for continuous-time Markov chains // Stat. Probab. Lett., 2014. Vol. 88. P. 66–72.
15. Zeifman A. I., Isaacson D. On strong ergodicity for nonhomogeneous continuous-time Markov chains // Stoch. Proc. Appl., 1994. Vol. 50. P. 263–273.
16. Zeifman A. I. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // Stoch. Proc. Appl., 1995. Vol. 59. P. 157–173.
17. Zeifman A., Leorato S., Orsingher E., Satin Ya., Shilova G. Some universal limits for nonhomogeneous birth and death processes // Queueing Syst., 2006. Vol. 52. P. 139–151.
18. Зейфман А. И., Королев В. Ю., Коротышева А. В., Шоргин С. Я. Общие оценки устойчивости для нестационарных марковских цепей с непрерывным временем // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 106–117.

Поступила в редакцию 12.08.14

ON THE BOUNDS OF THE RATE OF CONVERGENCE AND STABILITY FOR SOME QUEUEING MODELS

A. I. Zeifman^{1,2,3}, A. V. Korotysheva², K. M. Kiseleva², V. Yu. Korolev^{1,4}, and S. Ya. Shorgin¹

¹Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

³Institute of Socio-Economic Development of Territories, Russian Academy of Sciences, 56A Gorkogo Str., Vologda 160014, Russian Federation

⁴Department 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 generalization of the famous Erlang loss system has been considered, namely, a class of Markovian queueing systems with possible simultaneous arrivals and group services has been studied. Necessary and sufficient conditions of weak ergodicity have been obtained for the respective queue-length process and explicit bounds on

the rate of convergence and stability have been found. The research is based on the general approach developed in the authors' previous studies for nonhomogeneous Markov systems with batch arrival and service requirements. Also, specific models with periodic intensities and different maximum size of number of arrival customers are discussed. The main limiting characteristics of these models have been computed and the effect of the maximum size of the group of arrival customers on the limiting characteristics of the queue has been studied.

Keywords: nonstationary Markovian queue; Erlang model; batch arrivals and group services; ergodicity; stability; bounds

DOI: 10.14375/19922264140303

Acknowledgments

This work was financially supported by the Russian Science Foundation (grant No. 14-11-00397).

References

1. Erlang, A. K. 1917. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler. *Elektrotekniker* 13:5–13.
2. Gnedenko, B. V., and I. P. Makarov. 1971. Svoystva resheniy zadachi s poteryami v sluchae periodicheskikh intensivnostey [Properties of a problem with losses in the case of periodic intensities]. *Diff. Uravneniya* [Differential Equations] 7:1696–1698.
3. Zeifman, A. I. 1989. Properties of a system with losses in the case of variable rates. *Autom. Rem. Contr.* 50:82–87.
4. Kijima, M. 1990. On the largest negative eigenvalue of the infinitesimal generator associated with $M/M/n/n$ queues. *Oper. Res. Lett.* 9:59–64. doi: 10.1016/0167-6377(90)90041-3.
5. Massey, W. A., and W. Whitt. 1994. On analysis of the modified offered-load approximation for the nonstationary Erlang loss model. *Ann. Appl. Probab.* 4:1145–1160. doi:10.1214/aoap/1177004908.
6. Fricker, C., P. Robert, and D. Tibi. 1999. On the rate of convergence of Erlang's model. *J. Appl. Probab.* 36:1167–1184. doi: 10.1239/jap/1032374763.
7. Voit, M. 2000. A note of the rate of convergence to equilibrium for Erlang's model in the subcritical case. *J. Appl. Probab.* 37:918–923. doi: 10.1239/jap/1014842847.
8. Granovsky, B. L., and A. I. Zeifman. 2004. Nonstationary queues: Estimation of the rate of convergence. *Queueing Syst.* 46:363–388. doi: 10.1023/B:QUES.0000027991.19758.b4.
9. Van Doorn, E. A., and A. I. Zeifman. 2009. On the speed of convergence to stationarity of the Erlang loss system. *Queueing Syst.* 63:241–252. doi: 10.1007/s11134-009-9134-9.
10. Zeifman, A. I. 2009. On the nonstationary Erlang loss model. *Autom. Rem. Contr.* 70:2003–2012. doi: 10.1134/S000511790912008X.
11. Zeifman, A., A. Korotysheva, Ya. Satin, G. Shilova, and T. Panfilova. 2013. On a queueing model with group services. *Modern probabilistic methods for analysis of telecommunication network*. Eds. A. Dudin, V. Klimenok, G. Tsarenkov, and S. Dudin. Communications in computer and information science ser. 356:198–205. doi: 10.1007/978-3-642-35980-4_22.
12. Zeifman, A., A. Korotysheva, G. Shilova, and V. Korolev. 2014. On perturbation bounds for a queueing model with group services. Preprint.
13. Zeifman, A., A. Korotysheva, Y. Satin, V. Korolev, and V. Bening. 2014. Perturbation bounds and truncations for a class of Markovian queues. *Queueing Syst.* 76:205–221. doi: 10.1007/s11134-013-9388-0.
14. Zeifman, A. I., and V. Yu. Korolev. 2014. On perturbation bounds for continuous-time Markov chains. *Stat. Probab. Lett.* 88:66–72. doi: 10.1016/j.spl.2014.01.031.
15. Zeifman, A. I., and D. Isaacson. 1994. On strong ergodicity for nonhomogeneous continuous-time Markov chains. *Stoch. Proc. Appl.* 50:263–273. doi: 10.1016/0304-4149(94)90123-6.
16. Zeifman, A. I. 1995. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stoch. Proc. Appl.* 59:157–173. doi: 10.1016/0304-4149(95)00028-6.
17. Zeifman, A. I., S. Leorato, E. Orsingher, Y. Satin, and G. Shilova. 2006. Some universal limits for nonhomogeneous birth and death processes. *Queueing Syst.* 52:139–151. doi: 10.1007/s11134-006-4353-9.
18. Zeifman, A. I., V. Yu. Korolev, A. V. Korotysheva, and S. Ya. Shorgin. 2014. Obshchie otsenki ustoychivosti dlya nestatsionarnykh markovskikh tsepey s nepriyvnym vremenem [General bounds for nonstationary continuous-time Markov]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):106–117. doi: 10.14357/19922264140111.

Received August 12, 2014

Contributors

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Head of Department, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation;

principal scientist, Institute of Socio-Economic Development of Territories, Russian Academy of Sciences, 56A Gorkogo Str., Vologda 160014, Russian Federation; a_zeifman@mail.ru

Korotysheva Anna V. (b. 1988) — senior lecturer, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; a_korotysheva@mail.ru

Kiseleva Ksenia M. (b. 1992) — PhD student, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; a_zeifman@mail.ru

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; 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; vkorolev@cs.msu.su

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; SShorgin@ipiran.ru

СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ СОСТОЯНИЙ В СИСТЕМЕ ОБСЛУЖИВАНИЯ С ИНВЕРСИОННЫМ ПОРЯДКОМ ОБСЛУЖИВАНИЯ И ОБОБЩЕННЫМ ВЕРОЯТНОСТНЫМ ПРИОРИТЕТОМ*

Л. А. Мейханаджян¹, Т. А. Милованова², А. В. Печинкин³, Р. В. Разумчик⁴

Аннотация: Рассматривается однолинейная система массового обслуживания (СМО), в которую поступает поток заявок, называемый здесь потоком пуассоновского типа. Отличие этого потока от пуассоновского заключается в том, что интенсивность поступления заявок равна λ , если на приборе имеется заявка, и $\tilde{\lambda}$, если система пуста. Если заявка поступает в систему, в которой на приборе имеется заявка, то исходное распределение времени обслуживания поступающей заявки является произвольным с функцией распределения (ФР) $B(x)$, в противном случае — произвольным с ФР $\tilde{B}(x)$. В системе реализован инверсионный порядок обслуживания с обобщенным вероятностным приоритетом, заключающийся в следующем. Предполагается, что в любой момент времени известна остаточная длина каждой заявки в системе. В момент поступления в систему новой заявки ее исходная длина сравнивается с остаточной длиной заявки на приборе и в зависимости от результатов сравнения одна из них становится на прибор, а другая — на первое место в очередь, причем каждая заявка приобретает новую (случайную) длину и даже может покинуть систему. В статье предложены математические соотношения для вычисления основных показателей функционирования системы, связанных со стационарным распределением числа заявок в ней.

Ключевые слова: система массового обслуживания; специальные дисциплины; инверсионный порядок обслуживания; вероятностный приоритет

DOI: 10.14375/19922264140304

1 Введение

В текущих условиях активного развития инфотелекоммуникационной отрасли необходимы аналитические средства создания и анализа объектов и их совокупностей, используемых для передачи, хранения и обработки информации. Традиционно применяемые для описания функционирования инфотелекоммуникационных систем (ИТС) СМО и их комбинации позволяют получать точные и/или приближенные оценки для ключевых показателей производительности ИТС, хотя анализ зачастую связан с определенными вычислительными трудностями, что может приводить к бессодержательным результатам. Другой подход к анализу показателей производительности ИТС заключается в моделировании реальной системы с помощью простой модели СМО и применении в ней сложных дисциплин обслуживания. Например, предоставление приоритета обслуживания заявкам с наи-

меньшей остаточной длиной (дисциплина SRPT — shortest remaining processing time) [1] дает оптимальную стратегию с точки зрения минимизации числа заявок в системе. Но при расчетах показателей функционирования СМО с дисциплиной SRPT необходимо знание времени обслуживания (длины) каждой поступающей в систему заявки и использование информации о всех остаточных длинах. Подобные ограничения свойственны многим специальным дисциплинам.

На практике потоки, циркулирующие в ИТС, являются неоднородными, и требования к обслуживанию у сообщений различных классов могут быть различными. Кроме того, могут происходить сбои систем ввиду флуктуации нагрузки или появления дестабилизирующих факторов. Вместо введения в рассмотрение большого числа потоков можно воспользоваться следующим приемом: предположить, что длины заявок, а также эффекты (события), которые связаны с поступлением заявок

* Работа выполнена при частичной поддержке РФФИ (проект 13-07-00223).

¹ Российский университет дружбы народов, lameykanadzhyan@gmail.com

² Российский университет дружбы народов, tmiivanova77@mail.ru

³ Институт проблем информатики Российской академии наук, apetchinkin@ipiran.ru

⁴ Институт проблем информатики Российской академии наук; Российский университет дружбы народов, rrazumchik@iee.org

той или иной длины, известны лишь с некоторой вероятностью. Поэтому представляет интерес рассмотренный в [2] инверсионный порядок обслуживания с вероятностным приоритетом (дисциплина LCFS PP), при котором при поступлении в систему новой заявки место на приборе и первое место в очереди разыгрывают между собой вновь поступившая заявка и заявка, находившаяся ранее на приборе, причем с вероятностью, являющейся произвольной функцией от длины поступившей заявки и остаточной длины заявки на приборе.

По данной проблематике опубликовано и продолжает появляться значительное число работ как теоретического, так и прикладного характера (см., например, [3–8]). В частности, в [4] на примере одноканальной СМО с пуассоновским входящим потоком была рассмотрена дисциплина, предоставляющая преимущество заявкам с меньшими остаточными длинами. Эта дисциплина является в определенном смысле промежуточной между обычной дисциплиной без прерывания обслуживания и дисциплиной SRPT. Считаются известными остаточные времена обслуживания (длины) всех находящихся в системе заявок. При поступлении новой заявки ее длина сравнивается с длиной заявки на приборе, и та из них, длина которой минимальна, становится на прибор, оставляя за второй первое место в очереди.

В настоящей статье на основе развития идей вышеприведенных работ вводится новая дисциплина (инверсионный порядок обслуживания с обобщенным вероятностным приоритетом, или LCFS GPP), согласно которой при поступлении новой заявки система принимает решение не только о том, что с ней делать (поставить в очередь, поменять местами с заявкой на приборе, удалить из системы и т. д.), но и как долго обслуживать оставшиеся в системе заявки. Решение принимается путем задания вероятностей соответствующих событий. Исследуется СМО с такой дисциплиной, входящим потоком пуассоновского типа и произвольным распределением времени обслуживания заявки. На основе метода исключения состояний получены в терминах вычислительных алгоритмов и производящих функций выражения для стационарного распределения числа заявок в системе. Как будет показано, вычисление стационарного распределения (а также его моментов) связано с решением уравнения Фредгольма 2-го рода.

В следующем разделе дается подробное описание системы, затем выводятся формулы для нахождения стационарных вероятностных характеристик, и в заключении приводятся некоторые примеры численных расчетов, выполненных по полученным соотношениям.

2 Описание системы

Рассмотрим СМО с входящим потоком заявок, который для простоты будем называть здесь потоком пуассоновского типа. Отличие этого потока от пуассоновского заключается в следующем: интенсивность поступления заявки равна λ , если на приборе имеется заявка, и $\tilde{\lambda}$, если система пуста. Если в момент поступления заявки в систему на приборе имеется заявка, то исходное распределение времени обслуживания поступающей заявки является произвольным с ФП $B(x)$. Если же заявка поступает в систему в тот момент, когда система пуста, то исходное распределение времени обслуживания поступающей заявки является произвольным с ФП $\tilde{B}(x)$.

Далее для простоты изложения будем считать, что ФП $B(x)$ и $\tilde{B}(x)$ имеют непрерывные ограниченные плотности распределения $b(x) = B'(x)$ и $\tilde{b}(x) = \tilde{B}'(x)$.

Обобщенный инверсионный порядок обслуживания с вероятностным приоритетом (LCFS GPP) заключается в следующем. Предполагается, что в любой момент времени известна остаточная длина (далее будем говорить просто длина) каждой заявки в системе. В момент поступления в систему новой заявки ее исходная длина u сравнивается с (остаточной) длиной v заявки на приборе. С вероятностью $D(x, y|u, v)$, зависящей только от u и v , обслуживавшаяся ранее заявка продолжает обслуживаться, причем ее длина становится меньше y , а вновь поступившая становится на первое место в очереди и ее длина становится меньше x . Кроме того, с вероятностью $D^*(x, y|u, v)$, зависящей только от u и v , вновь поступившая заявка занимает прибор, вытесняя обслуживавшуюся ранее на первое место в очереди, причем длина заявки, бывшей ранее на приборе, становится меньше y , а вновь поступившей — меньше x .

Если на приборе находится заявка остаточной длины v и в систему поступает заявка длины u , то с вероятностью $D_0(x|u, v)$ заявка, находящаяся на приборе, покидает систему, а поступившая заявка становится на прибор, причем ее длина становится меньше x . Кроме того, с вероятностью $D_0^*(y|u, v)$ поступившая заявка сразу же покидает систему, а заявка, находящаяся на приборе, продолжает обслуживаться, причем ее длина становится меньше y . Введем также обозначение:

$$D(x|u, v) = D_0(x|u, v) + D_0^*(x|u, v).$$

Здесь $D(x|u, v)$ — вероятность того, что одна из двух заявок покинет систему, а вторая встанет на прибор и примет длину меньше x .

Наконец, предполагается, что с вероятностью $d_0(u, v)$ обе заявки покидают систему, а на прибор становится первая заявка из очереди.

Будем считать для удобства изложения, что все ФР $D(x, y|u, v)$, $D^*(x, y|u, v)$, $D_0(x|u, v)$, $D_0^*(y|u, v)$, $D(y|u, v)$ и $D_0(u, v)$ имеют непрерывные ограниченные плотности $d(x, y|u, v) = \partial^2 D(x, y|u, v)/(\partial x \partial y)$, $d^*(x, y|u, v) = \partial^2 D^*(x, y|u, v)/(\partial x \partial y)$, $d_0(x|u, v) = \partial D_0(x|u, v)/\partial x$, $d_0^*(y|u, v) = \partial D_0^*(y|u, v)/\partial y$ и $d(x|u, v) = \partial D(x|u, v)/\partial x$.

Естественно, для любых u и v выполнено условие:

$$\int_0^\infty \int_0^\infty [d(x, y|u, v) + d^*(x, y|u, v)] dx dy + \int_0^\infty d(x|u, v) dx + d_0(u, v) = D(\infty, \infty|u, v) + D^*(\infty, \infty|u, v) + D(\infty|u, v) + d_0(u, v) = 1. \quad (1)$$

Если длина заявки на приборе становится равной нулю, то она мгновенно покидает систему и на прибор переходит первая заявка из очереди. Остальная очередь сдвигается на единицу.

Далее будем предполагать, что система функционирует в стационарном режиме. К сожалению, для рассматриваемой СМО не удастся выписать общее необходимое и достаточное условие существования стационарного режима функционирования. Это условие зависит от конкретных параметров системы и в каждом отдельном случае нуждается в специальном исследовании. Здесь приведем простое достаточное условие, вытекающее из сравнения суммарной имеющейся работы в описанной СМО и суммарной работы в стандартной СМО $M/G/1/\infty$.

Условие состоит из выполнения следующих соотношений:

1. $\tilde{b} = \int_0^\infty x \tilde{b}(x) dx < \infty$.
2. $\rho = \lambda \tilde{b} = \lambda \int_0^\infty x b(x) dx < 1$.
3. $d(x, y|u, v) = 0$ при всех u, v и $y > v$ или $x > u$.
4. $d^*(x, y|u, v) = 0$ при всех u, v и $y > v$ или $x > u$.
5. $d(x|u, v) = 0$ при всех u, v и $x > u$.
6. $d^*(y|u, v) = 0$ при всех u, v и $y > v$.

Соотношения 3–6 соответствуют тому факту, что после поступления новой заявки измененные длины заявок не превышают те длины, которые были до поступления. Отметим, что здесь параметр $\rho = \lambda \tilde{b}$ не является загрузкой в традиционном смысле и может существенно от нее отличаться.

3 Система уравнений равновесия

Обозначим через $\nu(t)$ число заявок в системе в момент t , а через $\vec{\xi}(t) = (\xi_1(t), \dots, \xi_{\nu(t)}(t))$ — вектор, координатой $\xi_1(t)$ которой является (остаточное) время обслуживания заявки, находящейся в этот момент на приборе, $\xi_2(t)$ — первой заявки в очереди, \dots , $\xi_{\nu(t)-1}(t)$ — последней, $(\nu(t) - 1)$ -й заявки в очереди. При $\nu(t) = 0$ вектор $\vec{\xi}(t)$ не определяется. Тогда $\eta(t) = (\nu(t), \vec{\xi}(t))$ представляет собой марковский процесс, описывающий поведение числа заявок в рассматриваемой системе.

Положим

$$p_0(t) = \mathbf{P}\{\nu(t) = 0\};$$

$$P_n(t; x_1, \dots, x_n) = \mathbf{P}\{\nu(t) = n, \xi_1(t) < x_1, \dots, \xi_n(t) < x_n\}, \quad n \geq 1.$$

Обозначим через

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

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

стационарное распределение процесса $\eta(t)$. В силу сделанных в предыдущем разделе предположений относительно параметров системы существуют (см., например, [9]) непрерывные ограниченные плотности

$$p_n(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} P_n(x_1, \dots, x_n), \quad n \geq 1.$$

Выпишем систему интегродифференциальных уравнений, которой удовлетворяют стационарные плотности $p_n(x_1, \dots, x_n)$ и которую для краткости по аналогии с простейшими СМО будем называть системой уравнений равновесия (СУР). Для этого рассмотрим вспомогательную систему с $(n - 1)$ местами ожидания, отличающуюся от исходной системы только тем, что если в очереди находится $(n - 1)$ заявок, заявка на приборе имеет остаточную длину v и поступает новая заявка длины u , то с вероятностью $d(x, y|u, v)$ на приборе остается вновь поступившая заявка, длина которой становится равной x , а обслуживавшаяся ранее заявка покидает систему и, наоборот, с вероятностью $d^*(y, x|u, v)$ систему покидает вновь поступившая заявка, а находившаяся ранее на приборе заявка продолжает обслуживаться, но ее длина становится равной x .

В силу метода исключения состояний [9] стационарные вероятности состояний в исходной и вспомогательной системах отличаются лишь на постоянный множитель. Это дает возможность при

составлении уравнений для $p_n(x_1, \dots, x_n)$, $n \geq 1$, воспользоваться вспомогательной системой и получить следующие соотношения:

$$-p_1'(x) = \tilde{\lambda} \tilde{b}(x) p_0 - \lambda p_1(x) + \lambda \left(\int_0^\infty \int_0^\infty d(x|u, v) b(u) p_1(v) du dv + \int_0^\infty \int_0^\infty \int_0^\infty [d(x, y|u, v) b(u) p_1(v) + d^*(y, x|u, v) b(u) p_1(v)] dy du dv \right); \quad (2)$$

$$-p_n'(x_1, \dots, x_n) = \lambda \left(\int_0^\infty \int_0^\infty [d(x_2, x_1|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n) + d^*(x_1, x_2|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n)] du dv \right) - \lambda p_n(x_1, \dots, x_n) + \lambda \left(\int_0^\infty \int_0^\infty d(x_1|u, v) b(u) p_n(v, x_2, \dots, x_n) du dv + \int_0^\infty \int_0^\infty \int_0^\infty [d(x_1, y|u, v) b(u) p_n(v, x_2, \dots, x_n) + d^*(y, x_1|u, v) b(u) p_n(v, x_2, \dots, x_n)] dy du dv \right), \quad n \geq 2. \quad (3)$$

Остановимся подробнее на выводе уравнения (3). Рассмотрим моменты времени t и $(t + \Delta)$. Тогда для того чтобы в момент времени $(t + \Delta)$ в системе находилось n , $n \geq 2$, заявок, причем на приборе заявка длины x_1 , а в очереди заявки длин x_2, \dots, x_n , нужно, чтобы произошло одно из следующих событий:

- в момент t в системе находилось $(n - 1)$ заявок, причем заявка на приборе имела длину v , первая заявка в очереди имела длину x_3, \dots , последняя заявка в очереди имела длину x_n (с плотностью вероятностей $p_{n-1}(t; v, x_3, \dots, x_n)$), и за время Δ поступила заявка (с вероятностью $\lambda\Delta$) длины u (с плотностью вероятностей $b(u)$). Заявка на приборе продолжает обслуживаться, но ее длина становится равной x_1 , а вновь поступившая заявка

занимает первое место в очереди и ее длина становится равной x_2 (с плотностью вероятностей $d(x_2, x_1|u, v)$);

- в момент t в системе находилось $(n - 1)$ заявок, причем заявка на приборе имела длину v , первая заявка в очереди имела длину x_3, \dots , последняя заявка в очереди имела длину x_n (с плотностью вероятностей $p_{n-1}(t; v, x_3, \dots, x_n)$), и за время Δ поступила заявка (с вероятностью $\lambda\Delta$) длины u (с плотностью вероятностей $b(u)$). Поступившая заявка занимает прибор, и ее длина становится равной x_1 , а заявка, обслуживавшаяся до поступления новой заявки, занимает первое место в очереди, и ее длина становится равной x_2 (с плотностью вероятностей $d^*(x_1, x_2|u, v)$);
- в момент t в системе находилось n заявок, причем заявка на приборе имела длину $x_1 + \Delta$, первая заявка в очереди имела длину x_2, \dots , последняя заявка в очереди имела длину x_n (с плотностью вероятностей $p_n(t; x_1 + \Delta, x_2, \dots, x_n)$), и за время Δ не поступили заявки (с вероятностью $(1 - \lambda\Delta)$);
- в момент t в системе находилось n заявок, причем заявка на приборе имела длину v , первая заявка в очереди имела длину x_2, \dots , последняя заявка в очереди имела длину x_n (с плотностью вероятностей $p_n(t; v, \dots, x_n)$), и за время Δ поступила заявка (с вероятностью $\lambda\Delta$), имеющая длину u (с плотностью вероятностей $b(u)$). Заявка, находившаяся на приборе, покидает систему, а поступившая заявка становится на прибор, причем ее длина становится равной x_1 , или, наоборот, поступившая заявка сразу же покидает систему, а заявка, находившаяся на приборе, продолжает обслуживаться, причем ее длина становится равной x_1 (с плотностью вероятностей $d(x_1|u, v)$);
- в момент t в системе находилось n заявок, причем заявка на приборе имела длину v , первая заявка в очереди имела длину x_2, \dots , последняя заявка в очереди имела длину x_n (с плотностью вероятностей $p_n(t; v, x_2, \dots, x_n)$), и за время Δ поступила заявка (с вероятностью $\lambda\Delta$) длины u (с плотностью вероятностей $b(u)$). Заявка, находившаяся на приборе, покинула систему, а на прибор встала поступившая заявка, длина которой стала x_1 (с плотностью вероятностей $d(x_1, y|u, v)$);
- в момент t в системе находилось n заявок, причем заявка на приборе имела длину v , первая заявка в очереди имела длину x_2, \dots , последняя заявка в очереди имела длину x_n (с плотностью

вероятностей $p_n(t; v, x_2, \dots, x_n)$, и за время Δ поступила заявка (с вероятностью $\lambda\Delta$) длины u (с плотностью вероятностей $b(u)$). Вновь поступившая заявка покидает систему, на приборе продолжает обслуживаться заявка, находившаяся на приборе до поступления новой заявки, но длина ее становится равной x_1 (с плотностью вероятностей $d^*(y, x_1|u, v)$).

Вероятности других событий равны $o(\Delta)$.

Применяя формулу полной вероятности, имеем

$$\begin{aligned}
 p_n(t + \Delta; x_1, \dots, x_n) = & \\
 = \lambda\Delta & \left(\int_0^\infty \int_0^\infty [d(x_2, x_1|u, v)b(u)p_{n-1}(t; v, x_3, \dots, x_n) + \right. \\
 & \left. + d^*(x_1, x_2|u, v)b(u)p_{n-1}(t; v, x_3, \dots, x_n)] du dv \right) + \\
 & + (1 - \lambda\Delta)p_n(t; x_1 + \Delta, x_2, \dots, x_n) + \\
 + \lambda\Delta & \left(\int_0^\infty \int_0^\infty d(x_1|u, v)b(u)p_n(t; v, x_2, \dots, x_n) du dv + \right. \\
 & + \int_0^\infty \int_0^\infty \int_0^\infty [d(x_1, y|u, v)b(u)p_n(t; v, x_2, \dots, x_n) + \\
 & \left. + d^*(y, x_1|u, v)b(u)p_n(t; v, x_2, \dots, x_n)] dy du dv \right) + \\
 & + o(\Delta), \quad n \geq 2,
 \end{aligned}$$

откуда, перенося слагаемое $p_n(t; x_1 + \Delta, x_2, \dots, x_n)$ в левую часть равенства, деля на Δ , устремляя Δ к нулю и учитывая стационарный режим функционирования системы, получаем уравнение (3).

Уравнение (2) получается аналогично.

К системе уравнений (2), (3) нужно добавить начальные условия, которые удобно записать в виде:

$$p_1(\infty) = \lim_{X \rightarrow \infty} p_1(X) = 0; \quad (4)$$

$$p_n(\infty, x_2, \dots, x_n) = \lim_{X \rightarrow \infty} p_n(X, x_2, \dots, x_n) = 0, \quad n \geq 2. \quad (5)$$

Покажем, как получаются соотношения (4), (5) на примере (4). Интегрируя равенство (2) в пределах от 0 до X , приходим к формуле

$$\begin{aligned}
 p_1(0) - p_1(X) = & \tilde{\lambda}\tilde{B}(X)p_0 - \lambda P_1(X) + \\
 & + \lambda \int_0^X \left(\int_0^\infty \int_0^\infty d(x|u, v)b(u)p_1(v) du dv + \right.
 \end{aligned}$$

$$\left. + \int_0^\infty \int_0^\infty \int_0^\infty [d(x, y|u, v)b(u)p_1(v) + d^*(y, x|u, v)b(u)p_1(v)] dy du dv \right) dx,$$

которую перепишем в виде:

$$\begin{aligned}
 p_1(X) = p_1(0) - \tilde{\lambda}\tilde{B}(X)p_0 + \lambda P_1(X) - \\
 - \lambda \int_0^\infty \int_0^\infty b(u)p_1(v) du dv \int_0^X \left(d(x|u, v) + \right. \\
 \left. + \int_0^\infty [d(x, y|u, v) + d^*(y, x|u, v)] dy \right) dx. \quad (6)
 \end{aligned}$$

Правая часть (6) имеет предел (и даже с учетом (1) конечный) при $X \rightarrow \infty$. Поэтому $p_1(X)$ также стремится при $X \rightarrow \infty$ к пределу, который для плотности вероятностей не может быть ничем иным, кроме нуля, что доказывает справедливость (4).

Оставшаяся неизвестной стационарная вероятность p_0 отсутствия заявок в системе находится, как обычно, из условия нормировки:

$$\sum_{n=0}^\infty p_n = 1, \quad (7)$$

где $p_n = P_n(\infty, \dots, \infty)$, $n \geq 1$, — стационарная вероятность наличия в системе n заявок.

Полученные соотношения (2)–(7) позволяют (теоретически) последовательно по n находить стационарные плотности вероятностей $p_n(x_1, \dots, x_n)$ (один из методов решения СУР (2)–(7) будет рассмотрен в следующем разделе на примере маргинальных плотностей). Но практическое применение такого подхода невозможно, поскольку при $n \rightarrow \infty$ число аргументов x_i стационарных плотностей вероятностей $p_n(x_1, \dots, x_n)$ стремится к бесконечности.

Однако в большинстве практических случаев достаточно знать только маргинальные плотности

$$p_n(x) = \int \dots \int_{x_2, \dots, x_n > 0} p_n(x, x_2, \dots, x_n) dx_2 \dots dx_n, \quad n \geq 2.$$

Интегрируя (3) по x_2, \dots, x_n в пределах от нуля до бесконечности и вспоминая равенство (2), получаем следующее интегродифференциальное уравнение для $p_n(x)$, $n \geq 1$:

$$\begin{aligned}
 -p'_n(x) = a_n(x) - \lambda p_n(x) + \int_0^\infty K_n(x, v)p_n(v) dv, \\
 n \geq 1, \quad (8)
 \end{aligned}$$

где

$$a_1(x) = \tilde{\lambda} \tilde{b}(x) p_0;$$

$$K_1(x, v) = \lambda \int_0^\infty b(u) du \left(d(x|u, v) + \int_0^\infty [d(x, y|u, v) + d^*(y, x|u, v)] dy \right);$$

$$a_n(x) = \lambda \left(\int_0^\infty p_{n-1}(v) dv \int_0^\infty b(u) du \int_0^\infty [d(y, x|u, v) + d^*(x, y|u, v)] dy \right), \quad n \geq 2;$$

$$K_n(x, v) = \lambda \int_0^\infty b(u) du \left(d(x|u, v) + \int_0^\infty [d(x, y|u, v) + d^*(y, x|u, v)] dy \right), \quad n \geq 2.$$

Начальное условие для уравнения (8) по аналогии с (4) запишем в виде:

$$p_n(\infty) = \lim_{X \rightarrow \infty} p_n(X) = 0, \quad n \geq 1. \quad (9)$$

4 Метод численного решения системы уравнений равновесия

Приведем один из возможных методов решения интегродифференциального уравнения (8) с начальным условием (9).

Решение будем искать в виде:

$$p_n(x) = e^{\lambda x} q_n(x). \quad (10)$$

Подставляя в (8) вместо $p_n(x)$ ее выражение по формуле (10), получаем новое интегродифференциальное уравнение:

$$-q'_n(x) = e^{-\lambda x} a_n(x) + \int_0^\infty e^{\lambda v} e^{-\lambda x} K_n(x, v) q_n(v) dv.$$

Интегрируя последнее равенство по x в пределах от y до ∞ и учитывая начальное условие (9), получаем интегральное уравнение Фредгольма 2-го рода:

$$q_n(y) = b_n(y) + \int_0^\infty G_n(y, v) q_n(v) dv,$$

где

$$b_n(y) = \int_y^\infty e^{-\lambda x} a_n(x) dx;$$

$$G_n(y, v) = \int_y^\infty e^{\lambda(v-x)} K_n(x, v) dx. \quad (11)$$

Отметим, что свободный член $b_n(y)$ и ядро $G_n(y, v)$ интегрального уравнения (11) являются неотрицательными функциями.

Численные методы решения интегральных уравнений Фредгольма 2-го рода хорошо известны (см., например, [10–12]). Так, в данном случае хороший результат дает итерационный метод

$$q_n^{(s)}(y) = b_n(y) + \int_0^\infty G_n(y, v) q_n^{(s-1)}(v) dv,$$

причем в качестве начальной итерации необходимо взять нулевое приближение $q_n^{(0)}(v) \equiv 0$. Тогда итерации будут возрастать, что позволит контролировать скорость сходимости к точному решению.

5 Применение производящей функции

Введем производящую функцию (ПФ)

$$\pi(z, x) = \sum_{n=1}^{\infty} p_n(x) z^n.$$

Умножая уравнение (8) на z^n и суммируя по всем значениям $n \geq 1$, получаем:

$$-\frac{\partial \pi(z, x)}{\partial x} = z \tilde{\lambda} \tilde{b}(x) p_0 - \lambda \pi(z, x) + \lambda \int_0^\infty \pi(z, v) dv \int_0^\infty b(u) du \left(d(x|u, v) + \int_0^\infty [d(x, y|u, v) + d^*(y, x|u, v) + z d(y, x|u, v) + z d^*(x, y|u, v)] dy \right). \quad (12)$$

Граничное условие для интегродифференциального уравнения (12), как и прежде, имеет вид:

$$\pi(z, \infty) = \lim_{x \rightarrow \infty} \pi(z, x) = 0.$$

Трактуя z как параметр, для решения уравнения (12) можно применить метод, описанный в предыдущем разделе.

Однако ПФ $\pi(z, x)$, как правило, мало пригодна для вычисления стационарного распределения числа заявок в системе. Но с ее помощью можно найти моменты этого распределения.

Покажем, как это делается, на примере математического ожидания $E \nu$ стационарного распределения числа заявок в системе. При этом, не вдаваясь в математические тонкости, будем считать, что $E \nu$ существует и операции дифференцирования, которые будут применены далее, законны.

Сначала, решая уравнение (12) при $z = 1$, найдем значение $\pi(1, x)$.

Затем продифференцируем уравнение (12) по z в точке $z = 1$. При этом для сокращения записи применим обозначение:

$$\left. \frac{\partial \pi(z, x)}{\partial z} \right|_{z=1} = \pi'(1, x).$$

Имеем:

$$\begin{aligned} - \frac{\partial \pi'(1, x)}{\partial x} &= \tilde{\lambda} \tilde{b}(x) p_0 - \lambda \pi'(1, x) + \\ &+ \lambda \int_0^{\infty} \pi'(1, v) dv \int_0^{\infty} b(u) du \left(d(x|u, v) + \right. \\ &+ \left. \int_0^{\infty} [d(x, y|u, v) + d^*(y, x|u, v) + \right. \\ &+ \left. zd(y, x|u, v) + zd^*(x, y|u, v)] dy \right) + \\ &+ \lambda \int_0^{\infty} \pi(1, v) dv \int_0^{\infty} b(u) du \left(\int_0^{\infty} [d(y, x|u, v) + \right. \\ &+ \left. d^*(x, y|u, v)] dy \right). \quad (13) \end{aligned}$$

Граничное условие для интегродифференциального уравнения (13) определяется прежним равенством:

$$\pi'(1, \infty) = \lim_{x \rightarrow \infty} \pi'(1, x) = 0.$$

Уравнение (13) имеет точно такой же вид, что и уравнение (12), и может быть решено теми же самими методами.

Осталось заметить, что

$$E \nu = \sum_{n=1}^{\infty} n \int_0^{\infty} p_n(x) dx = \int_0^{\infty} \pi'(1, x) dx.$$

6 Примеры расчетов

На основе полученных результатов с помощью программных средств MATLAB была написана программа, позволяющая вычислять совместное стационарное распределение числа заявок в системе и остаточного времени обслуживания заявки на приборе и другие связанные с ним характеристики, а также исследовать поведение рассматриваемой СМО в зависимости от значений определяющих ее исходных параметров. Разработанная программа позволяет проводить исследование в широкой области изменения исходных параметров и при любых заданных аналитически функциях $d(x, y|u, v)$, $d^*(x, y|u, v)$, $d_0(x|u, v)$, $d_0^*(x|u, v)$ и $d_0(u, v)$, удовлетворяющих условию существования стационарного режима. Однако вид этих функций существенным образом влияет на эффективность работы программы.

Приведем некоторые результаты расчетов, проведенных с помощью разработанной программы. При этом будем предполагать, что $\tilde{B}(x) = B(x)$, и среднее время обслуживания b заявки на приборе положим равным 1. Параметр λ далее будет использоваться как аргумент при построении графиков, поскольку он совпадает с параметром $\rho = \lambda b = \lambda$.

6.1 Пример 1

Предположим, что $\lambda = \tilde{\lambda}$, функции $d^*(x, y|u, v)$, $d(x|u, v)$ и $d_0(u, v)$ тождественно равны нулю при всех x, y, u и v , а функция $d(x, y|u, v)$ при всех u и v имеет вид $d(x, y|u, v) = b(x)b(y)$. Стационарные вероятности состояний этой СМО совпадают со стационарными вероятностями состояний СМО $M|G|1$ при следующей дисциплине обслуживания: поступающая заявка становится на первое место в очереди, а обслуживавшаяся ранее заявка остается на приборе, но ее длина разыгрывается заново с той же функцией распределения $B(x)$.

Если же $d(x, y|u, v)$, $d(x|u, v)$ и $d_0(u, v)$ тождественно равны нулю, а $d^*(x, y|u, v) = b(x)b(y)$, то с точки зрения стационарных вероятностей состояний система эквивалентна СМО $M|G|1$ с дисциплиной, при которой поступающая заявка становится на прибор, а обслуживавшаяся ранее заявка переходит на первое место в очереди и ее длина разыгрывается заново с функцией распределения $B(x)$.

Для численных расчетов в примере 1 будем считать, что время обслуживания заявки распределено по экспоненциальному закону, т. е. $B(x) = 1 - e^{-x}$. Тогда при обоих вариантах выбора $d(x, y|u, v)$ и $d^*(x, y|u, v)$ стационарные вероятности состояний будут совпадать со стационарными вероятностями

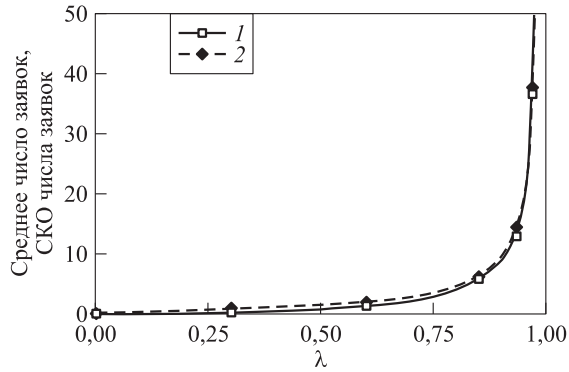


Рис. 1 Пример 1: среднее число заявок в системе (1) и СКО числа заявок (2)

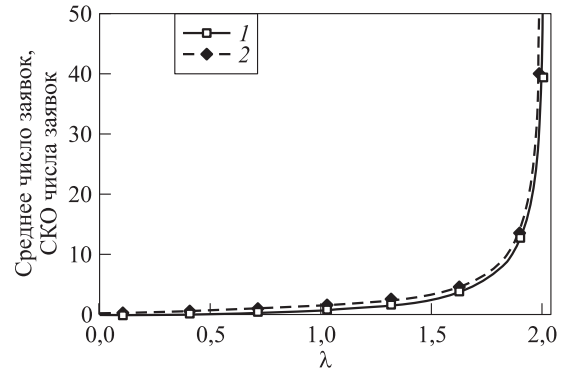


Рис. 2 Пример 2: среднее число заявок в системе (1) и СКО числа заявок (2)

состояний обычной СМО $M|M|1$. Графики среднего и среднего квадратичного отклонения (СКО) числа заявок в системе в зависимости от интенсивности входящего потока λ представлены на рис. 1. Как видно, эти данные полностью совпадают с хорошо известными для СМО $M|M|1$ результатами.

6.2 Пример 2

В следующем примере предположим, что функция $d(x|u, v)$ тождественно равна нулю при всех x, u и v , а функция $d_0(u, v) > 0$ при всех u и v . Это фактически означает, что с вероятностью $d_0(u, v)$ поступает отрицательная заявка, которая уничтожает находящуюся на приборе заявку и вместе с ней покидает систему. В частности, если интенсивность поступления в систему отрицательных заявок равна γ , то $d_0(u, v) = \gamma/\lambda$ при всех u, v . Теперь, если положить $\tilde{\lambda} = \lambda - \gamma$, функцию $d^*(x, y|u, v) = 0$ при всех x, y, u и v , функцию $d(x, y|u, v) = (\lambda - \gamma)b(x)b(y)/\lambda$ при всех u, v, x и y , то получим первый вариант СМО $M|G|1$ из примера 1, в которую, наряду с потоком интенсивности $\tilde{\lambda}$ обычных заявок, поступает поток отрицательных заявок интенсивности γ . Второй вариант СМО из примера 1 с дополнительным потоком отрицательных заявок получается, если, наоборот, положить $d^*(x, y|u, v) = (\lambda - \gamma)b(x)b(y)/\lambda$ и $d(x, y|u, v) = 0$ при тех же самых значениях остальных параметров.

Положим теперь $B(x) = 1 - e^{-x}$. При этом пусть $\gamma = 1$. Тогда приходим к СМО $M|M|1$ с инверсионным порядком обслуживания с (без) прерыванием обслуживания и отрицательными заявками. Графики среднего и СКО числа заявок в системе в зависимости от интенсивности входящего потока λ представлены на рис. 2. Нетрудно видеть, что данные результаты полностью совпадают с результатами, которые легко получить по аналитическим формулам для данной СМО.

6.3 Пример 3

Пусть задана функция распределения $R(x)$ неотрицательной случайной величины, определенная на интервале $(0, v)$, которая имеет ограниченную плотность распределения $r(x) = R'(x)$. Предположим, что функции $d^*(x, y|u, v)$, $d_0(x|u, v)$ и $d_0(u, v)$ тождественно равны нулю при всех x, y, u, v . При этом функция $d_0^*(y|u, v) = \gamma r(y)/\lambda$ при всех u, v и $y < v$ и $d_0^*(y|u, v) = 0$ иначе; функция $d(x, y|u, v) = (\lambda - \gamma)b(x)b(y)/\lambda$ при всех u, v и $x > v$ и $d(x, y|u, v) = 0$ иначе. Тогда при $\tilde{\lambda} = \lambda - \gamma$ систему можно трактовать как СМО $M|G|1$ с инверсионным порядком обслуживания, в которую с вероятностью γ/λ поступают заявки, приводящие к обслуживанию заявок на приборе заново с распределением $R(x)$. С вероятностью $(\lambda - \gamma)/\lambda$ в систему поступают обычные заявки.

В этом примере положим $B(x) = 1 - e^{-x}$ и $R(x) = 0,05(1 - e^{-x})/(1 - e^{-v})$. В результате получаем СМО $M|M|1$ с инверсионным порядком обслуживания, в которую с вероятностью 0,05 поступают заявки, приводящие к обслуживанию заявок на приборе заново. Графики среднего и СКО числа заявок в системе в зависимости от интенсивности входящего потока λ представлены на рис. 3.

6.4 Пример 4

Предположим, что функция $d^*(x, y|u, v) = 0$ при всех u, v, y, x , а функции $d(x, y|u, v)$, $d_0(x|u, v)$, $d_0^*(x|u, v)$ и $d_0(u, v)$ заданы следующим образом:

$$d(x, y|u, v) = 0,85e^{-(x+y)}(e^{-v} - 1)^{-1}(e^{-u} - 1)^{-1}$$

при всех u, v и $y < v, x < u$ и $d(x, y|u, v) = 0$ иначе;

$$d_0(x|u, v) = 0,05 \frac{1}{u+1} \frac{e^{-(1/(u+1))x}}{1 - e^{-(1/(u+1))u}}$$

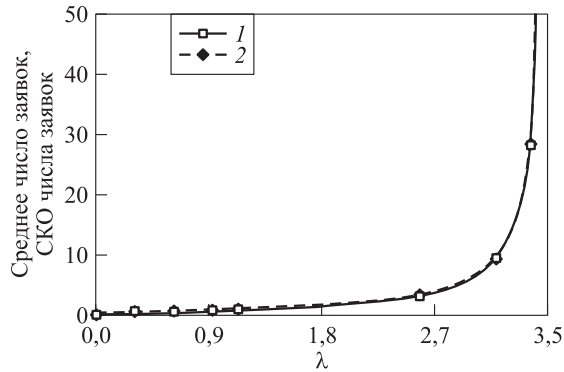


Рис. 3 Пример 3: среднее число заявок в системе (1) и СКО числа заявок (2)

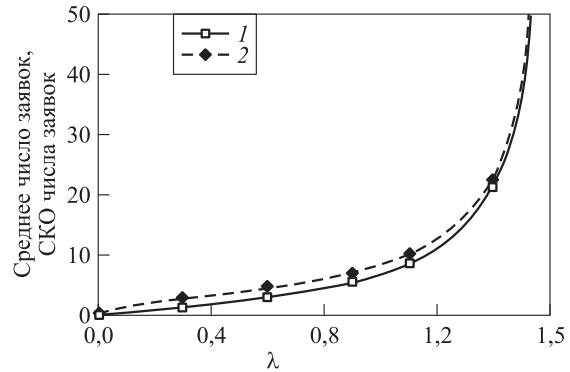


Рис. 4 Пример 4: среднее число заявок в системе (1) и СКО числа заявок (2)

при всех u, v и $x < u$ и $d_0(x|u, v) = 0$ иначе;

$$d_0^*(y|u, v) = 0,05 \frac{1}{v+1} \frac{e^{-(1/(v+1))y}}{1 - e^{-(1/(v+1))v}}$$

при всех u, v и $y < v$ и $d_0^*(x|u, v) = 0$ иначе; $d_0(u, v) = 0,05$ при всех u, v .

В этом случае имеет место СМО $M|G|1$ с инверсионным порядком обслуживания, причем допускается уход из системы вновь поступившей или недообслуженной заявки (с равными вероятностями 0,05), а также «выбивание» заявки с прибора (также с вероятностью 0,05). При данном определении функций $d_0(x|u, v)$ ($d_0^*(y|u, v)$) чем больше была длина вновь поступившей заявки (обслуживаемой заявки), тем меньше будет средняя длина заявки, которая останется обслуживаться на приборе. Отметим, что определенные выше функции подобраны таким образом, чтобы удовлетворять достаточному условию существования стационарного режима.

Для примера 4 графики среднего и СКО числа заявок в системе в зависимости от интенсивности входящего потока λ представлены на рис. 4.

7 Заключение

В настоящей работе получены интегродифференциальные уравнения для стационарных плотностей вероятностей СМО $M/G/1/\infty$ с входящим потоком пуассоновского типа и инверсионным порядком обслуживания с обобщенным вероятностным приоритетом (LCFS GPP) и предложен метод их решения. По полученным соотношениям была разработана программа, которая позволяет проводить исследование показателей производительности рассматриваемой СМО в широкой области изменения исходных параметров. Приводятся ре-

зультаты численных расчетов как для частных случаев, так и для общего случая рассматриваемой системы.

В дальнейших исследованиях предполагается обратиться к решению задачи вычисления характеристик, связанных с временем пребывания заявки в системе.

Литература

1. Schrage L. A proof of the optimality of the shortest remaining processing time discipline // Oper. Res., 1968. Vol. 16. P. 687–690.
2. Нагоненко В. А. О характеристиках одной нестандартной системы массового обслуживания. I, II // Изв. АН СССР. Технич. кибернет., 1981. № 1. С. 187–195; № 3. С. 91–99.
3. Нагоненко В. А., Печинкин А. В. О большой загрузке в системе с инверсионным обслуживанием и вероятностным приоритетом // Изв. АН СССР. Технич. кибернет., 1982. № 1. С. 86–94.
4. Печинкин А. В. Об одной инвариантной системе массового обслуживания // Math. Operationsforsch. und Statist. Ser. Optimization, 1983. Vol. 14. № 3. S. 433–444.
5. Нагоненко В. А., Печинкин А. В. О малой загрузке в системе с инверсионным порядком обслуживания и вероятностным приоритетом // Изв. АН СССР. Технич. кибернет., 1984. № 6. С. 82–89.
6. Печинкин А. В., Стальченко И. В. Система $MAR/G/1/\infty$ с инверсионным порядком обслуживания и вероятностным приоритетом, функционирующая в дискретном времени // Вестник Российского ун-та дружбы народов. Сер. Математика. Информатика. Физика, 2010. № 2. С. 26–36.
7. Касконе А., Манзо Р., Печинкин А. В., Салерно С. Система $MAR/G/1/\infty$ в дискретном времени с инверсионной вероятностной дисциплиной обслуживания // Автоматика и телемеханика, 2010. № 12. С. 57–69.

8. Милованова Т. А., Печинкин А. В. Стационарные характеристики системы обслуживания с инверсионным порядком обслуживания, вероятностным приоритетом и гистерезисной политикой // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 22–36.
9. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
10. Jerri A. Introduction to integral equations with applications. — N.Y.: John Wiley & Sons, 1999. 433 p.
11. Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. Numerical recipes: The art of scientific computing. 3rd ed. New York: Cambridge University Press, 2007. 1235 p.
12. Полянин А. Д., Манжиров А. В. Справочник по интегральным уравнениям. — М.: Физматлит, 2003. 608 с.

Поступила в редакцию 17.06.14

STATIONARY DISTRIBUTION IN A QUEUEING SYSTEM WITH INVERSE SERVICE ORDER AND GENERALIZED PROBABILISTIC PRIORITY

L. A. Meykhanadzhyan¹, T. A. Milovanova¹, A. V. Pechinkin², and R. V. Razumchik^{1,2}

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

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Consideration is given to $M|G|1$ type queueing system. Inverse service order with generalized probabilistic priority is implemented in the system. It is assumed that at any instant, the remaining service time of each customer residing in the system is known. Upon arrival of a new customer, the system finds out its service time and compares it with the remaining service time of the currently served customer. The result of this comparison leads to one of the cases: one of them enters the server and another occupies the first place in the queue; one of them leaves the system and another enters the server; or both leave the system. In each case when customer remains in the system, its remaining service time may be updated. An analytical method that allows computing stationary performance characteristics related to the number of customers in the system is presented. Numerical examples based on the developed mathematical relations are provided.

Keywords: queueing system; special discipline; LIFO; probabilistic priority; general service time

DOI: 10.14375/19922264140304

Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (grant No. 13-07-00223).

References

1. Schrage, L. 1968. A proof of the optimality of the shortest remaining processing time discipline. *Oper. Res.* 16:687–690.
2. Nagonenko, V. A. 1981. O kharakteristikakh odnoy nestandartnoy sistemy massovogo obsluzhivaniya [On the characteristics of one nonstandard queueing system]. I, II. *Izv. AN SSSR. Tekhnich. kibernet.* [Technical Cybernetics] 1:187–195; 3:91–99.
3. Nagonenko, V. A., and A. V. Pechinkin. 1982. O bol'shoy zagruzke v sisteme s inverсионным obsluzhivaniem i veroyatnostnym prioritetom [On high load in the system with an inversion procedure service and probabilistic priority]. *Izv. AN SSSR. Tekhnich. kibernet.* [Technical Cybernetics] (1):86–94.
4. Pechinkin, A. V. 1983. Ob odnoy invariantnoy sisteme massovogo obsluzhivaniya [On an invariant queueing system]. *Math. Operationsforsch. und Statist. Ser. Optimization* 14(3):433–444.
5. Nagonenko, V. A., and A. V. Pechinkin. 1984. O malo zagruzke v sisteme s inverсионным poryadkom obsluzhivaniya i veroyatnostnym prioritetom [On low load in the system with an inversion procedure service and probabilistic priority]. *Izv. AN SSSR. Tekhnich. kibernet.* [Technical Cybernetics] (6):82–89.
6. Pechinkin, A. V., and I. V. Stalchenko. 2010. Sistema MAP/G/1/∞ s inverсионным poryadkom obsluzhivaniya i veroyatnostnym prioritetom, funktsioniruyushchaya v diskretnom vremeni [The MAP/G/1/∞ discrete-time queueing system with inverse service order and probabilistic priority]. *Vestnik Rossiyskogo Un-ta družby na-*

- rodov. Ser. Matematika. Informatika. Fizika. [Bulletin of Peoples' Friendship University of Russia. Ser. Mathematics. Information Sciences. Physics] (2):26–36.
7. Cascone, A., R. Manzo, A. V. Pechinkin, and S. Salerno. 2010. Sistema MAP/G/1/∞ v diskretnom vremeni s inversionnoy veroyatnostnoy distsiplinoy obsluzhivaniya [Discrete-time MAP/G/1/∞ system with inversive probabilistic servicing discipline]. *Avtomat. i Telemekh.* [Automation Remote Control] (12):57–69.
 8. Milovanova, T. A., and A. V. Pechinkin. 2013. Statsionarnye kharakteristiki sistemy obsluzhivaniya s inversionnym poryadkom obsluzhivaniya, veroyatnostnym prioritetom i gisterizisnoy politikoy [Stationary characteristics of queueing system with an inversion procedure service probabilistic priority and hysteresis policy] *Informatika i ee Primeneniya — Inform. Appl.* 7(1):22–35.
 9. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
 10. Jerri, A. 1999. *Introduction to integral equations with applications*. John Wiley & Sons. 433 p.
 11. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. *Numerical recipes: The art of scientific computing*. 3rd ed. New York: Cambridge University Press. 1235 p.
 12. Polyanin, A. D., and A. V. Manzhirov. 2008. *Handbook of integral equations*. 2nd ed. Boca Raton – London: Chapman & Hall / CRC Press. 1144 p.

Received June 17, 2014

Contributors

Meykhanadzhyan Lusine A. (b. 1990) — postgraduate, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; lameykhanadzhyan@gmail.com

Milovanova Tatiana A. (b. 1977) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; tmilovanova77@mail.ru

Pechinkin Alexander V. (b. 1946) — Doctor of Science in physics and mathematics; principal scientist, Institute of Informatics Problems of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; apechinkin@ipiran.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, senior research fellow, Institute of Informatics Problems of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; assistant professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ieee.org

АНАЛИЗ СИСТЕМЫ ОБСЛУЖИВАНИЯ С ВХОДЯЩИМ ПОТОКОМ АВТОРЕГРЕССИОННОГО ТИПА*

Н. Д. Леонтьев¹, В. Г. Ушаков²

Аннотация: Рассматривается одноканальная система массового обслуживания с неограниченным числом мест для ожидания, в которую поступает пуассоновский поток групп требований. Особенностью системы является авторегрессионная зависимость размеров групп поступающих требований: размер n -й поступившей в систему группы требований либо с некоторой фиксированной вероятностью равен размеру $(n-1)$ -й поступившей в систему группы требований, либо с дополнительной вероятностью является независимой от него случайной величиной. Длительности обслуживания требований являются независимыми случайными величинами с произвольным распределением. Основным объектом изучения является длина очереди в произвольный момент времени. Получены соотношения, позволяющие найти преобразование Лапласа по времени производящей функции числа требований в системе в нестационарном режиме, а также ряд вспомогательных характеристик: время дообслуживания требования, находящегося на приборе в момент t , распределение размера последней группы требований, поступившей в систему до момента t .

Ключевые слова: теория массового обслуживания; нестационарный режим; системы с групповым поступлением требований

DOI: 10.14375/19922264140305

1 Введение

При разработке сложных информационно-вычислительных систем важную роль играют методы оценки интенсивности потоков запросов с целью выявления недостаточных или избыточных вычислительных ресурсов на базе соответствующих моделей. В данной работе рассматривается система массового обслуживания типа $M|G|1$ с групповым поступлением требований. При этом интенсивность потока запросов определяется регрессионной зависимостью между размерами групп требований.

Исследованию систем обслуживания с параметрами регрессионного типа посвящен целый ряд работ, опубликованных в специализированной литературе в последние годы. Отметим статьи [1, 2], где изучается система обслуживания с дискретным временем, в которой размеры пакетов поступающих требований связаны регрессионной зависимостью.

В основной теореме настоящей статьи приводятся формулы для распределения длины очереди в произвольный момент времени.

2 Описание системы

Рассмотрим систему обслуживания, состоящую из одного обслуживающего устройства, на которое поступает простейший поток групп требований с интенсивностью a . Будем считать, что число мест для ожидания не ограничено, а длительность обслуживания требований имеет распределение с плотностью $b(x)$ и преобразованием Лапласа $\beta(s)$. Поступающие в систему группы требований могут иметь размер $1, \dots, M$ с вероятностями соответственно h_1, \dots, h_M . При этом размер n -й поступившей в систему группы требований с вероятностью $0 \leq p < 1$ равен размеру $(n-1)$ -й, либо с дополнительной вероятностью $1-p$ является независимой от него случайной величиной.

Определим следующие случайные процессы:

- $L(t)$ — число требований в системе в момент t ;
- $X(t)$ — время, прошедшее с начала обслуживания требования, находящегося на обслуживании в момент t (в случае, когда система свободна, можно для определенности положить $X(t) = 0$);
- $N(t)$ — размер последней поступившей в систему до момента t группы требований.

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 12-07-00109а, 13-07-00223а).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ndleontyev@gmail.com

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук; vgushakov@mail.ru

Введем следующие обозначения:

$$P(n, k, x, t) = \frac{\partial}{\partial x} \mathbf{P}(L(t) = n, N(t) = k, X(t) < x);$$

$$P(n, k, t) = \mathbf{P}(L(t) = n, N(t) = k).$$

Обозначим

$$\pi(z, k, x, s) = \sum_{n=1}^{\infty} z^n \int_0^{\infty} e^{-st} P(n, k, x, t) dt;$$

$$\pi_0(k, s) = \int_0^{\infty} e^{-st} P(0, k, t) dt$$

при $|z| \leq 1, \operatorname{Re}(s) > 0$.

3 Основные результаты

Формула для определения $\pi(z, k, x, s)$ и ее вывод содержатся в следующей основной теореме.

Теорема. Функция $\pi(z, k, x, s)$ при $|z| < 1, k = 1, \dots, M, x \geq 0, \operatorname{Re}(s) > 0$ определяется по формуле:

$$\pi(z, k, x, s) = (1 - B(x)) \times \sum_{n=1}^M C_n(z) \frac{(1-p)h_k a z^k}{\tilde{\lambda}_n(z) - p a z^k} \exp\left(-\left(s + a - \tilde{\lambda}_n(z)\right)x\right),$$

где

$$C_n(z) = \frac{1}{1 - z^{-1} \beta \left(s + a - \tilde{\lambda}_n(z)\right)} \times \frac{\prod_{i=1}^M \left(\tilde{\lambda}_n(z) - p a z^i\right)}{\prod_{\substack{j=1 \\ j \neq n}}^M \left(\tilde{\lambda}_n(z) - \tilde{\lambda}_j(z)\right)} \sum_{m=1}^M \frac{b_m(z, s)}{\tilde{\lambda}_n(z) - p a z^m};$$

$$b_m(z, s) = -(s + a)\pi_0(m, s) + h_m + \left[p\pi_0(m, s) a z^m + (1-p) \sum_{n=1}^M \pi_0(n, s) h_n a z^m \right];$$

$\tilde{\lambda}_1(z), \dots, \tilde{\lambda}_M(z)$ определяются из уравнения

$$\prod_{i=1}^M (p a z^i - \tilde{\lambda}) + \sum_{i=1}^M (1-p) h_i a z^i \prod_{\substack{j=1 \\ j \neq i}}^M (p a z^j - \tilde{\lambda}) = 0,$$

а $\pi_0(1, s), \dots, \pi_0(M, s)$ — из системы

$$\sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - p a z_n^j) \left(-(s + a)\pi_0(m, s) + h_m + \tilde{\lambda}_n \pi_0(m, s) \right) = 0.$$

Доказательство. Функционирование системы описывается следующими уравнениями:

$$P(n, k, x + \Delta, t + \Delta) = P(n, k, x, t) [1 - (a + \eta(x))\Delta] + \mathbf{1}_{\{n > k\}} \left[p P(n - k, k, x, t) a \Delta + (1-p) \sum_{m=1}^M P(n - k, m, x, t) h_m a \Delta \right];$$

$$P(0, k, t + \Delta) = P(0, k, t) [1 - a\Delta] + \int_0^{\infty} P(1, k, x, t) \eta(x) dx \Delta;$$

$$\int_0^{\Delta} P(n, k, u, t + \Delta) du = \int_0^{\infty} P(n + 1, k, x, t) \eta(x) dx \Delta + \delta_{n,k} \left[p P(0, k, t) a \Delta + (1-p) \sum_{m=1}^M P(0, m, t) h_m a \Delta \right],$$

где $\eta(x) = b(x)/(1 - B(x))$.

Переходя к пределу при $\Delta \rightarrow 0$, получим:

$$\frac{\partial P(n, k, x, t)}{\partial t} + \frac{\partial P(n, k, x, t)}{\partial x} = -(a + \eta(x))P(n, k, x, t) + \mathbf{1}_{\{n > k\}} \left[p P(n - k, k, x, t) a + (1-p) \sum_{m=1}^M P(n - k, m, x, t) h_m a \right]; \quad (1)$$

$$\frac{\partial P(0, k, t)}{\partial t} = -a P(0, k, t) + \int_0^{\infty} P(1, k, x, t) \eta(x) dx; \quad (2)$$

$$P(n, k, 0, t) = \int_0^{\infty} P(n + 1, k, x, t) \eta(x) dx + \delta_{n,k} \left[p P(0, k, t) a + (1-p) \sum_{m=1}^M P(0, m, t) h_m a \right]. \quad (3)$$

Переходя в уравнениях (1)–(3) к производящим функциям и преобразованиям Лапласа по t , получим:

$$\begin{aligned} \frac{\partial \pi(z, k, x, s)}{\partial x} &= -(s + a + \eta(x))\pi(z, k, x, s) + \\ &+ \left[p\pi(z, k, x, s)az^k + (1 - p) \sum_{m=1}^M \pi(z, m, x, s)h_k az^k \right]; \\ (s + a)\pi_0(k, s) - h_k &= \\ &= \int_0^\infty \int_0^\infty P(1, k, x, t)\eta(x) dx e^{-st} dt; \quad (4) \\ \pi(z, k, 0, s) &= z^{-1} \int_0^\infty \pi(z, k, x, s)\eta(x) dx - \\ &- \int_0^\infty \int_0^\infty P(1, k, x, t)\eta(x) dx e^{-st} dt + \\ &+ \left[p\pi_0(k, s)az^k + (1 - p) \sum_{m=1}^M \pi_0(m, s)h_k az^k \right]. \quad (5) \end{aligned}$$

Подставляя (4) в (5), получим:

$$\begin{aligned} \frac{\partial \pi(z, k, x, s)}{\partial x} &= \\ &= -(s + a + \eta(x))\pi(z, k, x, s) + \left[p\pi(z, k, x, s)az^k + \right. \\ &\quad \left. + (1 - p) \sum_{m=1}^M \pi(z, m, x, s)h_k az^k \right]; \quad (6) \\ \pi(z, k, 0, s) &= z^{-1} \int_0^\infty \pi(z, k, x, s)\eta(x) dx - \\ &\quad - (s + a)\pi_0(k, s) + h_k + \\ &+ \left[p\pi_0(k, s)az^k + (1 - p) \sum_{m=1}^M \pi_0(m, s)h_k az^k \right]. \quad (7) \end{aligned}$$

Обозначим

$$\pi(z, k, x, s) = (1 - B(x))\tilde{\pi}(z, k, x, s).$$

В новых обозначениях (6) примет вид:

$$\begin{aligned} \frac{\partial \tilde{\pi}(z, k, x, s)}{\partial x} &= \\ &= -(s + a)\tilde{\pi}(z, k, x, s) + \left[p\tilde{\pi}(z, k, x, s)az^k + \right. \\ &\quad \left. + (1 - p) \sum_{m=1}^M \tilde{\pi}(z, m, x, s)h_k az^k \right]. \quad (8) \end{aligned}$$

Это линейная система дифференциальных уравнений первого порядка с постоянными коэффициентами. Принимая во внимание результат леммы, приведенной ниже, решение системы можно записать в виде:

$$\begin{aligned} \tilde{\pi}(z, k, x, s) &= \\ &= \sum_{n=1}^M C_n(z)u_{kn}(z) \exp(-(s + a - \tilde{\lambda}_n(z))x), \quad (9) \end{aligned}$$

где $\tilde{\lambda}_k(z) = \lambda_k(z, s) + (s + a)$, $k = 1, \dots, M$; $\lambda_1(z, s), \dots, \lambda_M(z, s)$ — собственные значения матрицы системы, а $u_1(z) = (u_{11}(z), \dots, u_{M1}(z))^T, \dots, u_M(z) = (u_{1M}(z), \dots, u_{MM}(z))^T$ — соответствующие собственные векторы.

Функции $\tilde{\lambda}_1(z), \dots, \tilde{\lambda}_M(z)$ являются решениями характеристического уравнения

$$\prod_{i=1}^M (paz^i - \tilde{\lambda}) + \sum_{i=1}^M (1-p)h_i az^i \prod_{\substack{j=1 \\ j \neq i}}^M (paz^j - \tilde{\lambda}) = 0. \quad (10)$$

Установим следующее вспомогательное утверждение.

Лемма. *Функции $\tilde{\lambda}_1(z), \dots, \tilde{\lambda}_M(z)$ обладают следующими свойствами:*

- (1) $\text{Re}(\tilde{\lambda}_m(z)) \leq a$ для всех m и $|z| \leq 1$;
- (2) функции $\tilde{\lambda}_i(z)$ и $\tilde{\lambda}_j(z)$ при $i \neq j$ могут совпадать лишь на конечном множестве точек z .

Доказательство леммы полностью аналогично доказательству леммы 1 из [3] (с. 181).

Вернемся к доказательству теоремы. Подставляя (9) в (8), находим:

$$u_{mn}(z) = \frac{(1-p)h_m az^m}{\tilde{\lambda}_n(z) - paz^m}. \quad (11)$$

Подставив (9) в (7), получим:

$$\begin{aligned} \sum_{n=1}^M C_n(z)u_{mn}(z) &= \\ &= z^{-1} \sum_{n=1}^M C_n(z)u_{mn}(z)\beta(s + a - \tilde{\lambda}_n(z)) - \\ &\quad - (s + a)\pi_0(m, s) + h_m + \\ &\quad + \left[p\pi_0(m, s)az^m + (1 - p) \sum_{n=1}^M \pi_0(n, s)h_n az^m \right]. \end{aligned}$$

Перепишем это уравнение в виде:

$$\begin{aligned} \sum_{n=1}^M C_n(z) \left(1 - z^{-1}\beta(s + a - \tilde{\lambda}_n(z)) \right) u_{mn}(z) &= \\ &= b_m(z, s), \quad (12) \end{aligned}$$

где

$$b_m(z, s) = -(s + a)\pi_0(m, s) + h_m + \left[p\pi_0(m, s)az^m + (1 - p) \sum_{n=1}^M \pi_0(n, s)h_n az^m \right].$$

Подставим (11) в (12) и поделим обе части полученного уравнения на $(1 - p)h_m az^m$:

$$\sum_{n=1}^M C_n(z) \left(1 - z^{-1}\beta \left(s + a - \tilde{\lambda}_n(z) \right) \right) \times \frac{1}{\tilde{\lambda}_n(z) - paz^m} = \frac{b_m(z, s)}{(1 - p)h_m az^m}.$$

Это система линейных алгебраических уравнений с матрицей Коши. Ее решение записывается в виде¹:

$$C_n(z) \left(1 - z^{-1}\beta \left(s + a - \tilde{\lambda}_n(z) \right) \right) = \frac{\prod_{i=1}^M \left(\tilde{\lambda}_n(z) - paz^i \right)}{\prod_{\substack{j=1 \\ j \neq n}}^M \left(\tilde{\lambda}_n(z) - \tilde{\lambda}_j(z) \right)} \times \prod_{\substack{i=1 \\ i \neq n}}^M \frac{\prod_{\substack{j=1 \\ j \neq m}}^M \left(paz^m - \tilde{\lambda}_i(z) \right)}{\prod_{\substack{j=1 \\ j \neq m}}^M \left(paz^m - paz^j \right)} \frac{b_m(z, s)}{(1 - p)h_m az^m}.$$

Далее, поскольку функции $\tilde{\lambda}_m(s)$, $m = 1, \dots, M$, являются решениями уравнения (10), можно записать:

$$\prod_{i=1}^M (paz^i - \tilde{\lambda}) + \sum_{i=1}^M (1 - p)h_i az^i \prod_{\substack{j=1 \\ j \neq i}}^M (paz^j - \tilde{\lambda}) = \prod_{j=1}^M \left(\tilde{\lambda}_j(z) - \tilde{\lambda} \right). \quad (13)$$

Подставляя в (13) $\tilde{\lambda} = paz^m$, получим:

$$(1 - p)h_m az^m \prod_{\substack{j=1 \\ j \neq m}}^M (paz^j - paz^m) = \prod_{j=1}^M \left(\tilde{\lambda}_j(z) - paz^m \right).$$

Отсюда

$$C_n(z) = \frac{1}{1 - z^{-1}\beta \left(s + a - \tilde{\lambda}_n(z) \right)} \times \frac{\prod_{i=1}^M \left(\tilde{\lambda}_n(z) - paz^i \right)}{\prod_{\substack{j=1 \\ j \neq n}}^M \left(\tilde{\lambda}_n(z) - \tilde{\lambda}_j(z) \right)} \sum_{m=1}^M \frac{b_m(z, s)}{\tilde{\lambda}_n(z) - paz^m}. \quad (14)$$

Остается найти $\pi_0(m, s)$, $m = 1, \dots, M$. Домножим уравнение (14) на $1 - z^{-1}\beta(s + a - \tilde{\lambda}_n(z))$:

$$C_n(z) \left(1 - z^{-1}\beta \left(s + a - \tilde{\lambda}_n(z) \right) \right) = \frac{\prod_{i=1}^M \left(\tilde{\lambda}_n(z) - paz^i \right)}{\prod_{\substack{j=1 \\ j \neq n}}^M \left(\tilde{\lambda}_n(z) - \tilde{\lambda}_j(z) \right)} \sum_{m=1}^M \frac{b_m(z, s)}{\tilde{\lambda}_n(z) - paz^m}. \quad (15)$$

Рассмотрим уравнение

$$z = \beta(s + a - \tilde{\lambda}_n(z)). \quad (16)$$

Обе части уравнения являются аналитическими в области $|z| \leq 1$ функциями. Имеем, с учетом доказанной леммы,

$$\left| \beta \left(s + a - \tilde{\lambda}_n(z) \right) \right| \leq \beta \left(\operatorname{Re} \left(s + a - \tilde{\lambda}_n(z) \right) \right) \leq \beta(\operatorname{Re}(s)) < 1 = |z|$$

при $|z| = 1$. В силу теоремы Руше отсюда следует, что функциональное уравнение (16) имеет единственное решение $z = z_n(s)$, причем функция $z_n(s)$ является аналитической в области $\operatorname{Re}(s) > 0$.

Подставляя $z = z_n(s)$ в уравнение (15), приходим после ряда преобразований к уравнению²:

$$\sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M \left(\tilde{\lambda}_n - paz_n^j \right) b_m(z_n, s) = 0. \quad (17)$$

Вспомним, что $b_m(z, s) = -(s + a)\pi_0(m, s) + h_m + \left[p\pi_0(m, s)az^m + (1 - p) \sum_{k=1}^M \pi_0(k, s)h_k az^m \right]$.

¹Про обращение матриц Коши см. [4].

²В дальнейшем будем для краткости писать z_n вместо $z_n(s)$ и $\tilde{\lambda}_n$ вместо $\tilde{\lambda}_n(z_n)$.

С учетом уравнения (10) будем иметь:

$$\begin{aligned} & \sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - paz_n^j) \left[p\pi_0(m, s)az_n^m + \right. \\ & \quad \left. + (1-p) \sum_{k=1}^M \pi_0(k, s)h_maz_n^m \right] = \\ & = \sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - paz_n^j) \left[(paz_n^m - \tilde{\lambda}_n)\pi_0(m, s) + \right. \\ & \quad \left. + \tilde{\lambda}_n\pi_0(m, s) + (1-p) \sum_{k=1}^M \pi_0(k, s)h_maz_n^m \right] = \\ & = - \prod_{j=1}^M (\tilde{\lambda}_n - paz_n^j) \sum_{m=1}^M \pi_0(m, s) + \\ & \quad + \sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - paz_n^j) \tilde{\lambda}_n\pi_0(m, s) + \\ & \quad + \prod_{j=1}^M (\tilde{\lambda}_n - paz_n^j) \sum_{m=1}^M \pi_0(m, s) = \\ & = \sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - paz_n^j) \tilde{\lambda}_n\pi_0(m, s). \end{aligned}$$

Возвращаясь к уравнению (17), получим:

$$\sum_{m=1}^M \prod_{\substack{j=1 \\ j \neq m}}^M (\tilde{\lambda}_n - paz_n^j) \left(-(s+a)\pi_0(m, s) + h_m + \tilde{\lambda}_n\pi_0(m, s) \right) = 0.$$

Там самым доказательство теоремы завершено.

Литература

1. *Hwang G. U., Choi B. D., Kim J.-K.* The waiting time analysis of a discrete-time queue with arrivals as an autoregressive process of order 1 // *J. Appl. Probab.*, 2002. Vol. 39. No. 3. P. 619–629.
2. *Hwang G. U., Sohraby K.* On the exact analysis of a discrete-time queueing system with autoregressive inputs // *Queueing Syst.*, 2003. Vol. 43. No. 1–2. P. 29–41.
3. *Матвеев В. Ф., Ушаков В. Г.* Системы массового обслуживания. — М.: МГУ, 1984. 242 с.
4. *Schechter S.* On the inversion of certain matrices // *Math. Tab. Aids Comput.*, 1959. Vol. 13. No. 66. P. 73–77.

Поступила в редакцию 14.07.14

ANALYSIS OF A QUEUEING SYSTEM WITH AUTOREGRESSIVE ARRIVALS

N. D. Leontyev¹ and V. G. Ushakov^{1,2}

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

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper studies a single server queueing system with infinite capacity and with the Poisson batch arrival process. A feature of the system under study is autoregressive dependence of the arriving batch sizes: the size of the n th batch is equal to the size of the $(n-1)$ st batch with a fixed probability and is an independent random variable with complementary probability. Service times are supposed to be independent random variables with a specified distribution. The main object of the study is the queue length at an arbitrary moment. The relations derived make it possible to find Laplace transform in time of the probability generating function of the transient queue length, and also a number of additional characteristics such as the residual service time and the distribution of the size of the last batch that arrived before time t .

Keywords: queueing theory; transient behavior; batch arrivals

DOI: 10.14375/19922264140305

Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (Grants Nos. 12-07-00109a and 13-07-00223a).

References

1. Hwang, G. U., B. D. Choi, and J.-K. Kim. 2002. The waiting time analysis of a discrete-time queue with arrivals as an autoregressive process of order 1. *J. Appl. Probab.* 39(3):619–629.
2. Hwang, G. U., and K. Sohraby. 2003. On the exact analysis of a discrete-time queueing system with autoregressive inputs. *Queueing Syst.* 43(1-2):29–41.
3. Matveev, V. F., and V. G. Ushakov. 1984. *Sistemy massovogo obsluzhivaniya* [Queueing systems]. Moscow: MSU. 242 p.
4. Schechter, S. 1959. On the inversion of certain matrices. *Math. Tab. Aids Comput.* 13(66):73–77.

Received July 14, 2014

Contributors

Leontyev Nikolai D. (b. 1988) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation; ndleontyev@gmail.com

Ushakov Vladimir G. (b. 1952) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vgushakov@mail.ru

МОДЕЛИ ДЛЯ СРАВНИТЕЛЬНОГО АНАЛИЗА МЕТОДОВ КЛАССИФИКАЦИИ В НЕКОТОРЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМАХ РАСПОЗНАВАНИЯ ОБРАЗОВ

Я. М. Агаларов¹

Аннотация: В работе рассматриваются системы распознавания, в которых классы заданы прямым указанием образцов, размещенных в распределенной базе данных (БД), и критерием распознавания (идентификации) является совпадение предъявленного образа хотя бы с одним из образцов. Ставится задача сравнительного анализа последовательного и параллельного методов классификации с точки зрения среднего времени ответа на запрос распознавания и необходимой мощности вычислительных ресурсов. Приведены результаты вычислительных экспериментов, полученных с использованием аналитических и компьютерных моделей сетей массового обслуживания (СеМО) с очередями на примерах распределенных мультибиометрических систем идентификации.

Ключевые слова: распределенные системы распознавания; последовательные и параллельные методы классификации; распределение вычислительных ресурсов; сеть массового обслуживания

DOI: 10.14375/19922264140306

1 Введение

Появление распределенных вычислительных систем и методов распределенных вычислений дает возможность более быстрой обработки больших объемов информации и выполнения трудоемких вычислений и значительно расширяет круг реализуемых на практике ресурсоемких вычислительных задач. Целесообразность применения распределенных вычислений в той или иной области определяется эффективностью распределенных вычислений при решении соответствующих задач, которую, в свою очередь, определяют степень возможного повышения пропускной способности оборудования и ускорения процесса вычислений [1].

Эффективность методов организации распределенных вычислений зависит как от аппаратных и программных особенностей вычислительной системы, так и от того, насколько сложна решаемая задача в смысле организации параллельных вычислений и алгоритм ее решения соответствует архитектуре распределенной вычислительной системы. Особенно эффективны распределенные вычисления в системах, в которых поступающие ресурсоемкие задания можно разбить на независимые (в смысле процесса выполнения) блоки любой длины (в смысле времени выполнения) и выполнять на различных вычислительных ресурсах одновременно. К таким системам относятся некоторые системы автоматического распознавания образов [2], среди которых отметим компьютерные системы

криптоанализа [3, 4], распределенные автоматизированные системы оперативного розыска личностей [5–7], распределенные электронные архивы документов с автоматической подсистемой массового распознавания изображений [8] и др.

Используемые в современных искусственных системах автоматического распознавания образов методы (процедуры распознавания) классификации делят, как правило, на два типа:

- (1) параллельные — принадлежность предъявленного объекта (образа) проверяется одновременно по всем признакам объекта;
- (2) последовательные — принадлежность объекта (образа) проверяется поочередно для каждого признака в зависимости от результатов уже проведенных для этого же образа проверок [9].

Каждый из этих методов имеет свои недостатки и достоинства, и их применимость в конкретной области зависит от предъявляемых к системе требований, от целесообразности их применения, от архитектуры вычислительной системы и др.

В работе рассматриваются распределенные системы распознавания, в которых классы заданы прямым указанием составляющих их образцов объектов, а критерием принадлежности классу предъявленного образа объекта служит наличие в классе образца, совпадающего с предъявленным образом. При этом образ (образец) объекта задается значением группы признаков. К таким системам относятся упомянутые выше в качестве примеров сис-

¹Институт проблем информатики Российской академии наук, agglar@yandex.ru

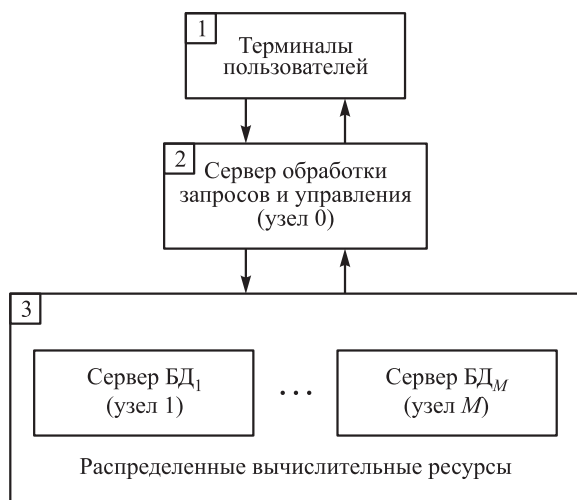


Рис. 1 Функциональная схема системы

темы распознавания. Как правило, такие системы представляют собой распределенные системы БД, в которых каждая БД содержит в виде записей цифровые образцы некоторой части объектов по одному или нескольким признакам. Условно упрощенную функциональную схему подобных систем можно представить в виде, приведенном на рис. 1. Блок 1 системы формирует модели объектов и генерирует задания в виде запросов на распознавание объектов. Блок 2 выполняет обработку заданий (преобразует модели объектов в образцы необходимого формата, транслирует запросы, получаемые от блока 1, на соответствующие подзапросы, направляемые на серверы БД), управляет процессами передачи заданий на серверы БД и приема от них значений целевых (решающих) функций, формирует ответы на запросы блока 1. В блоке 3 решается основная задача системы — проверяется принадлежность предъявленного образа к классам: каждый сервер БД в соответствии с запросом и образом объекта, полученными от блока 2, производит в БД поиск образца, совпадающего с предъявленным образом, и выдает соответствующее значение целевой функции в блок 2.

Поступающие в блок 2 запросы могут содержать различные требования: найти в указанных БД хотя бы один образец, совпадающий с предъявленным образом (именно для таких запросов ниже проводится анализ характеристик системы); найти в указанных БД все образцы, совпадающие с предъявленным образом; найти образец, содержащийся во всех указанных БД и совпадающий с предъявленным образом и т. д.

Важнейшей характеристикой любой системы распознавания является достоверность принимаемых ею решений, которая в основном определяет-

ся информативностью образа (образца). Чем больше он включает в себя информативных признаков, тем больше размер образа, но тем ниже скорость и эффективность поиска совпадающего с ним образца [2, 9]. Однако в системах оперативного принятия решений наряду с высокой достоверностью решений требуется и высокая производительность подсистемы распознавания. Так как мощность вычислительных ресурсов системы ограничена, то необходимо выбрать компромиссные значения показателей производительности и достоверности, отвечающие противоречивым требованиям к системе. Отсюда можно сделать вывод, что в больших распределенных системах распознавания, к которым предъявляются высокие требования по оперативности и достоверности, могут возникать очереди заданий к вычислительным ресурсам.

Хотя проведено множество исследований в направлении повышения производительности и достоверности распределенных систем распознавания рассматриваемого типа, все они, как правило, исходят из упрощенных моделей, в частности без учета очередей к вычислительным ресурсам [10].

Ниже приведены примеры использования компьютерной модели системы с очередями для оценки характеристик блока 3 (см. рис. 1) системы распознавания описанного выше типа и сравнительного анализа последовательного и параллельного методов классификации.

2 Модель системы

В качестве модели системы распознавания образов рассматривается СеМО [11], в которой узлы представляют сервер управления (обработки запросов и управления) и серверы БД (см. рис. 1), линии — каналы связи (сеть связи). Предполагается, что узлы пронумерованы числами $0, \dots, M$ и допускают неограниченные очереди, каналы связи имеют достаточную пропускную способность (очередями к ним можно пренебречь). В узел 0, представляющий сервер управления, извне поступают L пронумерованных числами $1, \dots, L$ потоков заданий, требующих выполнения на соответствующих множествах узлов $A_i, i = 1, \dots, L$, представляющих серверы БД. Задания, заставшие узел в занятом состоянии, становятся в очередь к узлу в порядке поступления и выполняются узлом в порядке поступления. Каждое задание во время его выполнения занимает узел полностью. Время пребывания задания в узле есть сумма времени ожидания в очереди и времени выполнения. Время пребывания задания в узле 0 считается близким к нулю, и ниже при расчете времени пребывания задания в системе узел 0 не рассматривается.

Процесс выполнения задания на узле заключается в следующем: в узле предъявленный образ поочередно сравнивается с хранящимися в нем образцами, пока не произойдет совпадение. Предполагается, что априорная вероятность совпадения образа с любым образцом одинаковая. Если произошло совпадение, то выполнение задания в этом узле завершается (произошло успешное завершение), иначе выполнение задания в узле завершается после сравнения со всеми образцами (завершение неуспешное). После завершения выполнения задания освобождает ресурсы узла (очередь и вычислительные ресурсы) и покидает его. Считается, что для каждого узла известна априорная вероятность успешного завершения для каждого типа задания.

Рассматриваются два типа процедур выполнения заданий:

- (1) параллельная процедура — i -задание (задание типа с номером i) из узла 0 направляется одновременно на все узлы соответствующего множества A_i и выполняется на них независимо;
- (2) последовательная процедура — i -задание из узла 0 направляется на узлы из множества A_i , выбираемые в узле 0 по одному в определенном порядке, причем задание направляется в последующий узел только после неуспешного выполнения на предыдущем.

В случае параллельной процедуры после успешного завершения выполнения i -задания в одном из узлов задание покидает все узлы из множества A_i , освободив занятые им ресурсы.

Введем обозначения:

- t_j — время сравнения образа с образцом в j -м узле;
- n_j — число образцов в j -м узле;
- ts_j — случайная величина времени сравнения в j -м узле при несовпадении образца, \overline{ts}_j — ее среднее значение;
- tu_j — случайная величина времени сравнения в j -м узле при совпадении образца, \overline{tu}_j — ее среднее значение;
- τ_{ji} — случайная величина времени выполнения в j -м узле i -задания, $\overline{\tau}_{ji}$ — ее среднее значение;
- \overline{W}_{ji} — среднее значение времени ожидания в j -м узле i -задания;
- \overline{Tzad}_i — среднее значение времени пребывания i -задания на вычислительных устройствах (в блоке 3);
- \overline{T}_j — среднее время пребывания произвольного задания в j -м узле;
- \overline{T} — среднее время пребывания произвольного задания в системе;

- \overline{N}_j — среднее число заданий в j -м узле;
- \overline{N} — среднее число заданий в системе;
- α_{ji} — априорная вероятность нахождения в j -м узле образца, совпадающего с образом i -задания;
- p_{ji} — вероятность того, что i -задание поступит для выполнения в j -й узел;
- λ_i — интенсивность потока i -заданий, поступающих извне;
- Λ_j — интенсивность суммарного потока заданий, поступающих в j -й узел.

Приведем выражения для некоторых вероятностно-временных характеристик рассматриваемой системы. Всюду ниже предполагаем, что система работает в установившемся режиме (в узлах системы конечные средние длины очередей).

Заметим, что среднее время ожидания в каждом узле системы для всех типов заданий будет одинаковой величиной, поэтому вместо величин \overline{W}_{ji} можно рассмотреть величины \overline{W}_j , $j = 1, \dots, M$, соответственно. Заметим также, что случайная величина времени выполнения в j -м узле i -задания при условии, что в узле произойдет совпадение образца с образцом, равна одной из случайных величин $ts_j, tu_j + ts_j, \dots, (n_j - 1)tu_j + ts_j$ с одинаковой вероятностью $1/n_j$. Тогда для среднего времени выполнения в j -м узле i -задания при условии, что в узле произойдет совпадение образца с образцом, справедлива формула $(n_j - 1)\overline{tu}_j/2 + \overline{ts}_j$. Для среднего времени выполнения в j -м узле i -задания справедлива формула:

$$\begin{aligned} \overline{\tau}_{ji} &= \alpha_{ji} \frac{(n_j - 1)\overline{tu}_j}{2} + \alpha_{ji}\overline{ts}_j + (1 - \alpha_{ji})n_j\overline{tu}_j = \\ &= \frac{\overline{tu}_j[n_j(2 - \alpha_{ji}) - \alpha_{ji}]}{2} + \alpha_{ji}\overline{ts}_j. \end{aligned} \quad (1)$$

Второй момент случайной величины τ_{ji} равен

$$\begin{aligned} \overline{\tau}_{ji}^2 &= \alpha_{ji} \frac{\overline{tu}_j^2}{n_j} \sum_{k=1}^{n_j-1} k^2 + \alpha_{ji}\overline{ts}_j^2 + (1 - \alpha_{ji})n_j^2\overline{tu}_j^2 = \\ &= \overline{tu}_j^2 \left[n_j - \frac{\alpha_{ji}}{6} (4n_j^2 + 3n_j - 1) \right] + \alpha_{ji}\overline{ts}_j^2. \end{aligned} \quad (2)$$

Легко заметить, что при большом значении параметра n_j и детерминированных величинах tu_j и ts_j функцию распределения времени выполнения задания в узле j можно аппроксимировать равномерным распределением, заданным на отрезке $[0, n_j tu_j]$, при этом погрешность первого момента случайной величины времени выполнения i -задания будет равна $\alpha_{ji}(2ts_j - tu_j)/(2n_j)$, а погрешность второго момента — $\alpha_{ji}(6ts_j^2 + tu_j^2)/(6n_j^2) - \alpha_{ji}tu_j^2/(2n_j)$.

Пусть в последовательной процедуре для узлов $i_k \in A_i, k = 1, \dots, |A_i|$, индекс k показывает порядок выполнения i -заданий на множестве узлов A_i , где i_1 — номер первого узла, на котором выполняется i -задание; i_r — номер узла, на который поступает i -задание после неуспешного завершения выполнения в узле с номером $i_{r-1}, r = 2, \dots, |A_i|$. Для среднего времени пребывания

$$\begin{aligned} \overline{Tzad}_i &= \alpha_{i_1 i} \left[\overline{W}_{i_1} + \frac{(n_{i_1} + 1) t_{i_1}}{2} \right] + \\ &+ (1 - \alpha_{i_1 i}) \alpha_{i_2 i} \left[\overline{W}_{i_1} + \overline{W}_{i_2} + t_{i_1} + \frac{(n_{i_2} + 1) t_{i_2}}{2} \right] + \dots \\ &\dots + \prod_{k=1}^{|A_i|-1} (1 - \alpha_{i_k i}) \left[\sum_{k=1}^{|A_i|-1} (\overline{W}_{i_k} + t_{i_k}) + \overline{W}_{i_{|A_i|}} + \right. \\ &\left. + \alpha_{i_{|A_i|} i} \frac{(n_{i_{|A_i|}} + 1) t_{i_{|A_i|}}}{2} + (1 - \alpha_{i_{|A_i|} i}) t_{i_{|A_i|}} \right] = \\ &= \overline{W}_{i_1} + \overline{\tau}_{i_1 i} + (1 - \alpha_{i_1 i}) (\overline{W}_{i_2} + \overline{\tau}_{i_2 i}) + \dots \\ &\dots + \prod_{k=1}^{|A_i|-1} (1 - \alpha_{i_k i}) (\overline{W}_{i_{|A_i|}} + \overline{\tau}_{i_{|A_i|} i}) = \\ &= \sum_{k=1}^{|A_i|} p_{i_k i} (\overline{W}_{i_k} + t_{i_k}) = \sum_{k=1}^{|A_i|} p_{i_k i} \overline{T}_{i_k}. \end{aligned}$$

Для среднего времени пребывания произвольного задания в системе справедливо соотношение:

$$\begin{aligned} \overline{T} &= \sum_{i=1}^L \frac{\lambda_i}{\lambda} \overline{Tzad}_i = \frac{1}{\lambda} \sum_{i=1}^L \lambda_i \sum_{k=1}^{|A_i|} p_{i_k i} \overline{T}_{i_k} = \\ &= \frac{1}{\lambda} \sum_{j=1}^M \overline{T}_j \sum_{i: j \in A_i} \lambda_i p_{ji} = \frac{1}{\lambda} \sum_{j=1}^M \overline{T}_j \Lambda_j, \end{aligned}$$

где $\lambda = \sum_{i=1}^L \lambda_i$.

Используя формулу Литтла [11], получим

$$\overline{T} = \frac{1}{\lambda} \sum_{j=1}^M \overline{N}_j = \frac{\overline{N}}{\lambda}. \quad (3)$$

Пусть R_j — мощность вычислительных ресурсов j -го узла, а суммарная мощность вычислительных ресурсов ограничена величиной R , т. е. $\sum_{j=1}^M R_j \leq R$.

Рассмотрим задачу выбора оптимальных значений величин $R_j, j = 1, \dots, M$, при заданных значениях остальных параметров системы при последовательной процедуре. Пусть tu_j и $ts_j, j = 1, \dots, M$, — величины, заданные для вычислительных ресурсов единичной мощности. Будем считать, что время

выполнения i -задания в j -м узле при мощности ресурсов R_j равно τ_{ji}/R_j (на практике это достигается путем фрагментации БД), а величина R такова, что существуют $R_j, j = 1, \dots, M$, при которых в узлах в установившемся режиме работы системы не могут возникать бесконечные очереди, т. е. $\rho_j = \Lambda_j (\overline{\tau}_{ji}/R_j) < 1, j = 1, \dots, M$. Предположим, что потоки заданий, поступающие на узлы системы, являются пуассоновскими. Тогда для среднего времени ожидания задания в узле справедлива следующая формула (вытекает из формулы Поллячека–Хинчина [11] и из (1), (2)):

$$\begin{aligned} \overline{W}_j &= \frac{\lambda_j}{2} \frac{\sum_{i: j \in A_i} (\lambda_{ji}/\Lambda_j) (\overline{\tau}_{ji}^2/R_j^2)}{1 - \Lambda_j \sum_{i: j \in A_i} (\lambda_{ji}/\Lambda_j) (\overline{\tau}_{ji}/R_j)} = \\ &= \frac{\sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}^2}{2R_j \left(R_j - \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji} \right)}, \quad (4) \end{aligned}$$

где $\lambda_{ji} = \sum_{i: j \in A_i} \lambda_i p_{ji}$; $\Lambda_j = \sum_{i: j \in A_i} \lambda_{ji}$; $\sum_{i: j \in A_i} (\lambda_{ji}/\Lambda_j) (\overline{\tau}_{ji}/R_j)$ — среднее время выполнения произвольного задания в j -м узле.

Среднее время пребывания вычисляется по формуле (следует из (4)):

$$\overline{T}_j = \frac{\sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}^2}{2R_j \left(R_j - \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji} \right)} + \sum_{i: j \in A_i} \frac{\lambda_{ji}}{\Lambda_j} \frac{\overline{\tau}_{ji}}{R_j}. \quad (5)$$

Среднее число заданий в узле вычисляется по формуле (следует из формулы Литтла и из (5)):

$$\begin{aligned} \overline{N}_j &= \frac{\Lambda_j \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}^2}{2R_j \left(R_j - \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji} \right)} + \\ &+ \frac{1}{R_j} \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}. \quad (6) \end{aligned}$$

Обозначим

$$a_j = \frac{\Lambda_j \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}^2}{2}; \quad b_j = \sum_{i: j \in A_i} \lambda_{ji} \overline{\tau}_{ji}.$$

Тогда (6) принимает вид:

$$\bar{N}_j = \frac{a_j}{R_j(R_j - b_j)} + \frac{b_j}{R_j}.$$

Предположив, что $R_j, j = 1, \dots, M$, — непрерывные величины, из (6) для производной функции \bar{N}_j по R_j получим:

$$\frac{\partial \bar{N}_j}{\partial R_j} = -\frac{a_j(2R_j - b_j)}{R_j^2(R_j - b_j)^2} - \frac{b_j}{R_j^2}. \quad (7)$$

Для второй производной получим:

$$\frac{\partial^2 \bar{N}_j}{\partial^2 R_j} = \frac{a_j[6(R_j - b_j) + 2b_j]}{R_j^3(R_j - b_j)^3} + \frac{2b_j}{R_j^3}. \quad (8)$$

Так как $R_j > b_j$ (следует из введенного выше предположения $\rho_j < 1, j = 1, \dots, M$), то из (7) и (8) следуют неравенства

$$\frac{\partial \bar{N}_j}{\partial R_j} < 0; \quad \frac{\partial^2 \bar{N}_j}{\partial^2 R_j} > 0$$

(т. е. среднее число заданий в узле является выпуклой функцией по переменной R_j) и $\partial^2 \bar{N}_j / \partial^2 R_j$ — непрерывные по $R_j > b_j, j = 1, \dots, M$, функции. Следовательно, существует единственный набор значений $\{R_j^*, j = 1, \dots, M\}$, такой что $\sum_{j=1}^M R_j^* = R, R_j^* > b_j, j = 1, \dots, M$, и при этих значениях среднее число заданий в системе \bar{N} достигает глобального минимума [12], причем единственного. Так как $\Lambda_j > 0$ не зависит от $R_i, j, i = 1, \dots, M$, и $\bar{N}_j = \Lambda_j \bar{T}_j$, то функция \bar{T} также достигает при наборе значений $\{R_j^*, j = 1, \dots, M\}$ единственного глобального минимума (так как верна формула (3)).

Необходимые и достаточные условия существования минимума для функции \bar{T} при непрерывных величинах $R_j, j = 1, \dots, M$, имеют вид [12]:

$$\frac{\partial \bar{N}_j}{\partial R_j} = \frac{\partial \bar{N}_k}{\partial R_k}, \quad j, k = 1, \dots, M. \quad (9)$$

Таким образом, как следует из (7)–(9), набор $\{R_j^*, j = 1, \dots, M\}$ является решением системы уравнений:

$$\left. \begin{aligned} & \frac{a_1(2R_1 - b_1)}{R_1^2(R_1 - b_1)^2} + \frac{b_1}{R_1^2} = \\ & = \frac{a_j(2R_j - b_j)}{R_j^2(R_j - b_j)^2} + \frac{b_j}{R_j^2}, \quad j = 2, \dots, M; \\ & \sum_{j=1}^M R_j = R. \end{aligned} \right\} \quad (10)$$

Как легко видеть из (10), распределение ресурсов $R_j = k\sqrt{b_j}, j = 1, \dots, M$, где $k = R / \sum_{j=1}^M \sqrt{b_j} > 1$, является хорошим приближением решения

системы уравнений (10), если $R_j \gg b_j, R_j \gg a_j$ (т. е. время ожидания близко к нулю). Иными словами, распределение ресурсов, балансирующее нагрузку в системе, дает хорошее приближение к оптимальному, если времена ожидания в узлах при этом распределении ресурсов близки к нулю, иначе следует искать более точное приближение решения системы (10).

3 Вычислительные эксперименты

Ниже приведены два примера системы распознавания, для которых с помощью компьютерных моделей проведены расчеты среднего времени выполнения произвольного задания и среднего числа заданий в системе при использовании параллельной и последовательной процедур распознавания.

В первом примере рассмотрена мультимодальная система идентификации по изображению лица и отпечаткам пальцев. Система состоит из трех узлов, соединенных коммуникационной средой: центра обработки запросов и управления системой (узел 0), узла идентификации по изображению лица (узел 1), узла идентификации по отпечаткам пальцев (узел 2). Основные параметры системы имеют следующие значения: размер базы образцов в каждом из узлов 1 и 2 равен 20 000 000 записей, скорость сравнения записи предъявленного образа с одной записью в БД на одном процессоре в первом узле 5000 сравнений/с, на одном процессоре во втором узле 250 000 сравнений/с, суточный поток запросов в узел 0 равен 60 000, вероятность обнаружения лица лежит в интервале 0,9–0,99, вероятность распознавания по отпечаткам пальцев равна 1,0 [7]. Предполагается, что в системе выполняются следующие условия:

- поиск в БД записи, совпадающей с записью предъявленного образа, осуществляется методом последовательного поиска;
- указанный в предыдущем пункте процесс поиска завершается либо по совпадению, либо после сравнения всех записей БД (т. е. осуществляется поиск по совпадению);
- в одном узле не может одновременно выполняться несколько заданий;
- задания, заставшие вычислительные ресурсы узла занятыми, становятся в очередь на выполнение в порядке поступления;
- задания в системе не теряются;
- времена пребывания в узле 0 и передачи заданий по коммуникационной среде не влияют на стационарные характеристики узлов (среднее

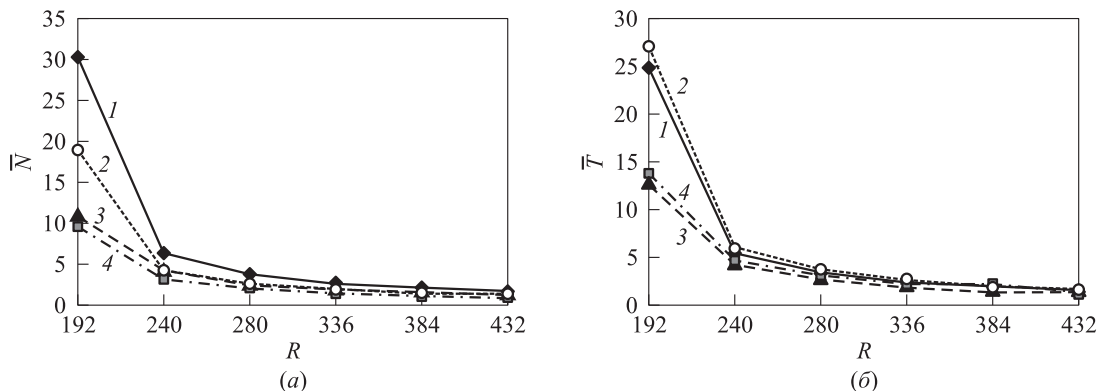


Рис. 2 Зависимость суммарного среднего числа заданий (а) и среднего времени пребывания заданий (б) в узлах 1 и 2 от общего объема процессоров

Распределение процессоров между узлами при последовательной процедуре

Вероятность обнаружения ($\alpha_{j1}, j = 2, 4, 6, 7$)	Среднее число заданий в системе	Среднее время пребывания задания в системе	Число процессоров в узлах		
			1	2	3
0,4	5,835192	4,167994	28	14	8
0,5	4,38	3,13	28	13	9
0,6	3,23	2,3	29	13	8
0,7	2,4	1,77	29	13	8
0,8	2,02	1,44	30	12	8
0,9	1,62	1,16	30	11	9

время пребывания заданий или среднее число заданий в узлах) и считаются равными нулю.

Для сокращения расчетов функция распределения вероятностей времени сравнения образцов в узле при условии, что в нем произойдет совпадение образа с одним из образцов, в примерах аппроксимирована равномерным распределением.

На рис. 2 приведены графики зависимости среднего числа заданий и среднего времени пребывания задания в системе от общего числа однородных процессоров в узлах 1 и 2 для параллельной (1 и 3) и

последовательной (2 и 4) процедур при вероятностях обнаружения лица 0,9 (1 и 2) и 0,99 (3 и 4). При этом процессоры распределены между узлами 1 и 2 так, чтобы величины нагрузок на эти узлы были примерно одинаковыми, а именно: объемы процессоров были выбраны в соотношении 1 : 8 (общее число однородных процессоров в указанных узлах меняется в пределах 192–432).

Во втором примере рассматривается система из четырех узлов (с номерами 0, 1, 2 и 3), где в трех узлах (с номерами 1, 2 и 3) размещены базы образцов (например, распределенные базы отпечатков

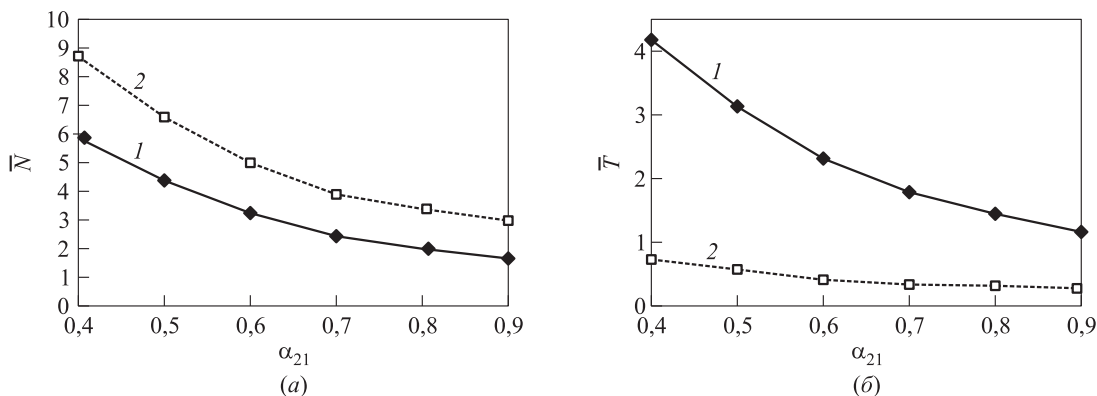


Рис. 3 Зависимость среднего числа заданий (а) и среднего времени пребывания задания (б) в системе от вероятности обнаружения $\alpha_{ji}, j = 1, \dots, 7, i = 1, 2, 3$ (1 — последовательная процедура, 2 — параллельная процедура)

пальцев), четвертый узел (с номером 0) выполняет функции центрального узла управления (как и в первом примере).

Считается, что в системе выполняются все условия, которые перечислены для системы в первом примере.

Также считается, что общее число однородных процессоров в узлах 1, 2 и 3 равно 50 и распределены они между узлами в таких соотношениях, чтобы среднее время пребывания произвольного задания в системе при последовательной процедуре было минимальным (см. таблицу).

На рис. 3 приведены графики зависимости суммарного среднего числа заданий и среднего времени пребывания задания в системе. Параметры системы имеют следующие значения: $M = 7$; $A_1 = \{1\}$; $A_2 = \{1, 2\}$; $A_3 = \{2\}$; $A_4 = \{2, 3\}$; $A_5 = \{3\}$; $A_6 = \{1, 3\}$; $A_7 = \{1, 2, 3\}$; $\lambda_1 = 0,2$; $\lambda_2 = 0,4$; $\lambda_3 = 0,8$; $\lambda_4 = 0,4$; $\lambda_5 = 1,6$; $\lambda_6 = 0,4$; $\lambda_7 = 2,6$; $\alpha_{j1} = 1$ для $j = 1, 3, 5$ и $\alpha_{j1} = \{0,4, 0,5, 0,6, 0,7, 0,8, 0,9\}$, $\alpha_{j2} = \alpha_{j1}/4$, $\alpha_{j2} = a_{j1}/16$ для $j = 2, 4, 6, 7$; $n_1\bar{u}_1 = 8$; $n_2\bar{u}_2 = 4$; $n_3\bar{u}_3 = 2$; время поиска образца в узлах 1, 2, 3 при условии, что там найдется образец, совпадающий с образом, имеет равномерное распределение на отрезках $[0, 8]$, $[0, 4]$, $[0, 2]$ соответственно.

4 Заключение

Сравнительный анализ результатов, полученных в данной работе при исследовании последовательной и параллельной процедур распознавания, показывает, что в системах распознавания образов рассмотренного выше типа при оптимальном (для последовательной процедуры) распределении вычислительных ресурсов:

- последовательная процедура предпочтительней, чем параллельная, в смысле среднего числа заданий, находящихся одновременно в системе;
- параллельная процедура предпочтительней, чем последовательная, в смысле среднего времени пребывания произвольного задания в системе;
- при малых нагрузках на вычислительные ресурсы (при $\bar{W}_{ji} \sim 0$) распределение вычислительных ресурсов, которое в системе с последовательной процедурой распознавания образов балансирует нагрузку, обеспечивает

среднее время пребывания в системе, близкое к оптимальному (в смысле выполнения условий (10)).

Рассмотренные в работе модели могут найти применение на этапах проектирования и эксплуатации распределенных систем распознавания рассмотренного выше типа для оценки оперативности системы и необходимой мощности вычислительных ресурсов в узлах системы при параллельной и последовательной процедурах распознавания.

Литература

1. Тель Ж. Введение в распределенные алгоритмы. — М.: МЦНМО, 2009. 616 с.
2. Горелик А.Л., Скрипкин В.А. Методы распознавания. — М.: Высшая школа, 2004. 261 с.
3. Хрулев А. Системы распознавания лиц. Состояние рынка. Перспективы развития // Системы безопасности, 2012. № 1. С. 70–72.
4. Бабенко Л.К., Ищукова Е.А., Сидоров И.Д. Параллельные вычисления в криптоанализе // Известия ЮФУ. Технические науки, 2012. Т. 137. № 12(137). С. 148–157.
5. Мурунин А.Б., Ковков Д.В., Лобанцов В.В., Маковкин К.А., Матвеев И.А., Десятников Д.Д., Чукупал В.Я. Мультимодальная биометрия — перспективное решение. Объединение алгоритмов для повышения надежности распознавания человека // Системы безопасности, 2006. № 6. С. 156–160.
6. Урмаев О.С., Босов А.В. Реализация концепции многофакторной биометрической идентификации в интегрированных аналитических системах // Бизнес и безопасность в России, 2008. № 49. С. 104–105.
7. Урмаев О.С. Проблемы распараллеливания биометрических вычислений в крупномасштабных идентификационных системах // Информатика и её применения, 2009. Т. 3. Вып. 1. С. 8–18.
8. Смирнов С.В. Подсистема массового распознавания изображений архивных документов // Тр. СПИИРАН, 2012. Вып. 3(22). С. 234–246.
9. Хант Э. Искусственный интеллект. — М.: Мир, 1978. 558 с.
10. Tamer Ezsü M., Valduriç P. Распределенные и параллельные системы баз данных // Системы управления базами данных, 1996. № 4. <http://www.osp.ru/dbms/1996/04/13031501>.
11. Клейнрок Л. Вычислительные системы с очередями. — М.: Мир, 1979. 600 с.
12. Корн Г., Корн Т. Справочник по математике. — М.: Наука, 1974. 831 с.

Поступила в редакцию 19.06.14

MODELS FOR COMPARATIVE ANALYSIS OF CLASSIFICATION METHODS IN DISTRIBUTED OBJECT RECOGNITION SYSTEMS

Ya. M. Agalarov

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper considers recognition systems where classes are defined by appropriate patterns located in distributed data base. Recognition criterion is full coincidence of the presented sample with at least one of the patterns. Parallel and sequential classification methods are compared in terms of mean response time to recognition request and performance requirements. The results of numerical experiments which were carried out for multibiometric recognition systems using analytical and simulation models of queueing networks are presented.

Keywords: distributed recognition system; parallel and sequential classification methods; resource allocation; queueing network

DOI: 10.14375/19922264140306

References

1. Tel, G. 1995. *Introduction to distributed algorithms*. Cambridge: Cambridge University Press. 546 p.
2. Gorelik, A. L., and V. A. Skripkin. 2004. *Metody raspoznavaniya* [Methods of recognition]. Moscow: Vysshaya Shkola. 261 p.
3. Khrulev, A. 2012. Sistemy raspoznavaniya lits. Sostoyanie rynka. Perspektivy razvitiya [Face recognition systems. State of the market. Prospects of development]. *Sistemy Bezopasnosti* [Security Systems] 1:70–72.
4. Babenko, L. K., E. A. Ishchukova, and I. D. Sidorov. 2012. Parallelnye vychisleniya v kriptanalize [Parallel computation in cryptanalysis]. *Izvestiya YuFU. Tekhnicheskie Nauki* [Gerald of South Federal University. Engineering Sciences] 12(137):148–157.
5. Murynin, A. B., D. V. Kovkov, V. V. Lobantsov, K. A. Makovkin, I. A. Matveev, D. D. Desyatnikov, and V. Ya. Chuchupal. 2006. Mul'timodal'naya biometriya — perspektivnoe reshenie. Ob"edinenie algoritmov dlya povysheniya nadezhnosti raspoznavaniya cheloveka [Multimodal biometrics is a promising solution. Merging algorithms to improve the reliability of recognition of a person]. *Sistemy Bezopasnosti* [Security Systems] 6:156–160.
6. Ushmaev, O. S., and A. V. Bosov. 2008. Realizatsiya konceptsii mnogofaktornoy biometricheskoy identifikatsii v integrirovannykh analiticheskikh sistemakh [Implementation of multimodal biometrics for integrated analytics]. *Biznes i Bezopasnost' v Rossii* [Business and Security in Russia] 49:104–105.
7. Ushmaev, O. S. 2009. Problemy rasparallelivaniya biometricheskikh vychisleniy v krupnomasshtabnykh identifikatsionnykh sistemakh [Parallel computing in large-scale multimodal biometric systems]. *Informatika i ee Primeneniya — Inform. Appl.* 3(1):8–18.
8. Smirnov, S. V. 2012. Podsystema massovogo raspoznavaniya izobrazheniy arkhivnykh dokumentov [Subsystem of mass image recognition of archival documents]. *Tr. SPIIRAN* [SPIIRAS Proceedings] 3(22):234–246.
9. Hunt, E. B. 1975. *Artificial intelligence*. New York: Academic Press. 468 p.
10. Tamer Ezsü, M., and P. Valduriez. 1996. Raspredelemnnye i parallelnye sistemy baz dannykh [Distributed and parallel database systems]. *Sistemy Upravleniya Bazami Danykh* [Databases Management Systems] 4. Available at: <http://www.osp.ru/dbms/1996/04/13031501> (accessed July 8, 2014).
11. Kleinrock, L. 1976. *Queueing systems: Volume II — Computer Applications*. New York: Wiley Interscience. 576 p.
12. Korn, G. A., and T. M. Korn. 1961. *Mathematical handbook for scientists and engineers*. New York: McGraw-Hill Book Co. 943 p.

Received June 19, 2014

Contributor

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; aglar@yandex.ru

ПРИМЕНЕНИЕ АЛГОРИТМОВ ОПТИМАЛЬНОЙ ФИЛЬТРАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ МОНИТОРИНГА ДОСТУПНОСТИ УДАЛЕННОГО СЕРВЕРА*

А. В. Борисов¹

Аннотация: Задача оперативного оценивания доступности удаленного сервера баз данных по протоколу http сформулирована в терминах оптимальной фильтрации состояния марковского процесса с конечным множеством состояний по наблюдениям конечномерного мультивариантного точечного процесса (МТП). Особенностью системы наблюдения является то, что случайная интенсивность наблюдаемого процесса является линейной функцией ненаблюдаемого марковского состояния. Оценка оптимальной фильтрации определяется с помощью решения некоторой замкнутой конечной системы дискретных рекуррентных соотношений и обыкновенных линейных дифференциальных уравнений со случайной правой частью. Применимость полученных теоретических результатов проиллюстрирована на примере мониторинга состояния пары «телекоммуникационный канал – сервер баз данных» с тремя допустимыми состояниями по наблюдениям ответов на запросы и сообщений об ошибках. Предложены возможные перспективы дальнейших научных исследований.

Ключевые слова: марковские модели; случайные скачкообразные процессы; оптимальная фильтрация; условное распределение вероятностей; системы массового обслуживания

DOI: 10.14375/19922264140307

1 Введение

Задачи определения состояний *систем массового обслуживания* (СМО) и их сетей в реальном масштабе времени по косвенным данным встречаются на практике в различных информационно-телекоммуникационных системах. Оценки состояний нужны для характеристики СМО и идентификации их параметров [1, 2]. Оценивание состояний СМО также является промежуточной процедурой при решении задачи оптимального управления потоком передачи данных [3–5] относительно различных критериев. Наконец, оперативное оценивание состояния отдельных элементов и систем обслуживания в целом может потребоваться при решении разнообразных задач балансировки нагрузки в децентрализованных информационных сетях и СМО с разделяемыми ресурсами [6, 7] при наличии только косвенных данных об их состоянии. Соответствующие алгоритмы должны служить основанием для создания эффективных программных продуктов.

При разработке специального программного обеспечения информационных веб-порталов [8, 9] необходимо рационально распределять пользовательские запросы, поступающие на портал, между различными информационными серверами, содер-

жащими идентичный контент. Сделать это можно, зная ресурсоемкость и алгоритмы обработки индивидуальных запросов, текущую загрузку и мощности серверов, а также характеристики и топологию телекоммуникационных каналов. На практике подобная детерминированная постановка задачи оптимизации не реализуема из-за отсутствия большинства перечисленных данных. Доступная информация подразделяется на две составляющие: априорные данные и статистическую информацию.

Прежде всего, СМО «телекоммуникационный канал – сервер» является разделяемым ресурсом: наряду с текущим запросом, она обслуживала и обслуживает также и множество других. Марковская модель смены состояний этой СМО описывает ее с достаточной степенью адекватности. Очевидно, что одним из состояний будет «отсутствие связи», а остальные соответствуют наличию связи и разной степени загруженности СМО. Интенсивности перехода СМО в различные состояния могут быть оценены по предварительно полученной статистической информации [1, 2].

Каждый индивидуальный запрос к исследуемой СМО обслуживается случайное время, по истечении которого пользователю выдается либо сообщение об ошибке, либо предметный ответ на за-

* Работа поддержана грантами РФФИ № 13-01-00406 и № 13-07-00408.

¹ Институт проблем информатики Российской академии наук; факультет прикладной математики и физики Московского авиационного института (национального исследовательского университета); A.Borisov@ipiran.ru

прос. Сообщения об ошибке бывают различными и обычно связаны с истечением какого-либо времени ожидания — тайм-аутом. Распределение времени получения ответа на запрос зависит от особенностей функционирования интернет-сервера [10], семантики самого запроса [11] и объема базы данных. Вероятности появления сообщений об ошибках и распределение времени получения ответа зависят от состояния СМО, в котором она пребывает в процессе всего времени обработки запроса. Все эти распределения можно предварительно идентифицировать по статистическим данным. При этом максимальное время ожидания ответа или сообщения об ошибке всегда ограничено известным тайм-аутом.

Запросы к серверу образуют процесс восстановления общего вида, не являющийся ни пуассоновским, ни даже процессом Кокса. Процесс исходящих предметных ответов или сообщений об ошибках порождается входным процессом *разнородных* запросов и также является обобщенным процессом восстановления.

Прикладная задача заключается в определении оценок состояния СМО «телекоммуникационный канал — сервер» по наблюдениям входного потока запросов и выходного потока предметных ответов и сообщений об ошибках.

Основной целью данной работы является формализация и решение поставленной практической задачи оценивания. В разд. 2 она сформулирована в терминах оценивания состояний динамических систем наблюдения, описываемых стохастическими дифференциальными уравнениями с мартингалами в правой части. Ненаблюдаемый процесс, определяющий состояние системы, предполагается марковским с конечным множеством состояний. Процесс наблюдений является мультивариантным и также принимает значения из конечного множества. Основной особенностью предложенной модели является представление наблюдений в форме мультивариантного точечного процесса, случайная интенсивность которого является функционалом от предыдущих состояний и наблюдений. При этом зависимость интенсивности от ненаблюдаемого состояния предполагается линейной. В разд. 3 проведен анализ предложенной системы наблюдения и представлены базовые выводы о распределении процесса наблюдений. Основной результат — уравнения оптимальной фильтрации — доказан в разд. 4. Раздел 5 содержит иллюстративный численный пример. В нем дано решение задачи оценивания доступности удаленного сервера баз данных по протоколу http по результатам запросов к нему. В задаче предполагается, что СМО «телекоммуникационный канал — сер-

вер» находится в одном из трех состояний: «связи нет», «сервер не загружен» и «сервер загружен», — а сообщения могут быть двух видов: «предметный ответ» и «сообщение об ошибке». Заключительные замечания и возможные перспективы развития предложенных исследований представлены в разд. 6.

2 Формальная постановка задачи

На полном вероятностном пространстве с фильтрацией $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \in \mathbb{R}_+})$ задан ненаблюдаемый *марковский скачкообразный процесс* (МСП) $\{X(t)\}_{t \in \mathbb{R}_+}$ [12]

$$X(t) = X_0 + \int_0^t \Lambda^\top(s)X(s) ds + M^X(t), \quad (1)$$

где $X(t) \in \mathbb{S}_N$ — состояние процесса ($\mathbb{S}_N \triangleq \{e_1, \dots, e_N\} \subset \mathbb{R}^N$ — множество единичных вектор-столбцов евклидова пространства \mathbb{R}^N); $\{\Lambda(t)\}_{t \in \mathbb{R}_+}$ — матрица интенсивностей МСП ($\Lambda(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}^{N \times N}$); $\{M^X(t)\}_{t \in \mathbb{R}_+}$ — \mathcal{F}_t -согласованный мартингал, $M^X(t) \in \mathbb{R}^N$. Здесь и далее для обозначения множества, по которому выполняется интегрирование, используется обозначение $\int_a^b \dots \triangleq \int_{(a,b)} \dots$, в противном случае множество будет указываться явным образом.

В качестве наблюдений выступает *мультивариантный точечный процесс* [13] $(\tau_n, Y_n)_{n \in \mathbb{N}}$:

- $\{\tau_n\}$ — возрастающая последовательность марковских моментов;
- $Y_n \in \mathbb{S}_M$ — значение наблюдения в момент τ_n ($\mathbb{S}_M \triangleq \{f_1, \dots, f_M\} \subset \mathbb{R}^M$ — множество единичных вектор-столбцов евклидова пространства \mathbb{R}^M).

С наблюдениями $\{(\tau_n, Y_n)\}_{n \in \mathbb{N}}$ связана неубывающая последовательность σ -подалгебр $\{\mathcal{G}(n)\}_{n \in \mathbb{N}}$: $\mathcal{G}(n) \triangleq \sigma\{(\tau_k, Y_k), 1 \leq k \leq n\}$; помимо этого полагается, что $\mathcal{G}(0) \triangleq \{\emptyset, \Omega\}$.

Наблюдениям в форме МТП однозначно соответствует случайный процесс $Y(t)$ в непрерывном времени со считающими компонентами

$$Y(t) \triangleq \sum_{n \in \mathbb{N}} Y_n \mathbf{I}_{[\tau_n, +\infty)}(t) :$$

j -я компонента $Y^j(t)$ ($j = \overline{1, M}$) равна числу наблюдений с исходом j -го типа ($Y_k = f_j, \tau_k \leq t$), появившихся на отрезке времени $[0, t]$.

По непрерывному процессу $Y(t)$ можно построить естественный поток σ -подалгебр $\mathcal{G}_t \triangleq$

$\triangleq \sigma\{Y(s), s \leq t\}$, при этом $\forall n \in \mathbb{N}$ верно тождество $\mathcal{G}_{\tau_n} \equiv \mathcal{G}(n)$ [14].

Для процесса $Y(t)$ существует мартингалное представление [15]. По условию рассматриваемой задачи оно имеет следующий вид (далее в изложении полагается $\tau_0 \triangleq 0$):

$$Y(t) = \sum_{n \in \mathbb{N}} \int_{\tau_{n-1} \wedge t}^{\tau_n \wedge t} (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \Phi_n(ds) + \sum_{n \in \mathbb{N}} M_n^Y(t). \quad (2)$$

Здесь $\phi_n^i(t) = \phi_n^i(\omega, t)$ ($i = \overline{0,1}$, $n \in \mathbb{N}$), $\phi_n^i(\cdot) : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^{M \times N} - \mathcal{G}(n-1)$ -измеримые векторные процессы; $\Phi_n(dt) = \Phi_n(\omega, dt) - \mathcal{G}(n-1)$ -измеримые неотрицательные стохастические σ -конечные меры; $M_n^Y(t) - \mathcal{F}_t$ -согласованные векторные мартингалы, $M_n^Y(t) \in \mathbb{R}^M$.

Задача оптимальной фильтрации состояния X по наблюдениям МТП $\{(\tau_n, Y_n)\}$ заключается в вычислении *условного математического ожидания* (УМО):

$$\widehat{X}(t) \triangleq \mathbf{E} \{X(t) | \mathcal{G}_t\}.$$

Благодаря предложенной модели наблюдений (2) данная система имеет ряд существенных особенностей.

Во-первых, в отличие от традиционного вероятностного определения наблюдаемого МТП с помощью семейства условных конечномерных распределений [13], в данной работе модель наблюдений задана в тех же терминах, что и динамика состояния, т. е. в форме мартингалного представления. Помимо единообразия в описании данные уравнения представляют «кинматику» наблюдений более наглядно. Подобное представление также иллюстративно для практиков. Действительно, j -я компонента подынтегрального выражения (2) определяет вероятность немедленного появления в момент времени s n -го наблюдения Y_n с j -м исходом при условии, что до момента s наблюдение Y_n не появилось [15], т. е.

$$f_j^\top [\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)] \Phi_n(ds) = \mathbf{P} \{ \tau_n \in [s, s + ds), Y_n = f_j | \mathcal{G}(n-1) \wedge \{ \tau_n \geq s \} \}.$$

Во-вторых, модель (2) является достаточно общей. Она включает классический случай нестационарного пуассоновского процесса, интенсивность которого зависит от текущего состояния системы X [16]. Она также описывает ситуацию, когда интенсивность появления наблюдений и их значения зависят от состояния системы только в момент предыдущего наблюдения [17].

В-третьих, первое слагаемое в (2) — компенсатор — является случайной разрывной функцией. Это свойство компенсатора не позволяет преобразовать наблюдения $Y(t)$ в процесс с независимыми приращениями ни с помощью гирсановской замены вероятностной меры, ни с помощью замены времени [14]. Это ведет к тому, что поставленная задача оценивания может быть решена только с помощью общей формулы фильтрации семимартингала по наблюдениям семимартингала [13].

3 Определение вероятностных характеристик наблюдений

Прежде чем приступить к решению задачи оптимальной фильтрации, необходимо формализовать свойства системы наблюдения. Далее в работе предполагается, что для (1), (2) выполнены следующие условия:

- (а) все траектории процессов $X(t)$ и $Y(t)$ непрерывны справа;
- (б) общий поток σ -алгебр $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ порожден только процессами $X(t)$ и $Y(t)$, т. е. $\mathcal{F}_t = \sigma\{X(s), Y(s) : 0 \leq s \leq t\}$; помимо этого поток $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ непрерывен справа, т. е. $\bigcap_{s: s > t} \mathcal{F}_s = \mathcal{F}_t \forall t > 0$;
- (в) для МСП $X(t)$ вектор начального распределения $p(0) \triangleq \mathbf{E} \{X_0\}$ и матрица интенсивностей переходов $\Lambda(t)$ известны; при этом все компоненты $\Lambda(t)$ являются кусочно-непрерывными функциями времени t ;
- (г) моменты поступления наблюдений $\{\tau_n\}_{n \geq 0}$ удовлетворяют условиям

$$0 \triangleq \tau_0 < \tau_1 < \tau_2 < \dots \text{ и } \lim_{k \rightarrow \infty} \tau_n = +\infty \quad \mathbf{P}\text{-п. н.};$$

при этом определено подмножество множества натуральных чисел $\mathbb{N}_1 \subset \mathbb{N}$, обладающее следующим свойством: $\forall n \in \mathbb{N}_1$ определена такая $\mathcal{G}(n-1)$ -измеримая случайная величина $T_n = T_n(\tau_1, \dots, \tau_{n-1})$, что

$$\mathbf{P} \{ \tau_{n-1} < \tau_n \leq T_n < +\infty | \mathcal{G}(n-1) \} = 1 \quad \forall n \in \mathbb{N}_1$$

или

$$\mathbf{P} \{ \tau_n > v | \mathcal{G}(n-1) \} = 1 \quad \forall n \in \mathbb{N} \setminus \mathbb{N}_1 \text{ и } v > 0;$$

- (д) мартингалы в уравнении состояния и наблюдениях сильно ортогональны, т. е.

$$[M^X, M_n^Y]_t \triangleq \sum_{s \leq t} \Delta M^X(s) (\Delta M_n^Y(s))^\top \equiv 0$$

\mathbf{P} -п. н. $\forall t \in \mathbb{R}_+$ и $n \in \mathbb{N}$;

(е) для меры $\Phi_n(\cdot)$ верно разложение

$$\Phi_n(\cdot) = \Phi_n^c(\cdot) + \Phi_n^d(\cdot),$$

где $\Phi_n^c(\cdot)$ — абсолютно непрерывная по мере Лебега, а $\Phi_n^d(\cdot)$ — дискретная (считающая) составляющие, а также $\Phi_n([0, \tau_{n-1}]) \equiv 0$ \mathbf{P} -п. н.; $\forall n \in \mathbb{N}$ меры $\Phi_n(\cdot)$ содержат лишь конечное множество скачков $\mathbb{U}_n \triangleq \{u_\ell^n\}_{\ell=\overline{1, K_n}} - \mathcal{G}(n-1)$ -измеримых моментов.

Для функций $\phi_n^i(\cdot)$ \mathbf{P} -п. н. выполняются условия:

$$\left. \begin{aligned} & (\phi_n^i(s))_{kj} \geq 0, \\ & \forall i = \overline{0, 1}, j = \overline{1, N}, k = \overline{1, M}, s \in \mathbb{R}_+; \\ & \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) > 0, \\ & \Phi_n - \text{п. в. на } (\tau_{n-1}, T_n] \forall n \in \mathbb{N}_1; \\ & \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) > 0, \\ & \Phi_n - \text{п. в. на } (\tau_{n-1}, +\infty) \forall n \in \mathbb{N} \setminus \mathbb{N}_1; \\ & \mathbf{1}_M^\top (\phi_n^0(u_\ell^n)X(\tau_{n-1}) + \\ & \quad + \phi_n^1(u_\ell^n)X(u_\ell^n-)) \Phi_n^d(\{u_\ell^n\}) \leq 1; \\ & \exp \left\{ - \int_{\tau_{n-1}}^{T_n} \mathbf{1}_M^\top (\phi_n^0(s)e_{k_1} + \right. \\ & \quad \left. + \phi_n^1(s)e_{k_2}) \Phi_n^c(ds) \right\} \times \\ & \quad \times \prod_{\tau_{n-1} < u_n^\ell \leq T_n} [1 - \mathbf{1}_M^\top (\phi_n^0(u_n^\ell)e_{k_3} + \\ & \quad + \phi_n^1(u_n^\ell)e_{k_4}) \Phi_n^d(\{u_n^\ell\})] = 0, \\ & \quad \forall k_i = \overline{1, N}, n \in \mathbb{N}_1; \\ & \exp \left\{ - \int_{(\tau_{n-1}, +\infty)} \mathbf{1}_M^\top (\phi_n^0(s)e_{k_1} + \right. \\ & \quad \left. + \phi_n^1(s)e_{k_2}) \Phi_n^c(ds) \right\} \times \\ & \quad \times \prod_{\tau_{n-1} < u_n^\ell} [1 - \mathbf{1}_M^\top (\phi_n^0(u_n^\ell)e_{k_3} + \\ & \quad + \phi_n^1(u_n^\ell)e_{k_4}) \Phi_n^d(\{u_n^\ell\})] = 0, \\ & \quad \forall k_i = \overline{1, N}, n \in \mathbb{N} \setminus \mathbb{N}_1, \end{aligned} \right\} (3)$$

где $\mathbf{1}_L$ — L -мерный вектор-столбец, составленный из единиц.

Далее, необходимо построить совместное распределение наблюдений $\{(\tau_n, Y_n)\}$ и состояния X . Распределение процесса X известно [18]. Пусть $\{\sigma_n\}_{n \in \mathbb{N}}$ — последовательность моментов скачков X , тогда $\forall t \in \mathbb{R}_+$ и $j \neq i$

$$\mathbf{P}\{\sigma_{n+1} \leq t, X(\sigma_{n+1}) = e_j | \mathcal{F}_{\sigma_n}\} = \mathbf{I}_{[\sigma_n, +\infty)}(t) \times \\ \times \sum_{i=1}^N e_i^\top X(\sigma_n) \int_{\sigma_n}^t \lambda_{ij}(s) \exp \left[\int_{\sigma_n}^s \lambda_{ii}(u) du \right] ds.$$

Осталось лишь построить условное распределение $\{(\tau_n, Y_n)\}$ относительно X .

Прежде всего, с каждым МТП $\{(\tau_n, Y_n)\}$ взаимно однозначным образом связана стохастическая мера μ :

$$\mu(\omega, dt, dy) \triangleq \sum_{n \in \mathbb{N}} \delta_{(\tau_n(\omega), Y_n(\omega))}(dt, dy).$$

Для вывода уравнений фильтрации необходимо определить компенсатор $\nu(\omega, dt, dy)$ меры $\mu(\omega, dt, dy)$, т. е. такую \mathcal{G}_t -предсказуемую стохастическую меру, для которой равенство

$$\mathbf{E} \left\{ \int_{\mathbb{R}_+ \times \mathbb{R}^M} \psi(\omega, t, y) \mu(\omega, dt, dy) \right\} = \\ = \mathbf{E} \left\{ \int_{\mathbb{R}_+ \times \mathbb{R}^M} \psi(\omega, t, y) \nu(\omega, dt, dy) \right\}$$

выполняется для любого неотрицательного ограниченного \mathcal{G}_t -предсказуемого случайного процесса $\psi(\omega, t, y)$. В дальнейшем изложении в обозначении процессов $\psi(\omega, t, y)$ и стохастических мер $\mu(\omega, dt, dy)$ для краткости зависимость от случайного события ω будет опускаться, т. е. $\psi(t, y) \triangleq \psi(\omega, t, y)$ и $\mu(dt, dy) \triangleq \mu(\omega, dt, dy)$. Будет также использоваться обозначение $\widehat{X}(t, u) \triangleq \mathbf{E}\{X(t) | \mathcal{G}_u\}$ ($u > t$) для оценки оптимального сглаживания состояния в фиксированной точке t по наблюдениям на отрезке $[0, u]$.

Как и в случае классического фильтра Калмана–Бьюси [19] и обобщения фильтра Вонэма [16], оценку оптимальной фильтрации в данной задаче удастся представить в конечномерном виде: в каждый момент времени t искомая оценка $\widehat{X}(t)$ выражается с помощью решения некоторой конечной системы дифференциальных и алгебраических уравнений. Это означает, что условное распределение $X(t)$ относительно \mathcal{G}_t характеризуется некоторым конечным набором достаточных статистик. В данной системе наблюдений такими статистиками являются матричнозначные функции $\overline{Z}_n(t) : [\tau_{n-1}, \tau_n) \rightarrow \mathbb{R}^{N \times N}$, $n \in \mathbb{N}$:

$$\overline{Z}_n(t) \triangleq \mathbf{E} \left\{ X(\tau_{n-1}) \widetilde{I}_n(t) X^\top(t) | \mathcal{G}(n-1) \right\},$$

где

$$\widetilde{I}_n(t) \triangleq \mathbf{E} \left\{ \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) | \mathcal{F}^X \vee \mathcal{G}(n-1) \right\}.$$

Следующее вспомогательное утверждение определяет искомое условное распределение $\{(\tau_n, Y_n)\}$ относительно его предыстории и МСП X , уравнение для процессов $\{\bar{Z}_n(\cdot)\}_{n \in \mathbb{N}}$, а также компенсатор ν меры μ .

Лемма 1.

1. Условная функция распределения момента скачка τ_n относительно $\mathcal{F}^X \vee \mathcal{G}(n-1)$ определяется формулой ($t > \tau_{n-1}$):

$$\begin{aligned} \mathbf{P} \{ \tau_n \leq t | \mathcal{F}^X \vee \mathcal{G}(n-1) \} &= \\ &= \mathbf{P} \{ \tau_n \leq t | \mathcal{F}_t^X \vee \mathcal{G}(n-1) \} = \\ &= 1 - \exp \left\{ - \int_{\tau_{n-1}}^t \mathbf{1}_M^\top (\phi_n^0(s) X(\tau_{n-1}) + \right. \\ &\quad \left. + \phi_n^1(s) X(s-)) \Phi_n^c(ds) \right\} \times \\ &\times \prod_{\tau_{n-1} < s \leq t} [1 - \mathbf{1}_M^\top (\phi_n^0(s) X(\tau_{n-1}) + \\ &\quad + \phi_n^1(s) X(s-)) \Phi_n^d(\{s\})] . \quad (4) \end{aligned}$$

2. Совместное условное распределение момента скачка τ_n и значения наблюдения Y_n относительно $\mathcal{F}^X \vee \mathcal{G}(n-1)$ определяется формулой ($t > \tau_{n-1}$):

$$\begin{aligned} \mathbf{P} \{ \tau_n \leq t, Y_n = f_i | \mathcal{F}^X \vee \mathcal{G}(n-1) \} &= \\ &= \mathbf{P} \{ \tau_n \leq t, Y_n = f_i | \mathcal{F}_t^X \vee \mathcal{G}(n-1) \} = \\ &= \int_{\tau_{n-1}}^t f_i^\top (\phi_n^0(s) X(\tau_{n-1}) + \phi_n^1(s) X(s-)) \times \\ &\quad \times \exp \left\{ - \int_{\tau_{n-1}}^s \mathbf{1}_M^\top (\phi_n^0(u) X(\tau_{n-1}) + \right. \\ &\quad \left. + \phi_n^1(u) X(u-)) \Phi_n^c(du) \right\} \times \\ &\times \prod_{\tau_{n-1} < u < s} [1 - \mathbf{1}_M^\top (\phi_n^0(u) X(\tau_{n-1}) + \\ &\quad + \phi_n^1(u) X(u-)) \Phi_n^d(\{u\})] \Phi_n(ds) . \end{aligned}$$

3. На каждом полуинтервале $[\tau_{n-1}, \tau_n)$ статистика $\{\bar{Z}_n(\cdot)\}_{n \in \mathbb{N}}$ является единственным решением системы обыкновенных линейных дифференциальных уравнений с разрывной случайной правой частью:

$$\begin{aligned} \bar{Z}_n(t) &= \text{diag}(\hat{X}(\tau_{n-1})) + \int_{\tau_{n-1}}^t \bar{Z}_n(s-) \Lambda(s) ds - \\ &- \int_{\tau_{n-1}}^t [\text{diag}(\mathbf{1}_M^\top \phi_n^0(s)) \bar{Z}_n(s-) + \\ &\quad + \bar{Z}_n(s-) \text{diag}(\mathbf{1}_M^\top \phi_n^1(s))] \Phi_n(ds) . \quad (5) \end{aligned}$$

4. Компенсатор ν меры μ равен

$$\begin{aligned} \nu(dt, dy) &= \\ &= \sum_{n \in \mathbb{N}} \sum_{j=1}^M \mathbf{I}_{(\tau_{n-1}, \tau_n]}(t) y^\top (\phi_n^0(t) \bar{X}_n(\tau_{n-1}, t-) + \\ &\quad + \phi_n^1(t) \bar{X}_n(t-)) \delta_{\{f_j\}}(dy) \Phi_n(dt) , \quad (6) \end{aligned}$$

где процессы

$$\begin{aligned} \bar{X}_n(\tau_{n-1}, t) &\triangleq \frac{1}{\mathbf{E} \{ \tilde{I}_n(t) | \mathcal{G}(n-1) \}} \times \\ &\times \mathbf{E} \{ X(\tau_{n-1}) \tilde{I}_n(t) | \mathcal{G}(n-1) \} \quad (7) \end{aligned}$$

и

$$\begin{aligned} \bar{X}_n(t) &\triangleq \frac{1}{\mathbf{E} \{ \tilde{I}_n(t) | \mathcal{G}(n-1) \}} \times \\ &\times \mathbf{E} \{ X(t) \tilde{I}_n(t) | \mathcal{G}(n-1) \} \quad (8) \end{aligned}$$

выражаются через статистики $\bar{Z}_n(t)$:

$$\left. \begin{aligned} \bar{X}_n(\tau_{n-1}, t) &= \frac{1}{\mathbf{1}_N^\top \bar{Z}_n(t) \mathbf{1}_N} \bar{Z}_n(t) \mathbf{1}_N ; \\ \bar{X}_n(t) &= \frac{1}{\mathbf{1}_N^\top \bar{Z}_n(t) \mathbf{1}_N} \bar{Z}_n^\top(t) \mathbf{1}_N . \end{aligned} \right\} \quad (9)$$

Доказательство леммы 1 приведено в приложении.

Лемма 1 позволяет получить важное следствие, необходимое в дальнейшем для обеспечения того, чтобы искомая оценка фильтрации являлась регулярной версией условного распределения. Для этого рассмотрим множества \mathcal{O}_n^i ($n \in \mathbb{N}$, $i = \overline{1, M}$)

$$\begin{aligned} \mathcal{O}_n^i &\triangleq \{ \omega \in \Omega : \tau_n \geq \tau_{n-1}, \\ Y_n = f_i, f_i^\top (\phi_n^0(\tau_n) X(\tau_{n-1}) + \phi_n^1(\tau_n) X(\tau_n-)) = 0, \\ &\quad \mathbf{1}_M^\top (\phi_n^0(\tau_n) X(\tau_{n-1}) + \phi_n^1(\tau_n) X(\tau_n-)) > 0 \} \end{aligned}$$

и вычислим условные вероятности $\mathbf{P} \{ \mathcal{O}_n^i | \mathcal{F}^X \vee \mathcal{G}(n-1) \}$. С использованием утверждения 2 леммы 1 это происходит следующим образом:

$$\begin{aligned}
 & \mathbf{P} \{ \mathcal{O}_n^i | \mathcal{F}^X \vee \mathcal{G}(n-1) \} = \\
 & = \mathbf{E} \{ I_{\mathcal{O}_n^i}(\omega) | \mathcal{F}^X \vee \mathcal{G}(n-1) \} = \\
 & = \int \mathbf{P} \{ \tau_n \in ds, \\
 & \left\{ \begin{array}{l} s \geq \tau_{n-1}: f_i^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) = 0; \\ \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) > 0 \end{array} \right\} \\
 & Y_n = f_i | \mathcal{F}^X \vee \mathcal{G}(n-1) \} = \\
 & = \int f_i^\top (\phi_n^0(s)X(\tau_{n-1}) + \\
 & \left\{ \begin{array}{l} s \geq \tau_{n-1}: f_i^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) = 0; \\ \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) > 0 \end{array} \right\} \\
 & + \phi_n^1(s)X(s-)) \exp \left\{ - \int_{\tau_{n-1}}^s \mathbf{1}_M^\top (\phi_n^0(u)X(\tau_{n-1}) + \right. \\
 & \left. + \phi_n^1(u)X(u-)) \Phi_n^c(du) \right\} \times \\
 & \times \prod_{\tau_{n-1} < u < s} [1 - \mathbf{1}_M^\top (\phi_n^0(u)X(\tau_{n-1}) + \\
 & + \phi_n^1(u)X(u-)) \Phi_n^d(\{u\})] \Phi_n(ds) = 0.
 \end{aligned}$$

Таким образом, доказано

Следствие 1. $\forall n \in \mathbb{N}$ и $i = \overline{1, M}$ \mathbf{P} -п. н. верно равенство

$$\mathbf{P} \{ \mathcal{O}_n^i | \mathcal{G}(n-1) \vee \mathcal{F}^X \} = 0.$$

С практической точки зрения смысл следствия вполне прозрачен: если в момент времени t появилось наблюдение Y_n и интенсивность появления исхода наблюдения f_i равна 0, т.е. $f_i^\top (\phi_n^0(t)X(\tau_{n-1}) + \phi_n^1(t)X(t-)) = 0$, то вероятность того, что $Y_n = f_i$, равна 0.

Утверждения леммы 1 также позволяют сделать важные выводы как относительно вида компенсатора ν , так и относительно искомой оценки фильтрации $\hat{X}(t)$.

Теорема 1.

1. Оценка $\hat{X}(t)$ оптимальной фильтрации состояния $X(t)$ имеет вид:

$$\begin{aligned}
 \hat{X}(t) &= \sum_{n \in \mathbb{N}} \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \times \\
 & \times \mathbf{E} \{ X(t) | \mathcal{G}(n-1) \vee \{ \tau_n > t \} \}, \quad (10)
 \end{aligned}$$

где

$$\begin{aligned}
 & \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \mathbf{E} \{ X(t) | \mathcal{G}(n-1) \vee \{ \tau_n > t \} \} = \\
 & = \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \overline{X}_n(t) \quad \mathbf{P} \text{ - п. н.}, \quad (11)
 \end{aligned}$$

а $\overline{X}_n(t)$, в свою очередь, определяется формулами (9) и (5).

2. Компенсатор ν меры μ равен

$$\begin{aligned}
 & \nu(dt, dy) = \\
 & = \sum_{n \in \mathbb{N}} \sum_{j=1}^M \mathbf{I}_{(\tau_{n-1}, \tau_n]}(t) y^\top \left(\phi_n^0(t) \hat{X}(\tau_{n-1}, t-) + \right. \\
 & \left. + \phi_n^1(t) \hat{X}(t-) \right) \delta_{\{f_j\}}(dy) \Phi_n(dt). \quad (12)
 \end{aligned}$$

Доказательство теоремы 1 приведено в приложении.

Утверждение 1 теоремы дает весьма важное представление о разнице между «дискретными» σ -алгебрами $\mathcal{G}(n)$ и их «непрерывными» аналогами \mathcal{G}_t , о которой упоминалось в [20]. Если до момента времени t произошло $n-1$ скачков процесса $Y(t)$, то \mathcal{G}_t — минимальная σ -алгебра, содержащая $\mathcal{G}(n-1)$, а также событие $\{\omega : \tau_n(\omega) > t\}$.

Из определений (7) и (8), уравнения (5) для вычисления $\overline{Z}_n(t)$ и вида (10) для УМО $\hat{X}(t)$ следует, что компенсатор (12) может быть записан более компактно. Для этого на промежутках $[\tau_{n-1}, \tau_n)$ введем в рассмотрение вспомогательные матричнозначные процессы с траекториями, \mathbf{P} -п. н. непрерывными слева:

$$\begin{aligned}
 \beta_n(t) &\triangleq \phi_n^0(t) \mathbf{E} \{ X(\tau_{n-1}) \tilde{I}_n(t-) X^\top(t-) | \mathcal{G}(n-1) \} + \\
 & + \phi_n^1(t) \mathbf{E} \{ X(t-) \tilde{I}_n(t-) X^\top(t-) | \mathcal{G}(n-1) \} = \\
 & = \phi_n^0(t) \overline{Z}_n(t-) + \phi_n^1(t) \text{diag}(\mathbf{1}_N^\top \overline{Z}_n(t-)); \quad (13)
 \end{aligned}$$

$$\begin{aligned}
 \hat{\beta}_n(t) &\triangleq \phi_n^0(t) \mathbf{E} \{ X(\tau_{n-1}) X^\top(t-) | \mathcal{G}_{t-} \} + \\
 & + \phi_n^1(t) \mathbf{E} \{ X(t-) X^\top(t-) | \mathcal{G}_{t-} \} = \\
 & = \frac{1}{\mathbf{1}_N^\top \overline{Z}_n(t-) \mathbf{1}_N} \beta_n(t). \quad (14)
 \end{aligned}$$

При этом выполняются равенства:

$$\begin{aligned}
 & \phi_n^0(t) \mathbf{E} \{ X(\tau_{n-1}) \tilde{I}_n(t-) | \mathcal{G}(n-1) \} + \\
 & + \phi_n^1(t) \mathbf{E} \{ X(t-) \tilde{I}_n(t-) | \mathcal{G}(n-1) \} = \beta_n(t) \mathbf{1}_N;
 \end{aligned}$$

$$\phi_n^0(t) \hat{X}(\tau_{n-1}, t-) + \phi_n^1(t) \hat{X}(t-) = \hat{\beta}_n(t) \mathbf{1}_N.$$

Таким образом, компенсатор ν также может быть записан в виде:

$$\begin{aligned}
 & \nu(dt, dy) = \\
 & = \sum_{n \in \mathbb{N}} \sum_{j=1}^M \mathbf{I}_{(\tau_{n-1}, \tau_n]}(t) \mathbf{1}_N^\top \hat{\beta}_n^\top(t) y \delta_{\{f_j\}}(dy) \Phi_n(dt).
 \end{aligned}$$

4 Решение задачи фильтрации

Следующее утверждение представляет основной результат — формулы для вычисления искомой оценки оптимальной фильтрации.

Теорема 2. Пусть для системы наблюдений (1), (2) выполнены условия (а)–(е). Тогда регулярная версия условного распределения $\hat{X}(t)$ вычисляется с помощью решения следующих рекуррентно связанных дифференциальных и алгебраических уравнений:

(1) при $t = 0$

$$\hat{X}(0) = p_0;$$

(2) при $t \in (\tau_{n-1}, \tau_n)$

$$\hat{X}(t) = \frac{1}{\mathbf{1}_N^T \bar{Z}_n(t) \mathbf{1}_N} \bar{Z}_n^T(t) \mathbf{1}_N,$$

где $\bar{Z}_n(t)$ является решением уравнения (5);

(3) при $t = \tau_n$

$$\hat{X}(\tau_n) = \begin{cases} \hat{X}(\tau_n-), & \text{если } \mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) Y_n = 0; \\ \frac{1}{\mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) Y_n} \hat{\beta}_n^T(\tau_n) Y_n, & \text{если } \mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) Y_n > 0 \\ \text{и } \mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) < 1; \\ \frac{1}{\mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) Y_n} \hat{\beta}_n^T(\tau_n) Y_n - \\ - \left[\hat{\beta}_n^T(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) - \hat{X}(\tau_n-) \right], & \text{если } \mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) Y_n > 0 \\ \text{и } \mathbf{1}_N^T \hat{\beta}_n^T(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) = 1, \end{cases} \quad (15)$$

где $\hat{\beta}_n(t)$ определяется формулами (13) и (14).

Доказательство теоремы 2 приведено в приложении.

5 Численный пример

В качестве иллюстративного примера рассматривается задача оперативного оценивания состояния удаленного сервера по результатам одного запроса к нему. Предполагается, что состояние пары «телекоммуникационный канал – сервер» $X(t)$ недоступно прямому наблюдению и описывается однородным МСП с тремя возможными состояниями:

- (1) $X(t) = e_1$ — с сервером отсутствует связь;
- (2) $X(t) = e_2$ — с сервером есть связь, сервер не загружен;
- (3) $X(t) = e_3$ — с сервером есть связь, сервер загружен.

Матрица интенсивности Λ переходов МСП $X(t)$ известна, в начальный момент времени $t = 0$ распределение $X(0)$ совпадает со стационарным.

О состоянии удаленного сервера имеется косвенная информация — результат отправленного в начальный момент времени запроса к данному серверу. Возможными исходами запроса являются:

- $Y_1 = f_1$ — «предметный» ответ сервера;
- $Y_1 = f_2$ — общее сообщение об ошибке.

Результат запроса к серверу появляется в случайный момент времени τ_1 , распределение которого зависит от состояния X . О модели результата запроса к серверу доступна следующая априорная информация. Сообщение об ошибке может появиться только в один из дискретных моментов истечения времен ожидания (тайм-аутов):

- t_1 — тайм-аут, связанный с временем ожидания установления сетевого соединения;
- t_2 — предельное время выполнения запроса системой управления базой данных на удаленном сервере;
- t_3 — предельное время формирования результатов запроса на интернет-сервере;
- t_4 — тайм-аут, связанный с временем передачи по сети результатов выполнения запроса.

Если в начальный момент времени связь с сервером отсутствует, то сервер выдает сообщение об ошибке. Первое слагаемое интенсивности наблюдений имеет вид:

$$\int_0^t \phi^0(s) X(0) \Phi(ds) = \begin{bmatrix} \int_0^t \frac{\sum_{j=1}^4 p_{1j} \delta_{t_j}(ds)}{\int_0^s \sum_{k=1}^4 p_{1k} \delta_{t_k}(du)} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} X(0).$$

Если удаленный сервер пребывает в состоянии e_i ($i = 2, 3$), то «предметный» ответ поступает через случайное время, имеющее плотность распределения $\psi_i(t)$. В этих же состояниях сервер также может выдавать общее сообщение об ошибках: это связано с превышением времени выполнения запросов,

подготовки их отправки и пр. Второе слагаемое интенсивности имеет вид:

$$\int_0^t \phi^1(s) X(0) \Phi(ds) = \int_0^t \begin{bmatrix} 0 & \frac{\pi_{21}\psi_2(s) ds}{\Psi_2(s)} & \frac{\pi_{31}\psi_3(s) ds}{\Psi_3(s)} \\ \frac{\pi_{22} \sum_{j=1}^4 p_{2j}\delta_{t_j}(ds)}{\Psi_2(s)} & & \frac{\pi_{32} \sum_{j=1}^4 p_{3j}\delta_{t_j}(ds)}{\Psi_3(s)} \end{bmatrix} X(s-),$$

где

$$\Psi_i(s) \triangleq \int_{[s, +\infty)} \left(\pi_{i1}\psi_2(u)du + \pi_{i2} \sum_{j=1}^4 p_{2j}\delta_{t_j}(du) \right), \quad i = 2, 3,$$

есть условные «функции дожития», соответствующие распределению момента τ_1 при условии $X(t) \equiv e_i$.

Численный эксперимент проводился для следующих вероятностных параметров и плотностей распределения:

$$\Lambda = \begin{bmatrix} -0,01 & 0,01 & 0 \\ 0,00002 & -0,00022 & 0,0002 \\ 0,00003 & 0,0008 & -0,00083 \end{bmatrix};$$

$$\pi_{21} = 0,99; \quad \pi_{22} = 0,01; \quad \pi_{31} = 0,85; \quad \pi_{32} = 0,15;$$

$\psi_i(t)$, $i = 2, 3$, — усеченные плотности распределения Парето:

$$\psi_i(t) \triangleq C(k_i, \sigma_i, \theta_i, T_i) \mathbf{I}_{[\sigma_i, T_i]}(t) \left(1 + k_i \frac{t - \theta_i}{\sigma_i} \right)^{-1-1/k_i},$$

где $C(\cdot)$ — нормировочные константы:

$$k_2 = 0,5; \quad \sigma_2 = 50; \quad \theta_2 = 1; \quad T_2 = 1000; \\ k_3 = 5; \quad \sigma_3 = 300; \quad \theta_3 = 100; \quad T_3 = 1000.$$

Временные и вероятностные характеристики дискретной части распределений приведены в таблице.

Характеристики дискретной части распределений

i	t_i	p_{1i}	p_{2i}	p_{3i}
1	200	0,6	0,01	0,01
2	600	0,4	0,79	0,09
3	900	0	0,1	0,7
4	1000	0	0,1	0,2

На рис. 1,а для сравнения представлены значения априорных вероятностей $\mathbf{P}\{X(t) = e_1\} = e_1^\top p(t)$ и условных вероятностей $\mathbf{P}\{X(t) = e_1 | \mathcal{G}(t)\} = e_1^\top \hat{X}(t)$. Для этих графиков предполагается, что в качестве результата обращения к серверу выступает «предметный» ответ $Y_1 = f_1$. Каждый график представляет собой эволюцию во времени априорной или условной вероятности, а разные графики соответствуют разным моментам τ_1 .

Рисунки 1,б и 1,в содержат соответствующие графики для состояний e_2 и e_3 .

Рисунки 2а – 2в аналогичны рисункам 1, а – 1, в, но соответствуют случаю, когда в качестве результата запроса к серверу выступает сообщение об ошибке, т. е. $Y_1 = f_2$. Каждый график представляет собой эволюцию во времени априорной или условной вероятности, а разные графики соответствуют разным моментам τ_1 .

Полученные результаты демонстрируют вариативность условного распределения по отношению к соответствующему априорному в зависимости от момента появления наблюдения τ_1 и его исхода Y_1 . Например, априорная вероятность того, что сервер свободен, равна $\mathbf{P}\{X(t) = e_1\} \equiv 0,801$. В то же время соответствующая условная вероятность колеблется в пределах от 0,5 до 1 в случае $Y_1 = f_1$ и от 0,17 до 0,82 в случае $Y_1 = f_2$. Этот факт означает, что имеющиеся косвенные неполные измерения, тем не менее, несут серьезную информацию о состоянии удаленного сервера. Их учет в алгоритмах управления распределением нагрузки между удаленными серверами может существенно увеличить эффективность их использования.

6 Заключение

Предложенная в работе концепция исследования системы наблюдений с марковским состоянием и наблюдением в форме МТП с компенсатором — линейной функцией состояния, представляется достаточно перспективной с точки зрения дальнейших теоретических и практических исследований. Прежде всего необходимо рассмотреть управляемый вариант предложенной в работе системы наблюдений и исследовать для нее мартигальную проблему. Во-вторых, важным результатом являлось бы построение некоторого «универсального» вероятностного пространства, канонического для всего рассматриваемого семейства управляемых марковских процессов. Интересным также представляется определение в этом пространстве условий, обеспечивающих непрерывную зависимость траекторий управляемых процессов от при-

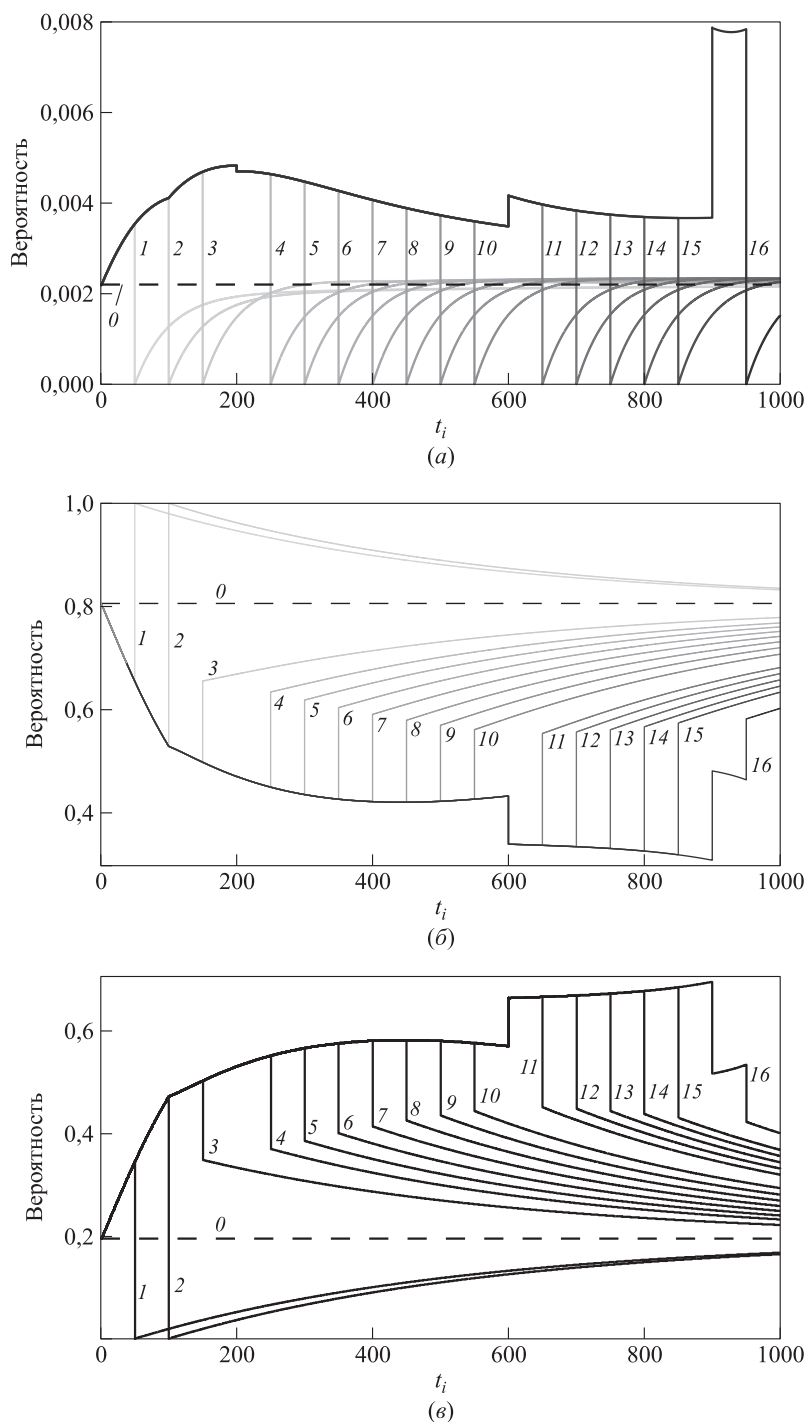


Рис. 1 Априорные (0) и условные вероятности событий $X(t) = e_1$ (а), e_2 (б) и e_3 (в) при $Y_1 = f_1$ и различных τ_1 : 1 — 50; 2 — 100; 3 — 150; 4 — 250; 5 — 300; 6 — 350; 7 — 400; 8 — 450; 9 — 500; 10 — 550; 11 — 650; 12 — 700; 13 — 750; 14 — 800; 15 — 850; 16 — 900

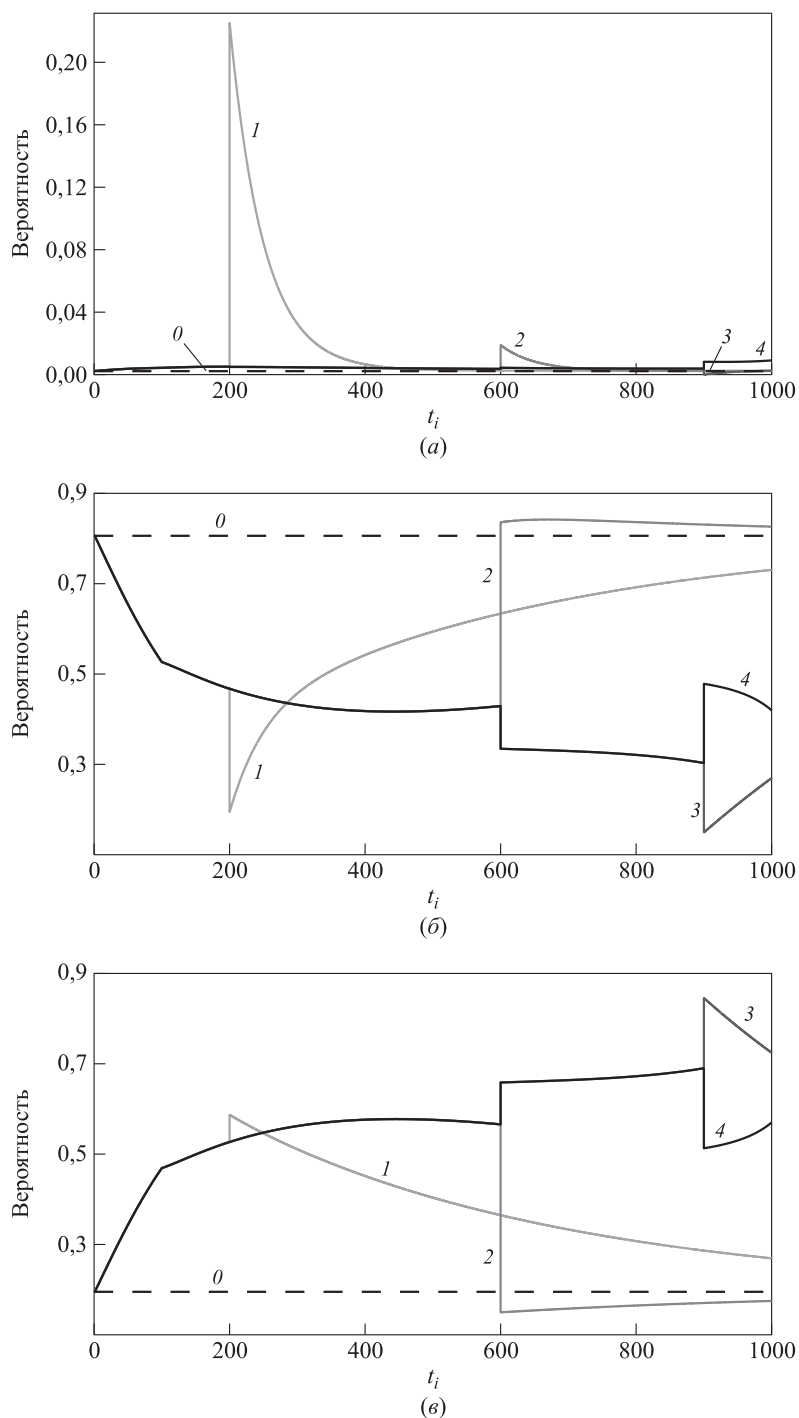


Рис. 2 Априорные (0) и условные вероятности событий $X(t) = e_1$ (а), e_2 (б) и e_3 (в) при $Y_1 = f_2$ и различных τ_1 : 1 — 200; 2 — 600; 3 — 900; 4 — 1000

меняемого управления. Эти результаты могли бы позволить корректно сформулировать и решить задачу оптимального управления марковским процессом по наблюдениям МТП.

Одновременно с этим для практики важной является разработка на основании полученных теоретических результатов эффективных численных субоптимальных и робастных алгоритмов балансировки нагрузки в распределенных инфотелекоммуникационных системах, а также решение для этих управляемых систем наблюдения задачи комплексной идентификации параметров.

Все перечисленные выше вопросы являются предметом дальнейших исследований.

Приложение

Доказательство леммы 1. Определим на $[\tau_{n-1}, +\infty)$ процесс с одним скачком, описывающий появление n -го наблюдения:

$$Y_n(t) = \begin{cases} 0, & t \in [\tau_{n-1}, \tau_n); \\ Y_n, & t \geq \tau_n. \end{cases}$$

Согласно (2) данный процесс является специальным семимартингалом и допускает мартингалное представление

$$Y_n(t) = \int_{\tau_{n-1} \wedge t}^{\tau_n \wedge t} (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \Phi_n(ds) + M_n^Y(t). \quad (16)$$

Сначала определим условное распределение τ_n относительно $\mathcal{F}_t^X \vee \mathcal{G}(n-1)$, для чего на $[\tau_{n-1}, +\infty)$ рассмотрим случайный процесс $I_n(t) \triangleq \mathbf{I}_{\{0\}}(\mathbf{1}^\top Y_n(t)) = \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t)$. В соответствии с этим определением, формулой (16) и свойствами меры $\Phi_n(\cdot)$ процесс $I_n(t)$ представим в виде:

$$\begin{aligned} I_n(t) &= 1 - \int_{\tau_{n-1}}^t I_n(s-) \mathbf{1}_M^\top dY_n(s) = \\ &= 1 - \int_{\tau_{n-1}}^t I_n(s-) \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \\ &+ \phi_n^1(s)X(s-)) \Phi_n(ds) - \int_{\tau_{n-1}}^t I_n(s-) \mathbf{1}_M^\top dM_n^Y(s). \end{aligned} \quad (17)$$

Тогда $\tilde{I}_n(t) \triangleq \mathbf{E} \{I_n(t) | \mathcal{F}_t^X \vee \mathcal{G}(n-1)\} = \mathbf{I}_{[\tau_{n-1}, +\infty)}(t) \mathbf{P} \{\tau_n > t | \mathcal{F}_t^X \vee \mathcal{G}(n-1)\}$ представляет собой так называемую «функцию дожития», зная которую, легко получить условную функцию распределения: $F_n(t) \triangleq \mathbf{P} \{\tau_n \leq t | \mathcal{F}_t^X \vee \mathcal{G}(n-1)\} = 1 - \tilde{I}_n(t)$.

Для нахождения $\tilde{I}_n(t)$ воспользуемся методом вывода уравнений оптимальной нелинейной фильтрации, предложенным в [16]. Беря УМО от обеих частей (17) относительно $\mathcal{F}_t^X \vee \mathcal{G}(n-1)$, легко проверить, что

$$\begin{aligned} \tilde{I}_n(t) &= 1 - \int_{\tau_{n-1}}^t \tilde{I}_n(s-) \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \\ &+ \phi_n^1(s)X(s-)) \Phi_n(ds) + \tilde{M}_n^I(t), \end{aligned} \quad (18)$$

где $\tilde{M}_n^I(t)$ — некоторый $\mathcal{F}_t^X \vee \mathcal{G}(n-1)$ -согласованный мартингал. Он представим в виде:

$$\tilde{M}_n^I(t) = \int_{\tau_{n-1}}^t \gamma(s) dM^X(s),$$

где $\gamma(s) \triangleq \gamma(\omega, s) : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^{1 \times N} - \mathcal{F}_s^X \vee \mathcal{G}(n-1)$ -предсказуемый процесс, подлежащий определению:

По обобщенному правилу Ито

$$\begin{aligned} X(t)\tilde{I}_n(t) &= X(\tau_{n-1}) + \int_{\tau_{n-1}}^t X(s-) d\tilde{I}_n(s) + \\ &+ \int_{\tau_{n-1}}^t dX(s)\tilde{I}_n(s-) + \sum_{\tau_{n-1} < s \leq t} \Delta X(s)\Delta\tilde{I}_n(s) = \\ &= X(\tau_{n-1}) - \int_{\tau_{n-1}}^t X(s-) \tilde{I}_n(s-) \mathbf{1}^\top (\phi_n^0(s)X(\tau_{n-1}) + \\ &+ \phi_n^1(s)X(s-)) \Phi_n(ds) + \\ &+ \int_{\tau_{n-1}}^t (\Lambda^\top(s)X(s-) \tilde{I}_n(s-) + \\ &+ \text{diag}(\gamma(s))\Lambda^\top(s)X(s-)) ds + M_1(t), \end{aligned} \quad (19)$$

где $M_1(t) - \mathcal{F}_t^X \vee \mathcal{G}(n-1)$ -согласованный мартингал.

Из условия (д) следует, что $[X, Y]_t \equiv 0$ п. н., а значит, и $[X, I_n]_t \equiv 0$, поэтому по правилу Ито

$$\begin{aligned} X(t)I_n(t) &= X(\tau_{n-1}) + \\ &+ \int_{\tau_{n-1}}^t X(s-) dI_n(s) + \int_{\tau_{n-1}}^t dX(s)I_n(s-) = \\ &= X(\tau_{n-1}) - \int_{\tau_{n-1}}^t X(s-) I_n(s-) \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \\ &+ \phi_n^1(s)X(s-)) \Phi_n(ds) + \\ &+ \int_{\tau_{n-1}}^t \Lambda^\top(s)X(s-) I_n(s-) ds + M_2(t), \end{aligned} \quad (20)$$

где $M_2(t) - \mathcal{F}_t$ -согласованный мартингал. Вычисляя УМО от обеих частей (20) относительно $\mathcal{F}_t^X \vee \mathcal{G}(n-1)$, получаем выражение

$$\begin{aligned} X(t)\tilde{I}_n(t) &= X(\tau_{n-1}) - \\ &- \int_{\tau_{n-1}}^t X(s-) \tilde{I}_n(s-) \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \end{aligned}$$

$$+ \phi_n^1(s)X(s-) \Phi_n(ds) + \int_{\tau_{n-1}}^t \Lambda^\top(s)X(s-)\tilde{I}_n(s-) ds + M_3(t), \quad (21)$$

где $M_3(t) - \mathcal{F}_t^X \vee \mathcal{G}(n-1)$ -согласованный мартингал. Формулы (19) и (21) представляют собой две записи мартингального разложения одного и того же специального семимартингала $X(t)\tilde{I}_n(t)$, и в силу единственности этого разложения верно равенство

$$\text{diag } \gamma(s)\Lambda^\top(s)X(s-) \equiv 0$$

Р-п. н. почти для всех $s \in [\tau_{n-1}, +\infty)$. Решение последнего уравнения не единственно, но версии полученного процесса $\tilde{I}_n(t)$ (18) при выборе разных γ будут неотличимыми, поэтому можно выбрать частный случай решения — тождественный ноль, т. е. положить $\gamma(s) \equiv 0, s \in [\tau_{n-1}, +\infty)$. Таким образом, функция дожития $\tilde{I}_n(t)$ удовлетворяет обыкновенному линейному дифференциальному уравнению с разрывной случайной правой частью

$$\tilde{I}_n(t) = 1 - \int_{\tau_{n-1}}^t \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \tilde{I}_n(s-) \Phi_n(ds). \quad (22)$$

Решение этого уравнения известно:

$$\begin{aligned} \tilde{I}_n(t) = \exp \left\{ - \int_{\tau_{n-1}}^t \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \Phi_n^c(ds) \right\} \times \\ \times \prod_{\tau_{n-1} < s \leq t} \left[1 - \mathbf{1}_M^\top (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \Phi_n^d(\{s\}) \right]. \quad (23) \end{aligned}$$

Рассмотрим теперь процесс $\tilde{I}_n(t, u) \triangleq \mathbf{E} \{ I_n(t) | \mathcal{F}_u^X \vee \mathcal{G}(n-1) \}$ как функцию времени u ($t \leq u$). Можно показать, что $\tilde{I}_n(t, u)$ представим в виде:

$$\tilde{I}_n(t, u) = \tilde{I}_n(t) + \int_t^u \epsilon(s) dM^X(s), \quad (24)$$

где $\epsilon(s) \triangleq \epsilon(\omega, s) : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^{1 \times N} - \mathcal{F}_s^X \vee \mathcal{G}(n-1)$ -предсказуемый процесс, вновь подлежащий определению. Действуя аналогичным образом как при выводе коэффициента $\gamma(s)$, можно показать, что $\epsilon(s) \equiv 0, s \in (t, +\infty)$, т. е. $\tilde{I}_n(t, u) \equiv \tilde{I}_n(t)$ при $u > t$. Этот факт совместно с формулой (24) доказывает истинность утверждения 1 леммы 1. При этом выполнение условий (3) гарантирует, что функция, определяемая формулой (23), действительно **Р**-п. н. обладает свойствами функции распределения.

Далее, процесс $Y_n(t)$ может быть представлен следующим образом:

$$\begin{aligned} Y_n(t) &= \int_{\tau_{n-1}}^t dY_n(s)I_n(s-) = \\ &= \int_{\tau_{n-1}}^t (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) I_n(s-) \Phi_n(ds) + \\ &\quad + \int_{\tau_{n-1}}^t dM_n^Y(s)I_n(s-). \end{aligned}$$

Обозначим $\tilde{Y}_n(t) \triangleq \mathbf{E} \{ Y_n(t) | \mathcal{F}_t^X \vee \mathcal{G}(n-1) \}$. Действуя аналогично доказательству утверждения 1 леммы 1 и используя (23) для $\tilde{I}_n(t)$, легко показать, что

$$\begin{aligned} \tilde{Y}_n(t) &= \mathbf{E} \left\{ Y_n(t) \middle| \mathcal{F}^X \vee \mathcal{G}(n-1) \right\} = \\ &= \int_{\tau_{n-1}}^t (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \tilde{I}_n(s-) \Phi_n(ds) = \\ &= \int_{\tau_{n-1}}^t (\phi_n^0(s)X(\tau_{n-1}) + \phi_n^1(s)X(s-)) \times \\ &\quad \times \exp \left\{ - \int_{\tau_{n-1}}^s \mathbf{1}_M^\top (\phi_n^0(u)X(\tau_{n-1}) + \phi_n^1(u)X(u-)) \Phi_n^c(du) \right\} \times \\ &\quad \times \prod_{\tau_{n-1} < u < s} \left[1 - \mathbf{1}_M^\top (\phi_n^0(u)X(\tau_{n-1}) + \phi_n^1(u)X(u-)) \Phi_n^d(\{u\}) \right] \Phi_n(ds). \quad (25) \end{aligned}$$

Формула (4) — утверждение 2 леммы 1 — непосредственно следует из (25) благодаря тому факту, что $\mathbf{P} \{ \tau_n \leq t, Y_n = f_i | \mathcal{F}_t^X \vee \mathcal{G}(n-1) \} = f_i^\top \tilde{Y}_n(t)$.

Рассмотрим на промежутке $[\tau_{n-1}, \tau_n)$ процесс $Z_n(t) \triangleq X(\tau_{n-1})\tilde{I}_n(t)X^\top(t)$. Используя правило Ито, формулы (2), (22) и свойства матриц, можно получить стохастическое дифференциальное уравнение для $Z_n(t)$:

$$\begin{aligned} Z_n(t) &= \text{diag}(X(\tau_{n-1})) + \int_{\tau_{n-1}}^t Z_n(s-)\Lambda(s) ds - \\ &\quad - \int_{\tau_{n-1}}^t \left[\text{diag}(\phi_n^{0\top}(s)\mathbf{1}_M)Z_n(s-) + \right. \\ &\quad \left. + Z_n(s-) \text{diag}(\phi_n^{1\top}(s)\mathbf{1}_M) \right] \Phi_n(ds) + \\ &\quad + \int_{\tau_{n-1}}^t X(\tau_{n-1})\tilde{I}_n(s-) d(M^X)^\top(s), \quad (26) \end{aligned}$$

где $M^X(t)$ — мартингал из разложения (1).

Беря УМО от левой и правой частей (26) относительно $\mathcal{G}(n-1)$, получаем искомую систему (5) для $\bar{Z}_n(t)$, т. е. доказываем истинность утверждения 3 леммы 1.

Согласно [14] компенсатор ν определяется формулой:

$$\nu(dt, dy) = \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau_{n-1}, \tau_n]}(t) \frac{\mathbf{P}_n(dt, dy)}{\mathbf{P}_n([t, +\infty) \times \mathbb{R}^M)}, \quad (27)$$

где $\mathbf{P}_n(dt, dy)$ — условное распределение пары (τ_n, Y_n) относительно $\mathcal{G}(n-1)$. Формула (27) определена корректно, так как на множестве $\{\omega : \tau_{n-1}(\omega) < t \leq \tau_n(\omega)\}$ случайная величина $\mathbf{P}_n([t, +\infty) \times \mathbb{R}^M) > 0$ \mathbf{P} -п. н. Формула (25) задает распределение относительно более широкой σ -алгебры $\mathcal{F}^X \vee \mathcal{G}(n-1)$. Используя свойства УМО, можно получить выражение для знаменателя в (27):

$$\begin{aligned} \mathbf{P}_n([t, +\infty) \times \mathbb{R}^M) &= \mathbf{P} \{ \tau_n \geq t | \mathcal{G}(n-1) \} = \\ &= \mathbf{E} \left\{ \mathbf{P} \{ \tau_n \geq t | \mathcal{F}^X \vee \mathcal{G}(n-1) \} | \mathcal{G}(n-1) \right\} = \\ &= \mathbf{E} \left\{ \tilde{I}_n(t-) | \mathcal{G}(n-1) \right\} \quad (28) \end{aligned}$$

и для числителя

$$\begin{aligned} \mathbf{P}_n(dt, dy) &= \mathbf{P} \{ \tau_n \in [t, t+dt), Y_n \in [y, y+dy) | \mathcal{G}(n-1) \} = \\ &= \mathbf{E} \left\{ \mathbf{P} \left\{ \tau_n \in [t, t+dt), Y_n \in [y, y+dy) | \mathcal{F}^X \vee \right. \right. \\ &\quad \left. \left. \vee \mathcal{G}(n-1) \right\} | \mathcal{G}(n-1) \right\} = \\ &= \mathbf{E} \left\{ \sum_{i=1}^M y^T (\phi_n^0(t) X(\tau_{n-1}) + \right. \\ &\quad \left. + \phi_n^1(t) X(t-) \tilde{I}_n(t-) \delta_{\{f_i\}}(dy) \Phi_n(dt) | \mathcal{G}(n-1) \right\} = \\ &= \sum_{i=1}^M y^T \left[\phi_n^0(t) \mathbf{E} \left\{ X(\tau_{n-1}) \tilde{I}_n(t-) | \mathcal{G}(n-1) \right\} + \right. \\ &\quad \left. + \phi_n^1(t) \mathbf{E} \left\{ X(t-) \tilde{I}_n(t-) | \mathcal{G}(n-1) \right\} \right] \delta_{\{f_i\}}(dy) \Phi_n(dt). \quad (29) \end{aligned}$$

Подставляя (28) и (29) в (27), получаем требуемую формулу (6). Все утверждения леммы доказаны.

Доказательство теоремы 1. По определению

$$\mathcal{G}(n-1) \vee \{ \tau_n > t \} \triangleq \sigma \{ \{ \omega : \tau_n(\omega) > t \}, A : A \in \mathcal{G}(n-1) \}$$

есть минимальная σ -алгебра, содержащая $\mathcal{G}(n-1)$ и $\{ \omega \in \Omega : \tau_n > t \}$.

Определим процессы

$$\begin{aligned} J_n(t) &\triangleq \mathbf{I}_{[\tau_n, +\infty)}(t); \\ \tilde{J}_n(t) &\triangleq \mathbf{E} \left\{ J_n(t) | \mathcal{F}^X \vee \mathcal{G}(n-1) \right\}; \\ \check{X}_n(t) &\triangleq \mathbf{I}_{[0, \tau_{n-1})}(t) \mathbf{E} \{ X(t) | \mathcal{G}(n-1) \} + \\ &\quad + \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \bar{X}_n(t) + \mathbf{I}_{[\tau_n, +\infty)}(t) \times \\ &\quad \times \mathbf{E} \left\{ \tilde{J}_n(t) | \mathcal{G}(n-1) \right\}^+ \mathbf{E} \left\{ X(t) \tilde{J}_n(t) | \mathcal{G}(n-1) \right\}. \end{aligned}$$

Выберем произвольное $A \in \mathcal{G}(n-1)$ и положим $B \triangleq A \cap \{ \tau_n > t \}$, тогда из свойств УМО следует цепочка равенств:

$$\begin{aligned} \mathbf{E} \left\{ \mathbf{I}_B(\omega) (X(t) - \check{X}_n(t)) \right\} &= \mathbf{E} \left\{ \mathbf{I}_A(\omega) (\mathbf{I}_{[0, \tau_{n-1})}(t) + \right. \\ &\quad \left. + \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \right\} (X(t) - \check{X}_n(t)) \Big\} = \\ &= \mathbf{E} \left\{ \mathbf{E} \left\{ \mathbf{I}_A(\omega) \mathbf{I}_{[0, \tau_{n-1})}(t) (X(t) - \right. \right. \\ &\quad \left. \left. - \mathbf{E} \{ X(t) | \mathcal{G}(n-1) \}) | \mathcal{G}(n-1) \right\} + \right. \\ &\quad \left. + \mathbf{E} \left\{ \mathbf{I}_A(\omega) \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \left(X(t) - \right. \right. \right. \\ &\quad \left. \left. - \frac{1}{\mathbf{E} \{ \tilde{I}_n(t) | \mathcal{G}(n-1) \}} \mathbf{E} \left\{ X(t) \tilde{I}_n(t) | \mathcal{G}(n-1) \right\} \right) \right\} \right\} = \\ &= \mathbf{E} \left\{ \mathbf{E} \left\{ \mathbf{I}_A(\omega) \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \left(X(t) - \right. \right. \right. \\ &\quad \left. \left. - \bar{X}_n(t) \right) | \mathcal{G}(n-1) \right\} \right\} = 0. \end{aligned}$$

Доказательство равенств $\mathbf{E} \left\{ \mathbf{I}_B(\omega) (X(t) - \check{X}_n(t)) \right\} = 0$ для $B = A \cup \{ \tau_n > t \}$, $B = A \cap \{ \tau_n \leq t \}$ и $B = A \cup \{ \tau_n \leq t \}$ проводится аналогично. Таким образом, доказано, что $\check{X}_n(t) = \mathbf{E} \{ X(t) | \mathcal{G}(n-1) \vee \{ \tau_n > t \} \}$ \mathbf{P} -п. н., из чего непосредственно следует истинность формулы (11).

Естественный поток σ -алгебр \mathcal{G}_t , порожденный процессом наблюдений в непрерывном времени $Y(t)$, совпадает с естественным потоком, порожденным стохастической мерой μ , т. е.

$$\mathcal{G}_t = \sigma \{ \mu(A \times B) : A \in \mathcal{B}([0, t]), B \in \mathbb{R}^M \}.$$

Поэтому для доказательства того, что оценка $\hat{X}(t)$, вычисленная по формуле (10), действительно является УМО относительно \mathcal{G}_t , достаточно показать, что $\mathbf{E} \left\{ (X(t) - \hat{X}(t)) \mu([a, b] \times \{f_j\}) \right\} \equiv 0$ для любых $0 \leq a < b \leq t$ и $j = \overline{1, M}$.

Свойства УМО позволяют получить следующую цепочку равенств:

$$\begin{aligned} \mathbf{E} \left\{ (X(t) - \hat{X}(t)) \mu([a, b] \times \{f_j\}) \right\} &= \\ &= \sum_{m \leq n-1} \mathbf{E} \left\{ \mathbf{I}_{[a, b)}(\tau_m) \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) (X(t) - \bar{X}_n(t)) Y_m^T \right\} f_j = \\ &= \sum_{m \leq n-1} \mathbf{E} \left\{ \mathbf{E} \left\{ \mathbf{I}_{[a, b)}(\tau_m) \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) (X(t) - \right. \right. \\ &\quad \left. \left. - \bar{X}_n(t)) Y_m^T | \mathcal{F}^X \vee \mathcal{G}(n-1) \right\} \right\} f_j = \\ &= \sum_{m \leq n-1} \mathbf{E} \left\{ \mathbf{I}_{[a, b)}(\tau_m) \tilde{I}_n(t) (X(t) - \bar{X}_n(t)) Y_m^T \right\} f_j = \\ &= \sum_{m \leq n-1} \mathbf{E} \left\{ \mathbf{E} \left\{ \mathbf{I}_{[a, b)}(\tau_m) \tilde{I}_n(t) (X(t) - \right. \right. \\ &\quad \left. \left. - \bar{X}_n(t)) Y_m^T | \mathcal{G}(n-1) \right\} \right\} f_j = \end{aligned}$$

$$= \sum_{m \leq n-1} \mathbf{E} \left\{ \mathbf{I}_{[a,b)}(\tau_m) \mathbf{E} \left\{ \tilde{I}_n(t) (X(t) - \bar{X}_n(t)) \middle| \mathcal{G}(n-1) \right\} Y_m^\top \right\} f_j = 0.$$

Истинность представления (10) для $\mathbf{E} \{X(t) | \mathcal{G}_t\}$ доказана. Формула (12) получается непосредственно из (6) и соотношения $\mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \bar{X}_n(t) = \mathbf{I}_{[\tau_{n-1}, \tau_n)}(t) \hat{X}(t)$. Теорема 1 доказана.

Доказательство теоремы 2. В силу включения $\mathcal{G}_t \subseteq \mathcal{F}_t$ из уравнения (1) следует, что найдется такой \mathcal{G}_t -согласованный мартингал $\widehat{M}^X(t)$, что

$$\hat{X}(t) = p_0 + \int_0^t \Lambda^\top(s) \hat{X}(s-) ds + \widehat{M}^X(t). \quad (30)$$

Общий вид $\widehat{M}^X(t)$ представлен в [13] (теорема 4.10.1), и с учетом специфических условий рассматриваемой задачи оценивания этот мартингал представим в виде интеграла по случайной мере

$$\widehat{M}^X(t) = \int_{[0,t] \times \mathbb{R}^M} H(s,y) (\mu(ds, dy) - \nu(ds, dy)), \quad (31)$$

где

$$\left. \begin{aligned} U(s,y) &\triangleq M_\mu^P \left\{ X | \tilde{\mathcal{P}}(\mathcal{G}) \right\} (s,y) - \hat{X}(s-); \\ H(s,y) &\triangleq U(s,y) + \frac{\mathbf{I}_{\{0 < \alpha(s) < 1\}}(s)}{1 - \alpha(s)} \int_{\mathbb{R}^M} U(s,z) \nu(\{s\}, dz); \\ \alpha(s) &\triangleq \nu(\{s\} \times \mathbb{R}^M). \end{aligned} \right\} \quad (32)$$

Ключевым в выводе уравнения оптимальной фильтрации является выражение случайного процесса $M_\mu^P \left\{ X | \tilde{\mathcal{P}}(\mathcal{G}) \right\} (s,y) - \text{УМО относительно } \sigma\text{-алгебры } \tilde{\mathcal{P}}(\mathcal{G})$. Это такая предсказуемая случайная функция, что

$$\begin{aligned} &\mathbf{E} \left\{ \int_{\mathbb{R}_+ \times \mathbb{R}^M} \psi(s,y) X(s) \mu(ds, dy) \right\} = \\ &= \mathbf{E} \left\{ \int_{\mathbb{R}_+ \times \mathbb{R}^M} \psi(s,y) M_\mu^P \left\{ X | \tilde{\mathcal{P}}(\mathcal{G}) \right\} (s,y) \nu(ds, dy) \right\} \quad (33) \end{aligned}$$

для любой ограниченной неотрицательной предсказуемой случайной функции $\psi(s,y)$. Можно проверить, что такая функция ψ может быть представлена в виде:

$$\psi(s,y) = \sum_{n \in \mathbb{N}} \bar{h}^\top y v_n(s) \mathbf{I}_{(\tau_{n-1}, \tau_n]}(s),$$

где $\bar{h} \triangleq \text{col}(h^1, \dots, h^M)$ — детерминированный вектор, $\{v_n(s)\}_{n \in \mathbb{N}}$ — последовательность $\mathcal{G}(n-1)$ -измеримых функций. Из определения интеграла по μ следует, что

$$\begin{aligned} &\mathbf{E} \left\{ \int_{\mathbb{R}_+ \times \mathbb{R}^M} \psi(s,y) X(s) \mu(ds, dy) \right\} = \\ &= \mathbf{E} \left\{ \sum_{n \in \mathbb{N}} \bar{h}^\top Y_n v_n(\tau_n) X(\tau_n) \right\}. \end{aligned}$$

Обозначим $\mathcal{L}_n \triangleq \mathbf{E} \left\{ \bar{h}^\top Y_n v_n(\tau_n) X(\tau_n) \right\}$ и выразим его в виде интеграла по мере ν подобно (33):

$$\begin{aligned} \mathcal{L}_n &= \mathbf{E} \left\{ \mathbf{E} \left\{ \bar{h}^\top Y_n v_n(\tau_n) X(\tau_n) \middle| \mathcal{F}^X \vee \mathcal{G}(n-1) \right\} \right\} = \\ &= \sum_{j=1}^M \bar{h}^\top f_j \mathbf{E} \left\{ \int_{\tau_{n-1}}^\infty v_n(s) X(s) f_j^\top (\phi_n^0(s) X(\tau_{n-1}) + \right. \\ &\quad \left. + \phi_n^1(s) X(s-)) \tilde{I}_n(s-) \Phi_n(ds) \right\} = \\ &= \sum_{j=1}^M \bar{h}^\top f_j \mathbf{E} \left\{ \int_{\tau_{n-1}}^\infty v_n(s) \beta_n^\top(s) f_j \Phi_n(ds) \right\} = \sum_{j=1}^M \bar{h}^\top f_j \times \\ &\times \mathbf{E} \left\{ \int_{\tau_{n-1}}^\infty v_n(s) \hat{\beta}_n^\top(s) f_j \mathbf{E} \left\{ \tilde{I}_n(s-) \middle| \mathcal{G}(n-1) \right\} \Phi_n(ds) \right\}. \end{aligned}$$

Из уравнения (22) для $\tilde{I}_n(t)$ следует, что $\tilde{I}_n(\tau_{n-1}) = 1$, а из условия (е) — $\lim_{t \rightarrow \infty} \tilde{I}_n(t) = 0$, поэтому выполняются равенства:

$$\begin{aligned} &\int_{(\tau_{n-1}, +\infty)} \mathbf{1}_M^\top (\phi_n^0(s) X(\tau_{n-1}) + \\ &\quad + \phi_n^1(s) X(s-)) \tilde{I}_n(s-) \Phi_n(ds) = 1; \end{aligned}$$

$$\begin{aligned} \tilde{I}_n(t) &= \int_{(t, +\infty)} \mathbf{1}_M^\top (\phi_n^0(s) X(\tau_{n-1}) + \\ &\quad + \phi_n^1(s) X(s-)) \tilde{I}_n(s-) \Phi_n(ds); \end{aligned}$$

$$\begin{aligned} \tilde{I}_n(t-) &= \int_{[t, +\infty)} \mathbf{1}_M^\top (\phi_n^0(s) X(\tau_{n-1}) + \\ &\quad + \phi_n^1(s) X(s-)) \tilde{I}_n(s-) \Phi_n(ds). \end{aligned}$$

Используя введенные процессы β_n и $\hat{\beta}_n$, свойства УМО и теорему Фубини, преобразования интеграла \mathcal{L}_n можно продолжить:

$$\begin{aligned} \mathcal{L}_n &= \sum_{j=1}^M \bar{h}^\top f_j \mathbf{E} \left\{ \int_{\tau_{n-1}}^\infty v_n(s) \hat{\beta}_n^\top(s) f_j \times \right. \\ &\quad \left. \times \left[\int_{[s, +\infty)} \mathbf{1}_M^\top \beta_n(u) \mathbf{1}_N \Phi_n(du) \right] \Phi_n(ds) \right\} = \\ &= \sum_{j=1}^M \bar{h}^\top f_j \mathbf{E} \left\{ \int_{\tau_{n-1}}^\infty \left[\int_{\tau_{n-1}}^u v_n(s) \hat{\beta}_n^\top(s) f_j \Phi_n(ds) \right] \times \right. \end{aligned}$$

$$\begin{aligned} & \times \mathbf{1}_M^\top \beta_n(u) \mathbf{1}_N \Phi_n(du) \Big\} = \sum_{j=1}^M \bar{h}^\top f_j \times \\ & \times \mathbf{E} \left\{ \mathbf{E} \left\{ \int_{\tau_{n-1}}^{\tau_n} v_n(s) \hat{\beta}_n^\top(s) f_j \Phi_n(ds) | \mathcal{G}(n-1) \right\} \right\} = \\ & = \mathbf{E} \left\{ \int_{(\tau_{n-1}, \tau_n] \times \mathbb{R}^M} \bar{h}^\top y v_n(s) \hat{\beta}_n^\top(s) y (\mathbf{1}_N^\top \hat{\beta}_n^\top(s) y)^\dagger \nu(ds, dy) \right\}. \end{aligned}$$

$$\hat{X}(\tau_n) = \begin{cases} (\mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) Y_n)^\dagger \hat{\beta}_n^\top(\tau_n) Y_n, & \text{если } \mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) < 1; \\ (\mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) Y_n)^\dagger \hat{\beta}_n^\top(\tau_n) Y_n - & \\ - [\hat{\beta}_n^\top(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) - \hat{X}(\tau_{n-})], & \\ \text{если } \mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) \mathbf{1}_M \Phi_n^d(\{\tau_n\}) = 1. & \end{cases} \quad (37)$$

Из последнего равенства следует, что

$$\begin{aligned} M_\mu^P \{X | \tilde{\mathcal{P}}(\mathcal{G})\}(t, y) = \\ = \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau_{n-1}, \tau_n]}(t) \hat{\beta}_n^\top(t) y (\mathbf{1}_N^\top \hat{\beta}_n^\top(t) y)^\dagger. \end{aligned} \quad (34)$$

Подставим (34) в (32):

$$\left. \begin{aligned} U(s, y) &= \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau_{n-1}, \tau_n]}(s) \times \\ & \times \left[\frac{1}{\mathbf{1}_N^\top \hat{\beta}_n^\top(s) y} \hat{\beta}_n^\top(s) y - \hat{X}(s-) \right]; \\ \alpha(s) &= \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau_{n-1}, \tau_n]}(s) \mathbf{1}_N^\top \hat{\beta}_n^\top(s) \mathbf{1}_M \Phi_n^d(\{s\}); \\ H(s, y) &= \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau_{n-1}, \tau_n]}(s) \times \\ & \times \left(\left[\frac{1}{\mathbf{1}_N^\top \hat{\beta}_n^\top(s) y} \hat{\beta}_n^\top(s) y - \hat{X}(s-) \right] + \right. \\ & \quad \left. + \left(1 - \mathbf{1}_N^\top \hat{\beta}_n^\top(s) \mathbf{1}_M \Phi_n^d(\{s\}) \right)^\dagger \times \right. \\ & \quad \left. \times \left[\hat{\beta}_n^\top(s) \mathbf{1}_M - \hat{X}(s-) \mathbf{1}_N^\top \hat{\beta}_n^\top(s) \mathbf{1}_M \right] \Phi_n^d(\{s\}) \right). \end{aligned} \right\} \quad (35)$$

Далее, комбинируя (35), (31) и (30), можно получить дифференциальную систему, описывающую эволюцию $\hat{X}(t)$ на полуинтервалах $[\tau_{n-1}, \tau_n)$ между наблюдениями:

$$\begin{aligned} \hat{X}(t) &= \hat{X}(\tau_{n-1}) + \int_{\tau_{n-1}}^t \Lambda^\top(s) \hat{X}(s-) ds - \\ & - \int_{\tau_{n-1}}^t \left[\hat{\beta}_n^\top(s) \mathbf{1}_M - \hat{X}(s-) \mathbf{1}_N^\top \hat{\beta}_n^\top(s) \mathbf{1}_M \right] \Phi_n(ds). \end{aligned} \quad (36)$$

Уравнение (36) не замкнуто относительно неизвестных: помимо искомой оценки фильтрации $\hat{X}(s)$ в правую часть (36) через матричнозначную функцию $\hat{\beta}_n(s)$ также входит оценка $\hat{X}(\tau_{n-1}, s)$ сглаживания в фиксированной точке τ_{n-1} . В то же время имеется уравнение (5), замкнутое относительно $\bar{Z}_n(\cdot)$. Из (5) уравнение (36) может быть легко получено подстановкой в него (14). Агрегируя (30), (31) и (35), можно также получить алгебраические соотношения, определяющие оценки фильтрации в моменты τ_n получения наблюдений:

Оценка $\hat{X}(\tau_n)$ (37) не является регулярной версией УМО в случае, если $\mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) Y_n = 0$ из-за нарушения условия нормировки. Тем не менее в силу следствия 1 $\mathbf{P} \{ \mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) Y_n = 0 \} = 0$, и на пренебрежимом множестве $\{ \omega : \mathbf{1}_N^\top \hat{\beta}_n^\top(\tau_n) Y_n = 0 \}$ искомая оценка может быть переопределена так, чтобы выполнялось условие регулярности: $\hat{X}(\tau_n) \triangleq \hat{X}(\tau_{n-})$. Тем самым показано, что в моменты τ_n появления наблюдений оценка фильтрации действительно определяется формулой (15). Теорема 2 доказана.

Литература

1. Gnedenko B. V., Kovalenko I. N. Introduction to queueing theory. — Boston: Birkhauser, 1989. 314 p.
2. Kalashnikov V. V. Mathematical methods for construction of queueing models. — N.Y.: Springer-Verlag, 1990. 431 p.
3. Bremaud P. Optimal thinning of a point process // SIAM J. Contr. Optim., 1979. Vol. 17. No. 2. P. 222–230. doi: 10.1137/0317017.
4. Миллер Б. М., Авраченко К. Е., Степанян К. В., Миллер Г. Б. Задача оптимального стохастического управления потоком данных по неполной информации // Проблемы передачи информации, 2005. Т. 41. № 2. С. 89–110. doi: 10.1007/s11122-005-0020-8.
5. Shen B., Wang Z., Shu H. Nonlinear stochastic systems with incomplete information: Filtering and control. — N.Y.: Springer Verlag, 2013. 248 p.
6. Anagnostopoulos A., Kirsch A., Upfal E. Load balancing in arbitrary network topologies with stochastic adversarial input // SIAM J. Comput., 2005. Vol. 17. No. 3. P. 616–639. doi: 10.1137/S0097539703437831.
7. Altman E., Ayesta U., Prabhu B. Load balancing in processor sharing systems // Telecommun. Syst., 2011. Vol. 47. No. 1-2. P. 35–48. doi: 10.4108/ICST.VALUETOOLS2008.4462.
8. Босов А. В. Моделирование и оптимизация процессов функционирования Информационного web-портала // Программирование, 2009. № 6. С. 53–66. doi: 10.1134/S0361768809060048.
9. Босов А. В. Задачи анализа и оптимизации для модели пользовательской активности. Часть 1. Анализ и прогнозирование // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 40–52.
10. Olshefski D., Nieh J., Agrawal D. Using CERTES to infer client response time at the web server // ACM Trans. Comput. Syst., 2004. Vol. 22. No. 1. P. 49–93. doi: 10.1145/511334.511355.

11. *Ozsu M. T., Valduriez P.* Principles of distributed database systems. — N.Y.: Springer-Verlag, 2011. 845 p.
12. *Elliott R. J., Aggoun L., Moore J. B.* Hidden Markov models: Estimation and control. — N.Y.: Springer-Verlag, 1994. 382 p.
13. *Liptser R. Sh., Shiriyayev A. N.* Theory of martingales. — N.Y.: Springer-Verlag, 1989. 812 p.
14. *Jacod J., Shiriyayev A. N.* Limit theorems for stochastic processes. — N.Y.: Springer-Verlag, 2003. 664 p.
15. *Elliott R. J.* Stochastic calculus and applications. — N.Y.: Springer-Verlag, 1982. 302 p.
16. *Wong E., Hajek B.* Stochastic processes in engineering systems. — N.Y.: Springer-Verlag, 1985. 361 p.
17. *Борисов А. В., Миллер Б. М., Семенухин К. В.* Фильтрация марковского скачкообразного процесса по наблюдениям мультивариантного точечного процесса // Автоматика и телемеханика, 2014 (в печати).
18. *Юшкевич А. А.* Управляемые марковские модели со счетным множеством состояний и непрерывным временем // Теория вероятностей и ее применения, 1977. Т. 22. № 2. С. 222–241. doi: 10.1137/1122029.
19. *Kalman R. E., Bucy R. S.* New results in linear filtering and prediction theory // Trans. ASME. J. Basic Eng., 1960. Vol. 83D. No. 1. P. 95–108.
20. *Cvitanić J., Liptser R., Rozovskii B.* A filtering approach to tracking volatility from prices observed at random times // Ann. Appl. Probab., 2006. Vol. 16. No. 3. P. 1633–1652. doi: 10.1214/105051606000000222.

Поступила в редакцию 27.06.14

MONITORING REMOTE SERVER ACCESSIBILITY: THE OPTIMAL FILTERING APPROACH

A. V. Borisov^{1,2}

¹Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Department of Probability Theory, School of Applied Mathematics and Physics, Moscow Aviation Institute, 4 Volokolamskoe Shosse, GSP-3, A-80, Moscow 125993, Russian Federation

Abstract: The online monitoring problem of a remote server, accessible via the http protocol, is formulated in the terms of optimal filtering. The unobservable server state is treated as a finite-dimensional Markov jump process, meanwhile the observation is supposed to be a multivariate point process with a finite set of possible values. The key point of the investigated observation system is that the random intensity of observations is a linear function of the unobservable Markov state. It is proved that the optimal filtering estimate is a solution to some closed finite system of recursive formulae and ordinary linear differential equations with a random right-hand side. The applicability of the obtained theoretical results is illustrated by an example of monitoring accessibility of the queueing system “communication channel – database server.” The unobservable state of this system consists of three possible values (no connection, low workload, high workload), meanwhile the possible observations belong to the set of two possible values (answer to the query, error message). The conclusion of the paper contains possible perspectives for the further research.

Keywords: Markov models; optimal filtering; stochastic jump processes; conditional probability distribution; queueing theory

DOI: 10.14375/19922264140307

Acknowledgments

The research is supported by the Russian Foundation for Basic Research (grants Nos. 13-01-00406 and 13-07-00408).

References

1. Gnedenko, B.V., and I. N. Kovalenko. 1989. *Introduction to queueing theory*. Boston: Birkhauser. 314 p.
2. Kalashnikov, V.V. 1990. *Mathematical methods for construction of queueing models*. N.Y.: Springer-Verlag. 431 p.
3. Bremaud, P. 1979. Optimal thinning of a point process. *SIAM J. Contr. Optim.* 17(2):222–230. doi: 10.1137/0317017.
4. Miller, B. M., K. E. Avrachenkov, K. V. Stepanyan, and G. B. Miller. 2005. Flow control as stochastic optimal control problem with incomplete information. *Problems Infor-*

- mation *Transmission* 41(2):150–170. doi: 10.1007/s11122-005-0020-8.
5. Shen, B., Z. Wang, and H. Shu. 2013. *Nonlinear stochastic systems with incomplete information: Filtering and control*. N.Y.: Springer Verlag. 248 p.
 6. Anagnostopoulos, A., A. Kirsch, and E. Upfal. 2005. Load balancing in arbitrary network topologies with stochastic adversarial input. *SIAM J. Comput.* 17(3):616–639. doi: 10.1137/S0097539703437831.
 7. Altman, E., U. Ayesta, and B. Prabhu. 2011. Load balancing in processor sharing systems. *Telecommun. Syst.* 47(1-2):35–48. doi: 10.4108/ICST.VALUETOOLS2008.4462.
 8. Bosov, A. V. 2009. Modeling and optimization of functioning of the Information Web Portal. *Programming Computer Software* 35(6):340–350 doi: 10.1134/S0361768809060048.
 9. Bosov, A. V. 2012. Zadachi analiza i optimizatsii dlya modeli pol'zovatel'skoy aktivnosti. Chast' 1. Analiz i prognozirovanie [Analysis and optimization problems for some users activity model. Part 1. Analysis and prediction]. *Informatika i ee Primeneniya — Inform. Appl.* 5(4):40–52.
 10. Olshefski, D., J. Nieh, and D. Agrawal. 2004. Using CERTES to infer client response time at the web server. *ACM Trans. Comput. Syst.* 22(1):49–93. doi: 10.1145/511334.511355.
 11. Ozsu, M. T., and P. Valduriez. 2011. *Principles of distributed database systems*. N.Y.: Springer-Verlag. 845 p.
 12. Elliott, R. J., L. Aggoun, and J. B. Moore. 1994. *Hidden Markov models: Estimation and control*. N.Y.: Springer-Verlag. 382 p.
 13. Liptser, R. Sh., and A. N. Shirayev. 1989. *Theory of martingales*. N.Y.: Springer-Verlag. 812 p.
 14. Jacod, J., and A. Shirayev. 2003. *Limit theorems for stochastic processes*. N.Y.: Springer-Verlag. 664 p.
 15. Elliott, R. J. 1982. *Stochastic calculus and applications*. N.Y.: Springer-Verlag. 302 p.
 16. Wong, E., and B. Hajek. 1985. *Stochastic processes in engineering systems*. N.Y.: Springer-Verlag. 361 p.
 17. Borisov, A. V., B. M. Miller, and K. V. Semenikhin. 2014 (in press). Filtering of Markov jump process given the observations the observations of multivariate point process. *Autom. Rem. Contr.*
 18. Yushkevich, A. A. 1978. Controlled Markov models with countable state space and continuous time. *Theory Probabl. Appl.* 22(2):215–235. doi: 10.1137/1122029.
 19. Kalman, R. E., and R. S. Bucy. 1960. New results in linear filtering and prediction theory. *Trans. ASME. J. Basic Eng.* 83D(1):95–108.
 20. Cvitanic, J., R. Liptser, and B. Rozovskii. 2006. A filtering approach to tracking volatility from prices observed at random times. *Ann. Appl. Probab.* 16:1633–1652. doi: 10.1214/105051606000000222.

Received June 27, 2014

Contributor

Borisov Andrey V. (b. 1965) — Doctor of Science in physics and mathematics, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Department of Probability Theory, School of Applied Mathematics and Physics, Moscow Aviation Institute, 4 Volokolamskoe Shosse, GSP-3, A-80, Moscow 125993, Russian Federation; A.Borisov@ipiran.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОЦЕДУР РЕГРЕССИОННОГО АНАЛИЗА

М. П. Кривенко¹

Аннотация: Рассмотрена задача прогнозирования значений одной переменной по значениям другой методами регрессионного анализа. В перечень сравниваемых процедур регрессионного анализа были включены следующие методы оценивания параметров модели: обычный наименьших квадратов, наименьших квадратов на основе знаковых статистик, нулевой корреляции на основе ранговых статистик. Реализация метода наименьших квадратов на основе знаковых статистик потребовала построения эффективной процедуры обработки данных, для чего был описан и исследован вариант метода наискорейшего спуска. В сложной ситуации кусочно-постоянной целевой функции получен работоспособный вариант соответствующего алгоритма. Проведенный сравнительный анализ процедур регрессионного анализа в реальных условиях, когда данные не подчиняются нормальному распределению, показал преимущества непараметрических методов. Принимая во внимание алгоритмические аспекты, в рассматриваемом случае предпочтение надо отдать процедурам, основанным на рангах, а не на знаках. Достоинства непараметрических методов позволяют повысить точность согласования двух способов измерения хромогранина А, широко используемого в качестве иммуногистохимического маркера нейроэндокринной дифференцировки.

Ключевые слова: регрессионный анализ; ранговые и знаковые процедуры; качество прогноза; согласование результатов измерений

DOI: 10.14375/19922264140308

1 Введение

Если в ходе регрессионного анализа появляются свидетельства того, что распределение данных существенно отличается от нормального (в частности, присутствуют аномальные наблюдения), появляются сомнения в эффективности общепринятых методов и приходится обращаться к робастным методам. Соответствующая мотивация дана во множестве работ (см., в частности, гл. 2 в [1] или разд. 1 в [2]). В данной работе рассматривается задача линейного регрессионного анализа двухмерных данных и использование результатов ее решения для прогнозирования значений одной переменной с помощью другой.

2 Методы линейного регрессионного анализа

Классический параметрический метод подбора вида зависимости между переменными основывается на методе наименьших квадратов, он обладает оптимальными свойствами при выполнении условий независимости и гомоскедастичности. Дополнительное предположение о нормальности распре-

деления данных приводит к простым статистическим процедурам описания получающихся оценок.

Для данных примем следующую модель:

$$Y = XA + E,$$

где $Y = (y_1, \dots, y_n)^T$ — зависимая переменная, $X = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}$ — независимая переменная (предиктор), $A = (a_1, a_2)^T$ — параметры модели, $E = (\varepsilon_1, \dots, \varepsilon_n)^T$ — ошибка представления Y с помощью X .

Обычные оценки наименьших квадратов (OLS-оценки) являются решением следующей задачи:

$$S_{\text{OLS}}(a_1, a_2) \xrightarrow{a_1, a_2} \min .$$

Здесь $S_{\text{OLS}}(a_1, a_2) = \sum_{i=1}^n (y_i - a_1 - x_i a_2)^2$ и имеют вид:

$$a_2^* = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad a_1^* = \bar{y} - a_2^* \bar{x},$$

где верхняя горизонтальная черта означает выборочное среднее. Здесь и далее аббревиатуры типа

¹Институт проблем информатики Российской академии наук, mkrivenko@ipiran.ru

OLS формировались на основе общепринятых англоязычных терминов.

Оценки параметров линейной регрессии с помощью непараметрических методов строятся путем применения идей OLS-оценки и использования аналогов обычных статистик (см., например, разд. 8 в [3]), причем объектом анализа становятся так называемые остатки, т. е. величины $y_i - a_1 - x_i a_2$, $i = 1, \dots, n$.

Регрессия на основе линейных ранговых статистик Манна–Уитни–Вилкоксона использует коэффициент корреляции общего вида с заменой значений остатков на их ранги, что приводит к уравнению для получения оценок параметров регрессионной модели, имеющему следующий вид:

$$S_{RZE}(a_1, a_2) = 0, \quad (1)$$

где

$$\begin{aligned} S_{RZE}(a_1, a_2) &= \\ &= \sum_{i=1}^n (x_i - \bar{x}) \operatorname{rank}(y_i - a_1 - a_2(x_i - \bar{x})). \end{aligned}$$

Принимая во внимание инвариантность рангов относительно сдвига, получаем, что (1) эквивалентно следующему уравнению для нахождения уже только оценки для a_2 :

$$T(a_2) = 0, \quad (2)$$

где

$$T(a_2) = \sum_{i=1}^n x_i \operatorname{rank}(y_i - a_2 x_i) - \frac{n(n+1)}{2} \bar{x}.$$

Трудности решения (2) возникают в силу того, что $T(u)$ — кусочно-постоянная функция. Если предположить, что $x_j \neq x_i$ для всех допустимых неравных значений i и j , и ввести обозначения $b_{ij} = (y_j - y_i)/(x_j - x_i)$, то оказывается, что функция $T(u)$ не меняет своего значения при любых u , лежащих между двумя последующими значениями b_{ij} . При этом в общем случае решения (2) может и не быть. В качестве выхода из создавшегося положения предлагалось рассматривать $T(u)$ как аппроксимацию некоторой непрерывной функции и использовать линейную интерполяцию между двумя соседними значениями u , для которых $T(u)$ принимает различные по знаку значения (см. разд. 8.1.2 в [3]). Таким образом может быть найдена оценка a_2^* для a_2 .

Но для задачи прогнозирования важна оценка и другого параметра регрессионной модели. Существуют различные варианты ее получения (см.

разд. 8.1.5 в [3]), один из них основывается на предположении относительно наблюдаемых значений и является действенным и простым. Для выбранной модели данных остатки вида $y_i - a_2 x_i$ для $i = 1, \dots, n$ имеют идентичные распределения с медианой, близкой к a_1 . Поэтому идея метода оценивания a_1 состоит в получении выборочной медианы для наблюдаемых значений $y_i - a_2^* x_i$. Это замечание завершает описание шагов по оцениванию параметров a_1 и a_2 .

Знаковые оценки параметров модели вводятся на основе знаковых критериев значимости и их статистик (см. гл. 3 в [4]). Они являются решением задачи

$$S_{SLS}(a_1, a_2) \xrightarrow{a_1, a_2} \min, \quad (3)$$

где

$$\begin{aligned} S_{SLS}(a_1, a_2) &= \left(\sum_{i=1}^n \operatorname{sign}(y_i - a_1 - x_i a_2) \right)^2 + \\ &+ \left(\sum_{i=1}^n x_i \operatorname{sign}(y_i - a_1 - x_i a_2) \right)^2. \end{aligned}$$

К сожалению, предложенные в [4] решения этой задачи даже применительно к простой двухпараметрической регрессионной модели далеки от завершения. Поэтому для того, чтобы провести сравнительный анализ методов оценивания, пришлось сначала предложить и исследовать приемы построения знаковых оценок.

Введем ряд функций от одной переменной u , зависящих от наборов векторных параметров r , v , w (всех или части из них), и исследуем их свойства. Они будут играть ключевую роль при решении задачи (3). Начнем с функции следующего вида:

$$\psi_0(r, v, w, u) = \sum_{i=1}^n r_i \operatorname{sign}(v_i - w_i u).$$

Выделим в сумме слагаемые с $w_i = 0$, что приведет к следующим представлениям:

$$\begin{aligned} \psi_0(r, v, w, u) &= \sum_{\substack{i=1, \dots, n \\ w_i=0}} r_i \operatorname{sign}(v_i) + \\ &+ \sum_{\substack{i=1, \dots, n \\ w_i \neq 0}} r_i \operatorname{sign}(v_i - w_i u) = \sum_{\substack{i=1, \dots, n \\ w_i=0}} r_i \operatorname{sign}(v_i) + \\ &+ \sum_{\substack{i=1, \dots, n \\ w_i \neq 0}} r_i \operatorname{sign}(w_i) \operatorname{sign}\left(\frac{v_i}{w_i} - u\right) = \\ &= \sum_{\substack{i=1, \dots, n \\ w_i=0}} r_i \operatorname{sign}(v_i) + \sum_{\substack{i=1, \dots, n \\ w_i \neq 0}} \tilde{r}_i \operatorname{sign}(\tilde{v}_i - u), \end{aligned}$$

где $\tilde{r} = r_i \operatorname{sign}(w_i)$; $\tilde{v}_i = v_i/w_i$.

Рассмотрим сумму, содержащую только слагаемые с $w_i \neq 0$, а именно:

$$\psi_1(\tilde{r}, \tilde{v}, u) = \sum_{i=1}^m \tilde{r}_i \text{sign}(\tilde{v}_i - u).$$

Упорядочим значения \tilde{v}_i по возрастанию, что даст последовательность $\{\tilde{v}_{(i)}\}$. После перестановки индексов получим эквивалентную запись суммы:

$$\psi_1(\tilde{r}, \tilde{v}, u) = \sum_{i=1}^m \tilde{r}_{i(\tilde{v}_{(i)})} \text{sign}(\tilde{v}_{(i)} - u),$$

где $i(\tilde{v}_{(i)})$ — индекс значения $\tilde{v}_{(i)}$ в наборе $\{\tilde{v}_i\}$ до упорядочивания. Тогда верны следующие утверждения:

- функции $\psi_1(\tilde{r}, \tilde{v}, u)$ и $\psi_0(r, v, w, u)$ являются кусочно-постоянными с точками разрыва 1-го рода $\tilde{v}_{(i)}$;
- $\psi_1(\tilde{r}, \tilde{v}, -\infty) = \sum_{\substack{i=1, \dots, n \\ w_i \neq 0}} \tilde{r}_i$ и $\psi_0(r, v, w, -\infty) = \sum_{\substack{i=1, \dots, n \\ w_i=0}} r_i \text{sign}(v_i) + \sum_{\substack{i=1, \dots, n \\ w_i \neq 0}} \tilde{r}_i$;
- переход через каждую точку $\tilde{v}_{(i)}$ при возрастании u дважды изменяет значение $\psi_1(\tilde{r}, \tilde{v}, u)$ и $\psi_0(r, v, w, u)$ на величину $\tilde{r}_{i(\tilde{v}_{(i)})}$.

Справедливость перечисленных утверждений следует непосредственно из вида рассматриваемых функций.

Объединим обозначения для множеств точек, где кусочно-постоянная функция принимает некоторое значение, следующим образом: $\langle \tilde{v}_{(i)}, \delta_i \rangle$, где $\delta_i = 0$, когда речь идет о точке, и $\delta_i = \tilde{v}_{(i+1)} - \tilde{v}_{(i)} > 0$, когда речь идет об интервале. Для пар $\langle \tilde{v}_{(i)}, \delta_i \rangle$ действует обычный лексикографический порядок, а для множества пар определен минимальный элемент.

Теперь можно перейти к описанию поведения квадратов введенных функций. Функции вида $\psi_0^2(r, v, w, u)$ или $\psi_1^2(\tilde{r}, \tilde{v}, u)$ достигают своего минимума либо в точке $\tilde{v}_{(i)}$, либо на интервале $(\tilde{v}_{(i)}, \tilde{v}_{(i+1)})$ для некоторого значения i ; этот минимум может быть найден с помощью перебора пар $\langle \tilde{v}_{(i)}, \delta_i \rangle$. Аналогичные выводы справедливы и для функции

$$\varphi_0(r, v, w, u) = \psi_0^2(1, v, w, u) + \psi_0^2(r, v, w, u),$$

где $1 = (1, \dots, 1)^T$.

Введем дополнительные ограничения на вектор r и рассмотрим следующую функцию:

$$\psi_2(r, v, u) = \sum_{i=1}^m r_i \text{sign}(v_i - u), \quad r_i > 0 \text{ для всех } i.$$

Для нее выполняется дополнительное свойство: переход через каждую точку $v_{(i)}$ при возрастании u уменьшает значение $\psi_2(r, v, u)$ дважды на величину $r_{i(v_{(i)})}$.

Далее понадобятся свойства функции $\psi_2(1, v, u)$. Она достигает нулевого значения при $\langle v_{(m/2)}, v_{(m/2)+1} - v_{(m/2)} \rangle$, если m четно, или при $\langle v_{\lfloor m/2 \rfloor + 1}, 0 \rangle$, если m нечетно.

Для $\varphi_2^2(r, v, u)$ поиск минимума перебором точек $v_{(i)}$ можно сделать более эффективным. Дело в том, что функция $\psi_2^2(r, v, u)$ не возрастает в области, где $\psi_2(r, v, u) > 0$, и не убывает, где $\psi_2(r, v, u) < 0$; этим можно воспользоваться для завершения последовательного перебора пар $\langle v_{(i)}, \delta_i \rangle$ для $i = 1, 2, \dots$

Функция $\psi_2^2(1, v, u)$ достигает своего нулевого минимального значения при $\langle v_{(m/2)}, v_{(m/2)+1} - v_{(m/2)} \rangle$, если m четно, или при $\langle v_{\lfloor m/2 \rfloor + 1}, 0 \rangle$, если m нечетно.

Как следствие, становится возможным сократить перебор и для функции вида:

$$\varphi_2(r, v, u) = \psi_2^2(1, v, u) + \psi_2^2(r, v, u).$$

Просмотр точек разрыва $v_{(i)}$ для $i = 1, 2, \dots$ следует прекратить для такого наименьшего $i = k$, при котором $\psi_2(1, v, v_{(k)}) < 0$ и $\psi_2(r, v, v_{(k)}) > \psi_2(r, v, v_{(k-1)})$.

Вернемся к задаче (3), решение которой будет формироваться каким-либо итерационным методом. Приближение на t -м шаге итерации обозначим как $(a_1^{(t)}, a_2^{(t)})$ и рассмотрим получение $(a_1^{(t+1)}, a_2^{(t+1)})$ на следующем шаге итерации. Зная $(a_1^{(t+1)}, a_2^{(t+1)})$, можно вычислить соответствующее значение $S_{SLS}(a_1^{(t+1)}, a_2^{(t+1)})$ и сравнить его с тем, что получено на предыдущем шаге итерации. Процесс уточнения оценок будет завершаться при многократном повторении одинаковых значений целевой функции.

Уже понятно, что решением оптимизационных задач может быть либо отдельное значение, либо интервал. Для того чтобы в дальнейшем обеспечить однозначность при использовании оценок параметров модели, во втором случае в качестве решения будем выдавать некоторое значение из интервала, полученное случайным образом, а именно: если минимум некоторой функции достигается при значениях $\langle v, \delta \rangle$, где $\delta > 0$, то итоговое решение $v^* = \text{Random}(v, v + \delta)$, где $\text{Random}(c, d)$ — случайная величина, равномерно распределенная на интервале (a, b) .

Руководствуясь условиями далее рассматриваемой реальной задачи, предположим, что $x_i > 0$ для $i = 1, \dots, n$. Теперь опишем возможные методы

решения (3), ограничившись заданием очередного шага итерации.

Метод поэлементно-покоординатного спуска. Будем сначала решать задачу

$$\left(\sum_{i=1}^n \text{sign} \left(y_i - a_1 - x_i a_2^{(t)} \right) \right)^2 \Rightarrow_{a_1} \min$$

или

$$\psi_2^2(1, v, u) \Rightarrow_u \min,$$

где $v_i = y_i - x_i a_2^{(t)}$.

Ранее было указано, что $\min_u \psi_2^2(1, v, u) = 0$ и он достигается при $a_1^{(t+1)} \in (v_{(n/2)+1} - v_{(n/2)})$, если n четно, или при $a_1^{(t+1)} = v_{\lfloor n/2 \rfloor + 1}$, если n нечетно.

После этого будем решать задачу

$$\left(\sum_{i=1}^n x_i \text{sign} \left(y_i - a_1^{(t+1)} - x_i a_2 \right) \right)^2 \Rightarrow_{a_2} \min$$

или

$$\psi_2^2(r, v, u) \Rightarrow_u \min,$$

где $r_i = x_i; v_i = (y_i - a_1^{(t+1)})/x_i$.

В итоге будет получено $(a_1^{(t+1)}, a_2^{(t+1)})$.

Такой вариант алгоритма был предложен в § 3.1 [4], где отмечается, что «теоретически его сходимость не доказана, но во всех многочисленных случаях его применений мы получили правильные результаты».

Метод покоординатного спуска. Будем сначала решать задачу

$$S_{\text{SLS}} \left(a_1, a_2^{(t)} \right) \Rightarrow_{a_1} \min,$$

где

$$S_{\text{SLS}} \left(a_1, a_2^{(t)} \right) = \left(\sum_{i=1}^n \text{sign} \left(y_i - a_1 - x_i a_2^{(t)} \right) \right)^2 + \left(\sum_{i=1}^n x_i \text{sign} \left(y_i - a_1 - x_i a_2^{(t)} \right) \right)^2,$$

или

$$\varphi_0(r, v, 1, u) \Rightarrow_u \min,$$

где $r_i = x_i; v_i = y_i - x_i a_2^{(t)}$. Ее решение даст значение $a_1^{(t+1)}$. После чего перейдем к задаче

$$S_{\text{SLS}} \left(a_1^{(t+1)}, a_2 \right) \Rightarrow_{a_2} \min,$$

где

$$S_{\text{SLS}} \left(a_1^{(t+1)}, a_2 \right) = \left(\sum_{i=1}^n \text{sign} \left(y_i - a_1^{(t+1)} - x_i a_2 \right) \right)^2 + \left(\sum_{i=1}^n x_i \text{sign} \left(y_i - a_1^{(t+1)} - x_i a_2 \right) \right)^2,$$

или

$$\varphi_0(r, v, 1, u) \Rightarrow_u \min,$$

где $r_i = x_i; v_i = (y_i - a_1^{(t+1)})/x_i$.

Таким образом будет получено значение для $a_2^{(t+1)}$.

Метод «наискорейшего» спуска. Возникающие при применении описанных методов проблемы побуждали к разработке процедур, обеспечивающих гарантированное получение минимума функции. Рассмотрим прямую, проведенную под углом α к горизонтальной оси и проходящую через точку $(a_1^{(t)}, a_2^{(t)})$. Положение точки (a_1, a_2) на этой прямой будет определяться относительно $(a_1^{(t)}, a_2^{(t)})$ с помощью величины u , т. е. координаты этой точки принимают вид $(u \cos \alpha + a_1^{(t)}, u \sin \alpha + a_2^{(t)})$. Тогда получаем:

$$\begin{aligned} y_i - a_1 - x_i a_2 &= \\ &= y_i - \left(u \cos \alpha + a_1^{(t)} \right) - x_i \left(u \sin \alpha + a_2^{(t)} \right) = \\ &= y_i - a_1^{(t)} - x_i a_2^{(t)} - (\cos \alpha + x_i \sin \alpha) u. \end{aligned}$$

В результате приходим к задаче

$$\varphi_0(r, v, w, u) \Rightarrow_u \min, \quad (4)$$

где $r_i = x_i; v_i = y_i - a_1^{(t)} - x_i a_2^{(t)}; w_i = \cos \alpha + x_i \sin \alpha$.

Решение (4) даст $(a_1^{(t+1)}, a_2^{(t+1)})$. При этом значение $S_{\text{SLS}} \left(a_1^{(t+1)}, a_2^{(t+1)} \right)$ может относительно $S_{\text{SLS}} \left(a_1^{(t)}, a_2^{(t)} \right)$ и не уменьшиться. Чтобы по возможности предотвратить это, найдем пары $(a_1^{(t+1)}, a_2^{(t+1)})$ для различных значений α и выберем среди них ту оценку параметров модели, которой отвечает наименьшее значение $S_{\text{SLS}}(a_1, a_2)$. В данной работе был опробован вариант с набором углов α вида $\{0, (1/m)\pi, \dots, ((m-1)/m)\pi\}$ для $m > 1$. Таким образом, появляется дополнительный параметр алгоритма поиска оценок параметров модели — число m перебираемых углов.

Была сформулирована и опробована еще одна процедура, реализующая условно названный метод опорных точек «наискорейшего» спуска. Далее описывается ее основной шаг. Искомое решение (a_1^*, a_2^*) задачи (3) есть некоторая прямая на плоскости, разделяющая точки (x_i, y_i) на три подмножества: точки, для которых $y_i - a_1^* - x_i a_2^* < 0$; точки, для которых $y_i - a_1^* - x_i a_2^* = 0$; точки, для которых $y_i - a_1^* - x_i a_2^* > 0$.

Рассмотрим ситуацию, когда некоторая прямая с коэффициентами (a_1, a_2) проходит по крайней мере через одну точку (x_l, y_l) , т. е. $y_l - a_1 - x_l a_2 = 0$ или $a_1 = y_l - x_l a_2$. В результате имеем:

$$\begin{aligned} y_i - a_1 - x_i a_2 &= y_i - (y_l - x_l a_2) - x_i a_2 = \\ &= (y_i - y_l) - (x_i - x_l) a_2. \end{aligned}$$

Тогда получаем задачу:

$$\varphi_0(r, v, w, u) \Rightarrow \min_u,$$

где $r_i = x_i$; $v_i = y_i - y_l$; $w_i = x_i - x_l$.

Таким образом, с помощью перебора всех пар (x_l, y_l) может быть получено решение задачи (3), правда, при дополнительном условии, что $y_l - a_1 - x_l a_2 = 0$ по крайней мере для одного из $l = 1, \dots, n$. Выяснить, нужно ли анализировать ситуации, когда прямая с коэффициентами (a_1, a_2) проходит между всеми точками (x_i, y_i) , $i = 1, \dots, n$, а также, если в этом есть необходимость, как искать минимум в этом случае, пока не удалось. При этом процедура продемонстрировала свою жизнеспособность, хотя и требует дальнейшего развития.

При найденных оценках параметров регрессии прогнозом для заданного значения x называется величина $y^* = a_1^* + a_2^* x$. При сравнении различных способов получения «новых» значений y^* необходимо определиться с понятием качества прогнозирования и методами оценивания выбранного показателя качества. Обычно для этих целей используется ошибка прогноза, измеренная как средний квадрат ошибки прогноза, где ошибка — отличие между предсказанным y^* и истинным значением y .

В данной работе интерес представляет случай, когда истинное значение y отсутствует, а имеется лишь результат его измерения с некоторой ошибкой. При этом предполагается, что при использовании переменных важны не столько их точные значения, сколько интервалы (диапазоны) значений, в которые они попадают. По этой причине в качестве показателя качества был выбран следующий. Разобьем диапазон возможных значений переменной y на k интервалов группировки (эти интервалы не пересекаются, их объединение дает

весь диапазон значений y) и представим отдельное y с помощью номера интервала, которому оно принадлежит. Считается, что прогноз осуществлен с ошибкой (индикатор ошибки равен 1), когда для некоторой пары (x, y) значения y^* и y попадают в различные интервалы группировки; в противном случае принимается, что прогноз безошибочен (индикатор — 0). Тогда ошибка прогноза есть среднее значение индикатора ошибки.

Простейший способ оценивания ошибки прогноза заключается в повторном использовании данных сначала для оценивания параметров регрессии, а затем для подсчета выборочного среднего индикатора ошибки. Понятно, что получающийся результат будет излишне оптимистичен и требует улучшения, которого можно достичь за счет использования различных частей исходных данных для подбора значений параметров модели и оценивания качества этого подбора. С учетом реально действующего условия на малый объем исходных данных приходим к следующим подходам:

- метод перепроверки [5], когда вся выборка делится на V непересекающихся групп приблизительно одинакового объема, далее удаляем одну группу и используем оставшиеся группы для подбора параметров модели, а удаленную группу — для оценивания качества подбора, повторяем эту процедуру V раз, каждый раз удаляя различные группы данных, находим итоговую оценку качества как среднее V получившихся оценок;
- бутстреп-метод [6], когда на основе исходных данных параметрическим или непараметрическим путем формируется бутстреп-выборка и с ее помощью осуществляется подбор регрессионной модели, которая используется для оценивания ошибки прогноза для новой бутстреп-выборки, подобная двухшаговая процедура повторяется необходимое число раз, что позволяет получить среднее значение ошибки прогноза.

Бутстреп-метод, безусловно, имеет более общий характер, нежели метод перепроверки, но для его реализации фактически остается только непараметрический подход, так как не существует работоспособных параметрических моделей для пары случайных переменных при отклонении от предположения о нормальном распределении (а иначе не имело бы смысла затевать исследование непараметрических методов линейного регрессионного анализа). Непараметрический же бутстреп-метод в случае выборок малого объема привносит дополнительные вопросы: например, что делать с повторяющимися данными в бутстреп-выборке. Поэтому в

данной работе использовался метод перепроверки при $V = n$ (осуществляется исключение единственного наблюдения для оценки ошибки прогноза).

3 Эксперименты

Совместно со специалистами ФГБУ «НИИ урологии» Минздрава России Сивковым А. В., Кешишевым Н. Г., Ковченко Г. А., Никоновой Л. М. решалась задача согласования двух способов определения хромогранина А (CgA) — иммуногистохимического маркера нейроэндокринной дифференцировки. В России официально зарегистрированы две соответствующие тест-системы: одна — компании ДАКО, а другая — компании Euro-Diagnostica (ED). Для исследования диагностической ценности двух тест-систем был проведен сравнительный статистический анализ показателей CgA при различных заболеваниях. Все пациенты были разделены на следующие группы: контрольную, в которую вошли здоровые мужчины (контрольная группа) и еще 7 групп с различными заболеваниями предстательной железы. Всего было проанализировано 113 мужчин, из которых 29 — это контрольная группа. Результаты исследований показали, что референсные значения принципиально различались. С помощью линейного регрессионного анализа была описана функциональная связь между измерениями, полученными тест-системами ДАКО и ED (далее — CgA_D и CgA_E соответственно). Наличие альтернативных методов измерения одного и того же показателя привело к необходимости их сравнительного анализа, при этом преследовались следующие цели:

- принять решение о применении в повседневной практике обоих методов или о предпочтении одного из них (например, с позиций доступности, точности, стоимости);
- иметь возможность одновременного использования результатов измерений, полученных различными методами.

Проведенные исследования показали, что эффективность использования обеих тест-систем практически одинакова. При этом референсные значения для системы ДАКО требуют корректировки, для чего можно использовать выявленную функциональную связь между измерениями, полученными различными способами. Эта связь также позволяет использовать в диагностике обе тест-системы.

При формировании статистических моделей для собранных данных было выяснено, в частности, следующее: для измерений CgA_D и CgA_E нет

предпосылок принять гипотезу о нормальном распределении, что побуждает к осторожности при использовании методов, ориентированных на эту модель данных.

В контексте данной работы становится интересным, можно ли повысить качество прогноза CgA_D по CgA_E, отказываясь от модели нормального распределения.

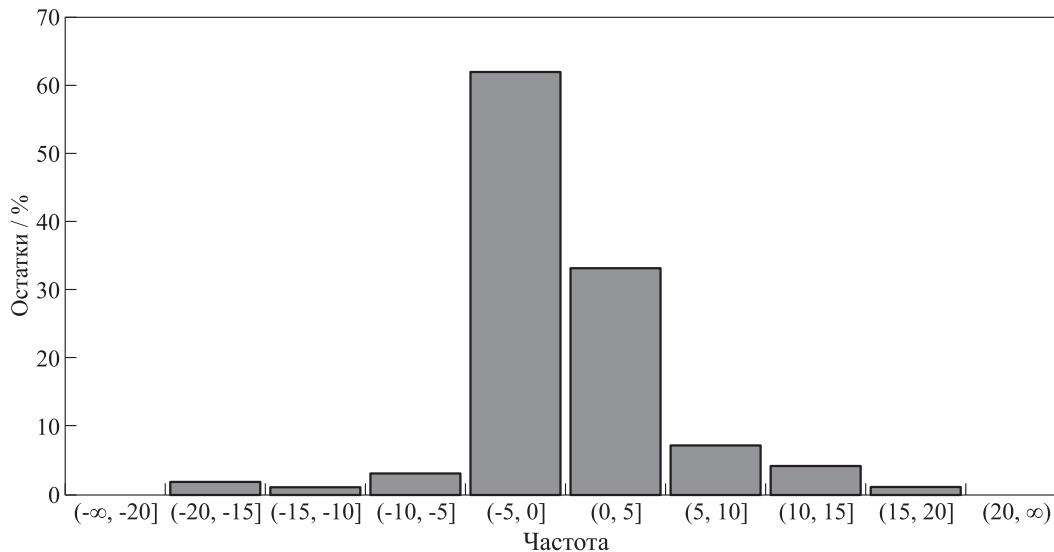
Для сравнительного анализа были взяты следующие методы оценивания:

- обычный наименьших квадратов (OLS);
- наименьших квадратов на основе знаковых статистик (SLS);
- нулевой корреляции на основе ранговых статистик (RZE).

Анализ остатков для OLS-оценки с помощью гистограммы, приведенной на рисунке, подтверждает сомнения в использовании модели нормального распределения (наличие явной асимметрии).

Наибольшую трудность в смысле реализации представляет получение SLS-оценок. Дело в том, что заверения авторов [4] о сходимости итерационного процесса не подтвердились: часто алгоритм сходил к решению, которое не было наилучшим, или его шаги сначала приводили к уменьшению целевой функции, а затем к обратному результату — росту значений $S_{SLS}(a_1, a_2)$. Отказ от поэлементной оптимизации и использование классического метода покоординатного спуска давал еще более плачевные результаты. Причина такого положения дел кроется в разрывности функции $S_{SLS}(a_1, a_2)$. В предлагаемом методе «наискорейшего» спуска за счет перебора углов α удалось преодолеть описанную трудность. При этом работоспособная реализация была достигнута при следующей схеме задания значений рассматриваемых углов: на первом шаге итерации $m = 11$, на втором — 101, на третьем — 997, на четвертом — 10 007, на пятом — $m = 100\,003$. Дальнейшего увеличения номера итерационного шага не понадобилось, так как при всех последующих экспериментах по крайней мере на двух последних итерациях целевая функция своего значения не меняла. Основные характеристики принятых значений m (простые числа в качестве значений, возрастание при увеличении номера шага итерации, увеличение на порядок) были установлены экспериментально. В качестве начального приближения при получении SLS-оценки бралась OLS-оценка.

Оценивание качества прогноза проводилось методом перепроверки следующим образом. Пусть из массива исходных данных исключено i -е наблюдение, $i = 1, \dots, n$. Найдем по оставшимся данным



Гистограмма остатков для OLS-оценки

Частота успешных прогнозов для различных оценок

k	OLS	SLS	RZE
3	97	98	98
4	91	91	91
5	85	84	84
6	76	77	77
7	68	70	70
8	62	62	62
9	64	64	64
10	58	57	57
11	50	52	51
12	45	50	50
13	48	49	50
14	39	41	41
15	35	38	38
16	38	40	41
17	32	35	36
18	27	30	32
19	27	31	31
20	30	30	30

одну из сравниваемых оценок зависимости и с ее помощью построим прогноз для исключенного наблюдения, полученный результат сравним с известным исключенным значением маркера. Если они попадают в один из k интервалов группировки, то пометим результат проведенного эксперимента как успешный, в противном случае — как неуспешный. По всем n подобным экспериментам посчитаем относительную частоту полученных успехов. Соответствующие результаты приведены в таблице.

Из полученных результатов видно, что прогноз на основе SLS-оценки лучше, чем с помощью OLS-оценки (частота случаев, когда SLS не хуже OLS, составляет 89%), а прогноз на основе RZE-оценки

лучше, чем с помощью SLS-оценки (частота случаев, когда RZE не хуже SLS, составляет 94%).

Для верификации полученного вывода о целесообразности отказа от модели нормального распределения в пользу предположений общего характера можно проделать следующее: предположить, что распределение данных есть нормальное с параметрами, совпадающими с соответствующими выборочными характеристиками (моментами и регрессионной связью), и смоделировать для него набор наблюдаемых значений, к которым применить описанную ранее процедуру обработки данных, а именно: нахождение OLS-, SLS-, RZE-оценок и сравнение качества прогноза с помощью метода перепроверки. Тогда если полученные выводы о нецелесообразности использования модели нормального распределения верны, то для смоделированных данных никаких преимуществ перехода к непараметрическим методам анализа не должно быть обнаружено. Это полностью подтвердилось, в частности частота случаев, когда SLS не хуже OLS, составляет 44%, а частота случаев, когда RZE не хуже OLS, составляет 56%.

Таким образом, обобщение модели данных приводит при ограниченном объеме наблюдений не к снижению, а к повышению качества принимаемых решений, что становится дополнительным доводом для отказа от использования нормального распределения.

4 Заключение

В перечень сравниваемых процедур регрессионного анализа были включены следующие ме-

тоды оценивания параметров модели: обычный наименьших квадратов, наименьших квадратов на основе знаковых статистик, нулевой корреляции на основе ранговых статистик. Реализация метода наименьших квадратов на основе знаковых статистик потребовала построения эффективной процедуры обработки данных, для чего был описан и исследован вариант метода «наискорейшего» спуска. В сложной ситуации кусочно-постоянной целевой функции получен работоспособный вариант соответствующего алгоритма.

Проведенный сравнительный анализ процедур регрессионного анализа в реальных условиях, когда данные не подчиняются нормальному распределению, показал преимущества непараметрических методов. Принимая во внимание алгоритмические аспекты, в рассматриваемом случае предпочтение надо отдать процедурам, основанным на рангах, а не на знаках.

Рассмотренные непараметрические процедуры содержат элементы, носящие скорее эвристический, нежели теоретически обоснованный характер (например, аналогия как принцип построения непараметрических методов, принцип подбора значений параметра a_1 в ранговых процедурах и т. п.). Это побуждает к дальнейшему развитию предлагаемых решений (например, в духе разд. 4 и 5 [7]),

что, в свою очередь, влечет за собой необходимость в проведении обширных экспериментов по сравнительному анализу получающихся модификаций методов и процедур обработки данных.

Литература

1. *Ryan T. P.* Modern regression methods. — 2nd ed. — Hoboken, NJ: Wiley, 2008. 672 p.
2. *Heritier S., Cantoni E., Copt S., Victoria-Feser M.-P.* Robust methods in biostatistics. — Chichester, U.K.: Wiley, 2009. 294 p.
3. *Sprent P., Smeeton N. C.* Applied nonparametric statistical methods. — 3rd ed. — London, U.K.: Chapman & Hall/CRC, 2001. 470 p.
4. *Болдин М. В., Симонова Г. И., Тюрин Ю. Н.* Знаковый статистический анализ линейных моделей. — М.: Наука, 1997. 288 с.
5. *Stone M.* Cross-validators choice and assessment of statistical predictions (with discussion) // *J. Roy. Statist. Soc. Ser. B*, 1974. Vol. 36. No. 2. P. 111–147.
6. *Efron B.* Bootstrap methods: Another look at the jack-knife // *Ann. Statist.*, 1979. Vol. 7. No. 1. P. 1–26.
7. *Maronna R. A., Martin D. R., Yohai V. J.* Robust statistics: Theory and methods. — Chichester, U.K.: Wiley, 2006. 436 p.

Поступила в редакцию 14.07.14

COMPARATIVE ANALYSIS OF REGRESSION ANALYSIS PROCEDURES

M. P. Krivenko

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article considers the problem of forecasting the values of one variable from the values of another variable using regression analysis techniques. The list of compared regression analysis procedures included the following parameter estimation methods: ordinary least squares, least squares based on the sign statistics, and zero correlation based on rank statistics. To implement the method of least squares on the basis of sign statistics, it is necessary to construct efficient data processing procedures. An efficient variant of the corresponding algorithm was realized for a difficult situation, when the goal function is piecewise constant. The comparative analysis of regression analysis procedures in real conditions, when data are not normally distributed, showed that nonparametric techniques are advantageous. In this case, taking into account algorithmic aspects, the preference should be given to procedures based on rank, not on signs. Advantages of nonparametric methods can improve the accuracy of measuring chromogranin A, widely used as an immunohistochemical marker of neuroendocrine differentiation.

Keywords: regression analysis; rank and sign-based procedures; prediction quality; adjustment of measurement results

DOI: 10.14375/19922264140308

References

1. Ryan, T. P. 2008. *Modern regression methods*. 2nd ed. Hoboken, NJ: Wiley. 672 p.
2. Heritier, S., E. Cantoni, S. Copt, and M.-P. Victoria-Feser. 2009. *Robust methods in biostatistics*. Chichester, U.K.: Wiley. 294 p.
3. Sprent, P., and N. C. Smeeton. 2001. *Applied nonparametric statistical methods*. 3rd ed. London, U.K.: Chapman & Hall/CRC. 470 p.
4. Boldin, M. V., G. I. Simonova, and Yu. N. Tyurin. 1997. *Sign-based methods in linear models*. Providence: AMS. 236 p.
5. Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Statist. Soc. ser. B*. 36(2):111–147.
6. Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7(1):1–26.
7. Maronna, R. A., D. R. Martin, and V. J. Yohai. 2006. *Robust statistics: Theory and methods*. Chichester, U.K.: Wiley. 436 p.

Received July 14, 2014

Contributor

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ КАК ИНСТРУМЕНТ ДВУХПАРАМЕТРИЧЕСКОГО АНАЛИЗА МАГНИТНО-РЕЗОНАНСНОГО ИЗОБРАЖЕНИЯ

Т. В. Яковлева¹, Н. С. Кульберг²

Аннотация: Рассмотрены методы анализа магнитно-резонансного изображения, основанные на решении так называемой двухпараметрической задачи. Разработанные методы обеспечивают совместное вычисление обоих статистических параметров — математического ожидания анализируемой случайной величины и ее дисперсии, т.е. одновременный расчет как полезного сигнала, так и шума. В рассматриваемых вариантах решения задачи используются методы математической статистики: метод максимума правдоподобия и варианты метода моментов. Существенное преимущество развитого двухпараметрического подхода состоит в том, что он обеспечивает эффективное решение нелинейных задач, к каковым относятся задачи шумоподавления в системах магнитно-резонансной визуализации. Оценивание искомым параметров основано исключительно на данных выборочных измерений и не ограничено какими-либо априорными предположениями. В работе проводится сопоставительный анализ вариантов рассматриваемой методологии, представлены результаты компьютерного моделирования, в ходе которого получены статистические характеристики смещения и разброса оцениваемых параметров при решении задачи различными методами. Представленные методы двухпараметрического анализа раисовского сигнала могут использоваться в составе новых информационных технологий на этапе обработки стохастических величин.

Ключевые слова: распределение Райса; метод максимума правдоподобия; метод моментов; двухпараметрический анализ; отношение сигнала к шуму

DOI: 10.14375/19922264140309

1 Введение

Методы фильтрации данных, традиционно используемые в составе информационных технологий при формировании, анализе и обработке изображений в системах магнитно-резонансной визуализации, можно объединить в следующие основные группы:

- методы анизотропной фильтрации магнитно-резонансных изображений, основанные на процессе диффузии [1, 2];
- методы фильтрации шума в магнитно-резонансных изображениях, основанные на вейвлет-преобразованиях [3, 4], а также на так называемых преобразованиях ridgelet и curvelet [5, 6], которые объединены единой концепцией разложения функций по определенным базисным функциям;
- методы фильтрации, основанные на принципах математической статистики, главным образом — на методе максимального правдоподобия [7–13]. Эти методы, как правило, направлены на оценивание параметра средней

величины или математического ожидания анализируемого сигнала магнитно-резонансного изображения. При решении задач анализа магнитно-резонансного изображения традиционно используются так называемые однопараметрические методы, основанные на предположении, что один из статистических параметров задачи — дисперсия шума — является известным априори. При этом задача решается в рамках так называемой однопараметрической модели, состоящей в предположении о том, что неизвестным является лишь один параметр задачи — параметр средней величины сигнала, в то время как второй значимый статистический параметр — дисперсия шума — предполагается априорно известной величиной. Эта однопараметрическая модель, используемая многими авторами при решении задачи, никогда не реализуется на практике, и поэтому ее применение является серьезным ограничением традиционно используемого однопараметрического подхода к обработке раисовских сигналов [8–11, 13]. Методы, рассматриваемые в настоя-

¹Вычислительный центр имени А. А. Дородницына Российской академии наук, yakovleva@ccas.ru

²Вычислительный центр имени А. А. Дородницына Российской академии наук, kulberg@yandex.ru

щей работе, основаны на использовании так называемой двухпараметрической модели для решения задачи, т. е. модели, предполагающей наличие двух неизвестных параметров анализируемого райсовского сигнала: как его средней величины, так и дисперсии шума.

Задачи анализа и обработки магнитно-резонансных изображений, как известно, относятся к кругу задач, в которых выходной сигнал представляет собой сумму искомого исходного сигнала и случайного шума, образованного многими независимыми нормально-распределенными слагаемыми с нулевым средним значением, а измеряемой и анализируемой величиной является амплитуда суммарного сигнала. Амплитуда, или огибающая, такого суммарного сигнала подчиняется распределению Райса [14]. Таким образом, задачи анализа и обработки магнитно-резонансного изображения адекватно описываются статистической моделью Райса [15].

Двухпараметрические методы анализа райсовской случайной величины, представленные в настоящей работе, обладают следующими существенными преимуществами по сравнению с упомянутыми выше традиционными методами:

- возможность применения к существенно нелинейным задачам, каковыми являются задачи шумоподавления в условиях райсовского распределения, что отличает эту технологию от первых двух упомянутых выше групп методов;
- отсутствие ограничений, характерных для однопараметрических методов, представленных третьей группой и связанных с априорными предположениями относительно величины шума.

Задачу анализа магнитно-резонансного изображения можно обобщенно отнести к кругу задач анализа состояний стохастической системы, когда начальное состояние системы характеризуется неким исходным «незашумленным» значением анализируемой величины, т. е. значением величины сигнала в отсутствие шума, а измеряются и анализируются данные, полученные в результате «зашумления» исходного сигнала гауссовским шумом. Ставшие предметом настоящей работы новые методы анализа и обработки данных на основе двухпараметрического анализа райсовской случайной величины можно рассматривать как варианты методологии, которая потенциально может использоваться в составе новых информационных технологий для обработки информации, характеризующей состояния вышеуказанной стохастической системы. Эта система изучается на основе измерения данных в «за-

шумленном» состоянии и восстановления ее исходного, не искаженного шумом значения на основе этих данных путем решения двухпараметрической задачи.

В работах [15, 16] рассматривается решение двухпараметрической задачи методом максимума правдоподобия, а в работах [17, 18] — вариантами метода моментов. Задача, поставленная и решаемая в настоящей работе, состоит в проведении сопоставления этих трех способов двухпараметрического анализа данных. В разд. 2 кратко изложены теоретические основы трех сопоставляемых методов, а в разд. 3 представлены результаты их сравнительного анализа, а именно: приведены данные, полученные в ходе численного моделирования сравниваемых методов, причем для выявления обеспечиваемых ими статистических характеристик смещения и разброса рассчитанных искомым параметром задачи при моделировании численного эксперимента проводится усреднение по очень большому числу (порядка 10^5) выборок измерений. Рассмотрены преимущества и ограничения указанных вариантов решения двухпараметрической задачи с точки зрения обеспечиваемой ими точности вычисления искомым параметром анализируемого сигнала. Теория представленных в работе методов двухпараметрического анализа райсовских данных развита в работах [15–18]. В результате проведенного теоретического исследования разработаны алгоритмы расчета неизвестных параметров, соответствующие трем рассмотренным методам математической статистики. Данные алгоритмы легли в основу созданных Н. С. Кульбергом программ расчета искомым параметром задачи с целью сопоставления представленных в работе методов [16, 17].

2 Теоретические основы двухпараметрического анализа данных

В данном разделе кратко изложены базовые положения теории двухпараметрического анализа райсовского сигнала на основе методов математической статистики, при этом опущены объемные выкладки и доказательства, представленные в работах [15–18].

2.1 Метод максимума правдоподобия

Ниже для краткости будем обозначать метод максимума правдоподобия как МП.

Как известно, при анализе данных магнитно-резонансного изображения измеряемой величиной

является амплитуда $x = \sqrt{x_{\text{Re}}^2 + x_{\text{Im}}^2}$ комплексной величины, действительная x_{Re} и мнимая x_{Im} части которой, характеризуемые величиной математического ожидания ν , искажаются нормально распределенным гауссовским шумом с дисперсией σ^2 . При этом амплитуда x имеет распределение Райса с плотностью вероятности

$$P(x|\nu, \sigma^2) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right),$$

где $I_\alpha(z)$ — модифицированная функция Бесселя первого рода порядка α ; x_i — величина сигнала i -й выборки; n — количество элементов в выборке (длина выборки). Поставленная задача состоит в том, чтобы определить неизвестные параметры ν и σ^2 на основе данных выборок измерений. Используя логарифмическую функцию правдоподобия и дифференцируя ее по ν и σ^2 с целью выявления ее максимума, после ряда преобразований получаем систему уравнений максимума правдоподобия [15, 16]:

$$\left. \begin{aligned} \nu &= \frac{1}{n} \sum_{i=1}^n x_i \tilde{I}\left(\frac{2x_i\nu}{\langle x^2 \rangle - \nu^2}\right); \\ \sigma^2 &= \frac{1}{2} (\langle x^2 \rangle - \nu^2), \end{aligned} \right\} \quad (1)$$

где

$$\tilde{I}(z) = \frac{I_1(z)}{I_0(z)}; \quad \langle x^2 \rangle = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

В [15] показано, что функция $\tilde{I}(t)$ представляет собой ограниченную монотонно возрастающую выпуклую вверх функцию на полуоси $(0, +\infty)$. Первое из уравнений системы (1) — уравнение для одной неизвестной ν , решив которое, определяем вторую неизвестную на основании второго уравнения из (1). Таким образом, решение системы двух уравнений для двух неизвестных сведено к решению одного уравнения для одной неизвестной, что существенно облегчает численное решение задачи. Вводя переменную $g(\nu) = 2\langle x \rangle \nu / (\langle x^2 \rangle - \nu^2)$, являющуюся однозначной функцией ν , первое из уравнений системы (1) можно представить в виде [16]:

$$\xi(g) = \frac{1}{n} \sum_{i=1}^n x_i \tilde{I}\left(\frac{x_i}{\langle x \rangle} g\right), \quad (2)$$

где левая часть представляет собой величину ν как обратную функцию аргумента g : $\nu = \xi(g(\nu))$. Доказано существование и единственность нетривиального решения уравнения (2) и, следовательно, существование и единственность решения системы (1) [15, 16]. Тем самым доказано существование и единственность нетривиального решения

рассматриваемой двухпараметрической задачи методом МП.

2.2 Метод, основанный на измерениях второго и четвертого моментов анализируемой величины

Ниже для краткости будем обозначать этот метод как ММ24.

Для 2-го и 4-го начальных моментов случайной величины x , подчиняющейся распределению Райса, известны следующие формулы [19]:

$$\left. \begin{aligned} \overline{x^2} &= 2\sigma^2 + \nu^2; \\ \overline{x^4} &= 8\sigma^4 + 8\sigma^2\nu + \nu^4. \end{aligned} \right\} \quad (3)$$

Принимая во внимание, что левые части уравнений (3), т. е. величины моментов анализируемого сигнала, представляют собой результаты выборочных измерений, формулы (3) можно рассматривать как простую систему двух уравнений для двух неизвестных ν и σ^2 . В решении данной системы и состоит метод ММ24.

Из формул (3) для определения параметров ν и σ^2 нетрудно получить:

$$\nu = \sqrt[4]{2\left(\overline{x^2}\right)^2 - \overline{x^4}}; \quad (4)$$

$$\sigma^2 = \frac{1}{2} \left[\overline{x^2} - \sqrt{2\left(\overline{x^2}\right)^2 - \overline{x^4}} \right]. \quad (5)$$

Важно отметить, что содержащиеся в формулах (4) и (5) величины моментов измеряемого сигнала изображения определяются в выборках измерений тем точнее, чем больше длина выборки. С помощью несложных выкладок (см. также [15, 17, 18]), можно сделать вывод о неотрицательности подкоренных выражений в формулах (4) и (5) при достаточно большой длине выборки измерений анализируемого сигнала и, следовательно, о существовании решения системы (3) для ν и σ^2 метода ММ24. Единственность этого решения следует из неотрицательности значений искомых параметров, в силу чего в решениях (4) и (5) остается лишь один из двух возможных знаков перед корнем.

Производя несложные преобразования формул (4) и (5), выражения для квадратов искомых параметров ν^2 и σ^2 можно представить в виде:

$$\left. \begin{aligned} \nu^2 &= \overline{x^2} \sqrt{1-t}; \\ \sigma^2 &= \frac{\overline{x^2}}{2} (1 - \sqrt{1-t}). \end{aligned} \right\} \quad (6)$$

В формулах (6) введено обозначение $t = \frac{\overline{x^4}}{\left(\overline{x^2}\right)^2} - 1$. Как нетрудно видеть, введенный

параметр t удовлетворяет соотношению: $0 < t \leq 1$. В частном случае распределения Рэлея, когда полезный сигнал отсутствует ($\nu = 0$), имеем $t = 1$. При этом второе из соотношений (6) дает очевидную для распределения Рэлея формулу для дисперсии: $\sigma^2 = \overline{x^2}/2$.

Таким образом, вариант метода моментов ММ24 позволяет рассчитать величины искоемых параметров математического ожидания ν и дисперсии σ^2 сигнала по формулам (4)–(6) на основе выборок измерений без затрат времени на численное решение задачи. Этот оригинальный метод является простым в практической реализации и позволяет существенно ускорить процесс обработки анализируемого сигнала.

2.3 Метод, основанный на измерениях первого и второго моментов анализируемой величины

Ниже для краткости будем обозначать этот метод как ММ12. Согласно данному методу искомые параметры ν и σ^2 рассчитываются на основе использования измеренных данных для 1-го и 2-го моментов райсовской величины, т. е. на основе метода ММ12.

Формула для 1-го момента $\bar{x} = \lim_{n \rightarrow \infty} \left((1/n) \sum_{i=1}^n x_i \right)$ райсовской величины x имеет вид [19]

$$\bar{x} = \sigma \sqrt{\frac{\pi}{2}} L_{1/2} \left(-\frac{\nu^2}{2\sigma^2} \right), \quad (7)$$

где $L_{1/2}$ — полином Лагерра. Используя соотношение между $L_{1/2}$ и модифицированными функциями Бесселя $I_0(z)$ и $I_1(z)$ [20], вместо (7) получаем:

$$\bar{x} = \sigma \sqrt{\frac{\pi}{2}} e^{-\nu^2/(4\sigma^2)} \left[\left(1 + \frac{\nu^2}{2\sigma^2} \right) I_0 \left(\frac{\nu^2}{4\sigma^2} \right) + \frac{\nu^2}{2\sigma^2} I_1 \left(\frac{\nu^2}{4\sigma^2} \right) \right]. \quad (8)$$

Формула (8) определяет аналитическую зависимость первого момента случайной величины x от искоемых параметров задачи ν и σ^2 и может использоваться в качестве одного из исходных уравнений для определения ν и σ^2 . В качестве второго уравнения в методе ММ12 используется выражение для второго момента (первая из формул (3)).

Вводя переменную $r = \nu^2/(2\sigma^2)$, характеризующую отношение сигнала к шуму, и преобразуя

уравнения для ν и σ^2 , получаем следующие уравнения для r и σ^2 [18]:

$$\sqrt{\frac{\pi}{4} \frac{\langle x^2 \rangle}{1+r}} e^{-r/2} \left[(1+r) I_0 \left(\frac{r}{2} \right) + r I_1 \left(\frac{r}{2} \right) \right] = \langle x \rangle; \quad (9)$$

$$\sigma^2 = \frac{\langle x^2 \rangle}{2(1+r)},$$

в которых была произведена замена моментов \bar{x} и $\overline{x^2}$ на соответствующие средние значения по выборке измерений $\langle x \rangle = (1/n) \sum_{i=1}^n x_i$ и $\langle x^2 \rangle = (1/n) \sum_{i=1}^n x_i^2$ в предположении, что длина выборки достаточно велика.

Уравнение (9) представляет собой уравнение с одной неизвестной r . Вопросы, связанные с существованием и свойствами его решения, рассмотрены в [18]. Не останавливаясь на деталях вычислений, отметим следующий важный результат: решение системы двух уравнений для двух неизвестных сведено к решению одного уравнения для одной неизвестной, что обеспечивает существенное сокращение объема вычислений.

В результате несложных преобразований из (9) получаем уравнение:

$$\sqrt{\frac{\pi}{4} \langle x^2 \rangle} \sqrt{1+r} e^{-r/2} I_0 \left(\frac{r}{2} \right) \left[1 + \frac{r}{1+r} \tilde{I} \left(\frac{r}{2} \right) \right] = \langle x \rangle, \quad (10)$$

где используется введенное выше обозначение $\tilde{I}(z) = I_1(z)/I_0(z)$.

Правомерность использования метода ММ12 для расчета параметров ν и σ^2 основывается на доказательстве существования решения уравнения (10) и, следовательно, существования решения исходной системы уравнений для ν и σ^2 . На основе найденного решения для r рассчитываются искомые ν и σ^2 анализируемого изображения:

$$\left. \begin{aligned} \nu &= \sqrt{\frac{r}{1+r}} \sqrt{\langle x^2 \rangle}; \\ \sigma^2 &= \frac{\langle x^2 \rangle}{2(1+r)}. \end{aligned} \right\} \quad (11)$$

Таким образом, определение искоемых параметров анализируемого магнитно-резонансного изображения методом ММ12, как и в случае метода МП, сводится к решению одного уравнения для одной неизвестной, что представляет собой одно из важных преимуществ анализируемых двухпараметрических методов и обуславливает возможное

эффективное использование данных методов в составе новых информационных технологий для целей магнитно-резонансной визуализации.

3 Сопоставление результатов численного решения двухпараметрической задачи различными методами

С целью сопоставления точности трех рассмотренных выше методов решения двухпараметрической задачи, которые могут использоваться в составе новых информационных технологий магнитно-резонансной визуализации, были проведены численные эксперименты, в которых по одним и тем же данным выборки измерений вычислялись искомые статистические параметры тремя алгоритмами, соответствующими методам МП, ММ24 и ММ12. Проведенное ниже сопоставление результатов численных экспериментов позволило выявить особенности каждого из рассматриваемых методов с точки зрения обеспечиваемой точности расчета параметров ν и σ^2 , характеризующих величину исходного сигнала и дисперсию шума в каждой анализируемой точке изучаемого магнитно-резонансного изображения.

Численные эксперименты проводились следующим образом. Для каждого из параметров ν и σ задавались массивы исходных значений, и в каждой точке такой двумерной сетки с помощью датчика случайных чисел, подчиняющихся распределению Райса, генерировались данные выборки случайного сигнала x : x_i ($i = 1, \dots, n$), причем параметры распределения Райса для данных выборки, генерируемых в каждой точке, соответствовали исходно заданным значениям средней величины ν и стандартного отклонения σ сигнала изображения в данной точке. Для каждой точки двумерной сетки в плоскости (ν, σ) , используя сгенерированные датчиком случайных чисел выборки значений сигнала x , вычислялись значения вышеуказанных параметров ν и σ на основе трех алгоритмов, соответствующих сопоставляемым методам решения задачи:

- (1) для метода МП: решение первого из уравнений (1) для параметра ν и последующий расчет параметра σ по второму из уравнений (1);
- (2) для метода ММ24: расчет параметров ν и σ по формулам (6) на основе рассчитанной по измеренным выборочным данным величины $t = \overline{x^4}/(\overline{x^2})^2 - 1$;

- (3) для метода ММ12: решение уравнения (10) для переменной $r = \nu^2/(2\sigma^2)$ и последующий расчет параметров ν и σ по формулам (11).

Результатом численного решения задачи каждым из методов являются две поверхности, образуемые рассчитанными значениями для каждого из параметров ν и σ , как функции исходно заданных значений этих параметров, соответствующих узлам двумерной сетки. Графические данные, представленные на рис. 1 и 2, являются сечениями этих поверхностей, при этом графики для ν представляют собой сечения поверхности, образованной расчетными значениями для ν , плоскостями $\sigma = const$, а графики, отображающие результаты расчетов параметра σ , являются сечениями поверхности, образованной расчетными значениями для σ , плоскостями $\nu = const$.

Представленные на рис. 1 и 2 графики соответствуют следующим значениям параметров: исходно заданные значения ν и σ , определяющие узлы двумерной сетки, на которой производились вычисления, изменялись в диапазоне от 0,2 до 3,0 с шагом 0,2; длина n выборки составляла 64; усреднение результатов проводилось по 50 выборкам.

На рис. 1 представлены графики, отображающие расчетные значения параметра ν в сечениях $\sigma = const$. На рис. 1, а – 1, в показаны зависимости отклонений расчетных значений ν от реальной, исходно заданной величины этого параметра (сплошные черные прямые) при различных исходных значениях параметра σ . Значения по оси абсцисс на графиках соответствуют точкам отсчета исходно заданных значений параметра ν , а по оси ординат — расчетным значениям данного параметра. На графике по оси абсцисс отложены цифры, соответствующие точкам отсчета для исходно заданного параметра ν . Отклонения ломаной кривой от прямой на графиках характеризуют точность расчетов при следующих значениях параметра стандартного отклонения σ : 1,6 (см. рис. 1, а), 0,8 (см. рис. 1, б) и 0,4 (см. рис. 1, в).

Из проведенного численного эксперимента, графические результаты которого представлены на рис. 1, можно сделать следующие выводы:

- точность вычисления искомого параметра ν при решении двухпараметрической задачи анализа магнитно-резонансного изображения всеми тремя рассматриваемыми методами заметно повышается с увеличением величины отношения сигнала к шуму, т. е. при перемещении слева направо вдоль расчетных кривых, а также по мере уменьшения значений «шумового» параметра σ ;

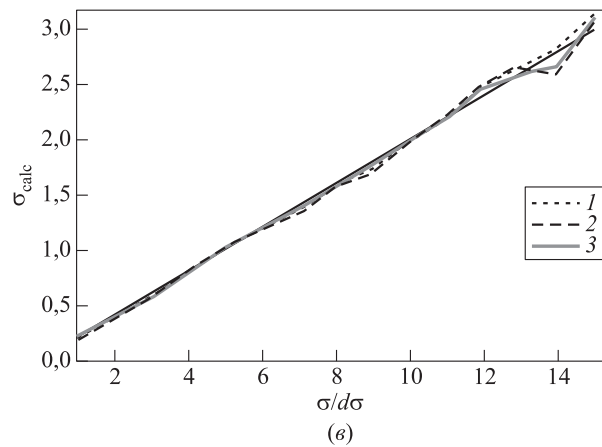
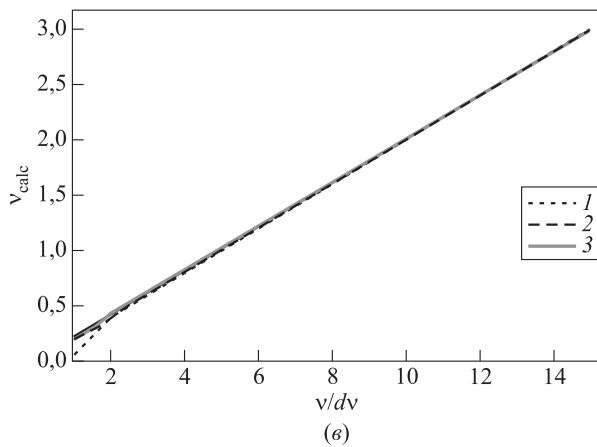
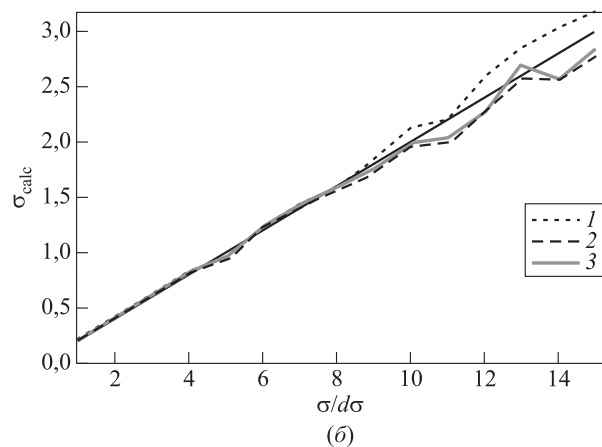
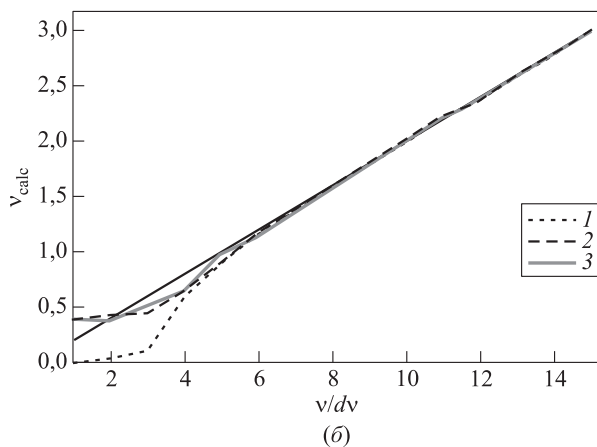
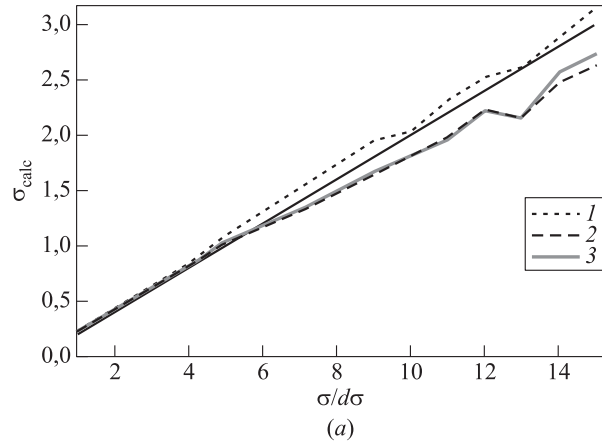
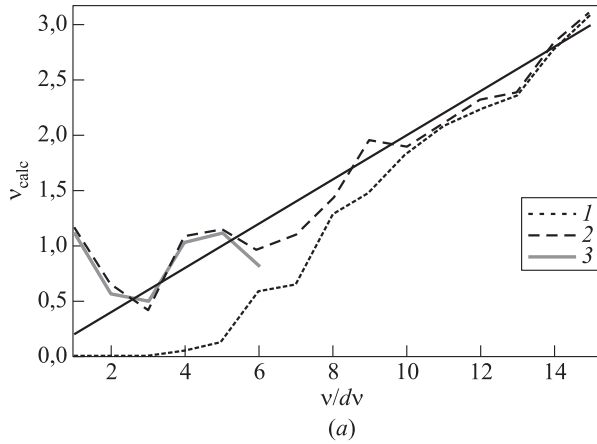


Рис. 1 Результаты численного расчета параметра ν при решении двухпараметрической задачи методами МП (1), ММ24 (2) и ММ12 (3): (а) $\sigma = 1,6$; (б) $0,8$; (е) $\sigma = 0,4$

Рис. 2 Результаты численного расчета параметра σ при решении двухпараметрической задачи методами МП (1), ММ24 (2) и ММ12 (3): (а) $\nu = 1,0$; (б) $\nu = 2,0$; (е) $\nu = 3,0$

- оба варианта метода моментов ММ24 и ММ12, будучи примерно одинаковыми по обеспечиваемой точности вычислений, заметно превосходят метод МП;
- все три рассматриваемых метода характеризуются существенным повышением точности

расчетов параметра ν и обеспечением поддержания ее на высоком уровне при достижении величиной ν/σ некоторого граничного значения: для метода МП это граничное значение близко к 2, т.е. удовлетворительная точность для параметра ν достигается при $\nu > 2\sigma$. Что

касается обоих вариантов метода моментов, то в данном случае аналогичное граничное значение для отношения ν/σ , начиная с которого обеспечивается существенное повышение точности расчетов параметра ν , составляет величину $\nu/\sigma \sim 1,5-2$. Другими словами, оба варианта метода моментов обеспечивают высокую точность расчетов в практически таком же диапазоне значений параметра ν , характеризующего полезную, незашумленную составляющую данных, формирующих магнитно-резонансное изображение, как и при использовании метода МП, в то время как вне этого диапазона точность вариантов метода моментов выше, чем точность, обеспечиваемая методом МП.

На рис. 2 представлены зависимости рассчитанной каждым из методов величины параметра стандартного отклонения σ , характеризующего уровень шума на изображении, показанные в сечениях $\nu = const$. Графики на рис. 2, а–2, в иллюстрируют зависимости отклонений расчетных значений σ от реальной величины этого параметра (сплошные черные прямые) при различных ν . На графике по оси абсцисс отложены цифры, соответствующие точкам отсчета для исходно заданного параметра σ . Отклонение ломаной линии от прямой на графиках характеризует точность расчетов. Графики построены для следующих значений ν : 1,0 (см. рис. 2, а), 2,0 (см. рис. 2, б) и 3,0 (см. рис. 2, в).

Анализируя графические результаты численного эксперимента, представленные на рис. 2 и характеризующие особенности расчета параметра стандартного отклонения σ , можно сделать следующие основные выводы:

- точность вычисления параметра σ , характеризующего уровень шума в анализируемом магнитно-резонансном изображении, при решении двухпараметрической задачи всеми тремя рассматриваемыми методами заметно падает с уменьшением величины отношения сигнала к шуму (т.е. при перемещении слева направо вдоль расчетных кривых), так как при этом увеличивается величина σ , являющаяся показателем стохастичности процесса. С ростом параметра ν отношение сигнала к шуму увеличивается и точность расчетов параметра σ возрастает, что демонстрируется сравнением графиков на рис. 2, а–2, в, соответствующих различным ν ;
- варианты метода моментов ММ24 и ММ12 характеризуются примерно одинаковой точностью расчета σ при $\nu = const$, но уступают методу МП;

– все три рассматриваемых метода характеризуются существенным повышением точности расчетов параметра σ , характеризующего степень зашумленности магнитно-резонансного изображения, при достижении величиной ν/σ некоторого граничного значения. В отличие от проанализированных выше результатов расчета параметра ν в сечениях $\sigma = const$ (см. рис. 1), графики на рис. 2, иллюстрирующие расчеты σ в сечениях $\nu = const$, показывают, что граничное значение отношения ν/σ , обеспечивающее существенное улучшение точности расчетов, примерно одинаково для всех трех методов и составляет величину $\nu/\sigma \sim 1,5$. Другими словами, удовлетворительная точность для расчета σ достигается при $\sigma < 0,7\nu$ и этот диапазон высокой точности расчетов параметра σ примерно одинаков для всех трех методов.

Для выявления статистических характеристик развитых методов решения двухпараметрической задачи были проведены численные эксперименты, в которых усреднение полученных данных производилось по большому количеству ($\sim 10^5$) выборок измерений, что позволило выявить и сопоставить характерные статистические зависимости смещения и разброса расчетных данных для параметров ν и σ^2 от условий экспериментов для методов МП, ММ24 и ММ12. Расчеты проводились для следующих значений длины выборки n : 16 (рис. 3, а и 4, а) и 64 (рис. 3, б и 4, б).

На рис. 3 представлены графики, характеризующие точность расчетов параметра ν методами ММ12 (1), ММ24 (2) и МП (3). По оси абсцисс отложена исходно заданная величина параметра ν в диапазоне от 0 до 3, в то время как величина параметра σ была принята равной 1, так что точки оси абсцисс фактически соответствуют величине отношения сигнала к шуму $SNR = \nu/\sigma$. В левом столбце рис. 3 приведены графики рассчитанных значений ν , их отклонение от прямой линии характеризует точность сопоставляемых методов. В правом столбце приведены графики среднеквадратичных отклонений (СКО) значений ν в зависимости от SNR.

На рис. 4 приведены графики, иллюстрирующие точность расчетов параметра σ методами ММ12 (1), ММ24 (2) и МП (3) при тех же условиях эксперимента, которые указаны в комментариях к рис. 3. Горизонтальная линия соответствует исходно заданному значению $\sigma = 1$. В левом столбце рис. 4 приведены графики рассчитанных значений σ , а в правом столбце — графики СКО значений σ .

Таким образом, данные в левых столбцах рис. 3 и 4 характеризуют смещения величин искомым па-

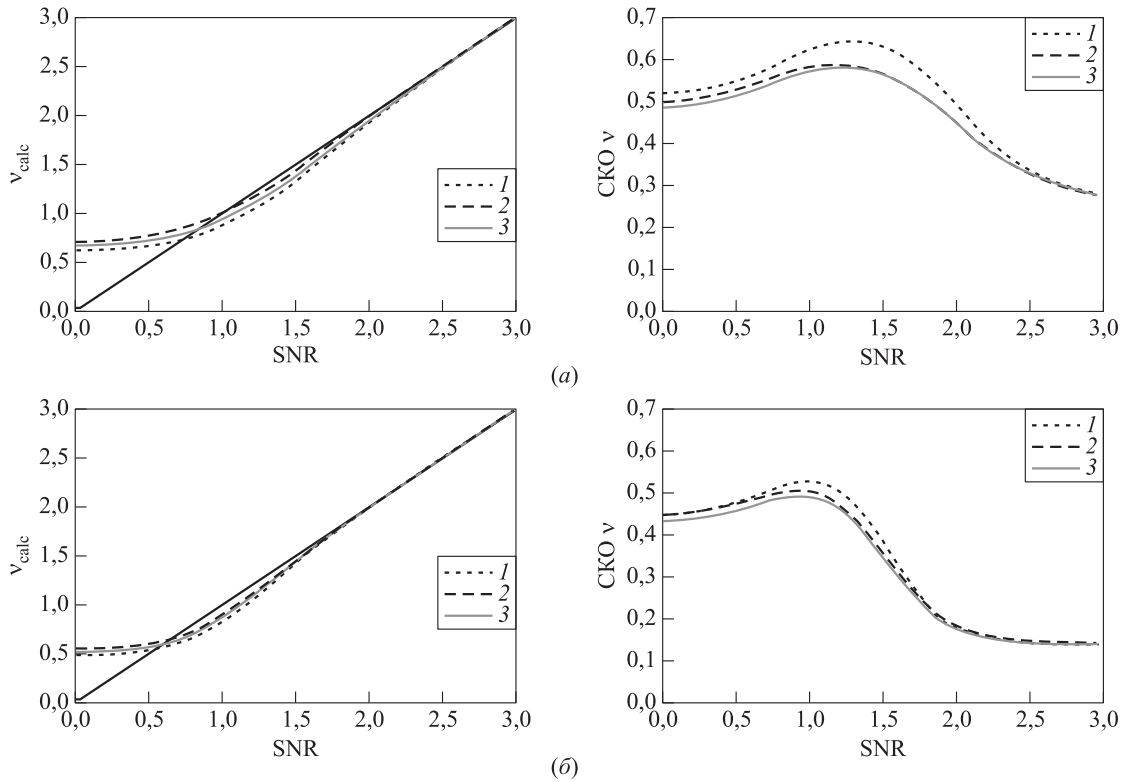


Рис. 3 Графики рассчитанных значений параметра ν (левый столбец) и его среднеквадратичного отклонения (правый столбец) в зависимости от отношения сигнала к шуму при двухпараметрическом анализе райсовского сигнала методами ММ12 (1), ММ24 (2) и МП (3) при длинах выборки $n = 16$ (а) и 64 (б)

раметров, а данные в правых столбцах характеризуют разброс результатов расчетов рассматриваемыми методами.

Графические результаты численных экспериментов, представленные на рис. 3 и 4, иллюстрируют следующие ожидаемые выводы:

- как смещение, так и разброс данных при расчете искомых параметров магнитно-резонансного изображения на основе рассматриваемых методов, т. е. посредством двухпараметрического анализа данных изображения, заметно уменьшаются с ростом отношения сигнала к шуму, а также с увеличением длины выборки n ;
- точность вычисления как ν , так и σ , методом ММ12 выше, чем другими методами при малом SNR и небольшой длине выборки, а с ростом длины выборки или увеличением SNR точность рассматриваемых методов выравнивается.

Представленные данные иллюстрируют возможность вычисления обоих искомых статистических параметров ν и σ^2 в условиях распределения Райса с высокой точностью в достаточно широком диапазоне значений отношения сигнала к шуму.

4 Заключение

В работе представлены теоретические основы, проведено численное исследование и сопоставление методов решения задачи анализа и обработки магнитно-резонансного изображения на основе нового подхода, состоящего в решении двухпараметрической задачи в условиях статистического распределения Райса. Рассматриваемые методы разработаны на основе использования так называемой двухпараметрической модели, в рамках которой анализируемые данные характеризуются не одним (как в традиционном однопараметрическом приближении), а двумя неизвестными параметрами: средней величиной сигнала и дисперсией шума. Рассматриваемые методы объединены концепцией двухпараметрического подхода к решению задачи анализа и обработки магнитно-резонансного изображения. Суть данной концепции заключается в обоснованной возможности математического расчета сразу обоих априорно неизвестных статистических параметров сигнала и шума, формирующих анализируемое изображение. Знание этих параметров позволяет, в свою очередь, восстановить исходное, не искаженное шумом изображе-

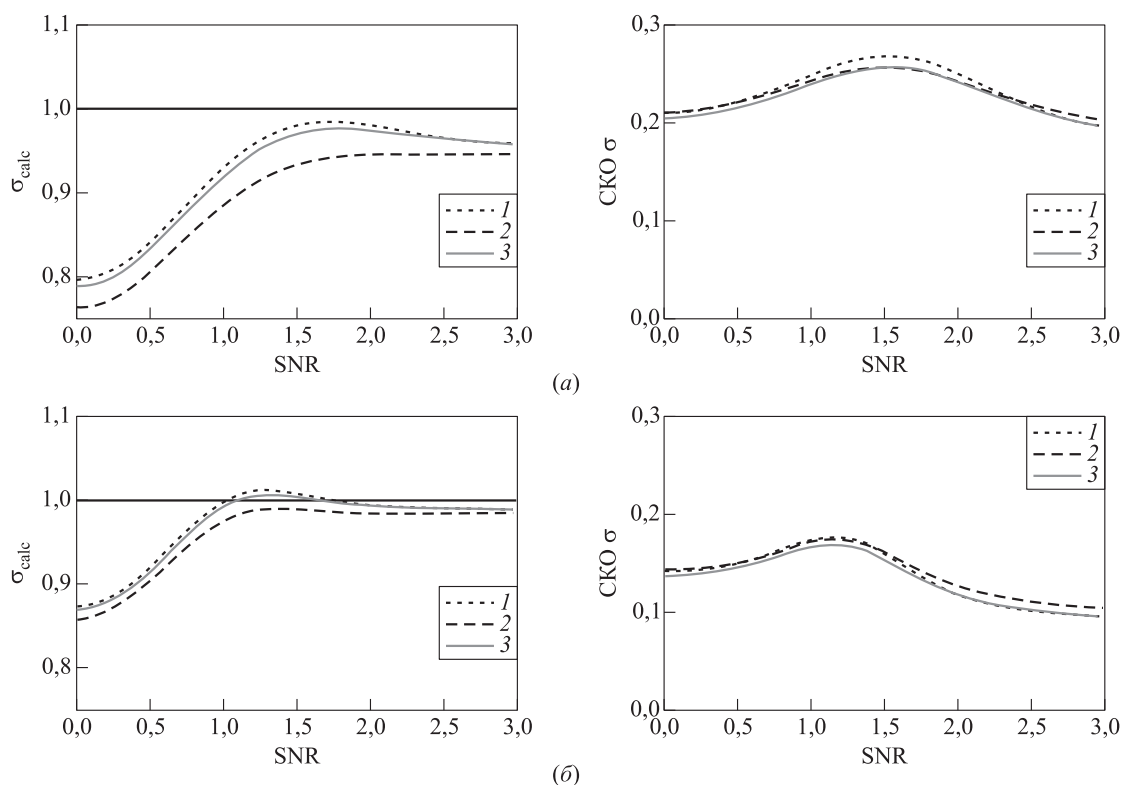


Рис. 4 Графики рассчитанных значений параметра σ (левый столбец) и его среднеквадратичного отклонения (правый столбец) в зависимости от отношения сигнала к шуму при двухпараметрическом анализе райсовского сигнала методами ММ12 (1), ММ24 (2) и МП (3) при длинах выборки $n = 16$ (а) и 64 (б)

ние в системах магнитно-резонансной визуализации. Разработаны различные алгоритмы расчета искомых величин сигнала и шума в зависимости от используемого статистического подхода к решению двухпараметрической задачи: метод максимума правдоподобия, метод моментов, основанный на измерении второго и четвертого моментов сигнала, метод моментов, основанный на измерении первого и второго моментов.

В работе проведен сравнительный анализ статистических характеристик рассматриваемых методов решения двухпараметрической задачи. В результате численного моделирования получены и сопоставлены данные для смещения и разброса искомых параметров, обеспечиваемые методами МП, ММ24 и ММ12.

Одним из важных преимуществ развитых методов является их применимость к решению существенно нелинейных задач, каковыми, в частности, являются задачи шумоподавления в магнитно-резонансной визуализации.

Результаты численного эксперимента демонстрируют возможность эффективного восстановления исходного, не искаженного шумом сигнала магнитно-резонансного изображения посредством

развитых методов, основанных на решении двухпараметрической задачи в условиях статистического распределения Райса.

Литература

1. Perona P., Malik J. Scale-space and edge detection using anisotropic diffusion // IEEE Trans. Pattern Anal. Machine Intelligence, 1990. Vol. 12. No. 7. P. 629–639.
2. Gerig G., Kubler O., Kikinis R., Jolesz F. A. Nonlinear anisotropic filtering of MRI data // IEEE Trans. Med. Imaging, 1992. Vol. 11. No. 6. P. 221–232.
3. Wood J. C., Johnson K. M. Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR // Magnet. Reson. Med., 1999. Vol. 41. No. 3. P. 631–635.
4. Delakis I., Hammad O., Kitney R. I. Wavelet based denoising algorithm for images acquired with parallel magnetic resonance imaging (MRI) // Phys. Med. Biol., 2007. Vol. 52. No. 13. P. 3741–3751.
5. Starck J.-L., Candès E. J., Donoho D. L. The curvelet transform for image denoising // IEEE Trans. Image Process., 2002. Vol. 11. No. 6. P. 670–684.
6. Ma J., Plonka G. The curvelet transform // IEEE Signal Proc. Mag., 2010. Vol. 27. No. 2. P. 118–133.

7. *Benedict T. R., Soong T. T.* The joint estimation of signal and noise from the sum envelope // *IEEE Trans. Inform. Theory*, 1967. Vol. IT-13. No. 3. P. 447–454.
8. *Henkelman R. M.* Measurement of signal intensities in the presence of noise in MR images // *Med. Phys.*, 1985. Vol. 12. No. 2. P. 232–233.
9. *Wang T., Lei T.* Statistical analysis of MR imaging and its application in image modeling // *IEEE Conference (International) Image Processing and Neural Networks Proceedings*, 1994. Vol. I. P. 866–870.
10. *Gudbjartsson H., Patz S.* The Rician distribution of noisy MRI data // *Magnet. Reson. Med.*, 1995. Vol. 34. P. 910–914.
11. *Sijbers J., den Dekker A. J., Scheunders P., van Dyck D.* Maximum-likelihood estimation of Rician distribution parameters // *IEEE Trans. Med. Imaging*, 1998. Vol. 17. No. 3. P. 357–361.
12. *Carobbi C. F. M., Cati M.* The absolute maximum of the likelihood function of the Rice distribution: Existence and uniqueness // *IEEE Trans. Instrum. Meas.*, 2008. Vol. 57. No. 4. P. 682–689.
13. *Sheil W. C.* Magnetic resonance imaging (MRI Scan). *MedicineNet.com*. Retrieved April 27, 2012.
14. *Rice S. O.* Mathematical analysis of random noise // *Bell Syst. Tech. J.*, 1944. Vol. 23. P. 282–322.
15. *Яковлева Т. В.* Условия применимости статистической модели Райса и расчет параметров райсовского сигнала методом максимума правдоподобия // *Компьютерные исследования и моделирование*, 2014. Т. 6. № 1. С. 13–25.
16. *Яковлева Т. В., Кульберг Н. С.* Особенности функции правдоподобия статистического распределения Райса // *Докл. РАН*, 2014. Т. 457. № 4. С. 394–397.
17. *Яковлева Т. В., Кульберг Н. С.* Noise and signal estimation in MRI: Two-parametric analysis of Rice-distributed data by means of the maximum likelihood approach // *Am. J. Theor. Appl. Stat.*, 2013. Vol. 2. No. 3. P. 67–79.
18. *Яковлева Т. В.* Обзор методов обработки магнитно-резонансных изображений и развитие нового двух-параметрического метода моментов // *Компьютерные исследования и моделирование*, 2014. Т. 6. № 2. С. 231–244.
19. *Park, Jr., J. H.* Moments of generalized Rayleigh distribution // *Q. Appl. Math.*, 1961. Vol. 19. No. 1. P. 45–49.
20. *Абрамовиц М., Стиган И.* Справочник по специальным функциям. — М.: Наука, 1979. 832 с.

Поступила в редакцию 09.06.14

MATHEMATICAL STATISTICS METHODS AS A TOOL OF TWO-PARAMETRIC MAGNETIC-RESONANCE IMAGE ANALYSIS

T. V. Yakovleva and N. S. Kulberg

Dorodnicyn Computing Center of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper considers the methods of the magnetic-resonance image analysis, based on the solution of the so-called two-parametric task. The elaborated methods provide joint calculation of both statistical parameters — the mathematical expectation of the random value being analyzed and its dispersion, i. e., simultaneous estimation of both the useful signal and the noise. The considered variants of the task solution employ the methods of mathematical statistics: the maximum likelihood method and variants of the method of moments. A significant advantage of the elaborated two-parametric approach consists in the fact that it provides an efficient solution of nonlinear tasks including the tasks of noise suppression in the systems of magnetic-resonance visualization. Estimation of the sought-for parameters is based upon measured samples' data only and is not limited by any *a priori* suppositions. The paper provides the comparative analysis of the considered methodology's variants and presents the results of the computer simulation providing the statistical characteristics of the estimated parameters' shift and scatter while solving the task by various methods. The presented methods of the Rician signal's two-parametric analysis can be used within new information technologies at the stage of the stochastic values' processing.

Keywords: Rice distribution; maximum likelihood method; method of moments; two-parametric analysis; signal-to-noise ratio

DOI: 10.14375/19922264140309

References

1. Perona, P., and J. Malik. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intelligence* 12(7):629–639.
2. Gerig, G., O. Kubler, R. Kikinis, and F. A. Jolesz. 1992. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. Med. Imaging* 11:221–232.
3. Wood, J. C., and K. M. Johnson. 1999. Wavelet packet denoising of magnetic resonance images: Importance of

- Rician noise at low SNR. *Magnet. Reson. Med.* 41(3):631–635.
4. Delakis, I., O. Hammad, and R. I. Kitney. 2007. Wavelet based denoising algorithm for images acquired with parallel magnetic resonance imaging (MRI). *Phys. Med. Biol.* 52(13):3741–3751.
 5. Starck, J. L., E. J. Candès, and D. L. Donoho. 2002. The curvelet transform for image denoising. *IEEE Trans. Image Process.* 11(6):670–684.
 6. Jianwei, M., and G. Plonka. 2010. The Curvelet transform. *IEE Signal Proc. Mag.* 27(2):118–133.
 7. Benedict, T. R., and T. T. Soong. 1967. The joint estimation of signal and noise from the sum envelope. *IEEE Trans. Inform. Theory* IT-13(3):447–454.
 8. Henkelman, R. M. 1985. Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.* 12(2):232–233.
 9. Wang T., and T. Lei. 1994. Statistical analysis of MR imaging and its application in image modeling. *IEEE Conference (International) Image Processing and Neural Networks Proceedings*. I:866–870.
 10. Gudbjartsson, H., and S. Patz. 1995. The Rician distribution of noisy MRI data. *Magnet. Reson. Med.* 34:910–914.
 11. Sijbers, J., A. J. den Dekker, P. Scheunders, and D. Van Dyck. 1998. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Trans. Med. Imaging* 17(3):357–361.
 12. Carobbi, C. F. M., and M. Cati. 2008. The absolute maximum of the likelihood function of the Rice distribution: Existence and uniqueness. *IEEE Trans. Instrum. Meas.* 57(4):682–689.
 13. Sheil, W. C. 2012. Magnetic resonance imaging (MRI Scan). MedicineNet.com. Retrieved April 27, 2012.
 14. Rice, S. O. 1944. Mathematical analysis of random noise. *Bell Syst. Tech. J.* 23:282–322.
 15. Yakovleva, T. V. 2014. Usloviya primenimosti statisticheskoy modeli Raysa i raschet parametrov raysovskogo signala metodom maksimuma pravdopodobiya [Conditions of Rice statistical model applicability and estimation of the Rician signal's parameters by maximum likelihood technique]. *Komp'yuternye issledovaniya i modelirovanie* [Computer Research and Simulation] 6(1):13–25.
 16. Yakovleva, T. V., and N. S. Kulberg. 2014. Osobennosti funktsii pravdopodobiya statisticheskogo raspredeleniya Raysa [The likelihood function's peculiarities for Rice statistical distribution]. *Dokl. RAS* 457(4):394–397.
 17. Yakovleva, T. V., and N. S. Kulberg. 2013. Noise and signal estimation in MRI: Two-parametric analysis of Rice-distributed data by means of the maximum likelihood approach. *Am. J. Theor. Appl. Stat.* 2(3):67–79.
 18. Yakovleva, T. V. 2014. Obzor metodov obrabotki magnitno-rezonansnykh izobrazheniy i razvitie novogo dvukhparametricheskogo metoda momentov [Review of MRI processing techniques and elaboration of a new two-parametric method of moments]. *Komp'yuternye Issledovaniya i Modelirovanie* [Computer Research and Simulation] 6(2):231–244.
 19. Park, Jr., J. H. 1961. Moments of generalized Rayleigh distribution. *Q. Appl. Math.* 19(1):45–49.
 20. Abramovits, M., and I. Stigan. 1979. *Spravochnik po spetsial'nym funktsiyam* [Reference book on special functions]. Moscow: Nauka. 832 p.

Received June 09, 2014

Contributors

Yakovleva Tatiana V. (b. 1956) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; yakovleva@ccas.ru

Kulberg Nickolas S. (b. 1970) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; kulberg@yandex.ru

ПРИМЕНЕНИЕ ПОЛУТОНОВЫХ ПРЕДСТАВЛЕНИЙ ПРИ АНАЛИЗЕ ИЗМЕНЕНИЙ ЦВЕТНЫХ ИЗОБРАЖЕНИЙ

О. П. Архипов, З. П. Зыкова

Аннотация: Чтобы сократить вычислительные и временные затраты при анализе цветных изображений, на практике часто используют их упрощенные представления. Например, применение полутоновых представлений цветных изображений в видеоаналитике позволяет уменьшить объем затрат в 2–3 раза. Часть информации об исходном изображении при этом теряется, что снижает точность анализа. Погрешности результатов существенно зависят от выбора функций преобразования цветных изображений в полутоновые. Предлагается алгоритм совместного использования различных полутоновых представлений цветных изображений, обеспечивающий повышение точности результатов. Указаны области практического применения, в которых использование предлагаемого алгоритма не влечет дополнительных временных затрат на соответствующие процедуры. Приведены иллюстрирующие примеры.

Ключевые слова: цветные изображения; полутоновые представления; погрешность анализа; видеоаналитика

DOI: 10.14375/19922264140310

1 Полутоновые представления изображений

При анализе цветных изображений в большинстве случаев применяют методы, основанные на использовании полутонового представления изображений. Для получения полутонового представления цветного изображения пиксели трехмерного RGB-пространства преобразуются в пиксели одномерного пространства, тем или иным способом характеризующего яркость RGB-пикселей.

Традиционно используемые полутоновые представления описываются функциями вида:

$$f_0 : f_0(r, g, b) = (c_0, c_0, c_0),$$

где $c_0 = 0,2126r + 0,7152g + 0,0722b$ (функция соответствует американскому стандарту аналогового телевидения NTSC — National Television Standards Committee);

$$f_1 : f_1(r, g, b) = (c_1, c_1, c_1),$$

где $c_1 = 0,21r + 0,72g + 0,07b$ (коэффициенты получены при округлении коэффициентов формулы значений функции f_0);

$$f_2 : f_2(r, g, b) = (c_2, c_2, c_2),$$

где $c_2 = 0,299r + 0,587g + 0,114b$ (функция соответствует американскому стандарту цифрового теле-

видения ATSC — Advanced Television Systems Committee);

$$f_3 : f_3(r, g, b) = (c_3, c_3, c_3),$$

где $c_3 = 0,3r + 0,59g + 0,11b$ (коэффициенты получены при округлении коэффициентов формулы значений функции f_2);

$$f_4 : f_4(r, g, b) = (c_4, c_4, c_4),$$

где

$$c_4 = \frac{r + g + b}{3};$$

$$f_5 : f_5(r, g, b) = (c_5, c_5, c_5),$$

где

$$c_5 = \frac{\max(r, g, b) + \min(r, g, b)}{2};$$

$$f_6 : f_6(r, g, b) = (c_6, c_6, c_6),$$

где

$$c_6 = \frac{255L(r, g, b)}{\rho(\max L(r, g, b) + \min L(r, g, b))},$$

L — координата из Lab-пространства, соответствующая пикселу (r, g, b) ;

$$f_7 : f_7(r, g, b) = (c_7, c_7, c_7),$$

где c_7 определяется алгоритмом PhotoShop.

Заметим, что вклад значений координат r, g и b в значения функций полутонового преобразования f_4 и f_5 одинаков, а для других функций, учитывающих цветовосприятие наблюдателя, различен.

¹Орловский филиал Института проблем информатики Российской академии наук, arhipov12@yandex.ru

²Орловский филиал Института проблем информатики Российской академии наук, zykzoia@yandex.ru



Рис. 1 Изображение I_1

В соответствии с потребностями телевидения в каждой из функций f_0-f_3 при получении серых пикселей применяются разные весовые коэффициенты при координатах r, g и b . В указанных формулах учитывается различная значимость коор-

динат r, g и b в формировании визуального восприятия человека. Так, больший вес зеленого цвета обусловлен тем, что зеленый цвет воспринимается более светлым, чем красный и синий. Меньший вес синего цвета используется, поскольку синий цвет воспринимается более темным, чем зеленый и красный.

Значения функций полутонового преобразования f_6 и f_7 нелинейно зависят от значений координат r, g и b , поскольку различный вклад этих координат определяется правилами построения равноконтрастных цветовых пространств, которые ориентированы на цветовосприятие стандартного наблюдателя.

Часть информации об исходном изображении в любом их полутоновом представлении теряется, поскольку происходит слияние цветов (одному серому пикселу соответствует множество RGB-пикселов).

Рассмотрим, например, изображение I_1 , состоящее из двенадцати разноцветных квадратов (рис. 1).

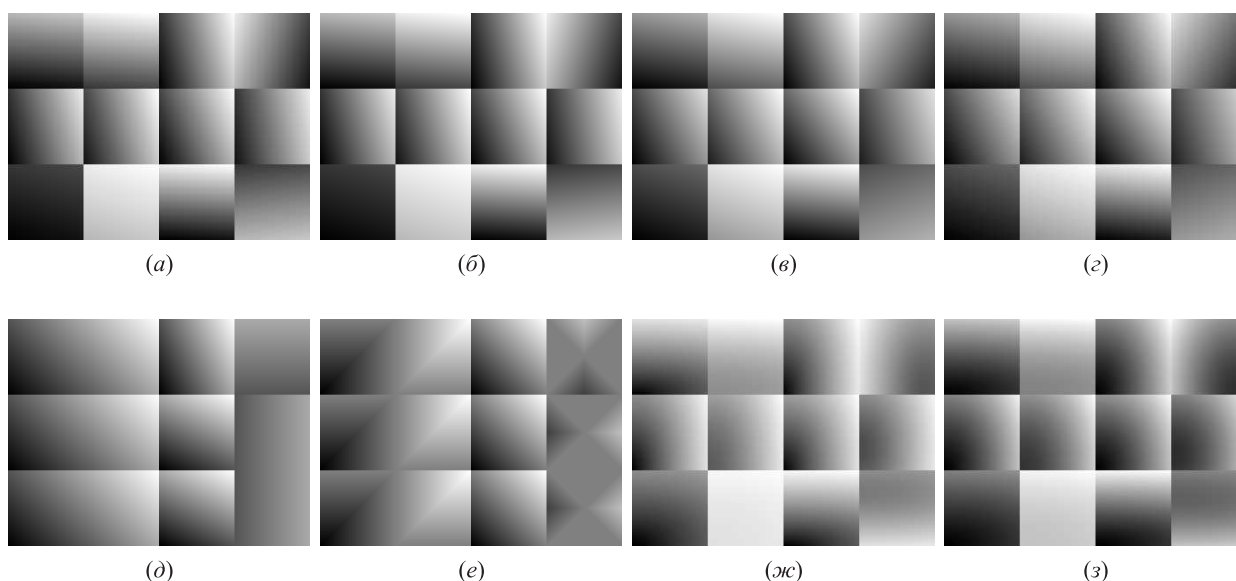


Рис. 2 Изображения $f_0(I_1)$ (а), $f_1(I_1)$ (б), $f_2(I_1)$ (в), $f_3(I_1)$ (г), $f_4(I_1)$ (д), $f_5(I_1)$ (е), $f_6(I_1)$ (ж) и $f_7(I_1)$ (з)

Таблица 1 Несовпадение пикселей изображений, приведенных на рис. 2

Рисунок	2, а	2, б	2, в	2, г	2, д	2, е	2, ж
2, б	41%						
2, в	97%	97%					
2, г	97%	97%	35%				
2, д	99%	99%	98%	98%			
2, е	99%	99%	99%	99%	98%		
2, ж	100%	100%	100%	100%	99%	99%	
2, з	98%	98%	97%	97%	99%	99%	100%

Полутоновые представления изображения I_1 имеют вид в соответствии с рис. 2 и различаются даже для близких по определению функций ($f_0(I_1)$ и $f_1(I_1)$, $f_2(I_1)$ и $f_3(I_1)$).

В табл. 1 приведены результаты подсчета несовпадений пикселей изображений, приведенных на рис. 2. Доля несовпадений представлена в процентах при округлении до целого.

Как видно из табл. 1, между значительным числом пикселей полутоновых представлений тестового изображения имеются существенные различия с количественной точки зрения. При этом, как и следовало ожидать, визуально сходные изображения ($f_0(I_1)$ и $f_1(I_1)$, $f_2(I_1)$ и $f_3(I_1)$) имеют меньшие различия по сравнению с общим случаем.

2 Характеристики полутоновых изображений

При автоматическом анализе используются различные количественные характеристики полутоновых изображений, которые могут значительно различаться даже для визуально неразличимых изображений.

Например, для обнаружения объектов на цветных изображениях и их особенностей часто применяют методы, основанные на поиске локальных экстремумов градиента (методы Робертса, Превитта, Собела и др.) в полутоновых представлениях изображений [1].

Например, градиенты $\nabla p = (G_x, G_y)$ во внутренних пикселах p полутоновых изображений, вычисленные с помощью оператора Собела, являются результатом линейной обработки фрагментов

полутоновых изображений масками 3×3 . Перед дальнейшим использованием вычисленные значения градиентов квантуются (минимум по восьми ячейкам: 0° , 45° , 90° , 135° , 180° , -45° , -90° , -135°).

Поля градиентов полутоновых представлений существенно зависят от вида функции полутонового преобразования. Несмотря на грубое квантование и малую выборку результатов, можно видеть существенное различие полей градиентов, например для полутоновых представлений $f_0(I_1)$ и $f_7(I_1)$, представленных на рис. 3 стрелками для пикселей, координаты которых по вертикали и горизонтали кратны десяти.

Для многих практически важных проектов анализа изображений необходимо и применяется квантование по большому числу ячеек. Например, в [2] описан автоматизированный программный метод анализа изображений аэрокосмических фотопланов, при реализации которого применяется 180 ячеек. Очевидно, что увеличение числа ячеек квантования приведет к еще большему разбросу результатов.

Аналогичные выводы можно сделать и при расширении базы сравнения результатов (рис. 4 и 5).

Различным направлениям на рис. 4 соответствуют векторы и соответственно пиксели, имеющие следующие цвета: 0° — красный; 45° — пурпурный; 90° — синий; 135° — циан; 180° — темно-зеленый; -45° — оранжевый; -90° — желтый; -135° — светло-зеленый.

Вследствие того, что для графического представления соотношения модулей градиентов использованы серые пиксели с координатами вида

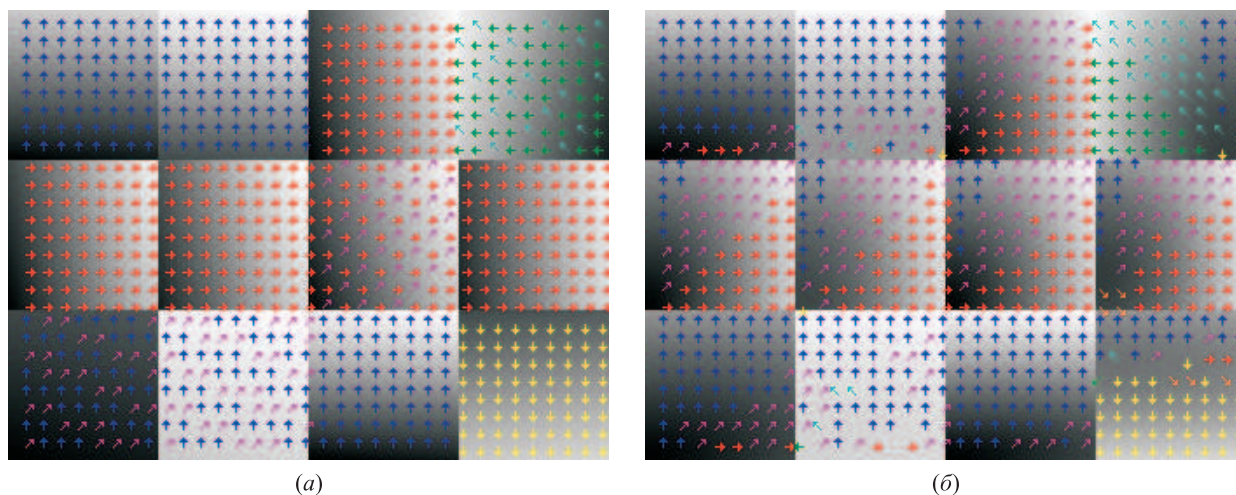


Рис. 3 Выборочные изображения полей градиентов для $f_0(I_1)$ (а) и $f_7(I_1)$ (б)

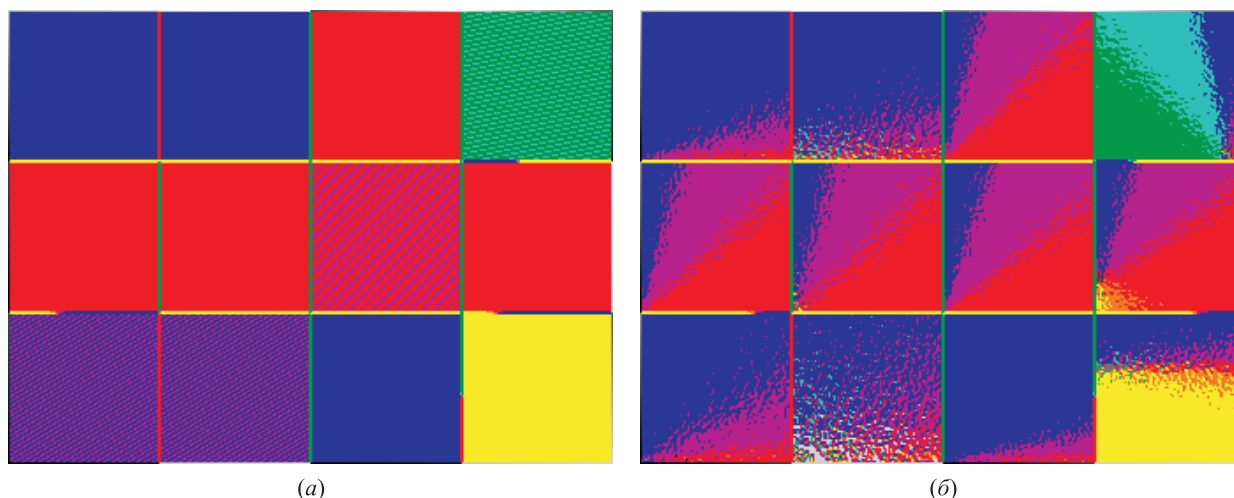


Рис. 4 Изображения полей градиентов для $f_0(I_1)$ (а) и $f_7(I_1)$ (б)



Рис. 5 Изображения полей модулей градиентов для $f_0(I_1)$ (а) и $f_7(I_1)$ (б)

$$m = \text{round} \left(255 \left(1 - \sqrt{\frac{G_x^2 + G_y^2}{2(4 \cdot 255)^2}} \right) \right), \quad (1)$$

при наименьшем значении модуля градиента ($G_x = G_y = 0$) $m = 255$, а при максимальном ($G_x = G_y = 4 \cdot 255$) $m = 0$.

меньшим значениям модуля градиента на рис. 5 соответствуют более светлые серые пиксели. Так,

В табл. 2 и 3 приведены результаты подсчета несовпадений пикселей полутоновых представле-

Таблица 2 Доля несовпадений пикселей полей направлений градиентов

Функция	$f_0(I_1)$	$f_1(I_1)$	$f_2(I_1)$	$f_3(I_1)$	$f_4(I_1)$	$f_5(I_1)$	$f_6(I_1)$	$f_7(I_1)$
$f_1(I_1)$	17%							
$f_2(I_1)$	47%	48%						
$f_3(I_1)$	47%	48%	15%					
$f_4(I_1)$	80%	80%	52%	51%				
$f_5(I_1)$	73%	74%	70%	70%	73%			
$f_6(I_1)$	38%	38%	50%	50%	72%	72%		
$f_7(I_1)$	44%	44%	48%	48%	72%	71%	24%	

Таблица 3 Доля несовпадений пикселей полей модулей градиентов

Функция	$f_0(I_1)$	$f_1(I_1)$	$f_2(I_1)$	$f_3(I_1)$	$f_4(I_1)$	$f_5(I_1)$	$f_6(I_1)$
$f_1(I_1)$	48%						
$f_2(I_1)$	81%	83%					
$f_3(I_1)$	81%	82%	40%				
$f_4(I_1)$	95%	96%	79%	79%			
$f_5(I_1)$	89%	90%	85%	85%	64%		
$f_6(I_1)$	83%	83%	82%	82%	83%	86%	
$f_7(I_1)$	85%	85%	86%	86%	90%	90%	81%

ний тестового изображения I_1 . Доля несовпадений представлена в процентах при округлении до целого.

Как видно из табл. 2 и 3, направления и модули локальных градиентов различаются для значительного числа пикселей полутоновых представлений тестового изображения. При этом, как и следовало ожидать, для визуально сходных изображений ($f_0(I_1)$ и $f_1(I_1)$, $f_2(I_1)$ и $f_3(I_1)$) поля направлений и модулей локальных градиентов имеют меньшие различия по сравнению с общим случаем.

го изображения будет использован один из многих возможных наборов функций преобразования, в котором для любой пары RGB-пикселей найдется функция, принимающая от этих аргументов разные значения: $f_0(I_1)$, $f_2(I_1)$, $f_4(I_1)$.

После применения пороговой обработки вида

$$n = \begin{cases} 255 & \text{при } m \geq 180; \\ 0 & \text{при } m < 180, \end{cases}$$

где m — нормализованные значения модуля градиентов из (1), к соответствующим полутоновым

3 Анализ полутоновых изображений

Направления и модули локальных градиентов применяются для вычисления различных особенностей цветных изображений. Например, на основе значений модуля градиентов могут быть вычислены контуры содержащихся в изображении объектов. Визуально в тестовом изображении I_1 выделяются контуры, представляющие равномерную сетку (рис. 6).

Чтобы хотя бы отчасти компенсировать погрешности, возникающие при полутоновом преобразовании цветных изображений, для совместного использования при определении контуров тестово-

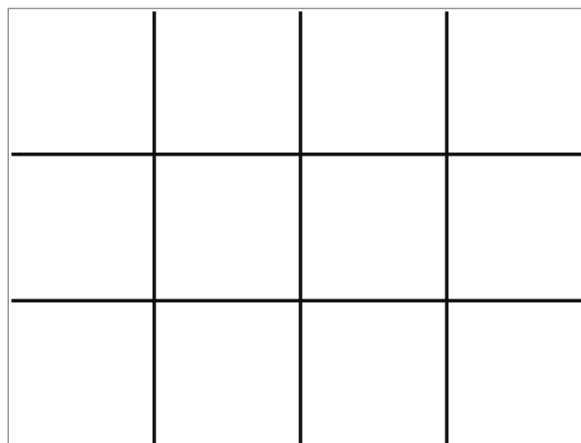


Рис. 6 Изображение визуально различных контуров I_1

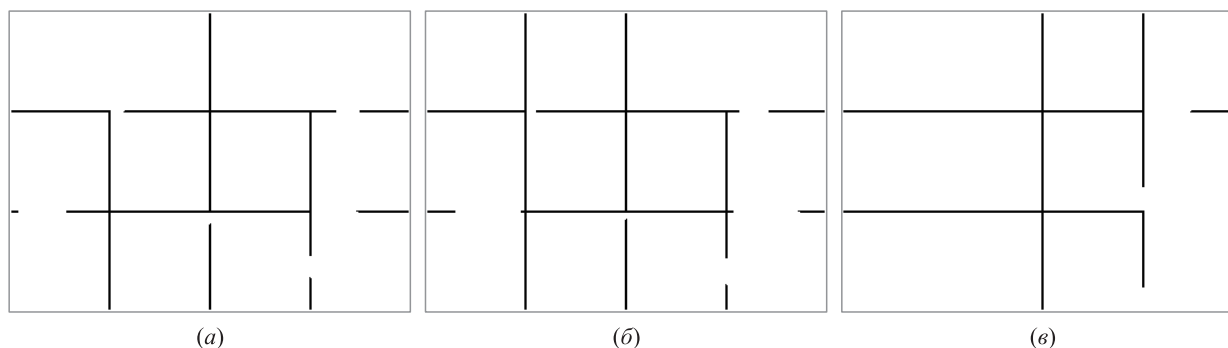


Рис. 7 Изображения фрагментов контуров для $f_0(I_1)$ (а), $f_2(I_1)$ (б) и $f_4(I_1)$ (в)

изображениям вычисленные совокупности фрагментов контуров тестового изображения имеют вид в соответствии с рис. 7.

В зависимости от используемой функции полутонового преобразования обнаруживаются разные совокупности фрагментов контуров тестового изображения. Ни одна из применяемых функций не позволила получить для тестового изображения точные результаты.

4 Совместный анализ полутоновых изображений

Точность результатов возрастает при совместном использовании нескольких полутоновых представлений цветного изображения для анализа цветных изображений [3, 4].

В рассматриваемом тестовом примере объединению данных (совокупностей фрагментов контуров, полученных при раздельном исследовании полутоновых изображений (см. рис. 7)), соответствует совокупность фрагментов контуров, изображенных на рис. 8.

Погрешность определения контуров значительно уменьшается, хотя и не устраняется полностью (два небольших фрагмента контуров не идентифицированы).

В рассмотренном примере причиной ненулевой погрешности определения контуров (см. рис. 8) при объединении отдельных результатов стало преувеличение значимости координаты g в двух из трех использованных функций по сравнению с другими координатами.

В связи с этим имеет смысл рассмотреть функции, которые в совокупности обеспечивают равнозначный вклад координат r , g и b в результат

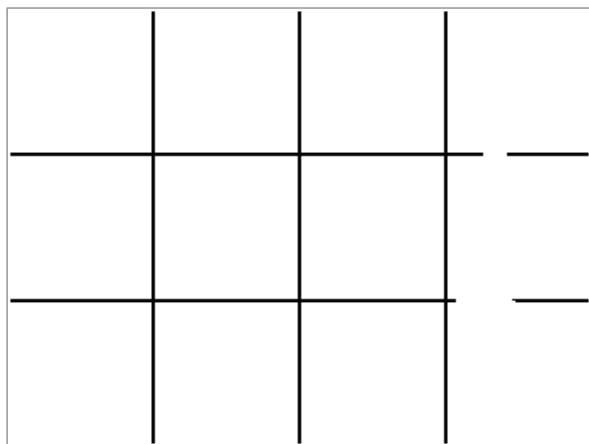


Рис. 8 Изображение объединенной совокупности фрагментов контуров

анализа. Этому требованию удовлетворяет набор функций вида

$$\left. \begin{aligned} f_8(r, g, b) &= (r, r, r); \\ f_9(r, g, b) &= (g, g, g); \\ f_{10}(r, g, b) &= (b, b, b), \end{aligned} \right\} \quad (2)$$

а также, например,

$$\left. \begin{aligned} f_{11}(r, g, b) &= (c_{11}, c_{11}, c_{11}); \\ f_{12}(r, g, b) &= (c_{12}, c_{12}, c_{12}); \\ f_{13}(r, g, b) &= (c_{13}, c_{13}, c_{13}), \end{aligned} \right\} \quad (3)$$

где

$$\begin{aligned} c_{11} &= 0,2r + 0,6g + 0,2b; \\ c_{12} &= 0,2r + 0,2g + 0,6b; \\ c_{13} &= 0,6r + 0,2g + 0,2b. \end{aligned}$$

Применение этих функций позволяет получить при объединении фрагментов точное описание контуров тестового изображения (рис. 9 и 10).

Затраты на проведение процедуры анализа складываются из суммы затрат на исследование каждого полутонового представления, на объединение данных и анализ интегрированных данных. При этом объем затрат увеличивается в несколько раз (по сравнению с затратами на реализацию аналитических процедур для единственного полутонового представления).

Это допустимо в исследовательских, но считается неприемлемым в инженерных проектах [5]. В связи с этим актуально выделение таких областей применения, которые допускают разработку более экономичных алгоритмов анализа на основе совместного исследования нескольких полутоновых представлений цветных изображений и учета особенностей их практического применения. В рамках данной работы рассматривается одна из таких областей — обнаружение изменений в последовательностях видеоизображений.

5 Совместное использование функций полутонового преобразования для обнаружения изменений в последовательностях видеоизображений

Совместное использование функций полутонового преобразования необходимо для уменьшения вероятности того, что возникновение новых или исчезновение имеющихся объектов цветного изображения останется незамеченным при автомати-

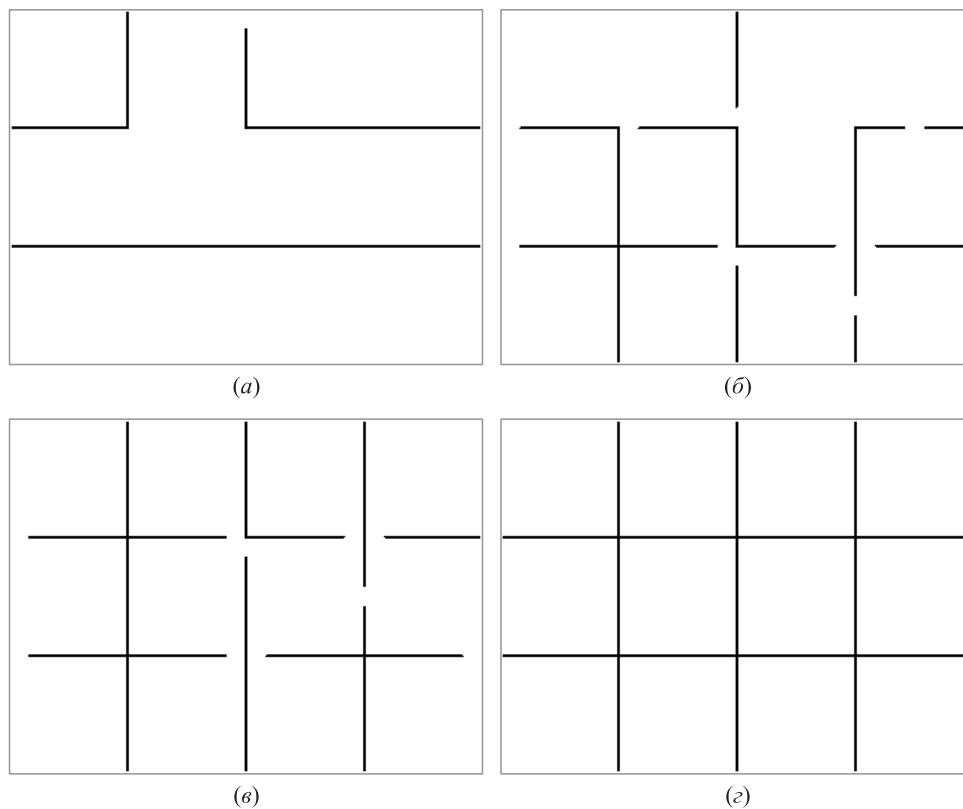


Рис. 9 Изображения фрагментов контуров для $f_8(I_1)$ (а), $f_9(I_1)$ (б), $f_{10}(I_1)$ (в) и их объединение (г)

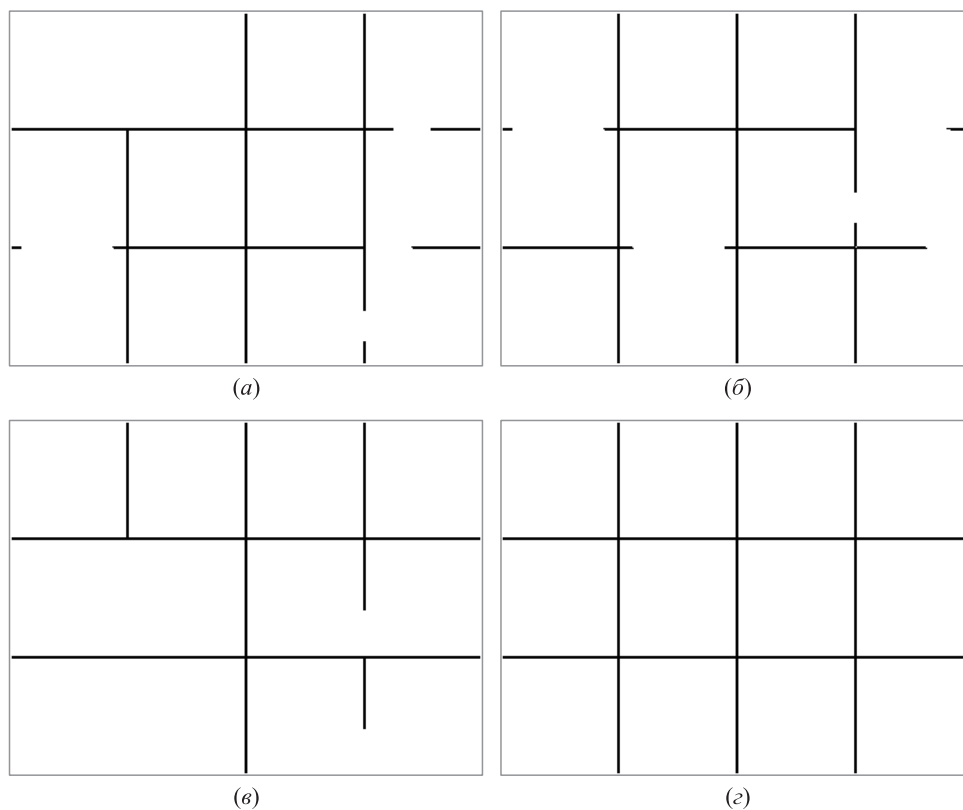


Рис. 10 Изображения фрагментов контуров для $f_{11}(I_1)$ (а), $f_{12}(I_1)$ (б), $f_{13}(I_1)$ (в) и их объединение (г)



Рис. 11 Изображения I_2 (а) и I_3 (б)

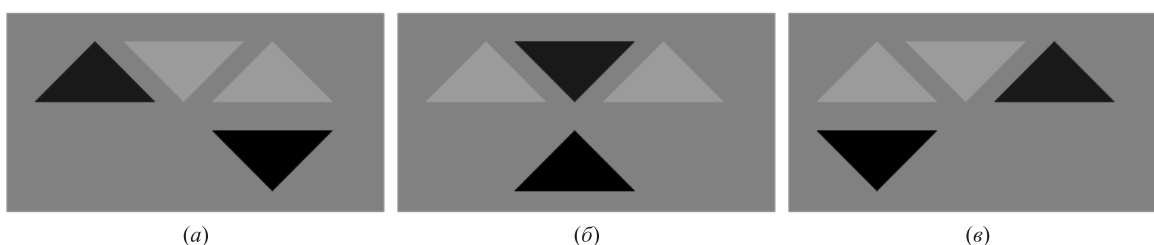


Рис. 12 Изображения $\varphi_1(V_1)$ (а), $\varphi_2(V_2)$ (б) и $\varphi_3(V_3)$ (в) при использовании набора функций полутонового преобразования (2)

ческом анализе последовательности полутоновых представлений видеоизображений.

Введем обозначения:

- $V_i, i = 0, 1, \dots$, — последовательность цветных видеоизображений;
- $\varphi_j, j = 1, 2, \dots, J$, — набор функций полутонового преобразования;
- F — функция анализа полутоновых изображений;
- ρ — мера близости значений функции F .

Два изображения $V_{i'}$ и $V_{i''}$ считаются различными, если для некоторой константы K выполнено соотношение:

$$\rho(F(\varphi_j(V_{i'})), F(\varphi_j(V_{i''}))) > K.$$

На начальном этапе предлагаемого алгоритма вычисляется J эталонных значений функции F :

$$F(\varphi_j(V_0)), j = 1, 2, \dots, J.$$

Затем при исследовании изображений из видеопоследовательности поочередно используются функции полутонового преобразования:

$$\rho(F(\varphi_j(V_0)), F(\varphi_j(V_i))) > K, i = 1, 2, \dots, \quad (4)$$

причем значение j определяется как остаток от деления i на $(J + 1)$.

Таким образом, изменения, не обнаруженные при использовании одной функции полутонового преобразования, могут быть обнаружены при использовании какой-либо другой. Затраты при этом не увеличиваются, поскольку для каждого видеоизображения используется только одно полутоновое представление.

Рассмотрим изображения I_2 и I_3 (рис. 11).

Пусть

$$V_0 = I_2; V_i = I_3, i = 1, 2, 3.$$

При использовании набора функций полутонового преобразования (2), когда

$$\begin{aligned} \varphi_1 = f_8; \varphi_2 = f_9; \varphi_3 = f_{10}; \\ \varphi_1(V_0) = \varphi_2(V_0) = \varphi_3(V_0) = V_0, \end{aligned}$$

последовательно идентифицируются все объекты, возникшие на изображениях видеопоследовательности (рис. 12).

При использовании набора функций полутонового преобразования (3), когда

$$\begin{aligned} \varphi_1 = f_{11}; \varphi_2 = f_{12}; \varphi_3 = f_{13}; \\ \varphi_1(V_0) = \varphi_2(V_0) = \varphi_3(V_0) = V_0, \end{aligned}$$

также последовательно идентифицируются все объекты, возникшие на изображениях видеопоследовательности (рис. 13).

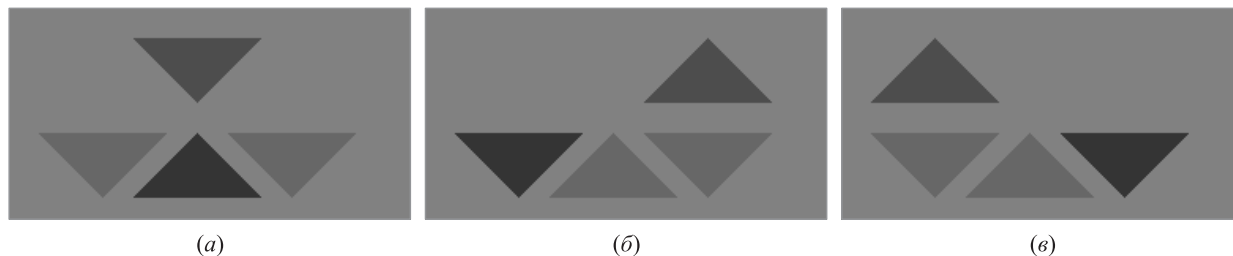


Рис. 13 Изображения $\varphi_1(V_1)$ (а), $\varphi_2(V_2)$ (б) и $\varphi_3(V_3)$ (в) при использовании набора функций полутонового преобразования (3)

Обнаружение изменений в рассмотренной тестовой последовательности изображений может быть установлено автоматически после вычисления нормы разности матриц, соответствующих полутоновым представлениям изображений тестовой последовательности и проверки выполнения критерия (4).

6 Заключение

Предложен алгоритм совместного использования полутоновых представлений при анализе цветных изображений, повышающий точность обнаружения изменений в видеопоследовательностях без увеличения затрат на выполнение соответствующих процедур.

Литература

1. Методы компьютерной обработки изображений / Под ред. В. А. Сойфера. — 2-е изд. — М.: Физматлит, 2003. 784 с.
2. Щепин М. В. Автоматизированный программный метод анализа изображений аэрокосмических фотопланов. Векторизация — анализ ландшафтных и тектонических структур // Современные проблемы дистанционного зондирования Земли из космоса (Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных явлений и объектов): Сб. докладов 2-й Открытой Всеросс. конф. Т. 2. С. 209–215. www.iki.rssi.ru/earth/articles/sec7_08.pdf.
3. Recog.ru — Распознавание образов для программистов. <http://recog.ru/blog/opencv>.
4. Van de Sande K. E. A., Gevers T., Shoek C. G. M. Evaluating color descriptors for object and scene recognition // IEEE Trans. Pattern Analysis Machine Intelligence, 2010. Vol. 32. No. 9. P. 1582–1596.
5. Конушин А. С. Введение в компьютерное зрение. Лекция 5. — Лекториум, 20.02.12. www.lektorium.tv/lecture/13541?id=13541.

Поступила в редакцию 24.06.14

APPLICATION OF GRAY-SCALE PRESENTATIONS IN THE CASE OF TREND ANALYSIS OF COLOR IMAGES

O. P. Arkhipov and Z. P. Zykova

Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation

Abstract: In order to cut computing and time expenses during analysis of color images, in practice often the simplified models are used. For example, application of gray-scale presentations of color images in video analysis allows reducing expenses by 2–3 times. At the same time, a part of information about the original image is lost which reduces the analysis accuracy. The error of the result depends essentially on choice of functions used for gray scaling. The algorithm of shared usage of multivariate gray-scale presentations of color images providing increase of accuracy of result is proposed. The regions of practical application are indicated, in which using the proposed algorithm does not result in additional temporary costs on corresponding procedures. Illustrative examples are given.

Keywords: color images; gray-scale presentations; analysis error; video analysis

DOI: 10.14375/19922264140310

References

1. Sojfer, V. A., ed. 2003. *Metody komp'yuternoy obrabotki izobrazheniy* [Methods of computer image processing]. 2nd ed. Moscow: Fizmatlit. 784 p.
2. Shchepin, M. V. 2004. *Avtomatizirovannyi programnyi metod analiza izobrazheniy aerokosmicheskikh fotoplanov. Vektorizatsiya — analiz landshaftnykh i tektonicheskikh struktur* [Automated image analysis software method aerospace photoplans. Vectorization — landscape analysis and tectonic structures]. Available at: www.iki.rssi.ru/earth/articles/sec7_08.pdf (accessed June 27, 2014).
3. Recog.ru — Raspoznavanie obrazov dlya programmistov [Pattern recognition for programmers]. Available at: <http://recog.ru/blog/opencv/> (accessed June 27, 2014).
4. Van de Sande, K. E. A., T. Gevers, and C. G. M. Shoek. 2010. Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Analysis Machine Intelligence* 32(9):1582–1596.
5. Konushin, A. C. 2012. *Vvedenie v komp'yuternoe zrenie* [Introduction to computer vision]. Lektsiya 5 [Lecture 5]. Moscow: Lektorium. Available at: www.lektorium.tv/lecture/13541?id=13541 (accessed June 27, 2014).

Received June 24, 2014

Contributors

Arkhipov Oleg P. (b. 1948) — Candidate of Science (PhD) in technology, Director, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; arkhipov12@yandex.ru

Zykova Zoya P. (b. 1953) — Candidate of Science (PhD) in physics and mathematics, Head of Laboratory, Oryol Branch of the Institute of Information Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; zykzoya@yandex.ru

ТЕКСТУРИРОВАНИЕ ВОКСЕЛЬНЫХ МОДЕЛЕЙ НА ОСНОВЕ ЦВЕТОВОЙ ИНФОРМАЦИИ ОБ ОПОРНЫХ ТОЧКАХ

О. П. Архипов¹, Ю. А. Маньяков²

Аннотация: Описана технология текстурирования воксельных моделей на основе имеющейся цветовой информации об опорных точках (ОТ). Опорные точки представляют собой особые точки в трехмерном пространстве, описывающие определенную часть поверхности объекта и содержащие обобщенную информацию о цвете соответствующей им области. Цвет вокселя в области, ограниченной ОТ, определяется цветом каждой из ОТ, ограничивающих эту область. Вес цветовой составляющей каждой ОТ при расчете цвета вокселя обратно пропорционален евклидову расстоянию от вокселя до нее. Представлен механизм расчета весовых коэффициентов цветовых компонент ОТ для вычисления цветов вокселей, зависящих от расстояния до каждой ОТ, ограничивающей область. Приведены описание и результаты экспериментов, в рамках которых осуществлялось текстурирование синтетических моделей и моделей натуральных объектов.

Ключевые слова: трехмерная реконструкция; текстурирование; воксельная модель

DOI: 10.14375/19922264140311

В современных условиях роста популярности медиатехнологий и востребованности компьютерной графики решение проблемы быстрого создания трехмерных моделей натуральных объектов и их анимации является одной из наиболее актуальных задач. Трехмерная реконструкция является одним из перспективных направлений в области компьютерной графики с точки зрения повышения скорости и уменьшения сложности процесса трехмерного моделирования.

Результатом применения трехмерной реконструкции являются трехмерные модели объектов, которые могут быть представлены в различных форматах. Одним из наиболее перспективных способов представления трехмерных моделей является воксельное представление. Воксельные модели, в частности, позволяют легко отображать деформации и разрушения объектов, а также их внутреннюю структуру, что может быть весьма полезно при создании анимации различных взаимодействий объектов.

Одним из этапов трехмерной реконструкции является текстурирование. В случае с воксельными моделями этот термин подразумевает присвоение цветов вокселям трехмерной модели. На сегодняшний день большинство технологий текстурирования подразумевают достаточно большое количество творческой, сложно автоматизируемой работы.

Перспективной является технология, основанная на понятии «мегатекстура», которая позволяет работать с воксельными объектами и используется в игровом графическом движке id Tech 4 и id Tech 5 [1]. Данная технология относится к методике распределения текстур, когда вся поверхность трехмерной сцены покрывается одной большой текстурой с заданным уровнем детализации вместо множества мелких. Однако такой подход используется в основном для текстурирования достаточно простых поверхностей, например ландшафтов, и не подходит для объектов со сложной геометрией.

Целью настоящей работы является разработка технологии текстурирования воксельной модели, позволяющей на основе цветовой информации, содержащейся в ОТ, производить перенос цветовой информации с изображений натурального объекта на его воксельную модель.

Опорную точку можно представить в виде кортежа $\langle x, y, z, r, g, b \rangle$, где (x, y, z) — координаты ОТ в трехмерном пространстве, (r, g, b) — цветовые координаты ОТ в пространстве RGB (red, green, blue).

Опорные точки генерируются на основе смешанных изображений, получаемых в результате круговой съемки объекта, что позволяет создать замкнутое описание поверхности модели объекта в трехмерном пространстве [2]. При этом цветом ОТ является цвет соответствующего ей сегмента, по-

¹Орловский филиал Института проблем информатики Российской академии наук, arhipov12@yandex.ru

²Орловский филиал Института проблем информатики Российской академии наук, maniakov_yuri@mail.ru

лученного в результате аппроксимации на основе сгенерированной палитры различных цветов.

Технология построения воксельной модели [3] подразумевает, что на основе известных координат ОТ в трехмерном пространстве производится интерполяция координат вокселей, заключенных между ОТ. Таким образом, единичной областью интерполяции как декартовых координат вокселей модели, так и их цветовых координат в пространстве RGB является область, ограниченная пространственным треугольником, образованным тремя взаимно ближайшими (ограничивающими) ОТ.

Цвет вокселя в области, ограниченной ОТ, определяется цветом каждой из ОТ, ограничивающих эту область. Вес цветовой составляющей каждой ОТ при расчете цвета вокселя обратно пропорционален евклидову расстоянию от вокселя до нее. Таким образом, необходимо вычислить коэффициенты пропорциональности цветовых значений каждого вокселя расстоянию до i -й ОТ, ограничивающей область:

$$k_i = \frac{1}{dd_i}, \quad d = \sum_{i=1}^3 \frac{1}{d_i},$$

где d_i — евклидово расстояние от рассматриваемого вокселя до i -й ОТ.

Сумма произведений соответствующих коэффициентов и цветовых составляющих каждой ограничивающей ОТ образует цвет каждого вокселя.

Таким образом, зависимость, описывающая способ вычисления цвета вокселя в области, ограниченной ОТ, может быть представлена в виде:

$$f(c) = \sum_{i=1}^3 c_i k_i,$$

где c_i — значение цвета i -й ограничивающей ОТ:

$$c = \begin{bmatrix} r \\ g \\ b \end{bmatrix}.$$

В результате получим текстурированную воксельную модель, которую можно формально представить в виде совокупности вокселей:

$$V_T = \bigcup_{i=1}^m v_i,$$

где m — количество вокселей модели; v — воксел модели; представляющий собой структуру вида $v = \langle x, y, z, r, g, b \rangle$, в которой (x, y, z) — координаты вокселя в трехмерном пространстве, (r, g, b) — координаты цвета вокселя в пространстве RGB.

Координаты ОТ контрольной модели

№ ОТ	X	Y	Z	R	G	B
1	10	-39	0	255	0	0
2	-10	17	0	0	255	0
3	14	17	0	0	0	255

На основе представленной модели был создан алгоритм и реализующий его программный модуль, а также были проведены экспериментальные исследования. Экспериментальные исследования проводились на совокупности тестовых воксельных моделей. В ходе исследования осуществлялось текстурирование данных моделей с использованием информации об ОТ, на основе которых они были построены, в соответствии с технологией пофрагментного анализа и представления натурального объекта и изменения его пространственного положения с последующей интеграцией полученных данных в трехмерные изображения [4].

На начальном этапе с целью повышения точности проверки корректности работы алгоритма использовалась контрольная воксельная модель, представляющая собой пространственный треугольник с расположенными в вершинах ОТ разных цветов, координаты которых представлены в таблице.

Результат текстурирования модели, представленный на рис. 1, показывает, что цвета составля-

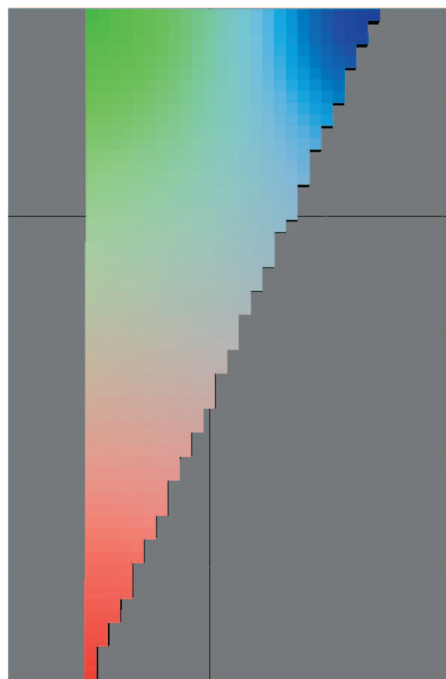


Рис. 1 Визуализация текстурированной контрольной модели

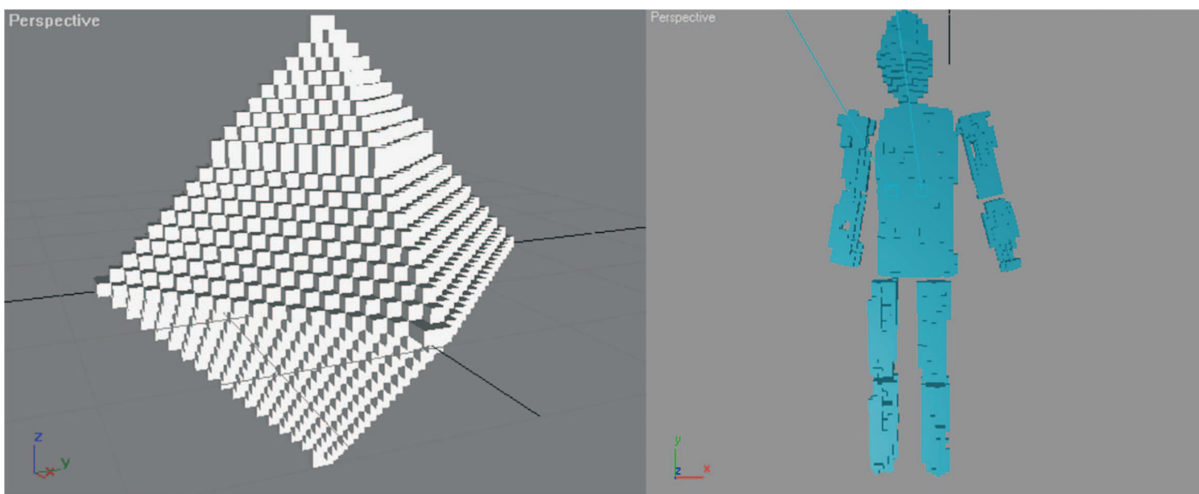
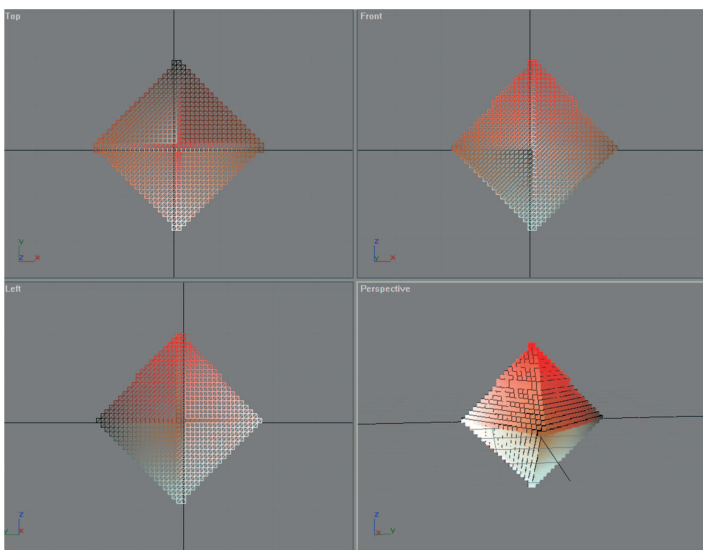


Рис. 2 Тестовые воксельные модели



(а)



(б)

Рис. 3 Визуализация текстурированных тестовых воксельных моделей: (а) текстурирование синтетической воксельной модели; (б) текстурирование модели натурального объекта

ющих ее вокселей образуют плавный градиентный переход между цветами ОТ, расположенными в вершинах треугольника.

В связи с тем, что алгоритм текстурирования работает с совокупностью пространственных треугольников, аппроксимирующих трехмерную модель, учитывая корректность текстурирования контрольной модели, можно сделать вывод о корректности работы алгоритма и модуля текстурирования воксельных моделей. Это подтверждают исследования с использованием других тестовых моделей, представленных на рис. 2.

В результате текстурирования рассмотренных моделей были получены результаты, представленные на рис. 3. При этом стоит уточнить, что источником исходной информации для генерации ОТ для тестовой модели натурального объекта (рис. 3, б) в соответствии с технологией представления натурального объекта и изменения его пространственного положения [4] являлись фотоснимки натурального объекта, один из которых представлен на рис. 4.

Таким образом, можно заключить, что рассмотренная технология текстурирования позволяет с необходимой точностью получить отображение



Рис. 4 Изображение исходного натурального объекта для создания тестовой модели (см. рис. 3, б)

цветов поверхности натурального объекта на воксельной модели при условии ограниченного объема

информации о цветах ОТ. Необходимая точность текстурирования достигается путем варьирования уровня детализации модели, который, в свою очередь, определяется количеством ОТ.

Литература

1. Использование мегатекстур (megatexture, clipmaps) // Gamedev.ru. 9 ноября 2007. <http://www.gamedev.ru/code/articles/Megatexture>.
2. Архипов О. П., Маньяков Ю. А., Сиротинин Д. О. Метод генерации виртуальной сетки опорных точек на цветных изображениях // Информационные технологии в науке, образовании и производстве (ИТНОП-2012): Мат-лы V Междунар. науч.-технич. конф. — Орел: Госуниверситет-УНПК, 2012. CD-ROM.
3. Маньяков Ю. А. Технология регистрации поведения объектов в трехмерном пространстве // Информационные технологии в науке, образовании и производстве (ИТНОП-2010): Мат-лы IV Междунар. науч.-технич. конф. — Орел: ОрелГТУ, 2010. Т. 3. С. 182–186.
4. Архипов О. П., Маньяков Ю. А., Сиротинин Д. О. Информационная модель технологии представления натурального объекта и изменения его пространственного положения // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 71–76.

Поступила в редакцию 03.06.14

VOXEL MODELS TEXTURING BASED ON REFERENCE POINTS' COLOR INFORMATION

O. P. Arkhipov and Y. A. Maniakov

Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation

Abstract: The article describes a technology for texturing voxel models based on color information contained in reference points. Reference points are special points in three-dimensional space which describe a definite part of an object surface and contain generalized color information about a corresponding area. The color of each voxel located in the area bounded by reference points is defined by each reference point bounding this area. The color components weights of each reference point are proportional to Euclidean distance from each voxel to these reference points. The article presents a method for calculating reference points color components weights based on dependence on the distance from reference points to voxels. This method is used for calculating colors of voxels. The description and the results of experiments are also provided. The experiments consisted in texturing generated models and real objects models.

Keywords: three-dimensional reconstruction; texture mapping; voxel model

DOI: 10.14375/19922264140311

References

1. Ispol'zovanie megatekstur (megatexture, clipmaps) [Using megatextures (megatexture, clipmaps)]. Available at: <http://www.gamedev.ru/code/articles/Megatexture> (accessed March 25, 2014).
2. Arkhipov, O. P., Yu. A. Maniakov, and D. O. Sirotnin. 2012. Metod generatsii virtual'noy setki opornykh tochek

- na tsvetnykh izobrazheniyakh [Method of virtual markers grid generation on color images]. *Mat-ly VI Mezhdunar. Nauch.-Tekhnich. Konf. Informatsionnye Tekhnologii v Nauke, Obrazovanii i Proizvodstve ITNOP-2012* [Information Technologies in Science, Education, and Production Science and Technology Conference (International) Proceedings]. Orel. CD-ROM.
3. Maniakov, Y. A. 2010. Tekhnologiya registratsii povedeniya ob"ektov v trekhmernom prostranstve [Three-dimensional space object behavior tracking technology]. *Informatsionnye Tekhnologii v Nauke, Obrazovanii i Proizvodstve (ITNOP)*. *Mater. Mezhdunar. Nauch.-Tekhnich. Konf.* [Information Technologies in Science, Education, and Production Science and Technology Conference (International) Proceedings]. Orel. 3:182–186.
4. Arhipov, O. P., Yu. A. Maniakov, and D. O. Sirotinin. 2014. Informatsionnaya model tekhnologii predstavleniya naturnogo ob"ekta i izmeneniya ego prostranstvennogo polozheniya [Information model of the full-scale object and its attitude changes representation technology]. *Informatika i ee Primeneniya — Inform. Appl.* 1:71–76.

Received June 03, 2014

Contributors

Arkhipov Oleg P. (b. 1948) — Candidate of Science (PhD) in technology, Director, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; arkipov12@yandex.ru

Maniakov Yuri A. (b. 1984) — Candidate of Science (PhD) in technology, scientist, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; maniakov-yuri@mail.ru

ПРОЕКТИРОВАНИЕ САМОСИНХРОННЫХ СХЕМ: СТРУКТУРНЫЕ МЕТОДЫ В ИЕРАРХИЧЕСКОМ АНАЛИЗЕ*

Л. П. Плеханов¹

Аннотация: Самосинхронные схемы (СС-схеиы) имеют уникальные свойства независимости от задержек и отказобезопасности. Рассмотрена одна из главных проблем проектирования таких схем — анализ самосинхронности больших схем. В традиционном подходе схемы анализируются событийными методами, по переключениям элементов. Сложность вычислений в таком подходе экспоненциально растет от размера и/или других параметров схем, что не позволяет анализировать большинство практически значимых схем. Решение проблемы предлагается в функциональном подходе — без использования переключений — и иерархическом описании схем. В иерархическом анализе самосинхронности наряду с анализом логических функций предлагается использовать структурные методы — исследование взаимосвязей элементов и фрагментов. Такой способ позволяет резко уменьшить трудоемкость вычислений и в итоге решить одну из главных проблем проектирования СС-схем — анализ схем любого размера. Эффективность предложенных методов подтверждена с помощью экспериментальных программных средств.

Ключевые слова: самосинхронные схемы; асинхронные схемы; проектирование схем; анализ самосинхронности

DOI: 10.14375/19922264140312

1 Введение

Самосинхронные схемы относятся к классу асинхронных схем, то есть схем, не имеющих тактовых генераторов. Русскоязычный термин «самосинхронная схема» предложен в книге [1] как аналог термина *speed-independent* [2] — «схема, правильность функционирования которой не зависит от величин задержек элементов». С практической точки зрения использование такого термина неудобно, так как он определен для замкнутых схем с одним начальным состоянием. (Другие близкие термины подробнее обсуждены в [3].) Поэтому для практических целей проектирования СС-схем было предложено более широкое определение [4, 5].

Самосинхронная схема — это разомкнутая или замкнутая схема, при всех реальных начальных состояниях и переходах между ними имеющая два свойства: отсутствие состязаний при любых конечных задержках элементов и отказобезопасность по отношению к константным залипаниям на 0 и 1 выходов элементов.

Так определенная схема удовлетворяет также и критериям [1, 2] (в случае разомкнутой схемы — при правильном замыкании). При реализации на чипе самосинхронность может зависеть от задержек в трассах. Поэтому на функционально-логическом уровне разработки свойства, указанные в опре-

делении, становятся необходимыми условиями и должны быть обеспечены схемотехническими решениями. Но и в случае возможного нарушения самосинхронности от задержек в трассах есть реальные способы (схемотехнические и топологические) обеспечения самосинхронности [5].

Уникальные свойства СС-схем, присущие им по определению, имеют и уникальные следствия. Они обеспечивают максимально широкий диапазон правильного функционирования, определяемый только физическими (не схемотехническими) возможностями переключений элементов, что недостижимо для других типов схем. Устойчивость работы СС-схем при малых напряжениях питания позволяет создавать схемы с малым энергопотреблением. А свойство отказобезопасности дает возможность получения высоконадежных схем, в том числе с самопроверкой и саморемонтом.

Разработка практических СС-схем началась в СССР группой В. И. Варшавского в 70-х гг. прошлого века [6] и с запозданием на 20 лет за рубежом [7]. Последняя ссылка указывает на единственную зарубежную методологию построения СС-схем *Null Convention Logic* (NCL). Эта методология была создана чисто интуитивно, по принципу «запрос—ответ». Методология имеет ряд существенных недостатков: ограниченный базис реализации

* Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований Президиума РАН № 1 за 2014 г. (проект № 43П) и РФФИ (проекты 13-07-12062 офи.м и 13-07-12068 офи.м).

¹Институт проблем информатики Российской академии наук, LPlekhanov@inbox.ru

(27 элементов), жесткая конвейерная структура построения схем, не допускающая альтернатив, и др. В результате схемы, получаемые по методологии NCL, имеют большие затраты в транзисторах, что и отмечено бывшими сотрудниками В. И. Варшавского [8].

Методология разработки СС-схем в ИПИ РАН опирается на математические методы Маллера и Варшавского и развивает методы группы В. И. Варшавского. Одним из результатов разработок ИПИ РАН являются 22 патента РФ и 2 США в области СС-схемотехники.

Методология ИПИ РАН позволяет создавать СС-схемы в тысячи элементов как на базовых матричных кристаллах (БМК), так и в заказном исполнении. Методология дает возможность сравнения множества вариантов и выбора оптимального. Подробный сравнительный анализ NCL-методологии, выполненный в ИПИ РАН [9], показывает, что схемы NCL уступают идентичным по функционированию схемам ИПИ РАН по всем показателям: затратам в транзисторах, быстродействию и энергопотреблению.

В настоящее время институтом в содружестве с другими учреждениями (НТЦ МИЭТ, НИИСИ РАН) разработаны, изготовлены и испытаны ряд изделий на БМК и в заказном формате [10–12]. Работы по развитию СС-схем продолжаются [9, 13–15].

Одна из главных проблем проектирования СС-схем заключается в необходимости применения специальных математических и других методов, требующих больших, а подчас огромных вычислительных ресурсов. Существующие классические методы ограничивают возможности анализа и синтеза СС-схем размером до нескольких десятков элементов. Поэтому актуальной является задача разработки и реализации новых методов, альтернативных существующим, с целью довести возможности проектирования до реальных потребностей, т. е. до схем любого размера. Одним из путей решения этой задачи могут служить методы анализа и синтеза, основанные на функциональном подходе [5, 14].

Данная статья продолжает развитие методов, приведенных в докладе на МЭС-2012 [16].

2 Проблемы анализа в функциональном подходе

Как показано в [5, 16], проблему полного анализа СС-схем можно решить только иерархическим способом. В этом способе на нижнем уровне

используются описания фрагментов в логических уравнениях, а на верхних уровнях — взаимосвязи фрагментов и результаты анализа более низких уровней.

В классическом подходе, основанном на событийных описаниях замкнутых схем (событие — изменение состояния схемы, в частности изменение одного сигнала), иерархического анализа пока не предложено.

В рамках функционального подхода (разомкнутое представление схем) анализ проводится иерархически, с разделением схем на фрагменты. Для каждого фрагмента должны проверяться оба требования определения СС-схем: отказобезопасность и отсутствие состязаний. При этом анализ фрагментов нижнего уровня и анализ на верхних уровнях делаются по-разному.

Основные проблемы анализа здесь возникают на нижнем уровне. На нем с неизбежностью необходимо использовать полные описания фрагментов в уравнениях. Уравнения должны учитывать, помимо функционирования, все возможные реальные состояния и переходы между ними.

Для учета всех состояний уравнения элементов записываются в зависимости от параметров анализа [15, 16]: независимых переменных информационных входов и переменных памяти. Для проверки отказобезопасности используется индизирование сигналов [1], для определения состязаний предложен способ проверочных функций для каждого элемента [5, 16].

Созданная по этому методу программа ФАЗАН [17] показала правильное выполнение анализа, а также выявила ряд трудностей. Выяснилось, что нахождение индизируемости сигналов выполняется достаточно быстро даже при большом числе параметров анализа и не представляет реальной проблемы. Основная трудность заключается в определении состязаний. Проверочные функции и часть уравнений в общем случае имеют увеличенное число аргументов, поскольку добавляются изоперемные, необходимые для выявления состязаний. В результате определение монотонности проверочных функций становится громоздким и ограничивает размер анализируемого фрагмента нижнего уровня.

Простым решением этой проблемы может быть построение нижнего уровня схем из фрагментов небольшого размера, что вполне реализуемо. Однако это не всегда удобно и ограничивает маневр разработчика.

Задачу радикального уменьшения трудоемкости анализа СС-схем в функциональном подходе можно решить привлечением структурных методов.

3 Структурные методы в иерархическом анализе

Идея структурных методов состоит в том, что при анализе состязаний можно отказаться от вычисления логических функций, а вместо этого выявить элементарные ячейки памяти и анализировать взаимосвязи этих ячеек и подключенных к ним сигналов.

Необходимым требованием отсутствия состязаний станет блокировка (запрет записи) запоминающих ячеек на время изменения их входов. Двухфазный характер работы СС-схем (чередование рабочей фазы и вспомогательной — спейсера) обеспечивает полную возможность для проверки этого требования.

Поскольку в СС-схемах допустимы любые конечные задержки элементов, то в пределах одной фазы нельзя допустить и блокировку, и изменение входов ячеек. Поэтому необходимое требование выполняется с разделением по фазам: в фазе, в которой входы запоминающих ячеек меняются, делается блокировка, в другой фазе, когда эти входы не изменяются, разрешается перезапись. Такой порядок и делает возможным анализ состязаний структурным методом.

3.1 Анализ на верхних уровнях

Описание схемы на любом верхнем уровне содержит только фрагменты, которые были успешно проверены ранее.

Иерархический анализ на верхних уровнях подробно изложен в [5, 16].

Предложенный метод был реализован и проверен с помощью экспериментальной программы ЛИМАН. Испытания программы на схемах средней сложности (8-битный микропроцессор «Микроядро») показали высокую эффективность метода. Анализ на любом из верхних уровней иерархии с полнотой по всем состояниям и переходам занимает сотые или десятые доли секунды.

Результаты анализа программой ЛИМАН в сравнении с событийным методом приведены в разд. 4.

3.2 Анализ на нижнем уровне

В отличие от верхних уровней, фрагмент на нижнем уровне описания содержит логические уравнения элементов и, возможно, другие фрагменты, ранее прошедшие анализ. Совокупность всех логических уравнений фрагмента будем называть его *логической частью*.

Предлагается описание логической части представить в виде взаимосвязанных фрагментов для иерархического анализа, полностью исключив при этом трудоемкий анализ состязаний нижнего уровня. Тем самым логическая часть может быть проанализирована как фрагмент верхнего уровня.

На первом шаге определяются необходимые для дальнейшего атрибуты интерфейса [5, 16] логической части по ее взаимосвязям: СС-типы ее входов и выходов (фазовые, нефазовые и др.), значения входных спейсеров.

На следующем шаге функциональным методом [3] проверяется индицирование всех внутренних сигналов логической части на ее фазовых выходах. Как упоминалось выше, это действие не представляет больших вычислительных трудностей. Попутно на этом шаге вычисляются и параметры интерфейса, необходимые для дальнейшего: значения спейсеров фазовых выходов, списки индицируемости.

Далее в логической части следует выделить комбинационную часть (КЧ) и отдельно все бистабильные ячейки (БСЯ). (Довольно редко используемые многостабильные ячейки для простоты не рассматриваются.) Выделение производится по взаимосвязям элементов, т. е. уже структурным способом.

Одну БСЯ будут составлять два элемента с перекрестными связями — с выхода одного элемента на один из входов другого. Элементы, не попавшие в БСЯ, будут относиться к КЧ.

И КЧ, и все БСЯ затем оформляются как фрагменты для иерархического анализа. Все требуемые при этом атрибуты интерфейсов новых фрагментов определяются либо непосредственно, либо на основании предыдущих шагов.

В соответствии с общим правилом анализа самосинхронности [5, 16] каждый новый фрагмент должен быть проверен на отказобезопасность и отсутствие состязаний внутренних сигналов.

Отказобезопасность фрагмента обеспечивается индицированием всех его внутренних сигналов на его внешних фазовых выходах. Ранее на одном из шагов была проверена индицируемость внутренних сигналов логической части. По свойству транзитивности индикации, если внутренний сигнал фрагмента индицируется на выходах логической части, то он с необходимостью индицируется и на выходах фрагмента, так как другого пути для его индикации нет.

Таким образом, после проверки индикации логической части все новые ее фрагменты также становятся проверенными на отказобезопасность.

Отсутствие состязаний внутренних сигналов в БСЯ априорно следует из их структуры: внутренних

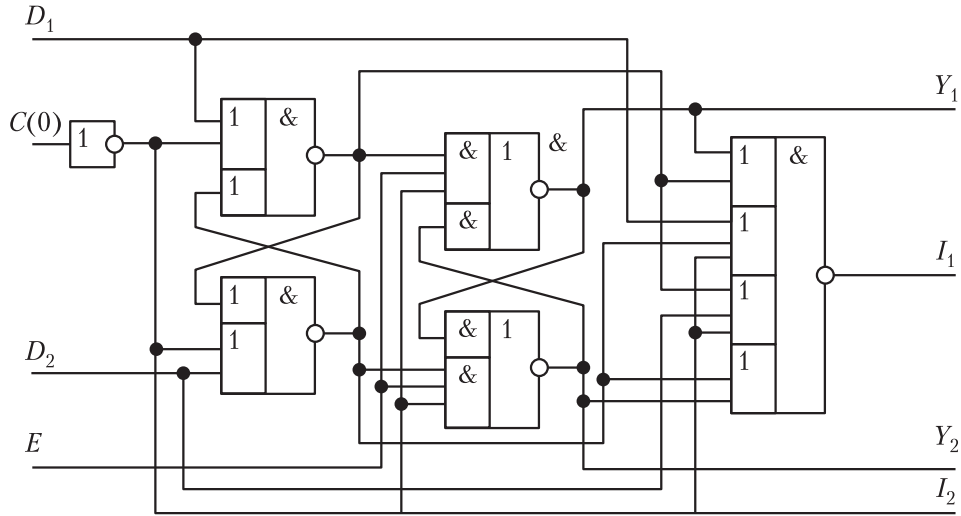


Рис. 1 Ячейка сдвигового регистра

сигналов всего два — это сигналы, связывающие последовательно два элемента в кольцо. Такие сигналы не могут вызвать состязаний. Состязания могут появиться из-за внешних сигналов БСЯ, но их проверка осуществляется уже на более высоком уровне иерархии.

Для КЧ отсутствие состязаний следует из серии доказанных в [1] теорем для разомкнутых схем, заканчивающихся следующим:

«Утверждение 4.4. Двухфазная комбинационная схема является аperiodической тогда и только тогда, когда она индицируема».

В цитируемой книге термин «аperiodическая» является синонимом термина «самосинхронная», а под индицируемой понимается схема, все внутренние элементы которой индицируются на внешних выходах. Поэтому КЧ не нуждается в отдельной проверке на отсутствие состязаний.

В результате описанной процедуры логическая часть будет представлена как соединение фрагментов, уже прошедших анализ.

Последним шагом предлагаемого метода будет иерархический анализ сначала логической части, затем целиком рассматриваемого фрагмента нижнего уровня.

Можно заметить, что вычислительная сложность этого метода линейно зависит от количества элементов и их взаимосвязей. Такая зависимость не порождает громоздких вычислений на ЭВМ рассматриваемой части задачи.

Рассмотрим пример схемы ячейки сдвигового регистра [18] (рис. 1). Здесь и далее знак « \wedge » означает отрицание, знак « \vee » — операцию ИЛИ, отсутствие знака — операцию И.

Схема задается уравнениями:

$$I_2 = \wedge C; \tag{1}$$

$$Q_1 = \wedge((D_1 \vee I_2)Q_2); \tag{2}$$

$$Q_2 = \wedge((D_2 \vee I_2)Q_1); \tag{3}$$

$$Y_1 = \wedge(Q_1EI_2 \vee Y_2); \tag{4}$$

$$Y_2 = \wedge(Q_2EI_2 \vee Y_1); \tag{5}$$

$$I_1 = \wedge((Y_1 \vee Q_1)(D_1 \vee Q_2 \vee I_2)(D_2 \vee Q_1 \vee I_2)(Y_2 \vee Q_2)). \tag{6}$$

Здесь C — управляющий сигнал; E — сигнал дополнительной блокировки.

В данном случае вся схема составляет логическую часть. Анализ индикации показывает, что все ее элементы индицируются: (2)–(5) — на индикаторном выходе I_1 , инвертор (1) — на выходе I_2 .

Схема разбивается на три фрагмента: БСЯ-1 — элементы (2) и (3), БСЯ-2 — элементы (4) и (5) и КЧ, в которую входят инвертор (1) и индикатор (6). Все фрагменты являются самосинхронными, так как прошли общую проверку на индицирование.

Проверка самосинхронности всей схемы будет состоять в анализе межсоединений фрагментов иерархическим методом [15, 16], для чего требуется вычислить параметры блокировки нефазовых сигналов Q_1, Q_2, Y_1, Y_2 . Здесь входы и выходы БСЯ-1 и нефазовые входы индикатора блокируются в фазе спейсера сигналом C с задержкой. Входы и выходы БСЯ-2 блокируются в рабочей фазе сигналом C также с задержкой и сигналом E без задержки.

Опуская подробности, можно сделать вывод, что все соединения фрагментов корректны по отсутствию состязаний. Сигнал E , вследствие задержки блокировки БСЯ-2, следует подключать к сигналам

лу I_2 последующей ячейки регистра, если таковая предусмотрена.

4 Сравнение результатов иерархического и событийного анализа

Для анализа была выбрана СС-схема 8-битного микропроцессора «Микроядро», ранее подробно проанализированная событийными методами [19]. Верхний уровень схемы показан на рис. 2 (на рисунке не показан управляющий блок, задающий режимные константы и не подлежащий анализу).

Иерархическая структура схемы по уровням и именам блоков выглядит следующим образом (в скобках даны размеры блоков в вентилях).

MicroCorg (892)

1 — ROM (81)

2 — GI2 (3), ROM_REG (28), PC_S (32), ROM_DC (18)

1 — DCF (11)

2 — INV (1), GI2 (3), R0RE10N (6), INV (1)

1 — CLK_MC (123)

1 — MUL (448)

2 — GI2 (3), INV (1), INV (1), GI2 (3), INV (1), GI2 (3), GI3 (4), SSMULT (337)

3 — SSSM (12), MINUSA (33), SSSM (12), SSSM (12), SSDC (48), SUM20 (23), SSSM1 (6), SUM21 (19), SUM22 (22), MULIND (51), PARTP0 (17), PARTP1 (34), PARTP2 (48)

2 — R1R10 (64)

3 — R1RE11 (6), R1RE11 (6), INV (1), GI4 (5), INV2 (1), GI2 (3), GI3 (4), INV2 (1), INV (1), R1RE11 (6), R1RE11 (6), R1RE11 (6), R1RE11 (6), R1RE11 (6), R1RE11 (6)

2 — MX22 (25)

1 — ROT (152)

2 — R010 (4), R010 (4), R010 (4), R010 (4), GI2 (3), GI2 (3), GI4 (5), MX21R (17), SHFT43_N (110)

3 — GI3M (3), GI2 (3), GI2 (3), INV (1), GI3 (4), INV3 (2), MX21N (25), PPTR (13), PPTR (13), PPTR (13), PPTR (13), MX310N (17)

1 — MX311 (29)

Суммарное время анализа всех блоков схемы программой ЛИМАН (без учета файловых операций) составило 0,84 с.

Иерархический анализ в функциональном подходе, как указывалось ранее, обеспечивает исчерпывающую полноту анализа, т. е. учитываются все реальные начальные состояния и все возможные

реальные сочетания значений входов блоков и переходов между ними.

Однократный анализ событийным методом этого же микропроцессора с одним начальным состоянием и одной из комбинаций входных значений [12] занимает 7 с.

Для обеспечения необходимой полноты требуется выполнить десятки тысяч таких сеансов анализа.

Приведенные данные показывают, какие практические трудности возникают при проектировании СС-схем классическими (событийными) методами.

5 Заключение

Структурные методы в иерархическом анализе самосинхронности схем состоят в том, что исследуются не уравнения элементов, а их взаимосвязи. Применение этих методов оказалось возможным в рамках функционального подхода, когда рассматриваются разомкнутые схемы, работающие по принципу чередования двух фаз.

Эффективность структурных методов следует из того, что их вычислительная сложность линейно зависит от количества элементов/фрагментов и их взаимосвязей.

На всех уровнях иерархического анализа выше нижнего проверки осуществляются исключительно структурным методом. Так, анализ 8-битного ядра микропроцессора «Микроядро» занимает десятки доли секунды.

На нижнем уровне одна из двух проверок — отсутствие состязаний — сводится к такому же структурному методу, что и анализ на верхних уровнях. Вторая проверка — индицирование — заключается в последовательном вычислении логических функций. На практике она обычно не вызывает затруднений, особенно с учетом того, что размер фрагмента нижнего уровня может быть выбран самим пользователем.

Таким образом, применение структурных методов в анализе СС-схем позволяет радикально уменьшить сложность вычислений и решить проблему анализа схем любого размера (не решенную в классическом событийном подходе).

Литература

1. Автоматное управление асинхронными процессами в ЭВМ и дискретных системах / Под ред. В. И. Варшавского. — М.: Наука, 1986. 398 с.
2. Muller D. E., Bartky W. C. A theory of asynchronous circuits // Symposium (International) on the Theory of

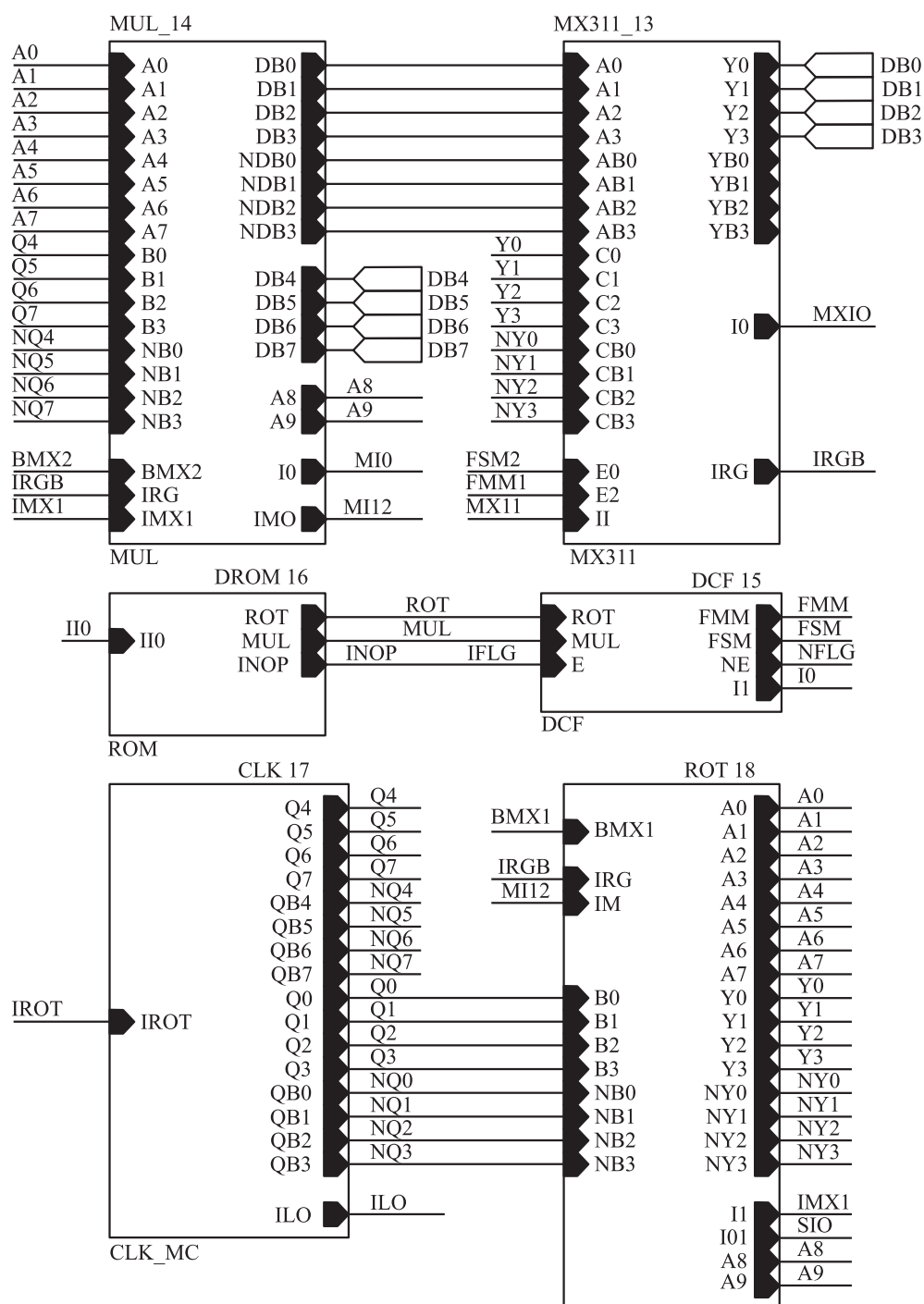


Рис. 2 Верхний уровень микропроцессора «Микроядро»

- Switching. — Harvard University Press, 1959. Part 1. P. 204–243.
3. Плеханов Л. П. Проектирование самосинхронных схем: функциональный подход // Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всеросс. науч.-технич. конф. (МЭС-2010): Сб. науч. тр. — М.: ИППМ РАН, 2010. С. 424–429.
 4. Плеханов Л. П. О свойстве самосинхронности цифровых электронных схем // Системы и средства информатики, 2011. Вып. 21. № 1. С. 84–91.
 5. Плеханов Л. П. Основы самосинхронных электронных схем. — М.: Бином. Лаборатория знаний, 2013. 208 с.
 6. Аperiodические автоматы / Под ред. В. И. Варшавского. — М.: Наука, 1976. 423 с.
 7. Fant K.M., Brandt S.A. NULL convention logic // Theseus Research: Technical Papers. P. 1–26. <http://www.theseusresearch.com/NCLPaper01.htm>.
 8. Taubin A., Cortadella J., Lavagno L., Kondratyev A., Peeters A. Design automation of real-life asynchronous devices and systems // Foundations and Trends in Electronic Design Automation, 2007. Vol. 2. No. 1. P. 1–133.
 9. Соколов И. А., Степченко Ю. А., Бобков С. Г. и др. Базис реализации супер-ЭВМ эксафлопного класса // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 45–70.
 10. Степченко Ю. А., Петрухин В. С., Дьяченко Ю. Г. Опыт разработки самосинхронного ядра на базовом матричном кристалле // Проблемы разработки перспективных микро- и нанoeлектронных систем: I Всеросс. науч.-технич. конф. (МЭС-2005): Сб. научных трудов. — М.: ИППМ РАН, 2005. С. 235–242.
 11. Степченко Ю. А., Дьяченко Ю. Г., Бобков С. Г. Квазисамосинхронный вычислитель: методологические и алгоритмические аспекты // Проблемы разработки перспективных микро- и нанoeлектронных систем: III Всеросс. науч.-технич. конф. (МЭС-2008): Сб. науч. тр. — М.: ИППМ РАН, 2008. С. 441–446.
 12. Степченко Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченко Д. Ю. Самосинхронный вычислитель для высоконадежных применений // Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всеросс. науч.-технич. конф. (МЭС-2010): Сб. науч. тр. — М.: ИППМ РАН, 2010. С. 418–423.
 13. Степченко Ю. А., Денисов А. Н., Дьяченко Ю. Г., Гринфельд Ф. И., Филимонов О. П., Морозов Н. В., Степченко Д. Ю. Библиотека элементов для проектирования самосинхронных полузаказных БМК микросхем серий 5503/5507 и 5508/5509. — М.: ИПИ РАН, 2013. 391 с.
 14. Бобков С. Г., Горбунов М. С., Дьяченко Ю. Г., Рождественский Ю. В., Степченко Ю. А., Сурков А. В. Использование самосинхронной логики для снижения потребляемой мощности и повышения надежности микропроцессоров // Проблемы разработки перспективных микро- и нанoeлектронных систем: VI Всеросс. науч.-технич. конф. (МЭС-2014): Сб. науч. тр. — М.: ИППМ РАН, 2014. С. 254–257.
 15. Степченко Ю. А., Рождественский Ю. В., Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю., Сурков А. В. Самосинхронное устройство умножения-сложения гигафлопсного класса: варианты реализации // Проблемы разработки перспективных микро- и нанoeлектронных систем: VI Всеросс. науч.-технич. конф. (МЭС-2014): Сб. науч. тр. — М.: ИППМ РАН, 2014. С. 258–263.
 16. Плеханов Л. П. Функциональный метод анализа самосинхронных схем любого размера // Проблемы разработки перспективных микро- и нанoeлектронных систем: V Всеросс. науч.-технич. конф. (МЭС-2012): Сб. науч. тр. — М.: ИППМ РАН, 2012. С. 107–112.
 17. Плеханов Л. П. Программа анализа самосинхронных схем функциональным методом (ФАЗАН). Свидетельство о государственной регистрации программы для ЭВМ № 2011611102 от 01.02.2011.
 18. Степченко Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Рождественскене А. В. Разряд самосинхронного регистра сдвига. Патент на изобретение 2319232 (РФ). Приоритет от 10.03.2008.
 19. Степченко Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченко Д. Ю. Разработка вычислителя, не зависящего от задержек элементов // Системы и средства информатики, 2010. Вып. 20. № 1. С. 5–23.

Поступила в редакцию 10.07.14

DESIGN OF SELF-TIMED CIRCUITS: STRUCTURAL METHODS IN HIERARCHICAL ANALYSIS

L. Plekhanov

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Self-timed circuits have unique properties of the delay-independence and fail-safe. One of the major problems of circuits design, self-timed analysis of large circuits, is considered. In the traditional approach, circuits

are analyzed by event methods with elements switches. Computational complexity in this approach increase exponentially with the size and/or other circuit parameters, which does not allow analyzing the most practically important circuits. The solution is proposed in the functional approach, without using switches, and in the hierarchical description of circuits. In the hierarchical analysis along with the analysis of logical functions, the author proposes to use structural methods, i. e., to study the interaction of elements and fragments. This method allows reducing the complexity of calculations dramatically and thus solves one of the major problems of self-timed circuits design — analysis of circuits of any size. Efficiency of the suggested methods is confirmed using the experimental software.

Keywords: self-timed circuits; asynchronous circuits; circuit design; self-timed analysis

DOI: 10.14375/19922264140312

Acknowledgments

The work was performed with partial financial support of the Program of Basic Research of the Presidium of the Russian Academy of Sciences No. 1 for 2014 (project No. 43P) and Russian Foundation for Basic Research (projects 13-07-12062 and 13-07-12068).

References

1. Varshavskiy, V. I., ed. 1986. *Avtomatnoe upravlenie asinkhronnymi protsessami v EVM i diskretnykh sistemakh* [Automata control of asynchronous processes in computers and discrete systems]. Moscow: Nauka. 398 p.
2. Muller, D. E., and W. C. Bartky. 1959. A theory of asynchronous circuits. *Symposium (International) on the Theory of Switching Proceedings*. Harvard University Press. 1:204–243.
3. Plekhanov, L. P. 2010. Proektirovanie samosinkhronnykh skhem: Funktsional'nyy podkhod [The design of self-timed circuits: Functional approach]. *Tr. IV Vseross. nauch.-tekhnich. konf. "Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem" (MES-2010)* [4th Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2010) Proceedings]. Moscow. 424–429.
4. Plekhanov, L. P. 2011. O svoystve samosinkhronnosti tsifrovyykh elektronnykh skhem [About the self-timed property of digital electronic circuits]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21(1):84–91.
5. Plekhanov, L. P. 2013. *Osnovy samosinkhronnykh elektronnykh skhem* [Basics of self-timed electronic circuits]. Moscow: Binom. Laboratoriya znaniy [Binom. Laboratory of knowledge]. 208 p.
6. Varshavskiy, V. I., ed. 1976. *Aperiodicheskie avtomaty* [Aperiodic automata]. Moscow: Nauka. 423 p.
7. Fant, K. M., and S. A. Brandt. NULL convention logic. *Theseus Research: Technical Papers*. 1–26. Available at: <http://www.theusresearch.com/NCLPaper01.htm> (accessed July 17, 2014).
8. Taubin, A., J. Cortadella, L. Lavagno, A. Kondratyev, and A. Peeters. 2007. Design automation of real-life asynchronous devices and systems. *Foundations and Trends in Electronic Design Automation*. 2(1):1–133.
9. Sokolov, I. A., Yu. A. Stepchenkov, S. G. Bobkov, et al. 2014. Bazis realizatsii super-EVM eksaflopnoyo klassa [The basis for the implementation of the super computer of exaflops class]. *Informatika i ee Primeneniya — Inform. Appl.* 25(1):5–34.
10. Stepchenkov, Yu. A., V. S. Petrukhin, and Yu. G. D'yachenko. 2005. Opyt razrabotki samosinkhronnogo yadra na bazovom matrichnom kristalle [The experience of developing self-timed kernel on basic matrix crystal]. *Tr. I Vseross. nauchno-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2005)* [1st Russian Scientific and Technical Conference "Problems of the Perspective micro- and nanoelectronic systems development" (MES-2005) Proceedings]. Moscow. 235–242.
11. Stepchenkov, Yu. A., Yu. G. D'yachenko, and S. G. Bobkov. 2008. Kvazisamosinkhronnyy vychislitel': Metodologicheskie i algoritmicheskie aspekty [Quasi-self-timed calculator: Methodological and algorithmic aspects]. *Tr. III Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2008)* [3rd Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2008) Proceedings]. Moscow. 441–446.
12. Stepchenkov, Yu. A., Yu. G. D'yachenko, Yu. V. Rozhdestvenskiy, N. V. Morozov, and D. Yu. Stepchenkov. 2010. Samosinkhronnyy vychislitel' dlya vysokonadezhnykh primeneniy [Self-timed calculator for high reliable applications]. *Tr. IV Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2010)* [4th Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2010) Proceedings]. Moscow. 418–423.
13. Stepchenkov, Yu. A., A. N. Denisov, Yu. G. D'yachenko, F. I. Grinfel'd, O. P. Filimonov, N. V. Morozov, and D. Yu. Stepchenkov. 2013. *Biblioteka elementov dlya proektirovaniya samosinkhronnykh poluzakaznykh BMK mikroskhem seriy 5503/5507 i 5508/5509* [Library of elements for designing self-timed semicustom VLSI 5503/5507 and 5508/5509]. Moscow: IPI RAN. 391 p.

14. Bobkov, S. G., M. S. Gorbunov, Yu. G. D'yachenko, Yu. V. Rozhdestvenskiy, Yu. A. Stepchenkov, and A. V. Surkov. 2014. Ispol'zovanie samosinkhronnoy logiki dlya snizheniya potrebyaemoy moshchnosti i povysheniya nadezhnosti mikroprotessorov [The use of self-timed logic to reduce power consumption and increase reliability of microprocessors]. *Tr. VI Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2014)* [6th Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2014) Proceedings]. Moscow. 254–257.
15. Stepchenkov, Yu. A., Yu. V. Rozhdestvenskiy, Yu. G. D'yachenko, N. V. Morozov, D. Yu. Stepchenkov, and A. V. Surkov. 2014. Samosinkhronnoe ustroystvo umnozheniya-slozheniya gigaflopsnogo klassa: Varianty realizatsii [Self-timed device for multiplication-addition of gigaflops class]. *Tr. VI Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2014)* [6th Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2014) Proceedings]. Moscow. 258–263.
16. Plekhanov, L. P. 2012. Funktsional'nyy metod analiza samosinkhronnykh skhem lyubogo razmera [The functional method of analysis of self-timed circuits of any size]. *Tr. V Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2012)* [5th Russian Scientific and Technical Conference "Problems of the Perspective Micro- and Nanoelectronic Systems Development" (MES-2012) Proceedings]. Moscow. 107–112.
17. Plekhanov, L. P. 2011. Programma analiza samosinkhronnykh skhem funktsional'nym metodom (FAZAN) [The program of the analysis of self-timed circuits with the functional method (FAZAN)]. Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM [Certificate of the State Registration of the Computer Program] No. 2011611102.
18. Stepchenkov, Yu. A., Yu. G. D'yachenko, Yu. V. Rozhdestvenskiy, and A. V. Rozhdestvenskene. 2008. Razryad samosinkhronnogo registra sdviga [Binary digit of self-timed shift register]. Patent RF No. 2319232.
19. Stepchenkov, Yu. A., Yu. G. D'yachenko, Yu. V. Rozhdestvenskiy, N. V. Morozov, and D. Yu. Stepchenkov. 2010. Razrabotka vychislitelya, ne zavisyashchego ot zaderzhek elementov [Design of speed-independent calculator]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 20(1):5–23.

Received July 10, 2014

Contributor

Plekhanov Leonid P. (b. 1943) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; LPlekhanov@inbox.ru

ИНДИКАТОРЫ ТЕМАТИЧЕСКИХ ВЗАИМОСВЯЗЕЙ НАУКИ И ТЕХНОЛОГИЙ: ОТ ТЕКСТА К ЧИСЛАМ*

В. А. Минин¹, И. М. Зацман², В. А. Хавансков³, С. К. Шубников⁴

Аннотация: Дано описание основных этапов вычисления индикаторов тематических взаимосвязей науки и технологий. Исходные данные для проведения расчетов представляют собой результаты обработки полных текстов описаний изобретений на естественном языке. Целью обработки является извлечение из полных текстов информации о научных публикациях, цитируемых в описаниях изобретений. Сопоставление тематики этой информации с индексами Международной патентной классификации (МПК) дает возможность экспертам выявлять взаимосвязи результатов направлений научных исследований (НИИ) с развитием технологий и оценивать их с помощью количественных индикаторов. В статье исследуются те технологические стадии определения индикаторов тематических взаимосвязей, на которых осуществляется переход от обработки текста к численным расчетам. Дано описание разработанных методов извлечения информации о научных публикациях из полных текстов изобретений и расчета значений индикаторов. Их применение дало возможность определить индикаторы взаимосвязей НИИ с информационными технологиями, запатентованными в РФ за период с 2000 по 2012 гг. Для отечественной научно-технической сферы значения этих индикаторов вычислены впервые.

Ключевые слова: взаимосвязи науки и технологий; методология определения индикаторов взаимосвязей; информационные технологии; обработка текста; расчет значений индикаторов

DOI: 10.14375/19922264140313

1 Введение

В 2012–2014 гг. в Институте проблем информатики РАН был создан действующий макет Аналитико-информационной системы (АИС), позволяющий выявлять тематические взаимосвязи НИИ с заданным видом патентуемых технологий и оценивать их с использованием количественных индикаторов. Основная цель статьи заключается в описании результатов, полученных с использованием действующего макета АИС, и тех технологических стадий вычисления индикаторов, с помощью которых осуществляется переход от обработки текстовых описаний запатентованных изобретений к численным расчетам значений индикаторов.

Методология вычисления индикаторов тематических взаимосвязей науки и технологий изложена в работе [1], в которой приводится обзор зарубежных работ по исследованию взаимосвязей науки и технологий [2–11]. На основе этой методологии были разработаны архитектурные решения АИС для вычисления индикаторов. Предложенные решения стали основой создания макета АИС, которая

не имеет аналогов в российской научно-технической сфере. Создание таких систем необходимо для мониторинга и оценивания программ научных исследований и принятия решений на всех этапах программной научно-технической деятельности.

В процессе создания макета был разработан метод извлечения из полных текстов изобретений ссылок на цитируемые научные публикации, которые и дают возможность определять тематические взаимосвязи науки и технологий. Разработанный метод извлечения учитывает тот факт, что ссылки могут размещаться не только в списке в конце описания изобретения, но и внутри текстовых абзацев описания на естественном языке [12].

С помощью созданного макета был проведен эксперимент, в рамках которого было обработано более 6000 полных текстов изобретений по информационным технологиям, запатентованных по индексам, входящим в класс⁵ G06 (Обработка данных; вычисление; счет) МПК, и опубликованных Федеральной службой по интеллектуальной собственности (Роспатент) за период с 2000 по 2012 гг.

* Работа выполнена в ИПИ РАН при частичной поддержке РФНФ (грант № 12-02-12019в).

¹ Российский фонд фундаментальных исследований, minin@rfbr.ru

² Институт проблем информатики Российской академии наук, iz_ipi@i170.ipi.ac.ru

³ Институт проблем информатики Российской академии наук, havanskov@i170.ipi.ac.ru

⁴ Институт проблем информатики Российской академии наук, sergeysh50@yandex.ru

⁵ Все индексы МПК образуют иерархическую структуру и делятся на 5 категорий: разделы (верхний уровень иерархии), классы, подклассы, группы и подгруппы (нижний уровень иерархии).

В текстовых описаниях отобранных изобретений с помощью макета были найдены ссылки на научные публикации, библиографические данные которых позволили соотнести их по тематике с теми или иными рубриками ННИ. В статье приводятся результаты вычисления двух индикаторов. Были вычислены интенсивности тематических взаимосвязей патентуемых информационных технологий с рубриками ННИ Государственного рубрикатора научно-технической информации (ГРНТИ) и распределение периода времени между цитируемыми научными публикациями и выдачей патентов на те изобретения, в которых они цитируются.

2 Операции обработки текстов изобретений

Архитектура АИС подробно изложена в работе [13], в которой перечислены ее входные и выходные информационные ресурсы, а также структурированные наборы данных, объединенные в базу данных (БД) АИС. В этой же работе приведена структурная схема АИС и входящие в нее подсистемы, которые взаимодействуют между собой через БД АИС.

Технология обработки текстов изобретений, поступающих на вход АИС, определяется с помощью сценария, в котором задаются:

- вид технологий, взаимосвязи которого с ННИ исследуются с помощью этого сценария, например информационные технологии;
- перечень индексов МПК, которые относятся к заданному виду технологий, например все индексы класса G06, относящиеся к информационным технологиям;
- временной интервал, на котором исследуются взаимосвязи заданного вида технологий с ННИ, например 2000–2012 гг.;
- поисковый запрос, с помощью которого с оптических носителей Роспатента будут отобраны изобретения заданного в сценарии вида технологий за выбранный период времени и сформирован массив рефератов изобретений как входных информационных ресурсов АИС с адресами полных описаний этих изобретений на сайте Роспатента;
- перечень тех индикаторов, значения которых должны быть вычислены для заданного вида технологий;
- автор(ы) сценария;
- имя сценария.

Приведенный список свидетельствует о том, что сценарий задает состав информационных ресурсов, обрабатываемых на технологических операциях, и перечень вычисляемых индикаторов. Основные технологические операции, реализованные в макете АИС, представлены на рис. 1.

Опишем их, обратив основное внимание на те стадии вычисления индикаторов тематических взаимосвязей, где осуществляется переход от обработки текста изобретений к численным расчетам значений индикаторов.

Как следует из приведенного выше списка, сценарий включает перечень тех индикаторов, значения которых должны быть вычислены в рамках сформированного сценария. Индикаторы выбираются из списка, содержание которого определяется спектром доступных ресурсов.

Формирование массива данных, необходимого для проведения расчетов, осуществляется подсистемой «Импорт», исходя из заданного сценария исследования. Сначала с оптических дисков Роспатента копируются библиографические описания изобретений, отобранные на основе критериев, которые заложены в сценарий. Для этого используется поисковый запрос из сценария.

Для отбора патентов по этому запросу используется информационно-поисковая система (ИПС) MIMOSA, разработанная в Федеральном институте промышленной собственности (ФИПС). Информационно-поисковая система MIMOSA поставляется вместе с библиографическими описаниями изобретений на оптических компакт-дисках, которые распространяются Роспатентом. Для поиска документов с помощью ИПС MIMOSA существует специально разработанный язык запросов, на котором написан поисковый запрос в сценарии. Ниже приводится запрос, который использовался в процессе проведения эксперимента:

$(ICA=G06* \text{ OR } ICAA=G06* \text{ OR } ICAI=G06*) \text{ AND } DP=2000* \text{ AND } \dots \text{ AND } DP=2012*$,

где символ «*» служит знаменителем, а многоточие обозначает пропущенные для краткости годы от 2001 до 2011 и операторы AND.

В этом запросе отбираются все изобретения, содержащие как основные, так и дополнительные индексы МПК, входящие в класс G06 «Обработка данных; вычисление; счет» МПК. При записи в БД АИС сохраняется информация о том, какой индекс помечен в изобретении как основной.

Обращение к ИПС MIMOSA происходит при выполнении каждого сценария. На первом его шаге определяется список номеров патентов, удовлетворяющих поисковому запросу, и формируется структурированный текстовый файл библиографических описаний изобретений, который пополня-

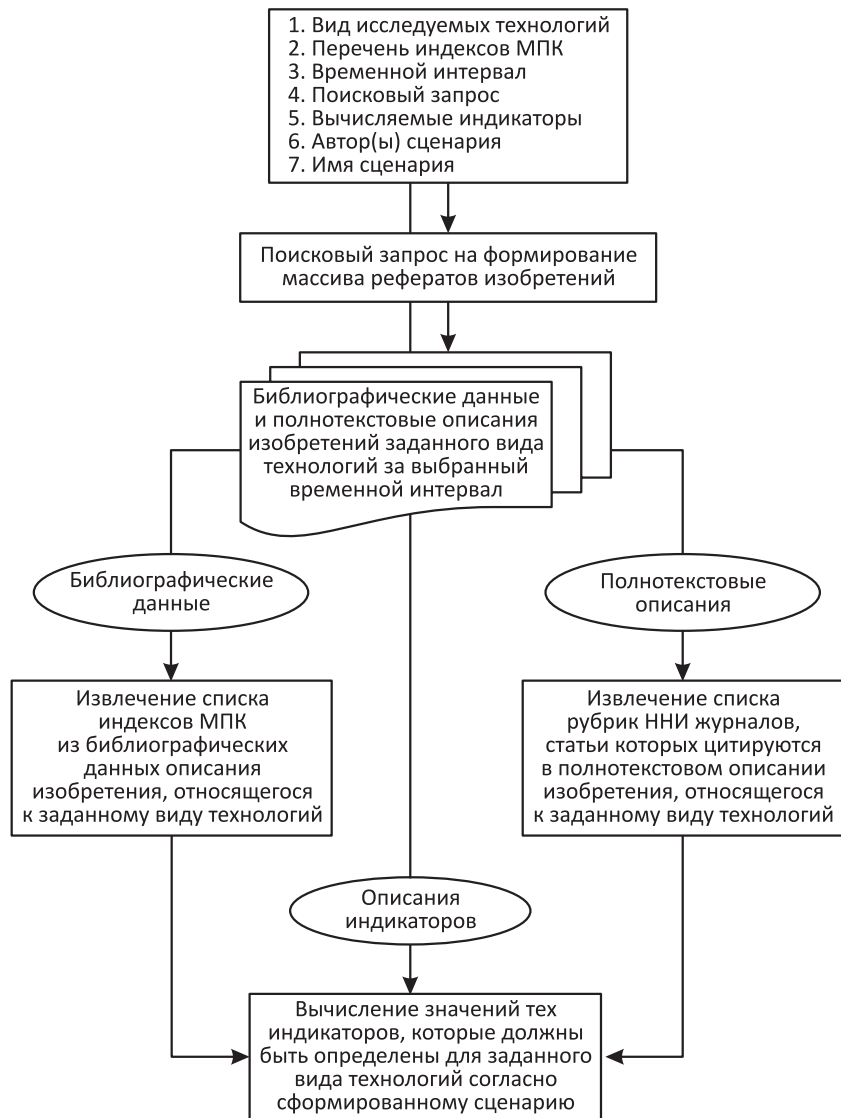


Рис. 1 Основные технологические операции: от текста до значений индикаторов

ет БД АИС (рис. 2). В правой части рис. 2 номер патента RU 02438579 является ссылкой на полное описание изобретения на сайте Роспатента.

В процессе проведения эксперимента по этому запросу было отобрано 6666 изобретений, распределение которых по годам показано на диаграмме (рис. 3). Из них 5243 имели индексы класса G06 в качестве основного.

Из диаграммы видно, что за период 2000–2012 гг. число выданных в РФ патентов по информационным технологиям выросло более чем в три раза: с 286 до 903.

Аналитико-информационная система, используя ссылки из всех отобранных библиографических описаний, копирует в свою БД полные описания изобретений. После завершения их копирования,

выполняются две технологические операции (см. рис. 1):

- (1) извлечение списка индексов МПК в каждом библиографическом описании изобретения заданного вида технологий;
- (2) извлечение списка рубрик ННИ журналов, статьи которых цитируются в полнотекстовом описании изобретения заданного вида технологий.

Эти операции не зависят друг от друга и могут выполняться в любой последовательности.

Первая операция выполняется программно с использованием алгоритма, разработанного согласно правилам написания индексов МПК на патентных документах и их записи в отдельном структури-

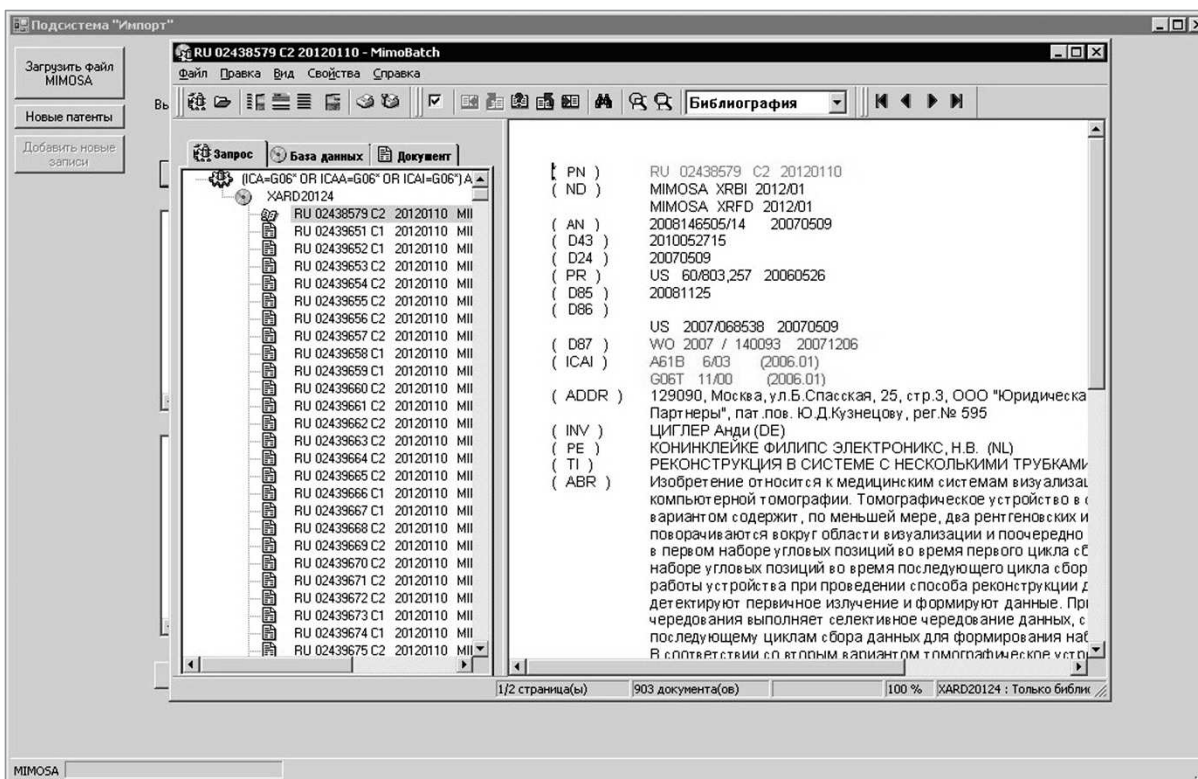


Рис. 2 Список патентов, полученных в результате выполнения запроса

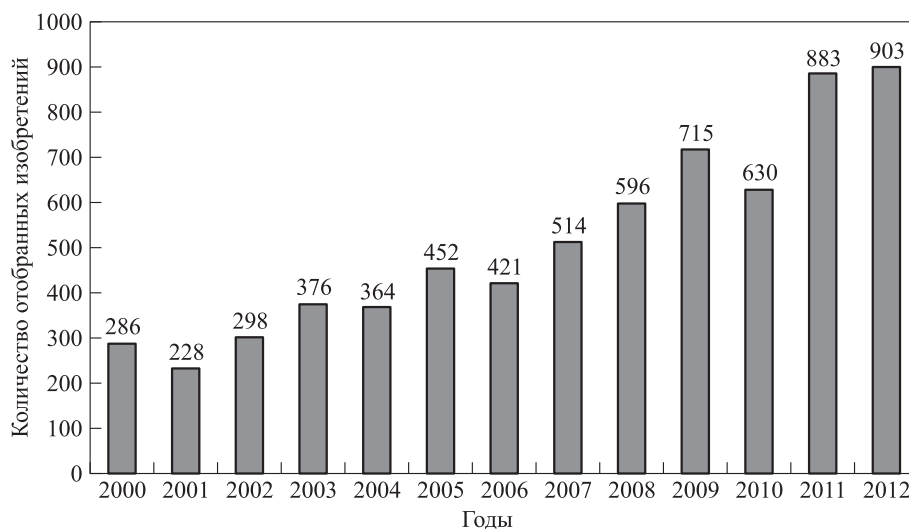


Рис. 3 Распределение отобранных изобретений по годам

рованном поле [14, 15]. Таблица 1 показывает в третьем столбце распределение по 11 подклассам класса G06 отобранных по основному индексу 5243 изобретений.

В процессе выполнения второй операции в текстах изобретений производится поиск и структурирование ссылок на научные публикации, цитиру-

емые в изобретениях. По окончании обеих операций завершается текстовая обработка и происходит переход к численным расчетам.

Задача поиска ссылок была подробно описана в работе [16], в которой было показано, что сложность ее решения является следствием несоблюдения требований стандартов Всемирной орга-

Таблица 1 Распределение изобретений и цитируемых статей по основному индексу МПК

Код подкласса	Название подкласса	Число изобретений	Число статей	Статей на 1 изобретение
G06C	Механические цифровые вычислительные машины	7	0	0
G06D	Гидравлические и пневматические цифровые вычислительные устройства	1	0	0
G06E	Оптические вычислительные устройства	52	8	0,15
G06F	Обработка цифровых данных с помощью электрических устройств	3415	107	0,03
G06G	Аналоговые вычислительные машины . . .	228	14	0,06
G06J	Гибридные вычислительные устройства	3	0	0
G06K	Распознавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации	681	64	0,09
G06M	Счетчики; способы и устройства для подсчета предметов, не отнесенные к другим подклассам	12	0	0
G06N	Компьютерные системы, основанные на специфических вычислительных моделях	107	8	0,07
G06Q	Системы обработки данных или способы, специально предназначенные для административных, коммерческих < . . . > целей	417	5	0,01
G06T	Обработка или генерация данных изображения	320	43	0,13
Всего по классу G06		5243	249	0,05

низации интеллектуальной собственности (ВОИС) к оформлению ссылок на цитируемые публикации [17, 18].

В отличие от принятой в научной сфере и нормативно закрепленной практики размещения списка используемой литературы в конце научной статьи или ее страниц, в описаниях изобретений ссылки на цитируемую публикацию могут встретиться в любом месте описания, включая любой его абзац. Поэтому известные методы извлечения библиографических ссылок из текстов статей [19, 20], по оценке авторов, применимы только к 10%–15% изобретений.

Для поиска ссылок на научные публикации в текстах изобретений и их структурирования был разработан метод, подробно описанный в работе [12]. Этот метод основан на следующей формализованной схеме описания ссылки как объекта ее поиска в тексте изобретения:

$$[автор\{S_1\}][название\ публикации] \\ \{S_2\}название\ источника \\ \{S_3\}атрибуты\ публикации .$$

Наличие квадратных скобок говорит о необязательности присутствия данного элемента схемы в реальной ссылке; $\{S_i\}$ обозначает множе-

ство возможных знаков-разделителей, которые стоят между элементами схемы. В отличие от библиографических элементов [*название публикации*] и [$\{S_2\}$ *название источника*], которые могут рассматриваться как простые множества слов и знаков пунктуации, элементы [*автор* $\{S_1\}$] и *атрибуты публикации* имеют свои собственные структурные особенности [12].

Разработанный метод был использован при поиске ссылок в массиве отобранных изобретений по классу G06 за период 2000–2012 гг. В результате из этого массива было выделено 2758 патентов, в которых найдены фрагменты текста, соответствующие шаблону поиска, построенному на основе приведенной выше схемы описания ссылок.

После анализа выделенных фрагментов текста обнаружилось около 30% случаев, когда используемые шаблоны выделили фрагменты текста описаний, не являющиеся ссылками на цитируемую публикацию, но отвечающие критериям поиска. Данное обстоятельство не влияет на конечный результат, так как при дальнейшей структуризации такие фрагменты текста исключаются из рассмотрения.

В то же время среди полнотекстовых описаний, в которых программой не обнаружено ссылок на цитируемые публикации, при визуальном про-

смотре были найдены пропущенные программой ссылки на цитируемые публикации, что говорит о необходимости дальнейшего развития системы шаблонов, используемых для поиска ссылок. Процедура добавления и редактирования шаблонов уже разработана и используется в макете АИС.

Из общего массива найденных ссылок на цитируемые публикации в описываемом эксперименте были отобраны только те, которые относятся к статьям в журналах или в трудах конференций. Само рубрицирование выполняется в несколько этапов:

- (1) выделение в ссылке на статью названия журнала или трудов конференции;
- (2) поиск названия в нормализованной БД;
- (3) присвоение ссылке на цитируемую статью кодов рубрик ННИ соответствующего журнала или трудов конференции.

Именно после выделения названия проходит граница между обработкой неструктурированных текстов изобретений и операциями подготовки данных для расчетов значений индикаторов.

3 От анализа текста изобретений к численным расчетам

На предыдущих этапах, предшествующих поиску названий журналов или трудов конференций в нормализованной БД, использовались методы анализа текстов изобретений. Эти тексты описывают технические решения изобретений и могут включать ссылки на цитируемые статьи. После этапа поиска названий в процессе вычисления значений индикаторов в макете АИС и в аналогичных зарубежных системах [6, 9, 21] используются не тексты ссылок, а коды рубрик ННИ, присвоенных тому изданию, в котором опубликована цитируемая статья. Использование этих кодов является необходимым условием проведения расчетов значений индикаторов. Отметим, что полнофункциональный вариант АИС будет отличаться от разработанного макета тем, что будут использоваться не все коды рубрик ННИ, присвоенные изданию, а только те, которые соответствуют тематике цитируемой статьи.

Это отличие является принципиальным. В макете и в аналогичных зарубежных системах для вычисления индикаторов используются все коды рубрик ННИ издания, среди которых могут оказаться те, которые не соответствуют тематике цитируемой статьи, что может повлиять на точность вычисления значений индикаторов тематических взаимосвязей

заданного вида технологий с ННИ. Степень этого влияния можно будет оценить только после создания полнофункционального варианта АИС.

Для выделения названия издания, в котором опубликована статья, используются шаблоны описания разделителей между библиографическими элементами структуры описания ссылки на цитируемую публикацию, а именно $\{S_2\}$ (разделитель — издание) и $\{S_3\}$ (разделитель — атрибуты). Если использовать действующий государственный стандарт «Библиографическая запись. Библиографическое описание» [22], то разделителем перед названием издания является двойной слеш «//». Как показал эксперимент, двойной слеш используется только в 60% ссылок. Поэтому в список разделителей $\{S_2\}$ были добавлены еще несколько признаков кроме двойного слеша.

После выделения издания осуществляется его поиск в нормализованной базе названий журналов и трудов конференций. В ней кроме полного официального названия журнала или трудов конференции указываются и его варианты, в частности сокращенные (так называемые псевдонимы). Именно по псевдонимам, как правило, ведется поиск издания после выделения его названия на предыдущем этапе, так как официальные названия в ссылках на цитируемые статьи встречаются редко. Если издание найдено, то его коды ННИ присваиваются цитируемой статье.

В случае отсутствия искомого названия в БД в работу включается оператор, задачей которого является привязка искомого названия к существующему периодическому изданию и добавление его в список псевдонимов. Если это издание отсутствует, то оператор вводит в БД новое официальное название издания вместе с перечнем его кодов ННИ, которые автоматически присваиваются цитируемой статье.

В результате обработки массива полнотекстовых описаний изобретений были автоматически идентифицированы 512 выделенных фрагментов текста как ссылки на статьи в журналах или трудах конференций. Отметим, что в нормализованном списке названий журналов и трудов макета АИС не все они имеют коды ННИ. Поэтому из 512 выделенных ссылок было зарубрицировано 249 цитируемых статей. Таблица 1 показывает в четвертом столбце распределение зарубрицированных статей по 11 подклассам класса G06.

В пятом столбце табл. 1 приведены относительные данные распределения зарубрицированных статей. Первые три позиции по интенсивности цитирования научных статей занимают «Оптические вычислительные устройства», «Обработка или генерация данных изображения», «Распознавание,

представление и воспроизведение данных». Аналогичные зарубежные данные по интенсивности цитирования приведены в работах [1, 10].

4 Расчет значений индикаторов

В проведенном эксперименте вычислялись значения двух индикаторов:

- (1) матрицы корреляций между индексами МПК и рубриками ННИ по ГРНТИ;
- (2) распределения времени отклика на статью (от момента ее публикации до момента публикации патента на изобретение, где она цитируется).

Вычисление первого индикатора как матрицы коэффициентов корреляции сводится к подсчету числа связей между индексами МПК и рубриками ГРНТИ, т. е. частоты встречаемости в массиве изобретений пар «индекс МПК изобретения – рубрика ГРНТИ цитируемой статьи». Эта частотность может быть вычислена с помощью разных алгоритмов вычисления частоты, каждому из которых соответствует свой индикатор.

Для описания используемого в проведенном эксперименте индикатора потребуется дополнительная информация о (1) уровне и (2) значении классификации, а также о (3) позициях индексов МПК [14], которая кодируется в отдельных полях для каждого индекса латинскими литерами следующим образом.

Во-первых, изобретения могут классифицироваться индексами по трем уровням: или только по подклассу, или только по основным группам, или по полному тексту МПК, иерархические структурные уровни которой перечислены в сноске 5 на с. 114, что помечается литерой *S*, *C* или *A* соответственно. В эксперименте не проводились различия между этими уровнями. Во-вторых, индексами могут классифицироваться сами изобретения или дополнительная информация к ним (это помечается литерой *I* или *N* соответственно), что также не различалось при расчете значений индикаторов. Кодирование с помощью этих двух литер в патентном деле называется определением «значения классификации». В-третьих, первый по расположению в изобретении индекс считается основным. Поэтому разные индикаторы соответствуют алгоритмам расчета их значений по основному и дополнительным (неосновным) индексам, что помечается литерой *F* и *L* соответственно. Возможен также расчет значений по их совокупности, который и использовался в проведенном эксперименте. Однако выше (см. табл. 1) для расчета распределения изобретений и

цитируемых статей использовался только основной индекс.

Таким образом, из всего спектра возможных индикаторов был выбран только один, алгоритм которого не учитывал литеры уровня и значения классификации. Позиции индексов МПК фиксировались, но не учитывались при вычислении значений этого индикатора. Иначе говоря, используемый в эксперименте индикатор характеризует связи между всеми индексами МПК и рубриками ГРНТИ, т. е. если n — число индексов некоторого изобретения с литерами *S*, *C*, *A*, *I*, *N*, *F* и *L*, а m_j — число рубрик ГРНТИ, присвоенных j -й статье, цитируемой в этом изобретении, то число его связей N , образуемых этими индексами и рубриками, равно

$$N = n \sum_j m_j.$$

Частотности связей между индексами МПК и рубриками ГРНТИ, вычисленные для всего массива изобретений по информационным технологиям, запатентованных в РФ за 2000–2012 гг., показаны в ячейках матрицы (табл. 2). В ее столбцах перечислены коды класса G06 и 8 его подклассов, названия которых содержит табл. 1. В строках матрицы приведены первые 9 ННИ в классификации ГРНТИ в порядке убывания числа связей с индексами по всему классу G06 (см. последний столбец табл. 2). Данные табл. 2 на первый взгляд выглядят парадоксально. С одной стороны, в описаниях технических решений широко используются математические формулы и методы, с другой стороны, рубрика «МАТЕМАТИКА» оказалась на 9-м месте в табл. 2. Этот парадокс был отмечен более 10 лет назад. Эксперты, которые изучали тексты изобретений и исследовали этот парадокс, отметили широко распространенную практику использования математических формул и методов без цитирования в изобретениях математических публикаций [11].

Есть еще одна причина, объясняющая этот парадокс. Согласно п. 5 ст. 1350 ГК РФ (часть четвертая) изобретениями не являются научные теории и математические методы [17]. Если бы такое ограничение отсутствовало, то в ряде классов МПК появились бы «математические» изобретения со ссылками на статьи, что привело бы к увеличению значений индикаторов интенсивностей «математических» взаимосвязей по этим классам.

Данные о частотности взаимосвязей между индексами МПК и рубриками ГРНТИ были вычислены для всех статей, цитируемых в описаниях изобретений, независимо от того, кем цитируется статья: экспертами, что обозначается в описании изобретения меткой 56, или авторами изобре-

Таблица 2 Частотности связей между индексами МПК и рубриками ГРНТИ (%)

Код рубрики ГРНТИ	Название рубрики	G06E	G06F	G06G	G06K	G06M	G06N	G06Q	G06T	G06
50.00.00	АВТОМАТИКА. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА	0	9,58	0,27	4,85	0	0,62	1,01	2,05	18,38
28.00.00	КИБЕРНЕТИКА	0	8,40	0,11	4,10	0	0,59	1,05	1,35	15,60
47.00.00	ЭЛЕКТРОНИКА. РАДИОТЕХНИКА	0,10	5,16	0,41	3,86	0	0,17	0	0,50	10,20
45.00.00	ЭЛЕКТРОТЕХНИКА	0	4,88	0,26	3,49	0	0,19	0	0,35	9,17
20.00.00	ИНФОРМАТИКА	0	4,41	0,02	3,25	0	0,12	0	0,05	7,85
30.00.00	МЕХАНИКА	0	4,19	0,00	3,23	0	0,11	0	0,04	7,57
29.00.00	ФИЗИКА	0	3,79	0,02	3,17	0	0,01	0	0	6,99
84.00.00	СТАНДАРТИЗАЦИЯ	0	3,50	0	3,11	0	0	0,01	0	6,62
27.00.00	МАТЕМАТИКА	0	2,08	0	0,71	0	0,52	1,01	1,30	5,62
	Остальные рубрики ГРНТИ	0,00	8,69	0,23	1,39	0,01	1,16	0,15	0,37	12,00

ний. Поэтому кроме вычисленного индикатора существуют еще два, которые учитывают авторство ссылки на статью. Эти индикаторы в статье не рассматриваются.

В рамках проведенного эксперимента авторство ссылок учитывалось при вычислении значений другого индикатора — распределения времени отклика на статьи (рис. 4). В процессе вычисления этого индикатора для каждой пары «индекс МПК – рубрика ГРНТИ» было определено время отклика как разность между годом публикации патента и годом публикации статьи. Отдельно отмечались статьи, цитируемые экспертами в отчетах о патентном по-

иске. Затем было построено распределение времени отклика с учетом авторства ссылок на статьи.

5 Заключение

Разработанный макет АИС и технология его применения впервые в отечественной практике дают возможность выявлять количественные взаимосвязи ННИ с заданным видом технологий. С помощью макета были вычислены значения индикатора тематических взаимосвязей информационных технологий, относящихся к классу G06 МПК, с ННИ в виде рубрик ГРНТИ.

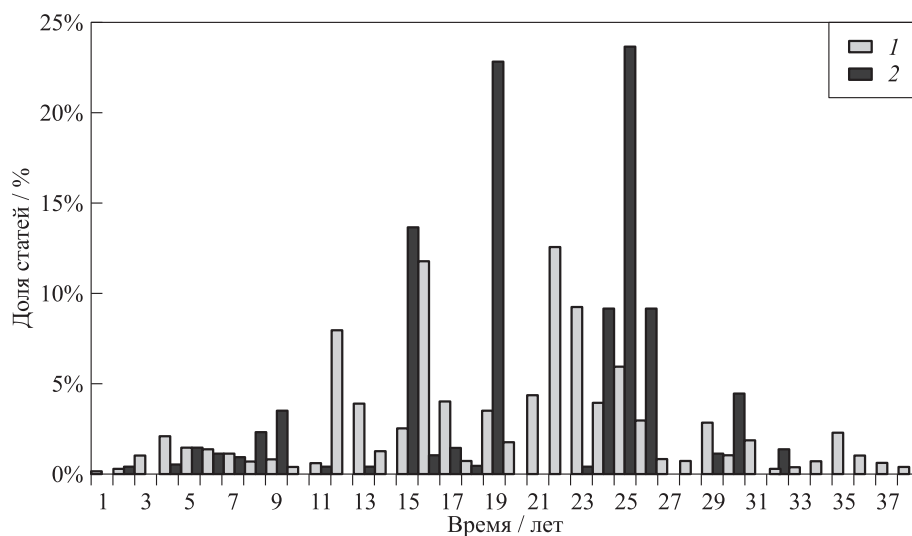


Рис. 4 Распределение времени между публикацией статьи и выдачей патента по классу G06: 1 — по полным описаниям изобретений; 2 — по спискам документов, цитируемых в отчетах об информационном поиске

Вычисленные значения показывают, что наиболее часто в изобретениях по информационным технологиям цитируются статьи по автоматике, вычислительной технике, кибернетике, электронике, радиотехнике, электротехнике и информатике (см. табл. 2). Таким образом, использование ННИ в виде рубрик ГРТИ дает в первую очередь прикладной разрез тематических взаимосвязей. Поэтому для получения более полной картины взаимосвязей научных дисциплин с технологиями необходимо также использовать рубрики фундаментальных наук, например классификатор РФФИ.

В макете АИС предусмотрена возможность выбора между разными рубриками ННИ, что является его ключевым отличием от имеющихся зарубежных аналогов, в которых используется только один рубрикатор без возможности его замены [9]. Однако при использовании любого рубрикатора ННИ в процессе принятия решений полученные значения могут использоваться только после их экспертной проверки, так как любые известные в настоящее время индикаторы тематических взаимосвязей науки и технологий являются косвенными. Отметим, что проведение экспертной верификации выходит за рамки выполненного проекта по гранту РГНФ. Его главная цель заключалась в создании действующего макета АИС и работоспособной технологии.

Кроме косвенного индикатора были определены значения одного прямого индикатора — распределение времени отклика на статьи. Рисунок 4 содержит экспериментальные данные, которые позволяют утверждать, что в период 2000–2012 гг. эксперты в отчетах о поиске и авторы изобретений по информационным технологиям наиболее часто цитировали статьи, опубликованные за 10, 20 и 30 лет до выдачи патентов на эти изобретения.

Разработанная технология включает операции обработки полных текстов описаний, на которых осуществляется переход от текстов к численным расчетам значений индикаторов. В процессе перехода одновременно использовались индексы МПК и рубрики ННИ. При использовании разных классификаторов ННИ результаты численных расчетов значений индикаторов для одного и того же массива изобретений будут отличаться. Это позволяет выявлять тематические взаимосвязи как в прикладном, так и в фундаментальном разрезе. Однако если на основе значений индикаторов (после их экспертной верификации) планируется формировать и реализовывать стратегию в сфере науки, в том числе распределять финансовые средства по ориентированному научным направлениям, то должен использоваться тот вариант классификатора

ННИ, который применяется для стратегического планирования.

Литература

1. Минин В. А., Зацман И. М., Кружков М. Г., Норецкая Т. П. Методологические основы создания информационных систем для вычисления индикаторов тематических взаимосвязей науки и технологий // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 70–81.
2. Narin F., Noma E. Is technology becoming science? // *Scientometrics*, 1985. Vol. 7. No. 3-6. P. 369–381.
3. Mansfield E. Academic research and innovation // *Res. Policy*, 1991. Vol. 20. No. 1. P. 1–12.
4. Schmoch U. Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. No. 1. P. 193–211.
5. Mansfield E. Academic research underlying industrial innovations: Sources, characteristics and financing // *Rev. Econ. Stat.*, 1995. Vol. 77. No. 1. P. 55–62.
6. Narin F., Olivastro D. Linkage between patents and papers: An interim EPO/US comparison // *Scientometrics*, 1998. Vol. 41. No. 1-2. P. 51–59.
7. Mansfield E. Academic research and industrial innovation: An update of empirical findings // *Res. Policy*, 1998. Vol. 26. No. 7-8. P. 773–776.
8. Tijssen R. J. W., Buter R. K., Van Leeuwen Th. N. Technological relevance of science: An assessment of citation linkages between patents and research papers // *Scientometrics*, 2000. Vol. 47. No. 2. P. 389–412.
9. Verbeek A., Debackere K., Luwel M., Andries P., Zimmermann E., Deleus D. Linking science to technology: Using bibliographic references in patents to build linkage schemes // *Scientometrics*, 2002. Vol. 54. No. 3. P. 399–420.
10. Van Looy B., Zimmermann E., Veugelers R., Verbeek A., Mello J., Debackere K. Do science–technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains // *Scientometrics*, 2003. Vol. 57. No. 3. P. 355–367.
11. European Commission. Third European Report on Science & Technology Indicators. — Luxembourg: Office for Official Publications of the European Communities, 2003. 451 p.
12. Зацман И. М., Хавансков В. А., Шубников С. К. Метод извлечения библиографической информации из полнотекстовых описаний изобретений // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 52–65.
13. Минин В. А., Зацман И. М., Хавансков В. А., Шубников С. К. Архитектурные решения для систем вычисления индикаторов тематических взаимосвязей науки и технологий // Системы и средства информатики, 2013. Т. 23. № 2. С. 260–283.
14. Стандарт ВОИС ST.8 // Справочник по информации и документации в области промышленной

- собственности. — М.: Роспатент, 2011. С. 3.8.2–3.8.4. http://www.rupto.ru/rupto/nfile/2381c8b7-1033-11e1-a520-9c8e9921fb2c/03_08_01.pdf.
15. Написание классификационных индексов и индексов кодирования на патентных документах. — М.: ФИПС, 2014. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/inform_resources/international_classification/inventions/mpk_begin/article_12.
 16. *Зацман И. М., Шубников С. К.* Принципы обработки информационных ресурсов для оценки инновационного потенциала направлений научных исследований // Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL'2007: Тр. IX Всеросс. научн. конф. — Переславль: Университет города Переславля, 2007. С. 35–44.
 17. Административный регламент исполнения Роспатентом приема заявок на изобретение, их рассмотрения и экспертизы. — М.: ФИПС, 2008. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/order_minobr/administrative_regulations/test_8.
 18. Стандарт ВОИС ST.14 // Справочник по информации и документации в области промышленной собственности. — М.: Роспатент, 2008. http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03_14_01.pdf.
 19. *Васильев А., Козлов Д., Самусев С., Шамина О.* Извлечение метаинформации и библиографических ссылок из текстов русскоязычных научных статей // Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL'2007: Тр. IX Всеросс. научн. конф. — Переславль: Университет города Переславля, 2007. С. 175–184.
 20. *Васильев А., Козлов Д., Самусев С., Шамина О.* Создание электронной библиотеки русскоязычных научных статей // Интернет-математика 2007: Сб. работ участников конкурса научных проектов по информационному поиску. — Екатеринбург: Изд-во Уральского университета, 2007. С. 37–45.
 21. *Van Looy B., Hansen W., Hollanders H., Tijssen R.* Using concordance tables to disentangle performance dynamics of HT manufacturing industries: An empirical assessment of national innovation systems // 10th Conference (International) on Science and Technology Indicators (STI'2008): Book of abstracts. — Vienna: ARC GmbH, 2008. P. 196–200.
 22. ГОСТ 7.1-2003 Библиографическая запись. Библиографическое описание. Общие требования и правила составления. — М.: Изд-во стандартов, 2004. 48 с.

Поступила в редакцию 17.06.14

INDICATORS OF THEMATIC SCIENCE–TECHNOLOGY LINKAGES: FROM TEXT TO NUMBERS

V. A. Minin¹, I. M. Zatsman², V. A. Havanskov², and S. K. Shubnikov²

¹RFFI, 32A Leninsky Prosp., Moscow 119991, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article describes the principal stages of calculation of indicators for thematic science–technology linkages. Source data for carrying out calculations are results of processing of full-text descriptions of inventions in natural language. The purpose of processing is extracting information about scientific publications cited in full-text descriptions of inventions. Comparing this information with indexes of the International Patent Classification helps experts to reveal science–technology linkages and to estimate them by means of quantitative indicators. In the article, the technological stages of calculating indicators for linkage assessment on which transition from text processing to numerical calculations is carried out are investigated. The article describes the techniques developed for extracting information about scientific publications from full-text descriptions of inventions and calculating indicator values. These techniques helped to define indicators of science–technology linkages for information technology research fields, which were patented in the Russian Federation in 2000–2012. For the domestic scientific and technical sphere, these indicators were calculated for the first time.

Keywords: science–technology linkages; methodology for calculation of linkage indicators; information technology; text processing; calculation of indicator values

DOI: 10.14375/19922264140313

Acknowledgments

The work was performed in the Institute of Informatics Problems of the Russian Academy of Sciences and was partially supported by the Russian Foundation for Humanities (grant No. 12-02-12019v).

References

1. Minin, V. A., I. M. Zatsman, M. G. Kruzhkov, and T. P. Norekjan. 2013. Metodologicheskie osnovy sozdaniya informatsionnykh sistem dlya vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologii [Methodological basis for the creation of information systems for the calculation of indicators of thematic linkages between science and technologies]. *Informatika i ee primeneniya — Inform. Appl.* 7(1):70–81.
2. Narin, F., and E. Noma. 1985. Is technology becoming science? *Scientometrics* 7(3–6):369–381.
3. Mansfield, E. 1991. Academic research and innovation. *Res. Policy* 20(1):1–12.
4. Schmoch, U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* 26(1):193–211.
5. Mansfield, E. 1995. Academic research underlying industrial innovations: Sources, characteristics and financing. *Rev. Econ. Stat.* 77(1):55–62.
6. Narin, F., and D. Olivastro. 1998. Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics* 41(1–2):51–59.
7. Mansfield, E. 1998. Academic research and industrial innovation: An update of empirical findings. *Res. Policy* 26(7–8):773–776.
8. Tijssen, R. J. W., R. K. Buter, and Th. N. Van Leeuwen. 2000. Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics* 47(2):389–412.
9. Verbeek, A., K. Debackere, M. Luwel, P. Andries, E. Zimmermann, and D. Deleus. 2002. Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(3):399–420.
10. Van Looy, B., E. Zimmermann, R. Veugelers, A. Verbeek, J. Mello, and K. Debackere. 2003. Do science–technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains. *Scientometrics* 57(3):355–367.
11. European Commission. 2003. Third European Report on Science & Technology Indicators. Luxembourg: Office for Official Publications of the European Communities. 451 p.
12. Zatsman, I. M., V. A. Havanskov, and S. K. Shubnikov. 2013. Metod izvlecheniya bibliograficheskoy informatsii iz polnotekstovyykh opisaniy izobreteniy [Extraction method of bibliographic information for full-text descriptions of inventions]. *Informatika i ee primeneniya — Inform. Appl.* 7(4):52–65.
13. Minin, V. A., I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov. 2013. Arkhitekturnye resheniya dlya sistem vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologii [Information system conceptual decisions for assessment of linkages between science and technologies]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):260–283.
14. Standart VOIS ST.8 [WIPO standard ST.8]. Available at: http://www.rupto.ru/rupto/nfile/2381c8b7-1033-11e1-a520-9c8e9921fb2c/03_08_01.pdf (accessed June 23, 2014).
15. Napisanie klassifikatsionnykh indeksov i indeksov kodirovaniya na patentnykh dokumentakh [Writing of classification indexes and coding indexes on patent documents]. Available at: http://www1.fips.ru/wps/wcm/connect/content_ru/ru/inform_resources/international_classification/inventions/mpk_begin/article_12 (accessed June 23, 2014).
16. Zatsman, I. M., and S. K. Shubnikov. 2007. Printsipy obrabotki informatsionnykh resursov dlya otsenki innovatsionnogo potentsiala napravleniy nauchnykh issledovaniy [Processing principles of information resources for an assessment of innovation potential of the scientific domains]. *Tr. IX Vseross. Nauchn. Konf. “Elektronnye Biblioteki”* [9th All-Russian Scientific Conference on Digital Libraries Proceedings]. Pereslavl’. 35–44.
17. Administrativnyy reglament ispolneniya Rospatentom priema zayavok na izobretenie, ikh rassmotreniya i ekspertizy [Rospatent administrative regulations for filing invention applications, their considerations, and examination]. Available at: http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/order_minobr/administrative_regulations/test_8/ (accessed June 23, 2014).
18. Standart VOIS ST.14 [WIPO standard ST.14]. Available at: http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03_14_01.pdf (accessed June 23, 2014).
19. Vasil’ev, A., D. Kozlov, S. Samusev, and O. Shamina. 2007. Izvlechenie metainformatsii i bibliograficheskikh ssylok iz tekstov russkoyazychnykh nauchnykh statey [Extraction of metainformation and bibliographic references from texts of Russian scientific articles]. *Tr. IX Vseross. Nauchn. Konf. “Elektronnye Biblioteki”* [9th All-Russian Scientific Conference on Digital Libraries Proceedings]. Pereslavl’. 175–184.
20. Vasil’ev, A., D. Kozlov, S. Samusev, and O. Shamina. 2007. Sozdanie elektronnoy biblioteki russkoyazychnykh nauchnykh statey [Creation of digital library of Russian scientific articles]. *Sb. rabot stipendiatov granta “Internet-matematika 2007”* [Paper collection of scholars of a grant “Internet-mathematics 2007”]. Ekaterinburg: Publishing house of Urals’ University. 37–45.
21. Van Looy, B., W. Hansen, H. Hollanders, and R. Tijssen. 2008. Using concordance tables to disentangle perfor-

- mance dynamics of HT manufacturing industries: An empirical assessment of national innovation systems. *10th Conference (International) on Science and Technology Indicators (STP2008): Book of abstracts*. Vienna: ARC GmbH. 196–200.
22. GOST 7.1-2003. 2004. Bibliograficheskaya zapis'. Bibliograficheskoe opisanie. Obshchie trebovaniya i pravila sostavleniya [Bibliographic record. Bibliographic description. General requirements and drawing up rules]. Moscow: Standardinform Pubs. 48 p.

Received June 17, 2014

Contributors

Minin Vladimir A. (b. 1941) — Doctor of Science in physics and mathematics, adviser, Russian Foundation for Basic Research, 32A Leninsky Prosp., Moscow 119991, Russian Federation; minin@rfbr.ru

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; iz_ipi@a170.ipi.ac.ru

Havanskov Valerij A. (b. 1950) — scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; havanskov@a170.ipi.ac.ru

Shubnikov Sergej K. (b. 1955) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sergeysh50@yandex.ru

АНАЛИТИЧЕСКИЕ АСПЕКТЫ ОЦЕНКИ ЭФФЕКТИВНОСТИ В ТЕХНОЛОГИИ ПОДДЕРЖКИ ДЕЯТЕЛЬНОСТИ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ

А. А. Зацаринный¹, А. П. Шабанов²

Аннотация: Рассматривается система нормированных и фактических показателей, используемая в технологии поддержки деятельности организационной системы — ведомства, предприятия. Приводятся критерии оценки, которые обеспечивают лицам, принимающим решения в отношении различных ситуаций, возникающих в среде деятельности организационной системы, классификационный подход к выбору сценария принятия решения. Классификация проводится по иерархии деятельности, осуществляемой в организационной системе: виды деятельности в подразделениях организационной системы, виды деятельности и деятельность организационной системы в целом. Основной акцент представленного методического подхода к оценке эффективности деятельности носит практический характер, учитывающий степень угрозы для деятельности организационной системы. Авторы выделяют три степени состояния деятельности: деятельность осуществляется в диапазоне заданных нормированных показателей, априорно существует угроза для продолжения деятельности и деятельность уже находится под воздействием проявлений угрозы.

Ключевые слова: организационная система; технология поддержки деятельности; система показателей; информация; принятие решений; эффективность; объекты наблюдения

DOI: 10.14375/19922264140314

1 Введение

В настоящее время существует *проблема* сокращения времени предоставления информации лицам, принимающим решения по предотвращению угроз, а в случаях их реализации — по быстрой ликвидации и минимизации потерь. Серьезность проблемы обусловлена возрастанием угроз, вызванных техногенными, природными и человеческими факторами. Другой проблемой является постоянный рост объемов информации, которую необходимо собрать, обработать и предоставить уполномоченным лицам для принятия решений по предотвращению проявления или по ликвидации угроз [1–3]. Существование указанных проблем обусловило необходимость разработки научно-обоснованной технологии поддержки деятельности организационных систем — ведомств и предприятий различных форм собственности и отраслей хозяйствования. Важным компонентом этой технологии является система показателей, предназначенная для оценки эффективности деятельности в организационной системе. В статье рассматривается система нормированных и фактических показателей, используемая в технологии

поддержки деятельности организационной системы [4].

2 Показатели эффективности деятельности

Технология информационной поддержки деятельности организационной системы [4] характеризуется тем, что содержит следующие этапы:

1. Формирование информации:

- о нормированных показателях и состояниях объектов наблюдения с учетом их влияния на деятельность организационной системы;
- о критических и допустимых показателях эффективности деятельности организационной системы;
- о сценариях принятия решений в их привязке к различным ситуациям, которые априорно могут произойти в среде деятельности организационной системы.

2. Установление объектов наблюдения в нормированные состояния.

¹Институт проблем информатики Российской академии наук, AZatsarinny@ipiran.ru

²Институт проблем информатики Российской академии наук, AShabanov@ibs.ru

3. Определение фактических показателей объектов наблюдения.
 4. Формирование информации о фактических показателях и состояниях объектов наблюдения.
 5. Проведение оценки эффективности деятельности организационной системы.
 6. Определение сценария принятия решения в зависимости от результата оценки эффективности и от ситуации, сложившейся в среде деятельности организационной системы.
 7. Применение сценария — установление объектов наблюдения, оказывающих влияние на деятельность организационной системы, в состоянии, соответствующие командам, содержащимся в сценарии.
- Объектами наблюдения являются материальные и нематериальные объекты организационной системы, внешней среды и субъекты, которые влияют на состояние деятельности организационной системы или отражают результаты этой деятельности.
- Объекты наблюдения размещаются в контролируемом пространстве и/или с возможностью удаленного управления и наблюдения над ними. В состав информации об объектах наблюдения входят

Таблица 1 Показатели эффективности деятельности

Показатель	Описание
N	Число видов деятельности организационной системы
$n = 1, 2, \dots, N$	Номер вида деятельности организационной системы
M	Число подразделений организационной системы
$m = 1, 2, \dots, M$	Номер подразделения организационной системы
D и D^*	Нормированный и фактический показатели состояния деятельности организационной системы в целом соответственно
D_n и D_n^*	Нормированный и фактический показатели состояния n -го вида деятельности организационной системы соответственно
S_{mn} и S_{nm}^*	Нормированный и фактический показатели состояния n -го вида деятельности в m -м подразделении соответственно
α_n и β_{nm}	Приоритет n -го вида деятельности организационной системы и приоритет n -го вида деятельности, осуществляемой в m -м подразделении организационной системы, соответственно
ΔD^* , ΔD_n^* и ΔS_{nm}^*	Фактические показатели эффективности деятельности организационной системы в целом, n -го вида деятельности организационной системы и n -го вида деятельности организационной системы в m -м подразделении соответственно
$\Delta D_{\text{крит}}$	Критический показатель эффективности деятельности организационной системы в целом, снижение по сравнению с которым фактического показателя ΔD^* эффективности означает существование угрозы для деятельности организационной системы в целом и необходимости принятия действий по ее ликвидации
$\Delta D_{\text{доп}}$	Допустимый показатель эффективности деятельности организационной системы в целом, снижение по сравнению с которым фактического показателя ΔD^* эффективности означает возможность появления угрозы для деятельности организационной системы в целом и необходимости принятия действий по недопущению ее появления
$\Delta D_{n\text{-крит}}$	Критический показатель эффективности n -го вида деятельности, снижение по сравнению с которым фактического показателя ΔD_n^* эффективности означает существование угрозы для этого вида деятельности и необходимости принятия действий по ее ликвидации
$\Delta D_{n\text{-доп}}$	Допустимый показатель эффективности n -го вида деятельности, снижение по сравнению с которым фактического показателя ΔD_n^* эффективности означает возможность появления угрозы для этого вида деятельности и необходимости принятия действий по недопущению ее появления
$\Delta S_{nm\text{-крит}}$	Критический показатель эффективности n -го вида деятельности в m -м подразделении, снижение по сравнению с которым фактического показателя ΔS_{nm}^* эффективности означает существование угрозы для этого вида деятельности в данном подразделении и необходимости принятия действий по ее ликвидации
$\Delta S_{nm\text{-доп}}$	Допустимый показатель эффективности n -го вида деятельности в m -м подразделении, снижение по сравнению с которым фактического показателя ΔS_{nm}^* эффективности означает возможность появления угрозы для этого вида деятельности в данном подразделении и необходимости принятия действий по недопущению ее появления

Таблица 2 Показатели объектов наблюдения и показатели их состояния

Показатель	Описание
K	Число объектов наблюдения в зоне ответственности организационной системы
$k = 1, 2, \dots, K$	Номер объекта наблюдения
L_k	Число показателей k -го объекта наблюдения
$l = 1, 2, \dots, L_k$	Номер показателя k -го объекта наблюдения
V_k^l и V_k^{*l} , V_{nmk}^l и V_{nmk}^{*l} , V_{nmk} и V_{nmk}^*	Нормированный и фактический l -й показатели k -го объекта наблюдения, нормированный и фактический l -й показатели k -го объекта наблюдения с учетом его влияния на n -й вид деятельности в m -м подразделении, нормированный и фактический показатели состояния k -го объекта наблюдения с учетом его влияния на n -й вид деятельности в m -м подразделении соответственно
ΔV_{nmk}^{*l}	Отклонение фактического l -го показателя от нормированного l -го показателя k -го объекта наблюдения, который оказывает влияние на n -й вид деятельности, осуществляемой в m -м подразделении организационной системы
μ_k^l и γ_{nmk}	Приоритет l -го показателя k -го объекта наблюдения и приоритет k -го объекта наблюдения с учетом его влияния на n -й вид деятельности в m -м подразделении организационной системы соответственно

сведения о физических, логических, информационных, территориальных, конструктивных, организационных и других типах связи.

Оценка фактической эффективности видов деятельности в подразделениях организационной системы, видов деятельности и деятельности организационной системы в целом (этапы 1–5 технологии) производится с применением нормированных, фактических, критических и допустимых показателей, приведенных в табл. 1.

Исходной информацией для определения показателей эффективности деятельности организационной системы являются фактические и нормированные показатели объектов наблюдения. Нормированные показатели объектов наблюдения определяются при проектировании ситуационных центров, уточняются в ходе испытаний и при их вводе в промышленную эксплуатацию. Фактические показатели объектов наблюдения определяются с помощью компонентов мониторинга, в частности центра мониторинга устойчивости информационных систем [5]. В табл. 2 приведены показатели объектов наблюдения и показатели их состояния.

3 Расчетные соотношения

Для определения показателей эффективности деятельности организационной системы в рассматриваемой технологии используются следующие соотношения:

- при оценке эффективности деятельности организационной системы в целом:

$$\Delta D^* = \frac{D^*}{D};$$

$$D^* = \alpha_1 D_1^* + \alpha_2 D_2^* + \dots + \alpha_N D_N^*;$$

$$D = \alpha_1 D_1 + \alpha_2 D_2 + \dots + \alpha_N D_N;$$

- при оценке эффективности видов деятельности организационной системы:

$$\Delta D_n^* = \frac{D_n^*}{D_n};$$

$$D_n^* = \beta_{n1} S_{n1}^* + \beta_{n2} S_{n2}^* + \dots + \beta_{nM} S_{nM}^*;$$

$$D_n = \beta_{n1} S_{n1} + \beta_{n2} S_{n2} + \dots + \beta_{nM} S_{nM};$$

- при оценке эффективности видов деятельности в подразделениях организационной системы:

$$\Delta S_{nm}^* = \frac{S_{nm}^*}{S_{nm}};$$

$$S_{nm}^* =$$

$$= \gamma_{nm1} V_{nm1}^* + \gamma_{nm2} V_{nm2}^* + \dots + \gamma_{nmK} V_{nmK}^*;$$

$$S_{nm} =$$

$$= \gamma_{nm1} V_{nm1} + \gamma_{nm2} V_{nm2} + \dots + \gamma_{nmK} V_{nmK}.$$

Для оценки показателей фактического состояния k -го объекта наблюдения, который оказывает влияние на n -й вид деятельности, осуществляемой в m -м подразделении организационной системы, используются следующие соотношения:

Таблица 3 Классификация сценариев принятия решений

Классификатор сценария	Описание содержания классификатора
$W_{nm-норм}$	Множество сценариев принятия решений, предназначенных для установления нормированного состояния n -го вида деятельности в m -м подразделении организационной системы; $n = 1, 2, \dots, N$; $m = 1, 2, \dots, M$
$W_{nm-норм}^{y_0}$	Множество команд управления, предназначенных для установления нормированного состояния n -го вида деятельности в m -м подразделении организационной системы; $y_0 = 1, \dots, Y_{nm-норм}$
$W_{крит}, W_{пред}$ и $W_{план}$	Множества соответственно критических, предупреждающих и плановых сценариев принятия решений, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности деятельности организационной системы в целом
$W_{крит}^{q_1}, W_{пред}^{q_2}$ и $W_{план}^{q_3}$	Множества соответственно критических, предупреждающих и плановых команд управления, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности организационной системы в целом; $q_1 = 1, \dots, Q_{крит}$; $q_2 = 1, \dots, Q_{пред}$; $q_3 = 1, \dots, Q_{план}$
$W_{n-крит}, W_{n-пред}$ и $W_{n-план}$	Множества соответственно критических, предупреждающих и плановых сценариев принятия решений, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности n -го вида деятельности организационной системы; $n = 1, 2, \dots, N$
$W_{n-крит}^{u_1}, W_{n-пред}^{u_2}$ и $W_{n-план}^{u_3}$	Множества соответственно критических, предупреждающих и плановых команд управления, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности n -го вида деятельности организационной системы; $u_1 = 1, \dots, U_{n-крит}$; $u_2 = 1, \dots, U_{n-пред}$; $u_3 = 1, \dots, U_{n-план}$
$W_{nm-крит}, W_{nm-пред}$ и $W_{nm-план}$	Множества соответственно критических, предупреждающих и плановых сценариев принятия решений, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности n -го вида деятельности в m -м подразделении организационной системы; $n = 1, 2, \dots, N$; $m = 1, 2, \dots, M$
$W_{nm-крит}^{y_1}, W_{nm-пред}^{y_2}$ и $W_{nm-план}^{y_3}$	Множества соответственно критических, предупреждающих и плановых команд управления, предназначенных соответственно для ликвидации угрозы, для предотвращения угрозы и для повышения эффективности n -го вида деятельности в m -м подразделении организационной системы; $y_1 = 1, \dots, Y_{nm-крит}$; $y_2 = 1, \dots, Y_{nm-пред}$; $y_3 = 1, \dots, Y_{nm-план}$

$$V_{nmk}^* = \mu_k^1 (V_{nmk}^1 - \Delta V_{nmk}^{*1}) + \mu_k^2 (V_{nmk}^2 - \Delta V_{nmk}^{*2}) + \dots + \mu_k^{L_k} (V_{nmk}^{*L_k} - V_{nmk}^{L_k});$$

$$\Delta V_{nmk}^{*l} = |V_{nmk}^l - V_{nmk}^{*l}|;$$

$$V_{nmk} = \mu_k^1 V_{nmk}^1 + \mu_k^2 V_{nmk}^2 + \dots + \mu_k^{L_k} V_{nmk}^{L_k}.$$

На основе результатов оценки эффективности деятельности организационной системы:

- определяются сценарии принятия решений;
- определяются, передаются на исполнение и исполняются команды, предназначенные для управления объектами, которые влияют на деятельность организационной системы (этапы 6 и 7 технологии).

В табл. 3 показана классификация сценариев принятия решений и команд управления.

4 Правила определения сценариев принятия решения

На рис. 1–3 приведены диаграммы, поясняющие подход в рассматриваемой технологии к определению состояния деятельности организационной системы и в дальнейшем к определению сценариев принятия решения, соответствующих данному состоянию и сложившейся ситуации, описываемой показателями объектов наблюдения.

Как видно из приведенных диаграмм, в основе применения рассматриваемой системы показате-

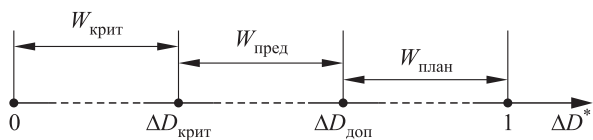


Рис. 1 Соотношения сценариев в зависимости от критических и допустимых показателей эффективности (пример 1). Выбор одного из множеств сценариев ($W_{крит}$, $W_{пред}$, $W_{план}$), из числа которых определяется сценарий принятия решения, производится в зависимости от области размещения показателя эффективности ($0-\Delta D_{крит}$; $\Delta D_{крит}-\Delta D_{доп}$; $\Delta D_{доп}-1$)

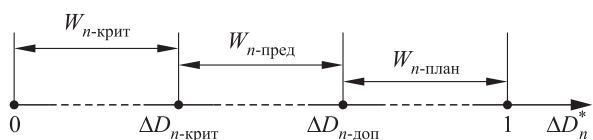


Рис. 2 Соотношения сценариев в зависимости от критических и допустимых показателей эффективности (пример 2). Выбор одного из множеств сценариев ($W_{n-крит}$, $W_{n-пред}$, $W_{n-план}$), из числа которых определяется сценарий принятия решения, производится в зависимости от области размещения показателя эффективности ($0-\Delta D_{n-крит}$; $\Delta D_{n-крит}-\Delta D_{n-доп}$; $\Delta D_{n-доп}-1$)

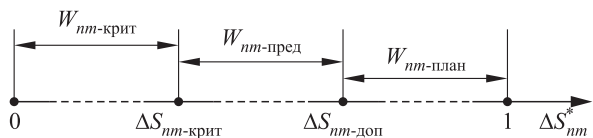


Рис. 3 Соотношения сценариев в зависимости от критических и допустимых показателей эффективности (пример 3). Выбор одного из множеств сценариев ($W_{nm-крит}$, $W_{nm-пред}$, $W_{nm-план}$), из числа которых определяется сценарий принятия решения, производится в зависимости от области размещения показателя эффективности ($0-\Delta S_{nm-крит}$; $\Delta S_{nm-крит}-\Delta S_{nm-доп}$; $\Delta S_{nm-доп}-1$)

лей лежат критерии оценки эффективности деятельности организационной системы в трехуровневом измерении:

- (1) оценка видов деятельности в подразделениях организационной системы;
- (2) оценка видов деятельности организационной системы;
- (3) оценка деятельности организационной системы в целом.

При выборе сценария принятия решения и соответствующих этому сценарию команд управления используются следующие правила:

1. В исходном состоянии из множества $W_{nm-норм}$ нормированных сценариев принятия решений определяют и исполняют команды $W_{nm-норм}^{y_0}$, предназначенные для установления нормированного состояния n -го вида деятельности в m -м подразделении организационной системы:

$$W_{nm-норм} = \{W_{nm-норм}^1; W_{nm-норм}^2; \dots; W_{nm-норм}^{Y_{nm-норм}}\}, \quad (1)$$

где $y_0 = 1, 2, \dots, Y_{nm-норм}$.

Принимают, что при выполнении этих команд показатели состояния видов деятельности и показатели состояния деятельности организационной системы в целом также будут нормированными.

2. В ходе деятельности организационной системы, если фактический показатель ΔD^* эффективности деятельности организационной системы в целом меньше соответствующего критического показателя $\Delta D_{крит}$:

$$0 \leq \Delta D^* < \Delta D_{крит},$$

то из множества $W_{крит}$ критических сценариев принятия решений определяют и исполняют команды $W_{крит}^{q_1}$, предназначенные для ликвидации последствий реализации угроз для деятельности организационной системы в целом:

$$W_{крит} = \{W_{крит}^1; W_{крит}^2; \dots; W_{крит}^{Q_{крит}}\},$$

где $q_1 = 1, 2, \dots, Q_{крит}$.

3. В ходе деятельности организационной системы, если фактический показатель ΔD^* эффективности деятельности организационной системы в целом не меньше соответствующего критического показателя $\Delta D_{крит}$ и меньше соответствующего допустимого показателя $\Delta D_{доп}$:

$$\Delta D_{крит} \leq \Delta D^* < \Delta D_{доп},$$

то из множества $W_{пред}$ предупреждающих сценариев определяют и исполняют команды $W_{пред}^{q_2}$, предназначенные для предотвращения угроз деятельности организационной системы в целом:

$$W_{пред} = \{W_{пред}^1; W_{пред}^2; \dots; W_{пред}^{Q_{пред}}\},$$

где $q_2 = 1, 2, \dots, Q_{пред}$.

4. В ходе деятельности организационной системы, если фактический показатель ΔD^* эффективности деятельности организационной системы в целом не меньше соответствующего допустимого показателя $\Delta D_{\text{доп}}$ и меньше единицы:

$$\Delta D_{\text{доп}} \leq \Delta D^* < 1,$$

то из множества $W_{\text{план}}$ плановых сценариев определяют и исполняют команды $W_{\text{план}}^{q_3}$, предназначенные для повышения эффективности деятельности организационной системы в целом:

$$W_{\text{план}} = \{W_{\text{план}}^1; W_{\text{план}}^2; \dots; W_{\text{план}}^{Q_{\text{план}}}\},$$

где $q_3 = 1, 2, \dots, Q_{\text{план}}$.

5. В ходе деятельности организационной системы, если фактический показатель ΔD_n^* эффективности n -го вида деятельности организационной системы меньше соответствующего критического показателя $\Delta D_{n\text{-крит}}$:

$$0 \leq \Delta D_n^* < \Delta D_{n\text{-крит}},$$

то из множества $W_{n\text{-крит}}$ критических сценариев определяют и исполняют команды $W_{n\text{-крит}}^{u_1}$, предназначенные для ликвидации последствий реализованных угроз n -му виду деятельности организационной системы:

$$W_{n\text{-крит}} = \{W_{n\text{-крит}}^1; W_{n\text{-крит}}^2; \dots; W_{n\text{-крит}}^{U_{n\text{-крит}}}\},$$

где $u_1 = 1, 2, \dots, U_{n\text{-крит}}$.

6. В ходе деятельности организационной системы, если фактический показатель ΔD_n^* эффективности n -го вида деятельности организационной системы не меньше соответствующего критического показателя $\Delta D_{n\text{-крит}}$ и меньше соответствующего допустимого показателя $\Delta D_{n\text{-доп}}$:

$$\Delta D_{n\text{-крит}} \leq \Delta D_n^* < \Delta D_{n\text{-доп}},$$

то из множества $W_{n\text{-пред}}$ предупреждающих сценариев определяют и исполняют команды $W_{n\text{-пред}}^{u_2}$, предназначенные для предотвращения угроз n -му виду деятельности организационной системы:

$$W_{n\text{-пред}} = \{W_{n\text{-пред}}^1; W_{n\text{-пред}}^2; \dots; W_{n\text{-пред}}^{U_{n\text{-пред}}}\},$$

где $u_2 = 1, 2, \dots, U_{n\text{-пред}}$.

7. В ходе деятельности организационной системы, если фактический показатель ΔD_n^* эффективности n -го вида деятельности организационной системы не меньше соответствующего

допустимого показателя $\Delta D_{n\text{-доп}}$ и меньше единицы:

$$\Delta D_{n\text{-доп}} \leq \Delta D_n^* < 1,$$

то из множества $W_{n\text{-план}}$ плановых сценариев определяют и исполняют команды $W_{n\text{-план}}^{u_3}$, предназначенные для повышения эффективности n -го вида деятельности организационной системы:

$$W_{n\text{-план}} = \{W_{n\text{-план}}^1; W_{n\text{-план}}^2; \dots; W_{n\text{-план}}^{U_{n\text{-план}}}\},$$

где $u_3 = 1, 2, \dots, U_{n\text{-план}}$.

8. В ходе деятельности организационной системы, если фактический показатель ΔS_{nm}^* эффективности n -го вида деятельности организационной системы в m -м подразделении меньше соответствующего критического показателя $\Delta S_{nm\text{-крит}}$:

$$0 \leq \Delta S_{nm}^* < \Delta S_{nm\text{-крит}},$$

то из множества $W_{nm\text{-крит}}$ критических сценариев определяют и исполняют команды $W_{nm\text{-крит}}^{y_1}$, предназначенные для ликвидации последствий реализованных угроз n -му виду деятельности организационной системы в m -м подразделении:

$$W_{nm\text{-крит}} = \{W_{nm\text{-крит}}^1; W_{nm\text{-крит}}^2; \dots; W_{nm\text{-крит}}^{Y_{nm\text{-крит}}}\},$$

где $y_1 = 1, 2, \dots, Y_{nm\text{-крит}}$.

9. В ходе деятельности организационной системы, если фактический показатель ΔS_{nm}^* эффективности n -го вида деятельности организационной системы в m -м подразделении не меньше соответствующего критического показателя $\Delta S_{nm\text{-крит}}$ и меньше соответствующего допустимого показателя $\Delta S_{nm\text{-доп}}$:

$$\Delta S_{nm\text{-крит}} \leq \Delta S_{nm}^* < \Delta S_{nm\text{-доп}},$$

то из множества $W_{nm\text{-пред}}$ предупреждающих сценариев определяют и исполняют команды $W_{nm\text{-пред}}^{y_2}$, предназначенные для предотвращения угроз n -му виду деятельности организационной системы в m -м подразделении:

$$W_{nm\text{-пред}} = \{W_{nm\text{-пред}}^1; W_{nm\text{-пред}}^2; \dots; W_{nm\text{-пред}}^{Y_{nm\text{-пред}}}\},$$

где $y_2 = 1, 2, \dots, Y_{nm\text{-пред}}$.

10. В ходе деятельности организационной системы, если фактический показатель ΔS_{nm}^* эф-

фективности n -го вида деятельности организационной системы в m -м подразделении не меньше соответствующего допустимого показателя $\Delta S_{nm\text{-доп}}$ и меньше единицы:

$$\Delta S_{nm\text{-доп}} \leq \Delta S_{nm}^* < 1,$$

то из множества $W_{nm\text{-план}}$ плановых сценариев определяют и исполняют команды $W_{nm\text{-план}}^{y_3}$, предназначенные для повышения эффективности n -го вида деятельности организационной системы в m -м подразделении:

$$W_{nm\text{-план}} = \{W_{nm\text{-план}}^1; W_{nm\text{-план}}^2; \dots; W_{nm\text{-план}}^{Y_{nm\text{-план}}}\},$$

где $y_3 = 1, 2, \dots, Y_{nm\text{-план}}$.

Заключение

1. Представленная в статье система показателей обеспечивает выполнение оценки эффективности технологии поддержки деятельности организационной системы на трех уровнях:
 - (1) при управлении деятельностью организационной системы в целом;
 - (2) при управлении видами деятельности организационной системы;
 - (3) при управлении деятельностью подразделений организационной системы.
2. Аналитическое обеспечение системы показателей эффективности деятельности учитывает влияние, оказываемое на деятельность организационной системы различных объектов,

входящих как в состав организационной системы, так и в состав других систем.

3. Предложена логика взаимосвязи нормированных и фактических показателей эффективности с критериями оценки, которая предоставляет возможность лицам, принимающим решения в отношении ситуаций, возникающих в среде деятельности организационной системы и во внешней среде, осуществлять обоснованный выбор сценария принятия решения (в диапазоне критических, предупреждающих или плановых сценариев).

Литература

1. Информатика: состояние, проблемы, перспективы / Под ред. И. А. Соколова. — М.: ИПИ РАН, 2009. 46 с.
2. Зацаринный А. А., Шабанов А. П. Системные аспекты эффективности ситуационных центров // Вестник Моск. ун-та им. С. Ю. Витте. Сер. 1: Экономика и управление, 2013. Вып. 2. С. 110–123.
3. Цымблер М. Big Data: несколько простых вопросов о сложном явлении // Суперкомпьютеры, 2014. № 1(17). С. 8–11. http://mzym.susu.ru/papers/Zymbler_Supercomputers-14a.pdf.
4. Зацаринный А. А., Сучков А. П., Шабанов А. П. Технология поддержки деятельности организационной системы. Заявка на изобретение RU2012148411A., Оpubл. 20.05.2014. Бюл. № 14.
5. Голяндин А. Н., Шабанов А. П. Центр мониторинга устойчивости информационных систем. Полезная модель. Патент RU130109U1, G06F21/50, 01.2013. Оpubл. 10.07.2013. Бюл. № 19.

Поступила в редакцию 31.07.14

ANALYTICAL ASPECTS OF EVALUATION OF EFFECTIVENESS OF TECHNOLOGICAL SUPPORT OF AN ORGANIZATIONAL SYSTEM

A. A. Zatsarinnyy and A. P. Shabanov

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article discusses the system consisting of normalized and actual indicators, which is used in technological support of activities of an organizational system. The article presents the evaluation criteria that are provided by the classification approach to the choice of scenario for decision for responsible persons for different situations arising in an environment of activities of an organizational system. Classification is performed according to the hierarchy of activities of an organizational system: (a) activities in the divisions of the organizational system; (b) activities in the organizational system; and (c) all activities of the organizational system in general. The main emphasis of the methodical approach presented in the article is practical. This allows one to take the degree

of threat to activities of an organizational system. The authors distinguish three degrees of state activities: (i) activities are carried out in a range of specified normalized indicators; (ii) *a priori*, there is a threat to continuation of activities of an organizational system; and (iii) activities of an organizational system are under threat.

Keywords: organizational system; technological support of activities of an organizational system; evaluation of the effectiveness; normalized and actual indicators; information; decision making; monitoring objects

DOI: 10.14375/19922264140314

References

1. Sokolov, I. A., ed. 2009. *Informatika: Sostoyanie, problema, perspektivy* [Informatics: State, problems, prospects]. Moscow: IPI RAN. 46 p.
2. Zatsarinnyy, A. A., and A. P. Shabanov. 2013. Sistemnye aspekty effektivnosti situatsionnykh tsentrov [System aspects of effectiveness of situation centers]. *Vestnik Mosk. un-ta im. S. Yu. Vitte. Ser. 1: Ekonomika i upravlenie* [Gerald of Moscow University named after S.J. Witte. Ser. 1: Economy and Management] 2:110–123.
3. Tsymbler, M. 2014. Big Data: neskol'ko prostykh voprosov o slozhnom yavlenii [Big Data: A few simple questions about complex phenomenon]. *Superkomp'yutery* [Supercomputers] 1(17):8–11. Available at: http://mzym.susu.ru/papers/Zymbler_Supercomputers-14a.pdf (accessed August 4, 2014).
4. Zatsarinnyy, A. A., A. P. Suchkov, and A. P. Shabanov. 2014. Tekhnologiya podderzhki deyatel'nosti organizatsionnoy sistemy [Technology of support of organizational system]. Zayavka na izobretenie [Application for invention] RU2012148411A. Published 20.05.2014.
5. Golyandin, A. N., and A. P. Shabanov. January 2013. Tsentri monitoringa ustoychivosti informatsionnykh sistem [Center for monitoring the sustainability of information systems]. Poleznaya model' [Utility model]. Patent RU130109U1, G06F21/50. Published 10.07.2013.

Received July 31, 2014

Contributors

Zatsarinnyy Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.ru

Shabanov Alexander P. (b. 1949) — Doctor of Science in technology, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AShabanov@ibs.ru

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент; ведущий научный сотрудник Института проблем информатики Российской академии наук

Архипов Олег Петрович (р. 1948) — кандидат технических наук, директор Орловского филиала Института проблем информатики Российской академии наук

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, ведущий научный сотрудник Института проблем информатики Российской академии наук; профессор факультета прикладной математики и физики Московского авиационного института (национального исследовательского университета)

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора Института проблем информатики Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Российской академии наук

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой Вологодского государственного университета; старший научный сотрудник Института проблем информатики Российской академии наук; главный научный сотрудник Института социально-экономического развития территорий Российской академии наук

Зыкова Зоя Петровна (р. 1953) — кандидат физико-математических наук, заведующая лабораторией Орловского филиала Института проблем информатики Российской академии наук

Киселева Ксения Михайловна (р. 1992) — аспирантка Вологодского государственного университета

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета; ведущий научный сотрудник

Института проблем информатики Российской академии наук

Коротышева Анна Владимировна (р. 1988) — старший преподаватель Вологодского государственного университета

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Российской академии наук

Кульберг Николай Сергеевич (р. 1970) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Леонтьев Николай Дмитриевич (р. 1988) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Маньяков Юрий Анатольевич (р. 1984) — кандидат технических наук, научный сотрудник Орловского филиала Института проблем информатики Российской академии наук

Мейханаджян Лусине Акобовна (р. 1990) — аспирантка Российского университета дружбы народов

Милованова Татьяна Александровна (р. 1977) — кандидат физико-математических наук, старший преподаватель Российского университета дружбы народов

Минин Владимир Александрович (р. 1941) — доктор физико-математических наук, советник РФФИ

Печинкин Александр Владимирович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник ИПИ РАН

Плеханов Леонид Петрович (р. 1943) — кандидат технических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН, доцент Российского университета дружбы народов

Синицын Владимир Игоревич (р. 1968) — доктор физико-математических наук, доцент, заведующий отделом Института проблем информатики Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Российской академии наук

Соколов Игорь Анатольевич (р. 1954) — академик Российской академии наук, доктор технических наук, директор Института проблем информатики Российской академии наук

Ушаков Владимир Георгиевич (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Российской академии наук

Хавансков Валерий Александрович (р. 1950) — научный сотрудник Института проблем информатики Российской академии наук

Шабанов Александр Петрович (р. 1949) — доктор технических наук, ведущий научный сотрудник Института проблем информатики Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора Института проблем информатики Российской академии наук

Шубников Сергей Константинович (р. 1955) — старший научный сотрудник Института проблем информатики Российской академии наук

Яковлева Татьяна Викторовна (р. 1956) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

ЮБИЛЕИ

К 70-летию заместителя директора ИПИ РАН,
члена редколлегии журнала «Информатика и её применения»
доктора технических наук В. И. Будзко



14 августа 2014 г. исполнилось 70 лет заместителю директора ИПИ РАН по научной работе доктору технических наук Владимиру Игоревичу Будзко.

Владимир Игоревич Будзко родился в г. Москве. Высшее образование получил на факультете электронно-вычислительных устройств в Московском инженерно-физическом институте (МИФИ), который он окончил в 1968 г., после чего был направлен для прохождения службы в одну из войсковых частей, где прошел путь от инженера до первого заместителя командира войсковой части.

С приходом В. И. Будзко в ИПИ РАН (2001 г.) в институте сформировалось новое научное направление теоретических исследований — «Построение информационно-телекоммуникационных систем высокой доступности». В рамках этого направления выполнен широкий круг фундаментальных исследований по поиску подходов и определению принципов построения средств обеспечения доступности, конфиденциальности и целостности современных крупномасштабных

информационно-телекоммуникационных систем (ИТС). Разработаны основные системно-технические принципы и базовые архитектурные решения построения перспективных для условий России ИТС с централизованной обработкой и хранением информации, сочетающих в себе свойства высокой доступности, отказо- и катастрофоустойчивости, информационной защищенности. Определены принципы, методы и математические основы рационального построения и оптимизации средств восстановления функционирования центров обработки данных (ЦОД) после возникновения отказов и катастроф, передачи и хранения данных, обеспечения информационной безопасности при достижении минимальной совокупной стоимости владения такими системами. Результаты нашли практическое воплощение при реализации проектов в интересах ряда отечественных государственных и негосударственных организаций, таких как Банк России (БР), Внешторгбанк, ОАО «ГМК «Норильский Никель», «Газпром», Минэкономразвития России, Правительство Москвы, а также ряд силовых ведомств.

Под руководством В. И. Будзко начиная с 2001 г. выполнен комплекс научно-исследовательских и опытно-конструкторских работ (свыше 100 проектов), направленных на развитие электронной информационной технологии БР. Разработаны концепции развития ИТС БР сначала до 2008 г., а затем до 2013 г., которые были приняты в качестве основы проведения технической политики. За реализацию проекта «Катастрофоустойчивая территориально-распределенная информационно-телекоммуникационная система централизованной обработки банковской информации» В. И. Будзко удостоен Премии Правительства РФ в области науки и техники за 2010 г.

В. И. Будзко возглавлял и возглавляет работы по ряду других прикладных проектов, связанных с созданием, совершенствованием и развитием крупномасштабных ИТС.

В. И. Будзко — генерал-майор, доктор технических наук, член-корреспондент Академии криптографии РФ, известный ученый в области информатики и применения информационных технологий при построении территориально распределенных ИТС различного назначения. Является автором свыше 250 научных работ, опубликованных в научно-технических и специальных изданиях.

В. И. Будзко уделяет большое внимание подготовке научных кадров. Под его руководством защищено 6 диссертаций на соискание ученой степени кандидата технических наук. Свыше 30 лет

он читает лекции в ИКСИ Академии ФСБ, профессор кафедры НИЯУ МИФИ. Является членом двух диссертационных советов, главным редактором журнала «Системы высокой доступности» и членом редколлегии журнала «Информатика и её применения».

Редакционный совет и Редакционная коллегия журнала «Информатика и её применения» сердечно поздравляют Владимира Игоревича Будзко с 70-летием и желают крепкого здоровья и новых научных достижений.

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 не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной 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 20 страниц указанного формата.

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.

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://translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты. . . » следует выбрать опцию BGN.

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 hozyaistvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’—kompozit” [Experimental study of the strength of joints “steel—composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Латыпов, А. Р., М. М. Хасанов, и В. А. Байков. 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 Publ. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ 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 publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

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.

9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.ru>, option BGN).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Article from the continuing publication (collection of works, proceedings):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy "stal'-kompozit" [Experimental study of the strength of joints "steel-composite"]. *Trudy MGTU "Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem"* [Bauman MSTU "Mathematical Modeling of Complex Technical Systems" Proceedings]. 593:125–130.

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publ. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author's name in the "Subject" field;
- an article and additional materials have to be attached using the "attach" function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. "Informatics and Applications" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

IPI RAN, Vavilova Str., 44, block 2, 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>