

Информатика и её применения

Том 10 Выпуск 2 Год 2016

СОДЕРЖАНИЕ

Моделирование и мониторинг состояния VoIP-соединения А. В. Борисов, А. В. Босов, Г. Б. Миллер	2
Нормальные условно-оптимальные фильтры и экстраполяторы Пугачёва для стохастических систем, линейных относительно состояния И. Н. Сеницын, Э. Р. Корепанов	14
Эллипсоидальные субоптимальные фильтры для нелинейных стохастических систем на многообразиях И. Н. Сеницын, В. И. Сеницын, Э. Р. Корепанов	24
Метрическая классификация временных рядов со взвешенным выравниванием относительно центроидов классов А. В. Гончаров, В. В. Стрижов	36
Метрическое обучение в задачах мультиклассовой классификации временных рядов Р. В. Исаченко, В. В. Стрижов	48
Распознавание зависимостей на основе обратного отображения А. Н. Тырсин, С. М. Серебрянский	58
Статистические свойства метода подавления шума, основанного на стабилизированной жесткой пороговой обработке О. В. Шестаков	65
Об оптимальном пороговом значении длины очереди в одной задаче максимизации дохода системы массового обслуживания типа $M/G/1$ Я. М. Агаларов, М. Я. Агаларов, В. С. Шоргин	70
Байесовская рекуррентная модель роста надежности: параболическое распределение параметров А. А. Кудрявцев, С. И. Палионная	80
Разложение типа Корниша–Фишера для распределений статистик, построенных по выборкам случайного размера А. С. Марков, М. М. Монахов, В. В. Ульянов	84
Система обслуживания с гиперэкспоненциальным входящим потоком и профилактиками прибора В. Г. Ушаков	92
Многомерное дробное движение Леви и его приложения Ю. С. Хохлов	98
Интенсивность цитирования научных публикаций в изобретениях по информационно-компьютерным технологиям, патентуемых в России отечественными и зарубежными заявителями В. А. Минин, И. М. Зацман, В. А. Хавансков, С. К. Шубников	107
Стационарные вероятности состояний в системе обслуживания конечной емкости с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом Л. А. Мейханаджян	123
Об авторах	132
Правила подготовки рукописей	134
Requirements for manuscripts	137

МОДЕЛИРОВАНИЕ И МОНИТОРИНГ СОСТОЯНИЯ VoIP-СОЕДИНЕНИЯ*

А. В. Борисов¹, А. В. Босов², Г. Б. Миллер³

Аннотация: В результате анализа транспортного протокола Real Time Transfer Protocol (RTP), используемого технологиями Voice over IP (VoIP) для передачи аудио- и видеoinформации, предложена математическая модель описания состояния VoIP-соединения. Она учитывает как формальные правила, лежащие в основе технологии VoIP, так и особенности сетевого взаимодействия, характерные для RTP: задержки, потери пакетов и пр. Математическое описание исследуемого процесса сетевого взаимодействия основано на аппарате скрытых марковских моделей. Предполагается, что ненаблюдаемое состояние соединения описывается марковским процессом с конечным множеством состояний, а наблюдению доступен поток принимаемых кадров, являющийся немарковским мультивариантным точечным процессом. Для предложенной системы наблюдения решена задача фильтрации состояния VoIP-соединения по имеющимся наблюдениям. Достоверность предложенной модели и работоспособность алгоритма фильтрации иллюстрируются примером обработки реального видеопотока, формируемого сервисом Linphone VoIP и передаваемого по сотовой сети стандарта 3G.

Ключевые слова: технологии VoIP; протокол RTP; сетевое соединение; скрытая марковская модель; мультивариантный точечный процесс; оптимальная фильтрация

DOI: 10.14357/19922264160201

1 Введение

Разработка и верификация математических моделей функционирования телекоммуникационных систем является одним из интенсивно развивающихся направлений научных исследований в последние полвека. Формальное описание случайных процессов, возникающих при работе телекоммуникационных и информационно-вычислительных устройств, обеспечивающих сетевой обмен по каналам связи под управлением различных сетевых протоколов, в настоящее время достаточно востребовано. Такие модели, с одной стороны, являются фундаментом для дальнейших исследований, более детального анализа и, в конечном итоге, реализации оптимизированных алгоритмов передачи и обработки информации в составе аппаратного или программного обеспечения телекоммуникационных средств. С другой стороны, разработка и совершенствование математических моделей стимулируются появлением новых средств связи и транспортных протоколов.

Качество моделей традиционно оценивается их способностью отражать природу описываемого ка-

нала связи, циркулирующих в нем потоков данных, в частности нестационарный характер процессов приема/передачи информационных пакетов, их задержки, перемешивание, потери и т. п.

По-видимому, пионерскими в данной области следует считать ставшие классическими модели [1, 2]. Изначально они были созданы для описания битовых потерь в телефонных сетях, но впоследствии достаточно успешно применялись для описания потерь пакетов в компьютерных сетях, управляемых протоколами стека TCP/IP. Будучи предельно простыми, они, тем не менее, хорошо описывали ряд принципиальных особенностей, свойственных сетям TCP/IP.

Для учета тонкой специфики сетевого взаимодействия предпринимались попытки детально описывать структуру и протоколы реальных систем и сетей пакетной передачи данных. Аппаратом, применяемым в таких исследованиях, как правило, выступали цепи Маркова с конечным множеством состояний [3–7].

С ростом числа возможных состояний, а также при исследовании процессов в условиях предельных сетевых нагрузок обоснованное применение

* Исследование частично поддержано РФФИ (проекты 16-07-00677 и 15-37-20611-мол_а_вед).

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

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

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

нашли жидкостные и диффузионные аппроксимации [8–11]. Попытки частично или полностью отказаться от марковского свойства моделей привели к использованию самоподобных процессов [12–14] и скрытых марковских моделей (СММ) [15–18].

Накопленный к настоящему времени опыт моделирования подкрепляет очевидный тезис о том, что успех в разработке математической модели достигается при компромиссе между полнотой учета технических деталей, разумными упрощениями, возможностями применяемого математического аппарата и устойчивостью к неточностям параметров модели.

В настоящее время существуют несколько стандартов, описывающих наборы протоколов, на базе которых разрабатываются и функционируют сервисы IP-телефонии. К наиболее распространенным можно отнести протокол установления сеанса SIP (Session Initiation Protocol), посредством которого осуществляется обмен служебной информацией, связанной с регистрацией и установкой сеансов связи между клиентами. Согласование медиаформатов и других параметров сессии после установления соединения между клиентами происходит с использованием протокола описания сессии SDP (Session Description Protocol).

Другой популярный VoIP-стандарт H.232 описывает отличный от SIP набор служебных протоколов, тем не менее наиболее часто используемым протоколом для непосредственной передачи аудио- и видеопотоков в IP-телефонии обоих стандартов является RTP (Real-Time Transport Protocol). В силу того что наибольший объем данных передается по этому протоколу, именно его естественно использовать для оперативной оценки состояния канала связи между клиентами на прикладном уровне сетевой модели.

Таким образом, объектом исследования в данной статье является RTP — протокол прикладного уровня, используемый для передачи аудио- и видеоданных между VoIP-клиентами. Для данного протокола предлагается новая математическая модель, учитывающая ряд его специфических черт, характерных именно для технологий VoIP.

Во-первых, RTP является протоколом, не гарантирующим доставку: в нем отсутствует механизм квитирования, т. е. поток передаваемых пакетов предполагает случайные потери. Кроме того, RTP допускает временное взрывное нарастание принимаемых пакетов (так называемое *bursting reception*) и нарушение порядка их приема. При этом если нарушение порядка пакетов выявляется непосредственно путем простого сравнения номеров пакетов, то определение факта их потери возможно только через определенный временной лаг, необ-

ходимый для того, чтобы убедиться в том, что пакет потерян вовсе, а не задерживается.

Во-вторых, на принимающей стороне существует возможность регистрации моментов приема пакетов. Эта временная последовательность является важным источником информации для определения характеристик задержек пакетов и последующей оценки состояния VoIP-соединения. При этом ясно, что это состояние недоступно прямым наблюдениям, поскольку определяется нагрузкой на программное обеспечение и средства телекоммуникаций, формирующие и обслуживающие канал связи, допустимыми размерами буферов в каналообразующей аппаратуре и т. п. Характеристики задержек пакетов позволяют судить о состоянии соединения лишь косвенно. Наконец, анализ реальных данных показывает, что время ожидания очередного RTP-пакета не является экспоненциально распределенной случайной величиной вне зависимости от состояния соединения, т. е. поток пакетов принимаемых RTP-пакетов не является марковским.

Перечисленные особенности предлагается включить в формальное описание с помощью аппарата СММ. Предполагается, что состояние VoIP-соединения описывается ненаблюдаемым («скрытым») марковским процессом с конечным множеством состояний, а наблюдению доступна последовательность пар «момент получения пакета — заголовок RTP-пакета». Эти заголовки позволяют правильно сформировать кадр, требующийся клиенту сервиса VoIP. Однако непосредственное использование такой информационной модели для решения последующих задач оценивания и оптимизации состояния соединения оказывается неэффективным в силу того, что эти «сырые» данные избыточны для решения вышеупомянутых задач. Тем не менее к ним удастся применить некоторую процедуру предобработки с фиксированным малым запаздыванием для извлечения значимой информации о состоянии соединения. Эта процедура опирается на наличие в видеопотоке неделимого элемента — отдельного кадра. Данное обстоятельство делает естественной агрегацию данных о моментах получения отдельных пакетов в последовательность моментов получения кадров. Дополнительно каждый кадр сопровождается информацией о его статусе, характеризующей качество кадра. Эта информация также является результатом агрегации информации из заголовков пакетов кадра. Таким образом, в качестве наблюдений в предлагаемой модели используется результирующая последовательность пар «момент получения кадра — статус кадра». Для математического описания этой последовательности предлагается использовать мультивариантные точечные процессы (МТП) [19].

Таким образом, предлагаемая модель сетевого взаимодействия относится к классу стохастических динамических систем наблюдения, для решения задач оценивания и оптимизации в которых может быть применен хорошо развитый аппарат стохастического анализа.

Статья организована следующим образом. Раздел 2 содержит неформальное описание процесса получения потока RTP-пакетов и аргументацию в пользу упомянутой процедуры предобработки. В результате предлагается математическая модель соединения в форме стохастической дифференциальной системы наблюдений.

В разд. 3 практическая задача онлайн-мониторинга состояния VoIP-соединения сформулирована в терминах оптимальной фильтрации состояния системы наблюдения. Ее решение может оказаться полезным, например, в качестве вспомогательной процедуры в задаче идентификации параметров модели (например, в EM-алгоритме (expectation-maximization algorithm) [20]) или для целей управления по неполной информации. Общее решение задачи фильтрации, включая подробное доказательство, приведено в [21], обсуждение модели и применение алгоритма фильтрации в сетях массового обслуживания — в [22].

Раздел 4 иллюстрирует представленные результаты. Качество предложенной модели и алгоритма оценивания продемонстрировано на примере обработки реальных статистических данных, полученных в результате работы сервиса Linphone VoIP по сети сотовой связи 3G. Выводы и заключительные замечания приведены в разд. 5.

2 Неформальное описание и математическая модель

Процесс получения кадров, передаваемых с помощью потока RTP-пакетов, предлагается описывать следующей стохастической дифференциальной системой наблюдения. Ее первый компонент определяет ненаблюдаемое состояние VoIP-соединения. Его значение зависит от множества случайных факторов: сформировавшейся топологии обеспечивающей физической сети, состояния и характеристик телекоммуникационного оборудования, состояния встроенного программного обеспечения и свойств реализуемых им алгоритмов передачи данных. Наиболее важным представляется фактор разделения сетевых ресурсов. Телекоммуникационное оборудование обслуживает одновременно большое число пользователей, поэтому общая текущая нагрузка на канал существенно влияет

на состояние конкретного соединения и служит дополнительным источником неопределенности.

Очевидно, что все упомянутые параметры и факторы, определяющие состояние соединения, недоступны прямым наблюдениям, поэтому его можно считать ненаблюдаемым («скрытым») процессом. В то же время значительное число независимых пользователей, постоянно запрашивая сетевые ресурсы, обеспечивают отсутствие последствий для этого процесса, так что предположение о наличии у состояния соединения марковского свойства представляется вполне реалистичным. В рассматриваемой модели будем предполагать, что состояние соединения описывается однородным марковским процессом X_t с начальным распределением p_0 , матрицей интенсивностей Λ и тремя возможными состояниями $\{e_1, e_2, e_3\} \triangleq \mathbb{S}^3$. Обозначение e_j здесь и далее применяется для j -го единичного вектора в евклидовом пространстве \mathbb{R}^3 . Значение e_1 используется для обозначения состояния соединения «свободно», e_2 — для состояния «умеренная нагрузка» и e_3 — для состояния «перегрузка и/или сбой соединения». Использование единичных векторов для формального обозначения значений марковского процесса обеспечивает возможность представить X_t в виде решения стохастического уравнения:

$$X_t = \xi + \int_0^t \Lambda^\top X_s ds + M_t^X, \quad (1)$$

где M_t^X — мартингал; ξ — случайный вектор с распределением p_0 [20]. Параметры (p_0, Λ) далее предполагаются известными или доступными идентификации по имеющимся статистическим данным (как это сделано, например, в разд. 4).

На принимающей стороне наблюдению доступны пары вида «момент получения пакета — заголовок RTP-пакета». Моменты получения пакетов могут фиксироваться клиентским программным обеспечением VoIP или сетевыми мониторами (см., например, [23, 24]). Формат заголовков пакетов (см., например, [25]) накладывает ряд важных ограничений на процесс передачи данных. Во-первых, факты потерь RTP-пакетов на приемнике не могут быть зарегистрированы непосредственно. Обнаружение потери пакета обеспечивается наличием в его заголовке информации, позволяющей определить, к какому кадру относятся полученные в пакете данные и какой по порядку пакет из кадра получен. Во-вторых, поток RTP-пакетов не является ординарным: большинство VoIP-серверов присваивают общую временную метку (timestamp) всем пакетам, образующим кадр, и отправляют их одновременно. На приемнике это свойство в целом также сохра-

няется: обычно пакеты, отправленные одновременно, регистрируются на стороне клиента также одновременно. В-третьих, отправитель использует некоторые низкоуровневые аппаратные протоколы для получения дополнительной информации о качестве физического соединения или наличии сбоев для последующей адаптации процедуры отправки. При отсутствии связи или ее неудовлетворительном качестве RTP-пакеты аккумулируются в серверном буфере, с тем чтобы быть отправленными, как только качество соединения улучшится. В таких ситуациях целые серии пакетов могут появляться на принимающей стороне одновременно и при этом может нарушаться их порядок.

Сказанное позволяет сделать вывод о том, что характер сетевого обмена, задающий в конечном итоге качество воспроизведения, определяется равномерностью доставки кадров и их качеством. Таким образом, изначально данные заголовков принимаемых пакетов оказываются избыточными для последующего оценивания состояния соединения, и требуется их предобработка для извлечения информации о событиях, связанных с кадрами. Результатом этой процедуры является МТП, т.е. последовательность пар $\{(\tau(n), Y(n))\}_{n \in \mathbb{N}}$, в которой $\tau(n)$ является случайным временем n -го наблюдаемого события, $Y(n)$ — типом события. «Алфавит» (множество возможных событий) задается множеством $\{f_1, f_2, f_3\} \triangleq \mathbb{S}^3$. Обозначение f_k здесь и далее используется для k -го единичного вектора евклидова пространства \mathbb{R}^3 . Значение f_1 присвоено событию «получен кадр большого размера», f_2 — событию «получен кадр малого размера», f_3 — событию «потери в составе кадра».

События первых двух типов регистрируются естественным образом в реальном масштабе времени при поступлении последнего пакета кадра. Наличие потерь в конкретном кадре фиксируется в том случае, если хотя бы один пакет, составляющий кадр, еще не получен, но уже получен пакет-маркер следующего кадра (в RTP для каждого кадра заголовок его последнего пакета маркируется битовым признаком). Тем не менее время регистрации события третьего типа определяется как момент получения последнего пакета, составляющего кадр с потерянными пакетами.

Наблюдения в форме МТП $\{(\tau(n), Y(n))\}_{n \in \mathbb{N}}$ эквивалентны следующему случайному процессу с непрерывным временем:

$$Y_t \triangleq \sum_{n \in \mathbb{N}} Y(n) \mathbf{I}(t - \tau(n)).$$

Здесь $\mathbf{I}(\cdot)$ — функция Хевисайда, непрерывная справа. Компоненты трехмерного процесса Y_t —

считающие: каждый компонент определяет общее число событий соответствующего типа, произошедших на интервале $[0, t]$.

Предположим, что известна условная плотность вероятности $\pi_{jkl}(\cdot)$:

$$\begin{aligned} \mathbf{P} \{ \tau(n) \leq t, Y(n) = f_j | \tau(n-1) = s, \\ Y(n-1) = f_k, X_u \equiv e_\ell, u \in [s, t] \} = \\ = \mathbf{I}(t-s) \int_s^t \pi_{jkl}(u-s) du, \end{aligned} \quad (2)$$

в противном случае ее предполагается предварительно идентифицировать (см. разд. 4).

Можно показать, что процесс Y_t допускает следующее представление:

$$Y_t = \int_0^t \phi(\omega, u) X_{u-} du + M_t^Y, \quad (3)$$

где M_t^Y — мартингал, а $\phi(\omega, u) : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^{3 \times 3}$ — предсказуемая матричнозначная функция мгновенных интенсивностей. Элементы $\phi_{j\ell}(\cdot)$ обладают ясной вероятностной интерпретацией: для любого $n \in \mathbb{N}$

$$\begin{aligned} \phi_{j\ell}(\omega, t) dt = \mathbf{P} \left\{ \tau(n) \in [t, t+dt), Y(n) = f_j | \right. \\ \left. \{(\tau(i), Y(i))\}_{i=1}^{n-1} \vee \{X_{t-} = e_\ell\} \vee \{\tau(n) \geq s\} \right\}. \end{aligned}$$

Матрица интенсивностей может быть вычислена с помощью условной плотности вероятности $\pi_{jkl}(\cdot)$ (здесь и далее предполагается, что $\tau_0 \triangleq 0$):

$$\phi(\omega, t) = \sum_{n \in \mathbb{N}} \mathbf{I}_{(\tau(n-1), \tau(n))}(t) \varphi(\tau(n-1), Y(n-1), t),$$

где $\varphi(\cdot) = \|\varphi_{j\ell}(\cdot)\|$ — матрично-значная функция соответствующей размерности с элементами

$$\begin{aligned} \varphi_{j\ell}(\tau(n-1), Y(n-1), t) \triangleq \\ \triangleq \sum_{k=1}^3 f_k^\top Y(n-1) \pi_{jkl}(t - \tau(n-1)) \times \\ \times \left(\sum_{i=1}^3 \int_t^{+\infty} \pi_{ikl}(u - \tau(n-1)) du \right)^+. \end{aligned}$$

Таким образом, уравнения (1) и (3) образуют исковую математическую модель функционирования VoIP-соединения. Подтверждению ее адекватности для описания функционирования сервиса Linphone VoIP посвящен разд. 4. В следующем разделе приведены постановка и решение задачи онлайн-оценивания (фильтрации) состояния рассматриваемой системы наблюдения.

3 Задача онлайн-оценивания состояния соединения

Предложенная математическая модель позволяет решить одну из важнейших практических задач мониторинга состояния VoIP-соединения по доступным наблюдениям. В терминах теории стохастических динамических систем эта задача эквивалентна построению условного распределения состояния относительно всех доступных на данный момент наблюдений. Помимо собственной практической и теоретической значимости представление условного распределения играет важную вспомогательную роль при решении различных задач управления по неполной информации.

Для рассматриваемой модели задача фильтрации состоит в нахождении условного математического ожидания (УМО) состояния VoIP-соединения по наблюдениям, доступным к текущему моменту времени t : $\hat{X}_t \triangleq \mathbf{E} \{X_t | Y_{[0,t]}\}$. Решение этой задачи определяет следующая теорема.

Теорема 1. *Условное математическое ожидание \hat{X}_t определяется как единственное решение следующей стохастической системы уравнений, имеющей единственное сильное решение:*

$$\begin{aligned} \hat{X}_t = & p_0 + \int_0^t \Lambda^\top \hat{X}_{s-} ds + \\ & + \int_0^t \left[\text{diag} \hat{X}_{s-} - \hat{X}_{s-} \hat{X}_{s-}^\top \right] \phi^\top(\omega, s-) \times \\ & \times \left[\text{diag} \left(\phi(\omega, s-) \hat{X}_{s-} \right) \right]^{-1} \times \\ & \times \left(dY_s - \phi(\omega, s-) \hat{X}_{s-} ds \right). \end{aligned} \quad (4)$$

Задача фильтрации семимартингала по наблюдению семимартингала в наиболее общей постановке была решена в [19], формулы частного случая фильтрации МСП по наблюдениям МТП с линейным компенсатором представлены в [21], а их обсуждение применительно к сетям массового обслуживания имеется в [22].

4 Численные эксперименты

Для проверки адекватности представленной модели сетевого взаимодействия и иллюстрации качества оценок фильтрации был проведен численный эксперимент по обработке реальных данных, полученных на клиенте видеоконференции, организованной с помощью VoIP-сервиса Liphone. Циркулирующие в рамках этой конференции пакеты

регистрировались на выбранном узле-получателе. Накопленные таким образом статистические данные использовались для одновременной идентификации параметров модели наблюдения (1), (3) и онлайн-вычисления оценок фильтрации.

Поскольку состояние сетевого соединения вычисляется по статистическим данным после некоторой предварительной их обработки, идентификация модели (определение матрицы интенсивности и условных плотностей $\pi_{jki}(\cdot)$) также выполняется с некоторым запаздыванием по отношению к поступающему потоку данных. Оценки фильтрации, хотя и вычисляются в онлайн режиме по всему накопленному объему наблюдений, строятся на основе модели, параметры которой получены с использованием лишь части наблюдений. Степень запаздывания, которая определяется разницей между количеством наблюдений (пакетов), используемых для идентификации модели и процесса фильтрации, соответствует размеру окна сглаживания. Оценки фильтрации сравниваются с «идеальной моделью», состояние которой вычисляется как результат сглаживания. Полученные качественные характеристики оценок фильтрации служат подтверждением как адекватности модели (1), (3), так и высокой точности самого алгоритма фильтрации (4).

Для проведения расчетов был зарегистрирован отрезок последовательности пакетов, переданных в рамках сессии обмена потоковым видео. Видеопоследовательность передавалась с использованием канала мобильной связи 3G с мобильного устройства на персональный компьютер, подключенный к сети Интернет. На этом компьютере перехватывался весь входящий трафик, из него извлекались пакеты, касающиеся конкретной видеосессии. В качестве средства проведения видеоконференции использовалось приложение Liphone VoIP. Этот широко распространенный кроссплатформенный программный клиент IP-телефонии имеет открытый исходный код и поддерживает версии как для мобильных, так и для настольных платформ. Наиболее важным обстоятельством для проводимых исследований является возможность отключения в данном приложении шифрования RTP-трафика, что обеспечивает прозрачность процессов его обработки. В частности, благодаря отсутствию шифрования удается легко разделять отдельные RTP-сессии и выделять аудио- и видеопоследовательности. В качестве сетевого монитора использовалось приложение Wireshark, выбор которого определялся простотой применения и совместимостью с операционной системой.

На рис. 1 приведены данные, полученные в течение примерно двадцатиминутной видеоконфе-

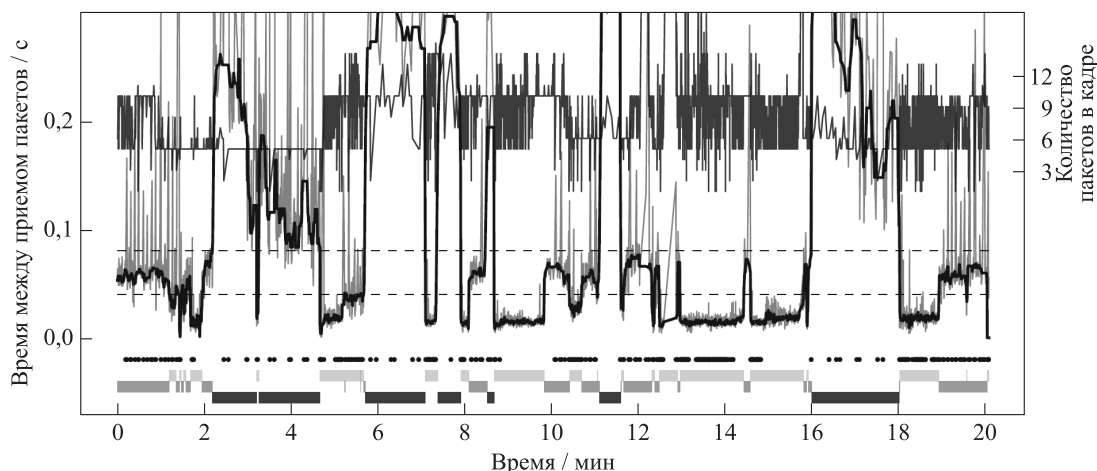


Рис. 1 Данные сессии RTP: тонкая темно-серая линия — размер кадра F_t ; тонкие серая и черная линии — скорость получения пакетов S_t и ее медиана $\text{med}(S_t)$; точки — событие «потери в составе кадра»; сплошные светло-серая, серая и темно-серая линии — состояние соединения e_1 , e_2 или e_3 ; пунктирные черные линии — пороговые значения 0,04 и 0,08

ренции. В качестве индикатора состояния VoIP-соединения использовалась скорость получения пакетов S_t , т. е. среднее время между приемом пакетов, относящихся к одному кадру. Отметим, что, несмотря на осреднение, этот процесс демонстрирует быстрые колебания, вызванные высокочастотной природой цифровых коммуникационных каналов связи. Для извлечения систематической составляющей, определяющей состояние сетевого соединения, траектория этого процесса подверглась сглаживанию путем вычисления скользящей двусторонней медианы $\text{med}(S_t)$. Состояния соединения («идеальная модель») определялись следующим образом:

- e_1 — состояние «свободно», если $\text{med}(S_t) \in [0; 0,04)$;
- e_2 — состояние «умеренная нагрузка», если $\text{med}(S_t) \in [0,04; 0,08)$;
- e_3 — состояние «перегрузка и/или сбой соединения», если $\text{med}(S_t) \in [0,08; +\infty)$.

Формируемые в результате предобработки наблюдения зависят от размера кадра F_t , который равен числу пакетов, составляющих кадр. Значение наблюдения в момент t определяется в соответствии с предложенной моделью, а именно:

- f_1 — событие «получен кадр большого размера», если $F_t \geq 6$;
- f_2 — событие «получен кадр малого размера», если $F_t < 6$;
- f_3 — событие «потери в составе кадра».

При обработке данных было отмечено соответствие, заключающееся в том, что наблюдения f_i появляются чаще, когда соединение находится в состоянии e_i . Дело в том, что применяемая версия программного обеспечения поддержки видеоконференции использует на стороне отправителя низкоуровневый протокол, позволяющий определять временные проблемы с мобильной связью. За счет этого программа при отправке предпринимает определенные действия, призванные предотвратить возможные потери: адаптирует число пакетов в кадре, ухудшая качество, накапливает пакеты в буфере, с тем чтобы отправить их незамедлительно при восстановлении канала связи, и пр.

Далее приведена оценка матрицы интенсивностей, построенная по полному объему собранных данных:

$$\Lambda = \begin{pmatrix} -2,171 & 1,964 & 0,207 \\ 2,442 & -3,161 & 0,719 \\ 1,665 & 0,666 & -2,331 \end{pmatrix} \cdot 10^{-3}.$$

Гистограммы времен пребывания в каждом состоянии показаны на рис. 2. На этом же рисунке для сравнения построены соответствующие экспоненциальные аппроксимации. Визуально экспоненциальная плотность вполне согласуется с реальными данными для первого и второго состояния. Значительная погрешность для третьего состояния может быть объяснена малым объемом выборки. Небольшое несоответствие первой гистограммы и ее экспоненциальной аппроксимации не является неожиданностью: небольшие отклонения на «хвостах» часто встречаются в задачах оценки плотностей.

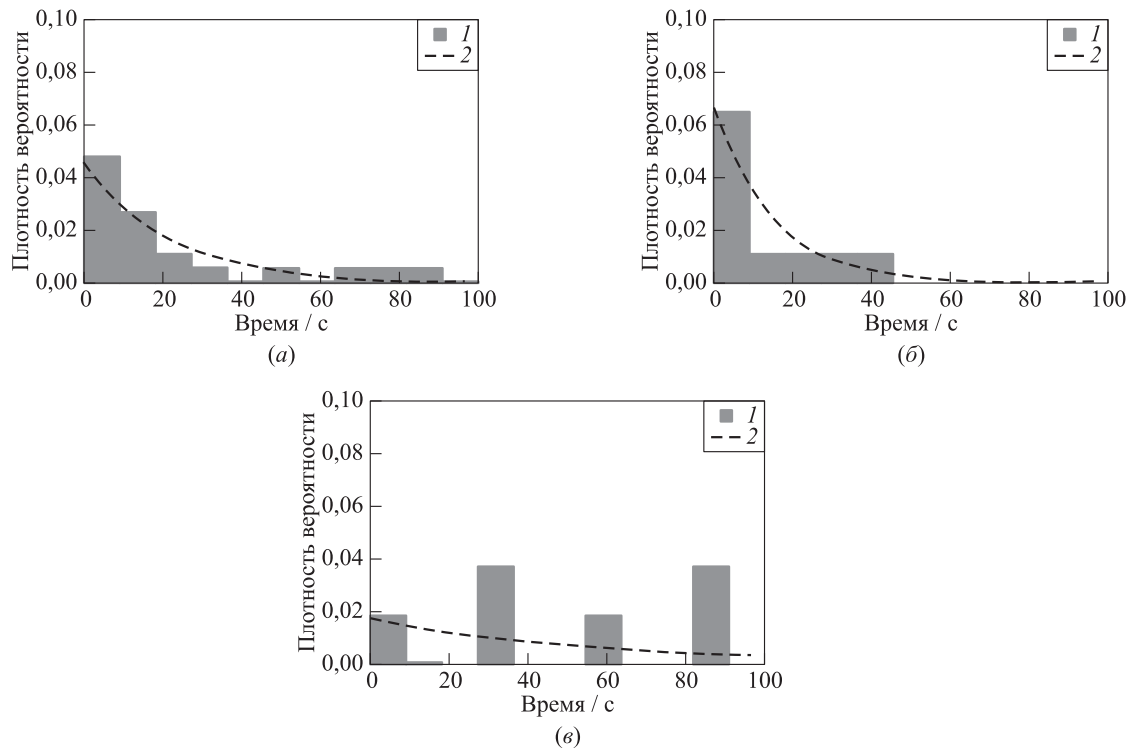


Рис. 2 Гистограммы для времени пребывания в состоянии e_i (1) и соответствующие им экспоненциальные плотности вероятности (2): (а) $i = 1$; (б) 2; (в) $i = 3$

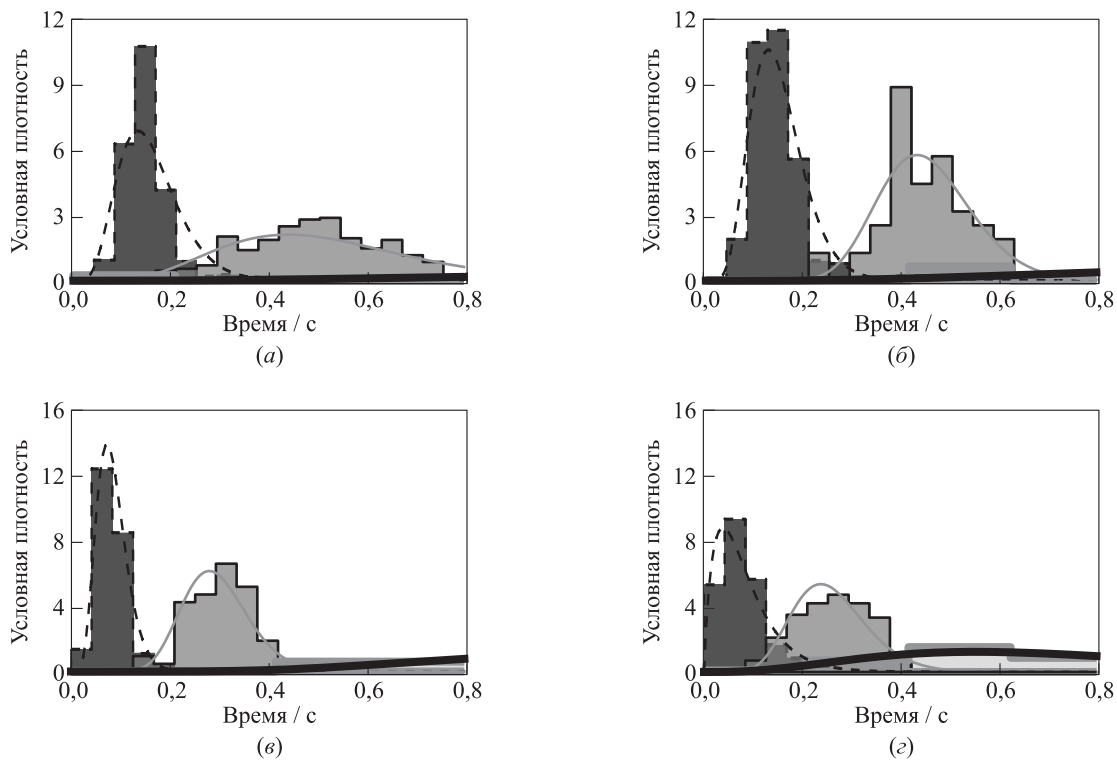


Рис. 3 Оценка условной плотности вероятности $\pi_{jkl}(\cdot)$ (2): (а) $Y_n = f_1, Y_{n-1} = f_1$; (б) $Y_n = f_1, Y_{n-1} = f_2$; (в) $Y_n = f_2, Y_{n-1} = f_1$; (г) $Y_n = f_2, Y_{n-1} = f_2$ для разных значений $X_{n-1} = e_i$; $X_{n-1} = e_1$ — цвет гистограммы темно-серый, оценка условной плотности — пунктирная линия; $X_{n-1} = e_2$ — цвет гистограммы серый, оценка условной плотности — тонкая линия; $X_{n-1} = e_3$ — цвет гистограммы светло-серый, оценка условной плотности — толстая линия

Плотности распределения вероятностей аппроксимировались гамма-распределениями с параметрами, вычисленными по методу максимального правдоподобия. На рис. 3 приведены некоторые гистограммы и соответствующие им оценки плотностей гамма-распределения, построенные по полному объему наблюдений. Все графики сгруппированы по следующим условиям: $Y_n = f_j$ при $Y_{n-1} = f_k$ и $X_{n-1} = e_\ell$ (т.е. j и k фиксированы и $\ell = 1, 2, 3$). Приведенные графики позволяют подтвердить хорошую согласованность оценок гамма-плотностей и соответствующих гистограмм и проиллюстрировать их различие при разных состояниях $X_{n-1} = e_\ell$.

Компоненты оценок фильтрации, являющиеся условными вероятностями событий $X_t = e_\ell$, представлены на рис. 4. Для сравнения там же приведены соответствующие компоненты «идеальной модели». Визуальное сравнение подтверждает высокое качество оценок фильтрации, а расчеты в целом — возможность их выполнения в режиме реального времени, что необходимо для целей онлайн-мониторинга состояния VoIP-соединения.

Качество оценивания, обеспечиваемое представленным алгоритмом фильтрации, сравнивалось с известным аналогом. Если условные плотности $\pi_{jkl}(\cdot)$ положить экспоненциальными, то наблюдаемый процесс Y_t будет представлять собой хорошо известный «маркированный» процесс Кокса, а пара (X_t, Y_t) будет обладать марковским свойством. Решение задачи фильтрации состояний частично наблюдаемых марковских процессов такого вида представлено в [20].

Класс систем наблюдения, рассматриваемых в предлагаемой статье, является более широким и содержит процессы, не обладающие марковским свойством. Алгоритм оптимальной фильтрации из [20] является частным случаем формул (4).

Для сравнения по имеющимся статистическим данным были вычислены оценки фильтрации состояния в предположении, что наблюдаемый процесс Y_t является «маркированным» процессом Кокса. Необходимые параметры экспоненциальных распределений для реализации алгоритма фильтрации вычислялись также с помощью метода максимального правдоподобия по нарастающей выборке

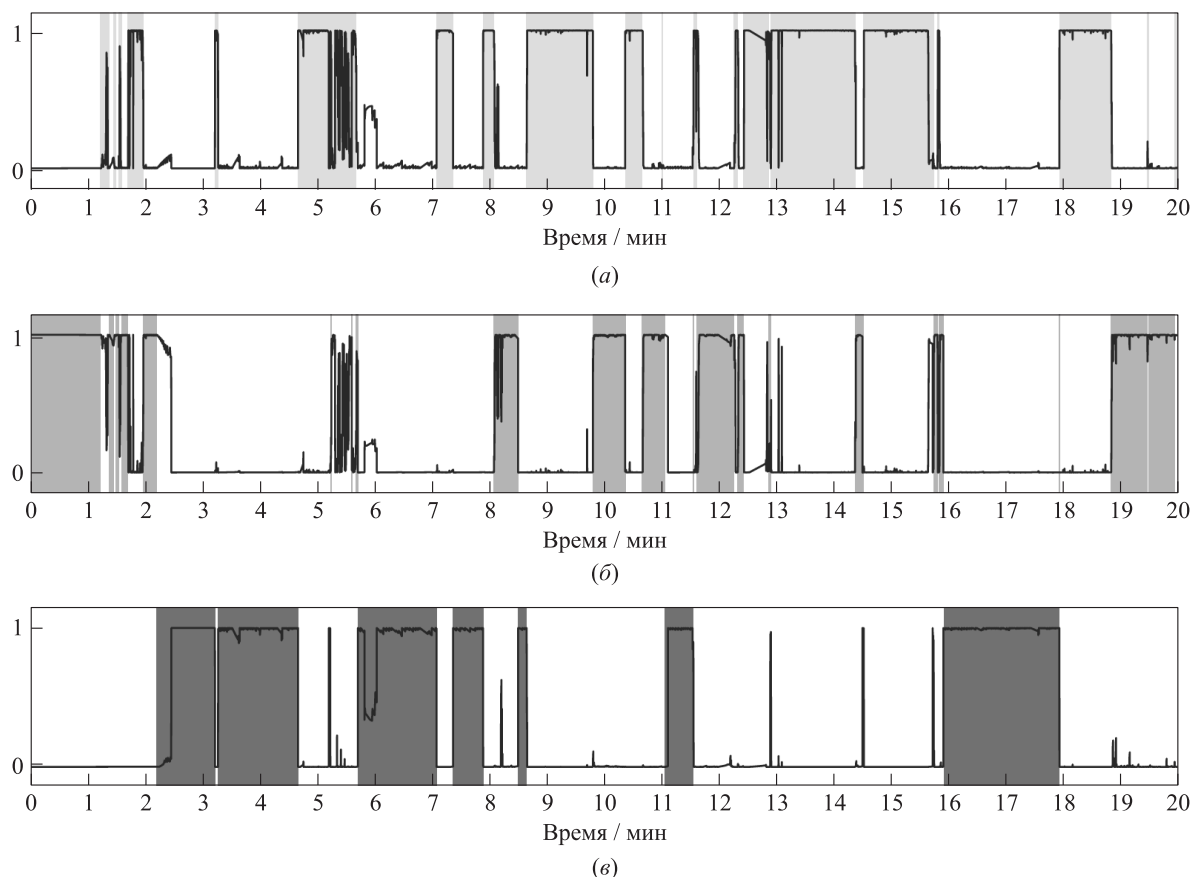


Рис. 4 Оценки фильтрации $P\{X_t = e_j | Y_{[0,t]}\}$ для $j = 1$ (а), 2 (б) и 3 (в) (тонкие черные линии) в сравнении с «идеальной моделью» (различные оттенки серого)

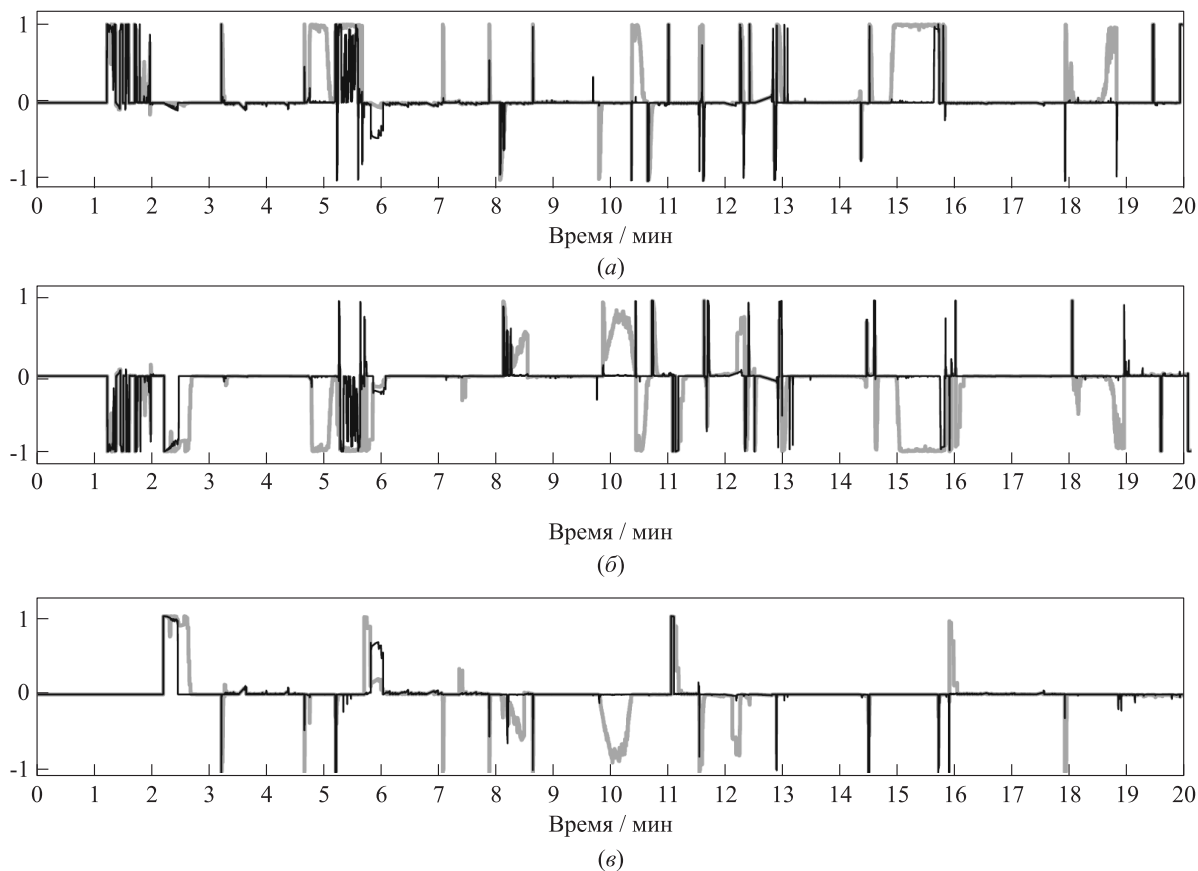


Рис. 5 Ошибки оценивания процесса X_t для $j = 1$ (а), 2 (б) и 3 (в): ошибка предложенного алгоритма фильтрации (черная линия); ошибка алгоритма фильтрации [20] (серая линия)

поступающих наблюдений. Результаты оценивания сравнивались с упомянутой выше «идеальной моделью». На рис. 5 представлены ошибки оценок сравниваемых алгоритмов фильтрации. Из рисунка видно, что на значительных промежутках времени алгоритм из [20] неверно идентифицирует истинное состояние канала: в районе 5-й, 11-й и 16-й минут.

По оценкам фильтрации для обоих алгоритмов были определены наиболее вероятные состояния, которые трактовались как точечные оценки текущего состояния соединения. Эти оценки далее сравнивались с «идеальной моделью». Оценка вероятности ошибки точечной оценки, построенной по предлагаемому фильтру, составила 0,069, в то время как соответствующее значение для фильтра из [20] равно 0,206.

Статистический анализ проведенного численного эксперимента с привлечением реальных телекоммуникационных данных демонстрирует хорошее согласование предлагаемой математической модели функционирования RTP-соединения с полученной измерительной информацией, а также

высокую точность алгоритма онлайн-оценивания состояния этого соединения.

5 Заключение

Основные результаты, представленные в работе, видятся в следующем.

1. Предложено новое феноменологическое описание процесса сетевого взаимодействия, реализуемого по технологии VoIP на базе протокола RTP. Предложена и подробно рассмотрена модель, базирующаяся на концепции СММ. Ненаблюдаемое («скрытое») состояние соединения представлено конечномерным марковским скачкообразным процессом, в то время как наблюдаемое время получения кадров — немарковский МТП. Предложенная модель обладает очевидными достоинствами, одно из которых — возможность применения к ним аппарата стохастического анализа.
2. В составе модели предложена процедура предобработки, преобразующая «сырой» поток па-

кетов в «очищенный» поток кадров, оказавшийся весьма удобным для онлайн-процедуры совместного решения задач идентификации параметров модели и оценивания состояния VoIP-соединения.

3. Предложенные для описания как состояния, так и наблюдений «алфавиты» демонстрируют хорошее соответствие модели реальным данным: соответствующие гистограммы оказываются унимодальными и хорошо аппроксимируются гамма-распределением.

Тем не менее, по мнению авторов, исследование нельзя считать законченным. Предложенная модель нуждается во всесторонней верификации. Далее модель должна быть развита с целью учета доступной информации, касающейся сетевой топологии, характеристик алгоритмов, реализуемых программным и техническим обеспечением, формирующим канал связи и реализующим протоколы информационного обмена, например данными, передаваемыми в рамках служебного протокола RTSP, который также используется на прикладном уровне VoIP. Представляет самостоятельный интерес возможность реализации робастной версии алгоритма фильтрации состояния. Наконец, финальной точкой могла бы стать постановка и решение задачи оптимального и/или робастного управления в предложенной модели по неполной информации.

Литература

1. *Gilbert E. N.* Capacity of a burst-noise channel // *Bell Syst. Tech. J.*, 1960. Vol. 39. P. 1253–1265.
2. *Elliott E. O.* Estimates of error rates for codes on burst-noise channels // *Bell Syst. Tech. J.*, 1963. Vol. 42. P. 1977–1997.
3. *Altman E., Avrachenkov K., Barakat C.* TCP in presence of bursty losses // *Perform. Evaluation*, 2000. Vol. 42. P. 129–147.
4. *Миллер Б. М., Авраченко К. Е., Степанян К. В., Миллер Г. Б.* Задача оптимального стохастического управления потоком данных по неполной информации // *Пробл. передачи информ.*, 2005. Т. 41. № 2. С. 89–110. doi: 10.1007/s11122-005-0020-8.
5. *Bruno R., Conti M., Gregori E.* Throughput analysis and measurements in IEEE 802.11 WLANs with TCP and UDP traffic flows // *IEEE Trans. Mobile Comput.*, 2008. Vol. 7. No. 2. P. 171–186.
6. *Haßlinger G., Hohfeld O.* The Gilbert–Elliott model for packet loss in real time services on the Internet // 14th GI/ITG Conference on Measurement, Modelling and Evaluation of Computer and Communication Systems (MMB) Proceedings. — Dortmund, Germany, 2008. P. 269–283.
7. *Malik M., Aydin M., Shah Z., Hussain S.* Stochastic model of TCP and UDP traffic in IEEE 802.11b/g. // *IEEE 9th Conference on Industrial Electronics and Applications (ICIEA) Proceedings.* — IEEE, 2014. P. 2170–2175.
8. *Whitt W.* Stochastic-process limits. An introduction to stochastic-process limits and their application to queues. — New York, NY, USA: Springer, 2002. 602 p.
9. *Bohacek S.* A stochastic model of TCP and fair video transmission // 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM) Proceedings. — IEEE, 2003. Vol. 2. P. 1134–1144.
10. *Liu Y., Gong W.* On fluid queueing systems with strict priority // *IEEE Trans. Automat. Contr.*, 2003. Vol. 48. No. 12. P. 2079–2088.
11. *Domańska J., Domański A., Czachórski T., Klamka J.* Fluid flow approximation of time-limited TCP/UDP/XCP streams // *Bull. Pol. Acad. Sci. Tech. Sci.*, 2014. Vol. 62. No. 2. P. 217–225.
12. *Leland W. E., Taqqu M. S., Willinger W., Wilson D. V.* On the self-similar nature of Ethernet traffic // *IEEE ACM Trans. Network.*, 1994. Vol. 2. No. 1. P. 1–15.
13. *Crovella M. E., Bestavros A.* Self-similarity in World Wide Web traffic: Evidence and possible causes // *IEEE ACM Trans. Network.*, 1997. Vol. 5. No. 6. P. 835–846.
14. *Tsybakov B., Georganas N.* Overflow and losses in a network queue with a self-similar input // *Queueing Syst.*, 2000. Vol. 35. No. 1–4. P. 201–235.
15. *Yariv E., Merhav N.* Hidden Markov processes // *IEEE Trans. Inform. Theory*, 2002. Vol. 48. No. 6. P. 1518–1569.
16. *Борисов А. В., Миллер Г. Б.* Анализ и фильтрация специальных марковских процессов в дискретном времени. II. Оптимальная фильтрация // *Автоматика и телемеханика*, 2005. № 7. С. 112–125. doi: 10.1007/s10513-005-0153-7.
17. *Anisimov V.* Switching processes in queueing models. — New York, NY, USA: Wiley, 2008. 352 p.
18. *Ellis M., Pezaros D. P., Kypraios T., Perkins C.* Modelling packet loss in RTP-based streaming video for residential users // 37th Conference on Local Computer Networks Proceedings. — New York, NY, USA: IEEE Press, 2012. P. 220–223.
19. *Лунцер П. Ш., Ширяев А. Н.* Теория мартингалов. — М.: Наука, 1986. 512 с.
20. *Elliott R. J., Aggoun L., Moore J. B.* Hidden Markov models: Estimation and control. — New York, NY, USA: Springer, 2008. 382 p.
21. *Борисов А. В.* Применение алгоритмов оптимальной фильтрации для решения задачи мониторинга доступности удаленного сервера // *Информатика и её применения*, 2014. Т. 8. Вып. 3. С. 53–69. doi: 10.14375/19922264140307.
22. *Borisov A.* Partially observable multivariate point processes with linear random compensators: Analysis and filtering with applications to queueing networks // 1st IFAC Conference on Modelling, Identification and Control of Nonlinear Systems (MICNON) Proceedings. — Saint Petersburg, 2015. P. 1119–1124.

23. Microsoft Message Analyzer. www.microsoft.com/en-us/download/details.aspx?id=44226.
24. Wireshark. www.wireshark.org/#learnWS.
25. Schulzrinne H., Casner S., Frederick R., Jacobson V. RTP: A transport protocol for real-time applications // RFC 3550, 2003. tools.ietf.org/html/rfc3550.

Поступила в редакцию 24.02.16

MODELING AND MONITORING OF VoIP CONNECTION

A. V. Borisov, A. V. Bosov, and G. B. Miller

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

Abstract: The Real Time Transfer Protocol (RTP), widely used in Voice over IP (VoIP) technologies for audio and video data transmission, is analyzed to design a mathematical model of VoIP connection. The model attempts to meet basic features of VoIP technologies as well as key features of the real link functioning like the frame delays, losses, etc. The proposed approach is based on the finite-state unobservable hidden Markov model. The unobservable state is assumed to be a finite-dimensional Markov process, whereas the observation is assumed to be a non-Markovian multivariate point process that indicates the heterogeneous frames reception. For the proposed model, the hidden link state optimal filtering problem given the packet/losses stream observations is formulated and its solution is provided. Proposed link model validity and filtering algorithm performance are illustrated by processing of captured real video streams delivered via 3G mobile network by Linphone VoIP services.

Keywords: VoIP technologies; RTP; network link; hidden Markov model; multivariate point process; optimal state filtering

DOI: 10.14357/19922264160201

Acknowledgments

The research is partly supported by the Russian Foundation for Basic Research (grants Nos. 16-07-00677 and 15-37-20611).

References

1. Gilbert, E. N. 1960. Capacity of a burst-noise channel. *Bell Syst. Tech. J.* 39:1253–1265.
2. Elliott, E. O. 1963. Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech. J.* 42:1977–1997.
3. Altman, E., K. Avrachenkov, and C. Barakat. 2000. TCP in presence of bursty losses. *Perform. Evaluation* 42:129–147.
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. *Probl. Inf. Transm.* 41(2):150–170. doi:10.1007/s11122-005-0020-8.
5. Bruno, R., M. Conti, and E. Gregori. 2008. Throughput analysis and measurements in IEEE 802.11 WLANs with TCP and UDP Traffic Flows. *IEEE Trans. Mobile Comput.* 7(2):171–186.
6. Haßlinger, G., and O. Hohlfeld. 2008. The Gilbert–Elliott model for packet loss in real time services on the internet. *14th GI/ITG Conference on Measurement, Modelling and Evaluation of Computer and Communication Systems (MMB) Proceedings*. Dortmund, Germany. 269–283.
7. Malik, M., M. Aydin, Z. Shah, and S. Hussain. 2014. Stochastic model of TCP and UDP traffic in IEEE 802.11b/g. *IEEE 9th Conference on Industrial Electronics and Applications (ICIEA) Proceedings*. 2170–2175.
8. Whitt, W. 2002. *Stochastic-process limits. An introduction to stochastic-process limits and their application to queues*. New York, NY: Springer. 602 p.
9. Bohacek, S. 2003. A stochastic model of TCP and fair video transmission. *22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM) Proceedings*. 2:1134–1144.
10. Liu, Y., and W. Gong. 2003. On fluid queueing systems with strict priority. *IEEE Trans. Automat. Contr.* 48(12):2079–2088.
11. Domańska, J., A. Domański, T. Czachórski, and J. Klamka. 2014. Fluid flow approximation of time-limited TCP/UDP/XCP streams. *Bull. Pol. Acad. Sci. Tech. Sci.* 62(2):217–225.
12. Leland, W. E., M. S. Taqqu, W. Willinger, and D. V. Wilson. 1994. On the self-similar nature of Ethernet traffic. *IEEE ACM Trans. Network.* 2(1):1–15.
13. Crovella, M. E., and A. Bestavros. 1997. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE ACM Trans. Network.* 5(6):835–846.

14. Tsybakov, B., and N. Georganas. 2000. Overflow and losses in a network queue with a self-similar input. *Queueing Syst.* 35(1-4):201–235.
15. Yariv, E., and N. Merhav. 2002. Hidden Markov processes. *IEEE Trans. Inform. Theory* 48(6):1518–1569.
16. Borisov, A. V., and G. B. Miller. 2005. Analysis and filtration of special discrete-time markov processes. II. Optimal filtration. *Automat. Rem. Contr.* 66(7):1125–1136. doi:10.1007/s10513-005-0153-7.
17. Anisimov, V. 2008. *Switching processes in queueing models*. New York, NY: Wiley. 352 p.
18. Ellis, M., D. P. Pizarro, T. Kypraios, and C. Perkins. 2012. Modelling packet loss in RTP-based streaming video for residential users. *37th Conference on Local Computer Networks Proceedings*. New York, NY: IEEE Press. 220–223.
19. Liptser, R. Sh., and A. N. Shiryaev. 1989. *Theory of martingales*. New York, NY: Springer-Verlag. 812 p.
20. Elliott, R. J., L. Aggoun, and J. B. Moore. 2008. *Hidden Markov models: Estimation and control*. New York, NY: Springer. 382 p.
21. Borisov, A. V. 2014. Primenenie algoritmov optimal'noy fil'tratsii dlya resheniya zadachi monitoringa dostupnosti udalennogo servera [Monitoring remote server accessibility: The optimal filtering approach]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):53–69. doi:10.14375/19922264140307.
22. Borisov, A. 2015. Partially observable multivariate point processes with linear random compensators: Analysis and filtering with applications to queueing networks. *Ist IFAC Conference on Modelling, Identification and Control of Non-linear Systems (MICNON 2015)*. St. Petersburg. 1119–1124.
23. Microsoft Message Analyzer. Available at: www.microsoft.com/en-us/download/details.aspx?id=44226 (accessed March 30, 2016).
24. Wireshark. Available at: www.wireshark.org/#learnWS (accessed March 30, 2016).
25. Schulzrinne, H., S. Casner., R. Frederick, and V. Jacobson. 2003. RTP: A transport protocol for real-time applications. *RFC 3550*. Available at: tools.ietf.org/html/rfc3550 (accessed March 30, 2016).

Received February 24, 2016

Contributors

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

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

Miller Grigoriy B. (b. 1980) — Candidate of Science (PhD) in physics and mathematics, scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gmliller@frcsc.ru

НОРМАЛЬНЫЕ УСЛОВНО-ОПТИМАЛЬНЫЕ ФИЛЬТРЫ И ЭКСТРАПОЛЯТОРЫ ПУГАЧЁВА ДЛЯ СТОХАСТИЧЕСКИХ СИСТЕМ, ЛИНЕЙНЫХ ОТНОСИТЕЛЬНО СОСТОЯНИЯ

И. Н. Сеницын¹, Э. Р. Корепанов²

Аннотация: Рассматривается теория аналитического синтеза непрерывных (дифференциальных) и дискретных (разностных) суб- и условно-оптимальных фильтров и экстраполяторов Пугачёва для обработки процессов в гауссовских и негауссовских стохастических системах (СтС), линейных относительно вектора состояния. Первые работы по фильтрации и экстраполяции для таких гауссовских систем были выполнены Липцером и Ширяевым, а для негауссовских — Пугачёвым и Сеницыным. Приведены алгоритмы нормальных суб- и условно-оптимальных фильтров для непрерывных и дискретных систем. Представлены алгоритмы нормальных суб- и условно-оптимальных экстраполяторов. Разработанные алгоритмы положены в основу программного обеспечения (StS-Filter, 2016). Результаты допускают развитие на случай авторегрессионных шумов в наблюдениях, а также систем с мультипликативными шумами.

Ключевые слова: дискретные СтС; дифференциальная СтС; метод нормальной аппроксимации (МНА) апостериорной плотности; метод статистической линеаризации (МСЛ); нормальный условно-оптимальный фильтр Пугачёва (НФП); нормальный условно-оптимальный экстраполятор Пугачёва (НЭП); стохастическая система (СтС); СтС, линейная относительно состояния; условия Липцера–Ширяева; фильтр Липцера–Ширяева (ФЛШ)

DOI: 10.14357/19922264160202

1 Введение

Многие практические задачи обработки информации в статистических научных исследованиях основаны на использовании теории фильтрации процессов в СтС, линейных относительно состояния [1–5].

Первые работы в этом направлении для гауссовских СтС выполнены Липцером и Ширяевым [1], а для негауссовских СтС на основе субоптимальной фильтрации — Пугачёвым и Сеницыным [2, 3]. В [5] рассмотрены вопросы синтеза алгоритмов нормальных условно-оптимальных фильтров Пугачёва (НФП) для обработки информации в дифференциальных гауссовских и негауссовских СтС, линейных относительно состояния (условия Липцера–Ширяева). Особое внимание уделено синтезу НФП для СтС при условиях Липцера–Ширяева на основе аппроксимации апостериорного распределения нормальным субоптимальным квазилинейным НФП, основанным на статистической линеаризации нелинейных функций, зависящих от наблюдений. Для СтС высокой размерности путем выбора структурных функций, отражающих аналитическую природу наблюдаемой системы, синте-

зированы НФП, простые в компьютерной реализации и способные работать в режиме реального времени. Алгоритмы положены в основу модуля инструментального программного обеспечения (StS-Filter, 2016).

Настоящая статья посвящена вопросам синтеза непрерывных и дискретных нормальных фильтров и экстраполяторов Пугачёва (НФП и НЭП) для непрерывных и дискретных СтС, линейных относительно состояния. В разд. 2 приведены уравнения, описывающие процессы дифференциальных и разностных СтС. Раздел 3 для гауссовских СтС посвящен нормальным фильтрам (НФ) на основе методов нормальной аппроксимации (МНА) и метода статистической линеаризации (МСЛ). В разд. 4 приводятся алгоритмы НФП на основе теории условно-оптимальной фильтрации Пугачёва. Для дифференциальных негауссовских СтС алгоритмы НФП даны в разд. 5. Обобщение НФ для дискретных СтС приведено в разд. 6. Алгоритмы нормальных условно-оптимальных экстраполяторов Пугачёва представлены в разд. 7. В заключении приведены основные выводы и возможные обобщения.

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

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

2 Непрерывные и дискретные стохастические системы, линейные относительно состояния

Рассмотрим сначала нелинейную дифференциальную СтС [5]:

$$\dot{X}_t = \varphi(X_t, Y_t, t) + \psi(X_t, Y_t, t)V, \quad X_{t_0} = X_0; \quad (1)$$

$$\dot{Y}_t = \varphi_1(X_t, Y_t, t) + \psi_1(X_t, Y_t, t)V, \quad Y_{t_0} = Y_0. \quad (2)$$

Здесь X_t и Y_t — векторы состояния и наблюдения размерности n_x и n_y ; $V = \dot{W}$, W — векторный стохастический процесс (СтП) с независимыми приращениями, состоящий из винеровской $W_0(t)$ и пуассоновской частей:

$$\left. \begin{aligned} W &= W_0(t) + \int c(\rho)P^0(t, d\rho); \\ \nu^W &= \nu^{W_0} + \int_{R_0^q} c(\rho)c(\rho)^T \nu_P(t, \rho) d\rho, \end{aligned} \right\} \quad (3)$$

где $c = c(\rho)$ — векторная функция (той же размерности q , что и W) аргумента ρ , а интеграл при любом $t \geq t_0$ представляет собой стохастический интеграл по централизованной пуассоновской мере $P^0(t, A)$, независимой от W_0 и имеющей независимые значения на попарно непересекающихся множествах; A — борелевское множество пространства R_0^q с выколотым началом 0 ; ν^W , ν^{W_0} и ν_P — интенсивности W , W_0 и P^0 ; $\varphi = \varphi(X_t, Y_t, t)$, $\psi = \psi(X_t, Y_t, t)$, $\varphi_1 = \varphi_1(X_t, Y_t, t)$ и $\psi_1 = \psi_1(X_t, Y_t, t)$ — известные функции размерности $(n_x \times 1)$, $(n_x \times n_v)$, $(n_y \times 1)$ и $(n_y \times n_v)$, удовлетворяющие следующим условиям Липцера–Ширяева [1]:

– функции φ и φ_1 линейны относительно состояния X_t :

$$\left. \begin{aligned} \varphi(X_t, Y_t, t) &= a_1(Y_t, t)X_t + a_0(Y_t, t); \\ \varphi_1(X_t, Y_t, t) &= b_1(Y_t, t)X_t + b_0(Y_t, t); \end{aligned} \right\} \quad (4)$$

– функции ψ и ψ_1 не зависят от состояния X_t :

$$\psi(X_t, Y_t, t) = \bar{\psi}(Y_t, t); \quad \psi_1(X_t, Y_t, t) = \bar{\psi}_1(Y_t, t). \quad (5)$$

Предполагается, что уравнения СтС (1) и (2) понимаются в смысле Ито и имеют решение в среднем квадратическом (с.к.) [2].

Систему (1)–(5) будем называть гауссовской, если $V = \dot{W}_0$, а X_0 и Y_0 — гауссовские случайные величины (с.в.).

Важный частный случай (1)–(5) составляют уравнения с аддитивными шумами, когда

$$\bar{\psi}(Y_t, t) = \psi_0(t); \quad \bar{\psi}_1(Y_t, t) = \psi_{10}(t). \quad (6)$$

Для случая, когда в уравнения (1), (2) входят независимые белые шумы V_1 и V_2 , следует принять:

$$V = [V_1^T V_2^T]^T, \quad \nu = \begin{bmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{bmatrix};$$

$$\psi V = \psi' V_1, \quad \psi_1 V = \psi'_1 V_2.$$

Для дискретных СтС, линейных относительно состояния, уравнения состояния и наблюдения имеют следующий вид [2]:

$$\begin{aligned} X_{k+1} &= \varphi_k(X_k, Y_k) + \psi'_k(X_k, Y_k)V_k^d = \\ &= a_k(Y_k)X_k + a_{0k}(Y_k) + c_{0k}(Y_k)V_k^d, \\ X_1 &= X_{10}; \end{aligned} \quad (7)$$

$$\begin{aligned} Y_k &= \varphi_{1k}(X_k, Y_k) + \psi_{1k}(X_k, Y_k)V_k^d = \\ &= b_k(Y_k)X_k + b_{0k}(Y_k) + l_{0k}(Y_k)V_k^d, \\ Y_1 &= Y_{10}. \end{aligned} \quad (8)$$

Здесь $n = n_x = n_y$; V_k^d — дискретный белый шум с известным в общем случае негауссовским распределением.

Для дискретных СтС с аддитивными шумами следует в (7), (8) принять:

$$c_{0k}(Y_k) = c_{0k0}; \quad l_{0k}(Y_k) = l_{0k0}.$$

3 Дифференциальные гауссовские стохастические системы. Фильтры Липцера–Ширяева и их субоптимальная аппроксимация

Для гауссовской СтС (1)–(5) известны следующие точные уравнения нелинейной фильтрации по критерию минимума с.к. ошибки [1, 2]:

$$\begin{aligned} \dot{\hat{X}}_t &= [a_1(Y_t, t)\hat{X}_t + a_0(Y_t, t)] + \\ &+ [R_t b_1(Y_t, t)^T + (\bar{\psi}\nu_0\bar{\psi}_1^T)(Y_t, t)] \times \\ &\times (\bar{\psi}_1\nu_0\bar{\psi}_1^T)^{-1}(Y_t, t) \left\{ \dot{Y}_t - [b_1(Y_t, t)\hat{X}_t + b_0(Y_t, t)] \right\}, \\ \hat{X}_{t_0} &= \hat{X}_0; \end{aligned} \quad (9)$$

$$\begin{aligned} \dot{R}_t = & a_1(Y_t, t) R_t + R_t a_1(Y_t, t)^T + \\ & + (\bar{\psi} \nu_0 \bar{\psi}^T)(Y_t, t) - [R_t b_1(Y_t, t)^T + \\ & + (\bar{\psi} \nu_0 \bar{\psi}_1^T)(Y_t, t)] (\bar{\psi}_1 \nu_0 \bar{\psi}_1^T)^{-1}(Y_t, t) \times \\ & \times [b_1(Y_t, t) R_t + (\bar{\psi}_1 \nu_0 \bar{\psi}^T)(Y_t, t)], \quad R_{t_0} = R_0, \quad (10) \end{aligned}$$

где \hat{X}_t — с. к. оценка; R_t — ковариационная матрица ошибки фильтрации $X_t - \hat{X}_t$.

Таким образом, имеет место утверждение.

Теорема 3.1. Пусть в гауссовской системе (1)–(5) диффузионная матрица $\sigma_1 = \sigma_1(Y_t, t) = \bar{\psi}_1 \nu_0 \bar{\psi}_1^T(Y_t, t)$ не вырождена. Тогда алгоритм с.к. оптимального фильтра определяется уравнением (9), а его точность оценивается согласно (10).

Как и в случае линейной фильтрации при аддитивных шумах [2], уравнения дифференциального фильтра Липцера–Ширяева (ФЛШ) представляют собой замкнутую систему уравнений, определяющую \hat{X}_t и R_t . Поэтому с.к. оптимальную оценку \hat{X}_t вектора состояния системы X_t и его апостериорную ковариационную матрицу R_t , характеризующую точность с.к. оптимальной оценки \hat{X}_t , можно вычислять по мере получения результатов наблюдений совместным интегрированием уравнений (9) и (10). Однако в противоположность линейной фильтрации для ФЛШ нельзя вычислить R_t заранее, до получения результатов наблюдений, так как от последних зависят коэффициенты уравнения (10). Поэтому ФЛШ в данном случае должен выполнять интегрирование обоих уравнений (9) и (10). Это приводит к существенному повышению порядка оптимального фильтра. Если линейный фильтр всегда описывается уравнениями порядка n_x , то в рассматриваемом более общем случае с.к. оптимальный фильтр описывается уравнениями порядка

$$Q_{\text{ФЛШ}} = n_x + \frac{n_x(n_x + 1)}{2} = \frac{n_x(n_x + 3)}{2}. \quad (11)$$

Очевидно, что ФЛШ будет совпадать с обобщенным (приближенным) фильтром Калмана–Бьюси, фильтрами второго порядка, гауссовыми фильтрами [2, 6].

Так как гауссовское (нормальное) распределение, аппроксимирующее апостериорное распределение вектора X_t , полностью определяется апостериорными математическим ожиданием \hat{X}_t и ковариационной матрицей R_t вектора X_t , то согласно теории нелинейной приближенной субоптимальной фильтрации при аппроксимации апостериорного распределения вектора X_t нормальным распределением будут иметь место следующие стохастические дифференциальные уравнения, определяющие \hat{X}_t и R_t [2, 5]:

$$\begin{aligned} \dot{\hat{X}}_t = & f(\hat{X}_t, Y_t, R_t, t) + \\ & + h(\hat{X}_t, Y_t, R_t, t) [\dot{Y}_t - f^{(1)}(\hat{X}_t, Y_t, R_t, t)]; \quad (12) \end{aligned}$$

$$\begin{aligned} \dot{R}_t = & \left\{ f^{(2)}(\hat{X}_t, Y_t, R_t, t) - h(\hat{X}_t, Y_t, R_t, t) \times \right. \\ & \times \bar{\psi}_1 \nu_0 \bar{\psi}_1^T(Y_t, t) h(\hat{X}_t, Y_t, R_t, t)^T \left. \right\} + \\ & + \sum_{r=1}^{n_y} \rho_r(\hat{X}_t, Y_t, R_t, t) [\dot{Y}_r - f_r^{(1)}(\hat{X}_t, Y_t, R_t, t)], \quad (13) \end{aligned}$$

где

$$\begin{aligned} f(\hat{X}_t, Y_t, R_t, t) = & M^N [\varphi(Y_t, X, t)] = \\ = & a_1(Y_t, t) \hat{X}_t + a_0(Y_t, t); \end{aligned}$$

$$\begin{aligned} f^{(1)}(\hat{X}_t, Y_t, R_t, t) = & \left\{ f_r^{(1)}(\hat{X}_t, Y_t, R_t, t) \right\} = \\ = & M^N [\varphi_1(Y_t, X, t)] = b_1(Y_t, t) \hat{X}_t + b_0(Y_t, t); \end{aligned}$$

$$\begin{aligned} h(\hat{X}_t, Y_t, R_t, t) = & M^N \left\{ [X \varphi_1(Y_t, X, t)^T + \bar{\psi} \nu_0 \bar{\psi}_1^T(Y_t, X, t)] - \right. \\ & \left. - \hat{X}_t f^{(1)}(\hat{X}_t, Y_t, R_t, t)^T \right\} (\bar{\psi}_1 \nu_0 \bar{\psi}_1^T)^{-1}(Y_t, t) = \\ = & [R_t b_1(Y_t, t)^T + (\bar{\psi} \nu_0 \bar{\psi}_1^T)(Y_t, t)] \times \\ & \times (\bar{\psi}_1 \nu_0 \bar{\psi}_1^T)^{-1}(Y_t, t); \end{aligned}$$

$$\begin{aligned} f^{(2)}(\hat{X}_t, Y_t, R_t, t) = & M^N \left\{ (X - \hat{X}_t) \varphi(Y_t, X, t)^T + \right. \\ & \left. + \varphi(Y_t, X, t) (X^T - \hat{X}_t^T) + \bar{\psi} \nu_0 \bar{\psi}^T(Y_t, X, t) \right\} = \\ = & [R_t b_1(Y_t, t)^T + (\bar{\psi} \nu_0 \bar{\psi}_1^T)(Y_t, t)] \times \\ & \times (\bar{\psi}_1 \nu_0 \bar{\psi}_1^T)^{-1}(Y_t, t) [b_1(Y_t, t) R_t + (\bar{\psi}_1 \nu_0 \bar{\psi}^T)(Y_t, t)]; \end{aligned}$$

$$\begin{aligned} \rho_r(\hat{X}_t, Y_t, R_t, t) = & M^N \left\{ (X - \hat{X}_t) (X^T - \hat{X}_t^T) a_r(Y_t, X, t) + \right. \\ & \left. + (X - \hat{X}_t) b_r(Y_t, X, t)^T (X^T - \hat{X}_t^T) + \right. \\ & \left. + b_r(Y_t, X, t) (X^T - \hat{X}_t^T) \right\} = 0 \quad (r = 1, \dots, n_y). \end{aligned}$$

Здесь функции φ , φ_1 , $\bar{\psi}$ и $\bar{\psi}_1$ определены (4) и (5); a_r — r -й элемент матрицы-строки $A_{\varphi_1} = (\varphi_1^T - \hat{\varphi}_n^T) (\bar{\psi}_1 \nu_0 \bar{\psi}_1^T)^{-1}$; $B_{kr}^{\bar{\psi} \bar{\psi}_1}$ — элемент k -й строки и r -го

столбца матрицы $B^{\bar{\psi}\bar{\psi}_1} = (\bar{\psi}\nu_0\bar{\psi}_1^T)(\bar{\psi}_1\nu_0\bar{\psi}_1^T)^{-1}$; b_r — r -й столбец матрицы $B^{\bar{\psi}\bar{\psi}_1}$, $b_r^{\bar{\psi}\bar{\psi}_1} = \left[b_{1r}^{\bar{\psi}\bar{\psi}_1} \dots b_{n_x r}^{\bar{\psi}\bar{\psi}_1} \right]^T$, M^N — символ математического ожидания по $X = X_t$.

Число уравнений МНА одномерного апостериорного распределения определяется по формуле (11).

За начальные значения \hat{X}_t и R_t при интегрировании уравнений (12) и (13), естественно, следует принять условные математическое ожидание \hat{X}_0 и ковариационную матрицу R_0 с. в. X_0 относительно Y_0 :

$$\begin{aligned} \hat{X}_0 &= M^N [X_0 | Y_0]; \\ R_0 &= M^N \left[(X_0 - \hat{X}_0) (X_0^T - \hat{X}_0^T) | Y_0 \right]. \end{aligned}$$

Если нет информации об условном распределении X_0 относительно Y_0 , то начальные условия можно взять в виде:

$$\begin{aligned} \hat{X}_0 &= M^N X_0; \\ R_0 &= M^N (X_0 - M^N X_0) (X_0^T - M^N X_0^T). \end{aligned}$$

Если же и об этих величинах нет никакой информации, то начальные значения \hat{X}_t и R_t приходится задавать произвольно.

Сравнивая уравнения (12), (13) с уравнениями ФЛШ (теорема 3.1), имеем следующий результат [5].

Теорема 3.2. Для гауссовской СтС (1)–(5) субоптимальные НФ на основе МНА для одномерной апостериорной плотности и ФЛШ совпадают.

4 Дифференциальные гауссовские стохастические системы. Нормальные фильтры Пугачёва

Следуя [2, 5], будем искать фильтр для оценки \hat{X}_t в виде следующего уравнения:

$$\begin{aligned} d\hat{X}_t &= \alpha_t \xi(\hat{X}_t, Y_t, t) dt + \\ &+ \beta_t \eta(\hat{X}_t, Y_t, t) dY_t + \gamma_t dt. \end{aligned} \quad (14)$$

Здесь $\xi = \xi(\hat{X}_t, Y_t, t)$ и $\eta = \eta(\hat{X}_t, Y_t, t)$ — некоторые известные структурные функции текущих значений оценки наблюдаемого процесса Y_t , \hat{X}_t и времени t ; α_t , β_t и γ_t — неизвестные функции времени. Если бы коэффициенты α_t , β_t и γ_t в (14) были известными функциями времени, то уравнение (14) определило бы фильтр того же порядка n_x , что

и уравнение (1), описывающее поведение системы. Поэтому, естественно, возникает мысль попытаться непосредственно определить коэффициенты α_t , β_t и γ_t в уравнении (14) как функции времени из условия минимума с.к. ошибки при всех $t > t_0$. Это приводит к теории условно-оптимального фильтра Пугачева (ФП), когда в уравнения ФП задаются заранее и оптимизируются только коэффициенты этого уравнения. Итак, приходим к идее нахождения оптимального фильтра в некотором классе допустимых фильтров, определяемом условием, что поведение фильтра описывается дифференциальным уравнением заданного порядка и заданной формы. Таким образом, мы отказываемся от абсолютной оптимизации и ограничиваемся условной оптимизацией в заданном ограниченном классе фильтров.

Определив класс допустимых фильтров, следует решить вопрос о том, какой фильтр в этом классе считается оптимальным. Следуя [2], будем считать оптимальным такой фильтр, который дает в известном смысле наилучшую оценку при всех $t > t_0$. Иными словами, задача оптимизации фильтра при всех $t > t_0$ является задачей многокритериальной оптимизации. Такие задачи, как правило, не имеют решения. Фильтр Калмана–Бьюси, дающий оптимальную линейную оценку состояния линейной системы в каждый момент $t > t_0$, является исключением [2]. Значит, надо определить такую оптимальность фильтра, при которой возможно решение задачи. Будем считать оптимальным такой допустимый фильтр, который на каждом бесконечно малом интервале времени совершает оптимальный переход из того состояния, в котором он был в начале этого интервала, в новое состояние. Такой допустимый фильтр будем называть условно-оптимальным. Тогда задачи фильтрации сведутся к нахождению оптимальных значений α_t , β_t и γ_t , в любой момент $t \geq t_0$ обеспечивающих минимум с.к. ошибки фильтрации в бесконечно близкий будущий момент $s > t$, $s \rightarrow t$.

Отметим, что ФП обладает тем свойством, что в данном классе допустимых фильтров не существует фильтра, который при данном начальном распределении Y_t , X_t и \hat{X}_t в момент t_0 был бы лучше условно-оптимального при всех $t > t_0$. Это значит, по терминологии теории многокритериальной оптимизации, что ФП представляет собой один из множества допустимых фильтров — оптимальный по Парето [2]. Общая теория ФП по с.к. критерию развита для уравнений (1), (2) и подробно изложена в [2]. Теория ФП обладает двумя несомненными преимуществами по сравнению с методами субоптимальной фильтрации. Во-первых, она позволяет получать фильтры более низкого порядка

и, следовательно, более простые в реализации. Во-вторых, она дает возможность получать фильтры не меньшей, а при желании даже большей точности, чем фильтры, даваемые методами субоптимальной нелинейной фильтрации.

Применяя теорию ФП [2] к нормальным процессам в гауссовской СтС (1)–(5) при $V = V_0$, приддем к нормальному ФП вида (14). Входящие в (14) коэффициенты α_t , β_t и γ_t определяются следующими уравнениями:

$$\alpha_t m_1 + \beta_t m_2 + \gamma_t = m_0; \quad (15)$$

$$m_0 = M^N[\varphi], \quad m_1 = M^N[\xi], \quad m_2 = M^N[\eta]; \quad (16)$$

$$\beta_t = \kappa_{02} \kappa_{22}^{-1}; \quad (17)$$

$$\kappa_{02} = M^N \left[\left(X_t - \hat{X}_t \right) \left(a_1 X_t + a_0 \right)^T \eta^T \right] + M^N \left[\bar{\psi} \nu_0 \bar{\psi}_1^T \eta^T \right]; \quad (18)$$

$$\kappa_{22} = M^N \left[\eta \bar{\psi}_1 \nu_0 \bar{\psi}_1^T \eta^T \right]; \quad (19)$$

$$\begin{aligned} \alpha_t \kappa_{11} + M^N \left[\left(\hat{X}_t - X_t \right) \left(\xi^T \alpha_t^T + \gamma_t^T \right) \frac{\partial \xi^T}{\partial \hat{X}_t} \right] = \\ = \kappa'_{01} - \beta_t \kappa'_{21}; \quad (20) \end{aligned}$$

$$\kappa'_{21} = M^N \left\{ \left[\eta \left(b_1 X_t + b_0 \right) - m_2 \right] \xi \right\}^T; \quad (21)$$

$$\begin{aligned} \kappa'_{01} = \kappa_{01} + M^N \left[\left(X_t - \hat{X}_t \right) \frac{\partial \xi^T}{\partial t} \right] + \\ + M^N \left\{ \left(X_t - \hat{X}_t \right) \left(b_1 X_t + b_0 \right)^T + \right. \\ \left. + \bar{\psi} \nu_0 \bar{\psi}_1^T - \beta_t \eta \bar{\psi}_1 \nu_0 \bar{\psi}_1^T \right\} \left(\frac{\partial}{\partial Y_t} + \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \xi^T + \\ + \frac{1}{2} M^N \left[\left(X_t - \hat{X}_t \right) \times \right. \\ \left. \times \left\{ \text{tr} \left[\bar{\psi}_1 \nu_0 \bar{\psi}_1^T \left(\frac{\partial}{\partial Y_t} + 2 \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \frac{\partial^T}{\partial Y_t} \right] + \right. \right. \\ \left. \left. + \text{tr} \left[\beta_t \eta \bar{\psi}_1 \nu_0 \bar{\psi}_1^T \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \frac{\partial^T}{\partial \hat{X}_t} \right] \right\} \xi^T \right], \quad (22) \end{aligned}$$

$$\left. \begin{aligned} \kappa_{11} &= M^N \left\{ \left[\xi - m_1 \right] \xi^T \right\}; \\ \kappa_{21} &= M^N \left\{ \left[b_1 X_t + b_0 - m_2 \right] \xi^T \right\}; \\ \kappa_{01} &= M^N \left\{ \left[a_1 X_t + a_0 - m_0 \right] \xi^T \right\}. \end{aligned} \right\} \quad (23)$$

Точность НФП определяется уравнением:

$$\begin{aligned} \dot{R}_t = M^N \left[\left(X_t - \hat{X}_t \right) \left(a_1 X_t + a_0 \right)^T + \right. \\ \left. + \left(a_1 X_t + a_0 \right) \left(X_t^T - \hat{X}_t^T \right) - \right. \\ \left. - \beta_t \eta \left(Y_t, \hat{X}_t, t \right) \bar{\psi}_1 \nu_0 \bar{\psi}_1^T \eta \left(Y_t, \hat{X}_t, t \right)^T \beta_t^T + \right. \\ \left. + \bar{\psi} \nu_0 \bar{\psi}^T \right], \quad R_{t_0} = R_0. \quad (24) \end{aligned}$$

Таким образом, справедливо следующее утверждение [5].

Теорема 4.1. Пусть для гауссовской системы (1), (2) при условиях Липцера–Ширяева выполнены условия невырожденности матрицы κ_{22} (19) и конечности величин κ_{ij} ($i, j = 0, 1, 2$). Тогда алгоритм НФП определяется уравнением (14), а коэффициенты α_t , β_t и γ_t – (15)–(24).

Отметим, что в качестве условно-оптимальных НФП могут служить субоптимальные фильтры, определяемые теоремой 3.2.

5 Дифференциальные негауссовские стохастические системы. Нормальный фильтр Пугачёва

Для негауссовской СтС в уравнениях теоремы 4.1 следует заменить ν_0 на ν согласно (3), а в выражении (22) учесть два дополнительных интегральных члена:

$$\begin{aligned} \kappa'_{01} = \kappa_{01} + M^N \left[\left(X_t - \hat{X}_t \right) \frac{\partial \xi^T}{\partial t} \right] + \\ + M^N \left\{ \left(X_t - \hat{X}_t \right) \left[\varphi_1^T - \int_{R_0^q} c(\rho)^T \nu_P(t, \rho) d\rho \psi_1^T \right] + \right. \\ \left. + \bar{\psi} \nu \bar{\psi}_1^T - \beta_t \eta \bar{\psi}_1 \nu \bar{\psi}_1^T \right\} \left(\frac{\partial}{\partial Y_t} + \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \xi^T + \\ + \frac{1}{2} M^N \left\{ \left[\left(X_t - \hat{X}_t \right) \right] \times \right. \\ \left. \times \left\{ \text{tr} \left[\bar{\psi}_1 \nu \bar{\psi}_1^T \left(\frac{\partial}{\partial Y_t} + 2 \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \frac{\partial^T}{\partial Y_t} \right] + \right. \right. \\ \left. \left. + \text{tr} \left[\beta_t \eta \bar{\psi}_1 \nu \bar{\psi}_1^T \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \frac{\partial^T}{\partial \hat{X}_t} \right] \right\} \xi^T \right\} + \\ + M^N \left\{ \int_{R_0^q} \left[X_t - \hat{X}_t + \left(\bar{\psi} - \beta_t \eta \bar{\psi}_1 \right) c(\rho) \times \right. \right. \\ \left. \left. \times \xi \left(Y_t + \bar{\psi} c(\rho), \hat{X}_t + \beta_t \eta \bar{\psi}_1 c(\rho), t \right) - \right. \right. \\ \left. \left. - \xi^T \right]^T \nu_P(t, d\rho) d\rho \right\}. \quad (25) \end{aligned}$$

Здесь функции φ , φ_1 , $\bar{\psi}$ и $\bar{\psi}_1$ удовлетворяют условиям (4)–(5). В результате имеем следующее утверждение.

Теорема 5.1. Пусть для негауссовской СтС (1)–(5) матрица κ_{22} (18) не вырождена, а интегралы (15),

(16), (19), (21), (23) и (25) конечны. Тогда алгоритм НФП определяется уравнением (14), а коэффициенты α_t , β_t и γ_t — (15)–(23).

Теория НФП не позволяет получить нормальные с.к. оптимальные фильтры. Можно получить только ФП, которые в общем случае хуже с.к. оптимальных, но зато легко реализуемы. Однако если с.к. оптимальная оценка \hat{X}_t вектора X_t удовлетворяет уравнению допустимого фильтра (14) при каких-либо коэффициентах времени α_t , β_t и γ_t , то уравнения теорем 4.1 и 5.1, конечно, определяют именно эти α_t , β_t и γ_t и НФП будет с.к. оптимальным (последний в данном случае будет допустимым и, следовательно, оптимальным в классе допустимых фильтров).

Как известно [2], теория НФП дает возможность оценивать не все компоненты вектора состояния системы (в общем случае расширенного), а только некоторые из них. Для этого достаточно взять структурные функции ξ и η в (14) зависящими лишь от соответствующих компонент вектора \hat{X}_t . К примеру, взяв ξ и η в (14) зависящими лишь от Y_t , t и оценок неизвестных параметров системы, можно оценивать только параметры системы, не оценивая ее состояния. В таких случаях будут получаться НФП, порядок которых меньше размерности n_x расширенного вектора состояния.

Особое практическое значение имеет случай (1)–(6) с аддитивными (в общем случае негауссовскими) шумами. Следуя [2], проведем статистическую линеаризацию нелинейных функций:

$$\begin{aligned} a_1(Y_t, t) X_t &\approx (k_{0x}^{a_1x} - k_{1x}^{a_1x}) m_t^x + \\ &+ (k_{0y}^{a_1x} - k_{1y}^{a_1x}) m_t^y + k_x^{a_1y} X_t + k_{0y}^{a_1x} Y_t; \\ b_1(Y_t, t) X_t &\approx (k_{0x}^{b_1x} - k_{1x}^{b_1x}) m_t^x + \\ &+ (k_{0y}^{b_1x} - k_{1y}^{b_1x}) m_t^y + k_x^{b_1y} X_t + k_{0y}^{b_1x} Y_t; \\ a_0(Y_t, t) &\approx (k_{0y}^{a_0} - k_{1y}^{a_0}) m_t^y + k_{0y}^{a_0} Y_t; \\ b_0(Y_t, t) &\approx (k_{0y}^{b_0} - k_{1y}^{b_0}) m_t^y + k_{0y}^{b_0} Y_t. \end{aligned}$$

Тогда (1)–(6) приводятся к эквивалентной гауссовской системе, линейной относительно X_t^0 , Y_t^0 и нелинейной относительно m_t^x , m_t^y :

$$\begin{aligned} \dot{X}_t &= \bar{a}Y_t + \bar{a}_1X_t + \bar{a}_0 + \bar{\psi}V; \\ \dot{Y}_t &= \bar{b}Y_t + \bar{b}_1X_t + \bar{b}_0 + \bar{\psi}_1V. \end{aligned}$$

Здесь введены обозначения:

$$\left. \begin{aligned} \bar{a} &= k_y^{a_1x} + k_y^{a_0}; & \bar{a}_1 &= k_x^{a_1x}; \\ a_0 &= (k_{0y}^{a_0} - k_{1y}^{a_0})m_t^y; \\ \bar{b} &= k_y^{b_1x}; & \bar{b}_1 &= k_x^{b_1x}; & \bar{b}_0 &= (k_{0y}^{b_0} - k_{1y}^{b_0})m_t^y. \end{aligned} \right\} (26)$$

Правые части уравнений (26) зависят от вероятностных моментов первого и второго порядка и определяются из следующей линейной дифференциальной системы для составного вектора $Z_t = [X_t^T Y_t^T]^T$:

$$\dot{m}_t^z = c m_t^z + c_0; \quad \dot{K}_t^z = c K_t^z + K_t^z c^T + l \nu l^T, \quad (27)$$

где

$$c = \begin{bmatrix} \bar{a}_1 & \bar{a} \\ \bar{b}_1 & \bar{b} \end{bmatrix}; \quad c_0 = \begin{bmatrix} \bar{a}_0 \\ \bar{b}_0 \end{bmatrix}; \quad l = \begin{bmatrix} \bar{\psi} \\ \bar{\psi}_1 \end{bmatrix}.$$

Применяя теорию квазилинейной фильтрации [2] к уравнениям (27), получим следующие уравнения субоптимального квазилинейного НФП:

$$\begin{aligned} \dot{\hat{X}}_t &= \bar{a}Y_t + \bar{a}_1\hat{X}_t + \bar{a}_0 + \\ &+ \beta_t \left[\dot{Y}_t - (\bar{b}Y_t + \bar{b}_1\hat{X}_t + \bar{b}_0) \right]; \quad (28) \end{aligned}$$

$$\beta_t = R_t \bar{b}_1^T + (\bar{\psi} \nu \bar{\psi}_1^T) (\bar{\psi}_1 \nu \bar{\psi}_1^T)^{-1}; \quad (29)$$

$$\begin{aligned} \dot{R}_t &= \bar{a}_1 R_t + R_t \bar{a}_1^T + \bar{\psi} \nu \bar{\psi}^T - \\ &- (R_t \bar{b}_1^T + \bar{\psi} \nu \bar{\psi}_1) (\bar{\psi}_1 \nu \bar{\psi}_1^T)^{-1} (\bar{b} R_t + \bar{\psi}_1 \nu \bar{\psi}_1^T). \quad (30) \end{aligned}$$

Таким образом, имеем следующий результат [5].

Теорема 5.2. Пусть уравнения негауссовской СтС (1)–(6) с аддитивными шумами допускают применение МСЛ. Тогда уравнения алгоритма квазилинейного НФП имеют вид (28)–(30).

6 Обобщения нормальных фильтров на случай дискретных стохастических систем

Теоремы разд. 3–5 допускают обобщение на случай дискретных СтС (7), (8). Для дискретных гауссовских СтС (7) и (8) при $V_k^d = V_{0k}^d$ алгоритм НФ на основе МНА определяется уравнениями (теорема 6.1):

$$\begin{aligned} \hat{X}_{k+1} &= f_k \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) + \\ &+ h_k \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) \times \\ &\times \left[Y_{k+1} - f_k^{(1)} \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) \right]; \end{aligned}$$

$$\begin{aligned} R_{k+1} &= \left\{ f_k^{(2)} \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) - \right. \\ &- h_k \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) (\psi_{1,k} \nu_{0k} \psi_{1,k}^T) \times \\ &\left. \times h \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right)^T \right\} + \end{aligned}$$

$$+ \sum_{r=1}^{n_y} \rho_r \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) \times \\ \times \left[Y_{r,k+1} - f_{r,k}^{(1)} \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) \right].$$

Здесь введены следующие обозначения:

$$f_k = f_k \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) = M^N [\varphi_k]; \\ f_k^1 = f_k^{(1)} \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) = M^N [\varphi_{1k}]; \\ h_k = h_k \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) = \\ = M^N \left[X \varphi_{1k}(Y_k, X) + \bar{\psi}_k \nu_{0k} \bar{\psi}_{1k}^T(Y_k, X) \right]; \\ f_k^{(2)} = f_k^{(2)} \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) = \\ = M^N \left[\left(X - \hat{X}_{k+1|k} \right) \varphi_k(Y_k, X)^T + \right. \\ \left. + \varphi_k(Y_k, X) \left(X^T - \hat{X}_{k+1|k}^T \right) + \bar{\psi}_k \nu_{0k} \bar{\psi}_{1k}^T(Y_k, X) \right]; \\ \rho_r \left(\hat{X}_{k+1|k}, Y_k, R_{k+1|k} \right) = \\ = M^N \left[\left(X - \hat{X}_{k+1|k} \right) \left(X^T - \hat{X}_{k+1|k}^T \right) a_r(Y_k, X) + \right. \\ \left. + \left(X - \hat{X}_{k+1|k} \right) b_r(Y_k, X)^T \left(X^T - \hat{X}_{k+1|k}^T \right) + \right. \\ \left. + b_r(Y_k, X) \left(X^T - \hat{X}_{k+1|k}^T \right) \right],$$

где a_r — r -й элемент матрицы-строки $(\varphi_{1k}^T - \hat{\varphi}_{1k}^T)(\bar{\psi}_{1k} \nu_{0k} \bar{\psi}_{1k}^T)^{-1}$; $b_r = [b_{1r} \dots b_{n_x r}]^T$; b_{lr} — элемент l -й строки и r -го столбца матрицы $(\bar{\psi}_k \nu_{0k} \bar{\psi}_{1k}^T)(\bar{\psi}_{1k} \nu_{0k} \bar{\psi}_{1k}^T)^{-1}$.

В качестве начальных условий принимаются

$$\hat{X}_{1|1} = \hat{X}_1 = M^N [X_1 | Y_1]; \\ R_{1|1} = R_1 = M^N \left[(X_1 - X_1^0) (X_1 - X_1^0)^T | Y_1 \right],$$

определяющие начальное нормальное распределение.

Алгоритм дискретного НФП для дискретных СтС (7), (8), основываясь на [2], представим в следующем виде (**теорема 6.2**):

$$\hat{X}_{k+1} = \alpha_k \xi_k \left(\hat{X}_k \right) + \beta_k \eta_k \left(\hat{X}_k \right) Y_k + \gamma_k.$$

Здесь $\xi_k = \xi_k(\hat{X}_k)$ и $\eta_k = \eta_k(\hat{X}_k)$ — известные структурные функции НФП; неизвестные коэффициенты фильтра α_k , β_k и γ_k определяются из уравнений:

$$\alpha_k \kappa_{11}^{(k)} + \beta_k \kappa_{21}^{(k)} = \kappa_{01}^{(k)}; \quad \alpha_k \kappa_{12}^{(k)} + \beta_k \kappa_{22}^{(k)} = \kappa_{02}^{(k)}; \\ \gamma_k = \rho_0^{(k)} - \alpha_k \rho_1^{(k)} - \beta_k \rho_2^{(k)};$$

$$\rho_k = \left[\rho_1^{(k)T} \rho_2^{(k)T} \right]^T; \\ \rho_1^{(k)} = M^N [\xi_k]; \quad \rho_2^{(k)} = M^N [\eta_k \varphi_{1k}]; \\ B_k = \begin{bmatrix} \kappa_{11}^{(k)} & \kappa_{12}^{(k)} \\ \kappa_{21}^{(k)} & \kappa_{22}^{(k)} \end{bmatrix}, \quad \det |B_k| \neq 0; \\ \kappa_{11}^{(k)} = M^N \left\{ \left[\xi_k - \rho_1^{(k)} \right] \xi_k^T \right\}; \\ \kappa_{12}^{(k)} = \kappa_{21}^{(k)T} = M^N \left\{ \left[\xi_k - \rho_1^{(k)} \right] \varphi_{1k}^T \eta_k^T \right\}; \\ \kappa_{22}^{(k)} = M^N \left\{ \left[\eta_k \varphi_{1k} - \rho_2^{(k)} \right] \varphi_{1k}^T \eta_k^T \right\} + \\ + M^N \left\{ \eta_k \bar{\psi}_{1k} \nu_k \bar{\psi}_{1k}^T \eta_k^T \right\}; \\ D_k = \left[\kappa_{01}^{(k)} \kappa_{02}^{(k)} \right]; \quad \kappa_{01}^{(k)} = M^N \left\{ \left[\varphi_k - m_{k+1} \right] \xi_k^T \right\}; \\ \kappa_{02}^{(k)} = M^N \left\{ \left[\varphi_k - m_{k+1} \right] \varphi_{1k}^T \eta_k^T \right\} + \\ + M^N \left\{ \bar{\psi}_k \nu_k \bar{\psi}_{1k}^T \eta_k^T \right\}; \\ m_{k+1} = \rho_0^{(k)}, \quad \rho_0^{(k)} = M^N [\varphi_k],$$

где функции φ_k , φ_{1k} , $\bar{\psi}_k$ и $\bar{\psi}_{1k}$ определены (7), (8); $M^N V_k^d = 0$, $M^N V_k^d V_k^{dT} = \nu_k$ — ковариация белого шума V_k^d .

7 Нормальные экстраполяторы Пугачёва

Будем считать условно-оптимальным (по Пугачёву [2]) такой экстраполятор из класса допустимых, который при любом совместном распределении величин X_t , \hat{X}_t и Y_t в момент $t \geq t_0$ в дифференциальной СтС (1)–(5) дает наилучшую оценку вектора $X_{s+\Delta}$, в бесконечно близкий момент $s > t$, $s \rightarrow t$, реализующую минимум с.к. ошибки. Тогда задача условно-оптимальной экстраполяции сведется к нахождению оптимальных значений α_t , β_t и γ_t в (14) в любой момент времени $t \geq t_0$, обеспечивающих минимум с.к. ошибки экстраполяции в бесконечно близкий будущий момент $s > t$, $s \rightarrow t$.

Для решения задачи экстраполяции необходимо ограничиться случаем (1)–(5), когда функции φ и ψ не зависят от наблюдаемого вектора Y_t , процесс $W(t)$ состоит из двух независимых блоков $W_1(t)$ и $W_2(t)$ и соответственно матрицы ψ и ψ_1 имеют блочную структуру $\psi = [\psi' 0]$; $\psi_1 = [0 \psi'_1]$, так что $\psi dW = \psi' dW_1$ и $\psi_1 dW = \psi'_1 dW_2$. Теперь, отбросив штрихи у функций ψ' и ψ'_1 , запишем уравнения (1) и (2) в следующем виде:

$$\left. \begin{aligned} \dot{X}_t &= \varphi(X_t, Y_t, t) dt + \psi(X_t, t) V_1; \\ V_1 &= \dot{W}_1; \quad X_{t_0} = X_0; \\ \dot{Y}_t &= \varphi(X_t, Y_t, t) dt + \psi_1(X_t, Y_t, t) V_2; \\ V_2 &= \dot{W}_2; \quad Y_{t_0} = Y_0, \end{aligned} \right\} \quad (31)$$

где $W_1 = W_1(t)$ и $W_2 = W_2(t)$ — независимые процессы с независимыми приращениями и нулевыми математическими функциями:

$$K_{w_i}(t_1, t_2) = k_i(\min(t_1, t_2));$$

$$k_i(t) = k_i(t_0) + \int_{t_0}^t \nu_i(\tau) d\tau.$$

Совершенно так же, как в случае условно-оптимальной нелинейной фильтрации, решается задача условно-оптимальной экстраполяции. Разница будет лишь в том, что в случае экстраполяции \hat{X}_t представляет собой оценку будущего состояния системы $X_{t+\Delta}$, $\Delta > 0$, которое определяется стохастическим дифференциальным уравнением следующего вида:

$$\dot{X}_{t+\Delta} = \varphi(X_{t+\Delta}, t + \Delta) + \psi(X_{t+\Delta}, t + \Delta)V(t + \Delta), \quad V = \dot{V}. \quad (32)$$

Заменяв этим уравнением второе уравнение (8) и повторив все выкладки ФП, получим уравнения, определяющие коэффициенты в уравнениях нормального условно-оптимального экстраполятора Пугачёва (НЭП). Приведем окончательные результаты:

$$\left. \begin{aligned} \kappa_{02} &= \\ &= M^N \left[(X_{t+\Delta} - \hat{X}_t) \varphi_1(X_t, Y_t, t)^T \times \right. \\ &\quad \left. \times \eta(Y_t, \hat{X}_t, t)^T \right]; \\ \kappa_{22} &= M^N \left[\eta(Y_t, \hat{X}_t, t) \psi_1(X_t, Y_t, t) \nu_1(t) \times \right. \\ &\quad \left. \times \psi_1(X_t, Y_t, t)^T \eta(Y_t, \hat{X}_t, t)^T \right]; \\ M^N [\hat{X}] &= M^N [X_{t+\Delta}]; \\ M^N \left[(\hat{X} - X_{t+\Delta}) \xi_t^T \right] &= 0; \\ m_0 &= M^N [\varphi(X_{t+\Delta}, t + \Delta)]; \\ \kappa_{01} &= M^N \left\{ [\varphi(X_{t+\Delta}, t + \Delta) - \right. \\ &\quad \left. - m_0] \xi(Y_t, \hat{X}_t, t)^T \right\}; \end{aligned} \right\} \quad (33)$$

$$\begin{aligned} \kappa'_{01} &= \kappa_{01} + M^N \left[(X_{t+\Delta} - \hat{X}_t) \frac{\partial \xi^T}{\partial t} \right] + \\ &+ M^N \left\{ (X_{t+\Delta} - \hat{X}_t) \left[\varphi_1^T - \int_{R_0^q} c_2(\rho)^T \nu_{sP}(t, \rho) d\rho \bar{\psi}_1^T \right] - \right. \\ &\quad \left. - \beta_t \eta \bar{\psi}_1 \nu_{20} \bar{\psi}_1^T \right\} \left(\frac{\partial}{\partial Y_t} + \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \xi^T + \\ &\quad + \frac{1}{2} M^N \left\{ (X_{t+\Delta} - \hat{X}_t) \times \right. \end{aligned}$$

$$\begin{aligned} &\times \left\{ \text{tr} \left[\bar{\psi}_1 \nu_{20} \bar{\psi}_1^T \left(\frac{\partial}{\partial Y_t} + 2\eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \frac{\partial^T}{\partial Y_t} \right] + \right. \\ &\quad \left. + \text{tr} \left[\beta_t \eta \bar{\psi}_1 \nu_{20} \bar{\psi}_1^T \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \frac{\partial^T}{\partial \hat{X}_t} \right] \right\} \xi^T \left. \right\} + \\ &\quad + \int_{R_0^q} M^N \left\{ [X_{t+\Delta} - \hat{X}_t - \beta_t \eta \bar{\psi}_1 c_2(\rho)] \times \right. \\ &\quad \times \left[\xi \left(X_t + \bar{\psi}_1 c_2(\rho), \hat{X}_t + \beta_t \eta \bar{\psi}_1 c_2(\rho), t \right)^T - \right. \\ &\quad \left. \left. - \xi^T \right] \right\} \nu_{2P}(t, \rho) d\rho. \quad (35) \end{aligned}$$

Здесь $c_2(\rho)$, ν_{20} и ν_{2P} — соответствующие величины в представлении интенсивности ν_2 процесса $W_2(t)$ формулой вида (3).

В случае винеровского процесса $W_2(t)$ имеем $c_2(\rho) = 0$, $\nu_{20} = \nu_2$ и формула (35) принимает вид:

$$\begin{aligned} \kappa'_{01} &= \kappa_{01} + M^N \left[(X_{t+\Delta} - \hat{X}_t) \frac{\partial \xi^T}{\partial t} \right] + \\ &+ M^N \left\{ \left[(X_{t+\Delta} - \hat{X}_t) \varphi_1^T - \beta_t \eta \bar{\psi}_1 \nu_2 \bar{\psi}_1^T \right] \times \right. \\ &\quad \times \left(\frac{\partial}{\partial Y_t} + \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \xi^T \left. \right\} + \\ &\quad + \frac{1}{2} M^N \left\{ (X_{t+\Delta} - \hat{X}_t) \times \right. \\ &\quad \times \left\{ \text{tr} \left[\bar{\psi}_1 \nu_2 \bar{\psi}_1^T \left(\frac{\partial}{\partial Y_t} + 2\eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \right) \frac{\partial^T}{\partial Y_t} \right] + \right. \\ &\quad \left. \left. + \text{tr} \left[\beta_t \eta \bar{\psi}_1 \nu_2 \bar{\psi}_1^T \eta^T \beta_t^T \frac{\partial}{\partial \hat{X}_t} \frac{\partial^T}{\partial \hat{X}_t} \right] \right\} \xi^T \right\}, \end{aligned}$$

при этом точность экстраполятора Пугачёва определяется, согласно [2], формулой:

$$\begin{aligned} \dot{R}_t &= M^N \left[(X_{t+\Delta} - \hat{X}) \varphi(X_{t+\Delta}, t + \Delta)^T + \right. \\ &\quad \left. + \varphi(X_{t+\Delta}, t + \Delta) (X_{t+\Delta}^T - \hat{X}^T) - \right. \\ &\quad \left. - \beta_t \eta(Y_t, \hat{X}_t, t) \psi_1(X_t, Y_t, t) \nu_2(t) \psi_1(X_t, Y_t, t)^T \times \right. \\ &\quad \left. \times \eta(Y_t, \hat{X}_t, t)^T \beta_t^T + \right. \\ &\quad \left. + \psi(X_{t+\Delta}, t + \Delta) \nu_1(t + \Delta) \psi(X_{t+\Delta}, t + \Delta)^T \right]. \quad (36) \end{aligned}$$

Для вычисления математических ожиданий в (33)–(36) недостаточно знать одномерное нормальное распределение составного случайного процесса $[X_t^T Y_t^T \hat{X}_t^T]^T$, необходимо также знать совместное нормальное распределение величин X_t , $X_{t+\Delta}$, Y_t и \hat{X}_t при каждом t .

Таким образом, приходим к следующим результатам.

Теорема 7.1. Пусть векторный стохастический процесс $[X_t^T Y_t^T]^T$ определяется негауссовскими уравнениями (31) и обладает конечными двумерными вероятностными моментами. Тогда алгоритм НЭП определяется формулами (32)–(35), (36).

Теорема 7.2. Для гауссовской СтС (31) алгоритм НЭП определяется (32)–(35), (36).

Аналогично получаются уравнения алгоритмов НЭП для дискретных СтС.

Таким образом, теория условно-оптимальной экстраполяции стохастических процессов дает возможность строить экстраполяторы Пугачёва для одновременного оценивания состояния и параметров системы и экстраполяции ее состояния на несколько различных интервалов времени в реальном масштабе времени. Все сложные расчеты, необходимые для проектирования таких экстраполяторов, не опираются на результаты наблюдения и могут быть выполнены по априорным данным в процессе проектирования. Практическое применение таких экстраполяторов сводится к одновременному интегрированию уравнений (14) для оценивания текущего и будущих состояний системы.

8 Заключение

Разработана теория аналитического синтеза непрерывных и дискретных нормальных суб- и условно-оптимальных фильтров и экстраполяторов Пугачёва для обработки процессов в гауссовских

и негауссовских СтС, линейных относительно состояния. Результаты допускают обобщение на случай автокоррелированных шумов в наблюдениях. Алгоритмы положены в основу программного обеспечения StS-Filter (version 2016) для стохастических информационных технологий [7].

Литература

1. Липцер Р. Ш., Шираев А. Н. Статистика случайных процессов. — М.: Наука, 1974. 476 с.
2. Сеницын И. Н. Фильтры Калмана и Пугачёва. — 2-е изд. — М.: Логос, 2007. 776 с.
3. Сеницын И. Н., Сеницын В. И. Лекции по нормальной и эллипсоидальной аппроксимации в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 476 с.
4. Сеницын 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.
5. Сеницын И. Н., Корепанов Э. Р. Нормальные условно-оптимальные фильтры Пугачёва для дифференциальных стохастических систем, линейных относительно состояния // Информатика и её применения, 2015. Т. 9. Вып. 2. С. 30–38.
6. Ройтенберг Я. Н. Автоматическое управление. — 3-е изд., перераб. и доп. — М.: Наука, 1992. 576 с.
7. Корепанов Э. Р. Стохастические информационные технологии на основе фильтров Пугачёва // Информатика и её применения, 2011. Т. 5. Вып. 2. С. 36–57.

Поступила в редакцию 02.02.16

NORMAL PUGACHEV CONDITIONALLY-OPTIMAL FILTERS AND EXTRAPOLATORS FOR STATE LINEAR STOCHASTIC SYSTEMS

I. N. Sinitsyn and E. R. Korepanov

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

Abstract: The analytical synthesis theory of continuous and discrete sub- and Pugachev conditionally optimal filters and extrapolators for information processing in linear state stochastic systems (StS) is presented. For Gaussian StS, Liptzer and Shiraev performed the first works for filters and extrapolators synthesis. For non-Gaussian StS, the first works belong to Pugachev and Sinitsyn. Stochastic equations for state and observation of continuous and discrete StS are given. Algorithms for continuous normal sub- and conditionally optimal filters and extrapolators are presented. The corresponding algorithms for discrete StS are also given. The developed algorithms are the basis of the software tool “StS-Filter, 2016.” The results may be developed for autocorrelated noises and multiplicative noises.

Keywords: Liptser–Shiraev filter (LSF); Liptser–Shiraev conditions; normal approximation method (NAM) for *a posteriori* density; normal conditionally optimal Pugachev filter (NPF); stochastic systems (StS); state linear StS; statistical linearization method (SLM)

DOI: 10.14357/19922264160202

References

1. Liptser, R. Sh., and A. N. Shiryaev. 1974. *Statistika sluchaynykh protsessov* [Statistics of stochastic processes]. — Moscow: Nauka. 476 p.
2. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd ed. Moscow: Logos. 776 p.
3. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po normal'noy i ellipsoidal'noy approksimatsii v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 476 p.
4. 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.
5. Sinitsyn, I. N., and E. R. Korepanov. 2015. Normal'nye uslovno optimal'nye fil'try Pugacheva dlya differentsial'nykh stokhasticheskikh sistem, lineynykh otnositel'no sostoyaniya [Normal Pugachev filters and extrapolators for state linear stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 9(2):30–38.
6. Roytenberg, Ya. N. 1992. *Avtomatskoe upravlenie* [Automatic control]. 3rd ed. Moscow: Nauka. 576 p.
7. Korepanov, E. R. 2011. Stokhasticheskie informatsionnye tekhnologii na osnove fil'trov Pugacheva [Stochastic informational technologies based on Pugachev filters]. *Informatika i ee Primeneniya — Inform. Appl.* 5(2):36–57.

Received February 2, 2016

Contributors

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

Korepanov Eduard R. (b. 1966) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ekorepanov@ipiran.ru

ЭЛЛИпсоИДАЛЬНЫЕ СУБОПТИМАЛЬНЫЕ ФИЛЬТРЫ ДЛЯ НЕЛИНЕЙНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ НА МНОГООБРАЗИЯХ*

И. Н. Сеницын¹, В. И. Сеницын², Э. Р. Корепапов³

Аннотация: Разработана теория аналитического синтеза эллипсоидальных субоптимальных фильтров (ЭСОФ) для нелинейных дифференциальных стохастических систем (СтС) на многообразиях (МСтС). Рассмотрены случаи гауссовских и негауссовских СтС. Алгоритмы положены в основу модуля экспериментального программного обеспечения StS-Filter (version 2016). Результаты допускают развитие на случай дискретных СтС. Теоретический и практический интерес представляет теория ЭСОФ на основе ненормированных распределений.

Ключевые слова: апостериорное одномерное распределение; винеровский шум; метод эллипсоидальной аппроксимации (МЭА); метод эллипсоидальной линеаризации (МЭЛ); пуассоновский шум; стохастическая система на многообразиях (МСтС); субоптимальный фильтр (СОФ); уравнения точности МЭА и МЭЛ; уравнения чувствительности МЭА и МЭЛ

DOI: 10.14357/19922264160203

1 Введение

В [1, 2] метод ортогональных разложений (МОР) был развит для аналитического моделирования одно- и многомерных распределений в МСтС и дано его применение для задач надежности и безопасности технических систем. В [3] представлена теория субоптимальных фильтров (СОФ) на базе методов нормальной аппроксимации (МНА) и статистической линеаризации (МСЛ), а также МОР для МСтС с винеровскими шумами в уравнениях наблюдения и винеровскими и пуассоновскими шумами в уравнениях состояния. В основу СОФ были положены точные нелинейные уравнения для апостериорного одномерного распределения.

Рассмотрим развитие [3] на случай, когда апостериорное одномерное распределение ошибки фильтрации допускает эллипсоидальную аппроксимацию (ЭА) [4–7]. В разд. 2 и 3 приведены точные фильтрационные уравнения, а также уравнения точности и чувствительности на основе МОР. Элементы эллипсоидального анализа распределений даются в разд. 4. Раздел 5 содержит уравнения ЭСОФ на основе методов ЭА (МЭА) и эллипсоидальной линеаризации (МЭЛ). Заключение содержит выводы и некоторые обобщения.

2 Точные фильтрационные уравнения

На практике часто возникают задачи непрерывного определения состояния системы по результатам непрерывных наблюдений. Так как наблюдения всегда сопровождаются случайными ошибками, то следует говорить не об определении состояния системы, а о его оценивании (фильтрации, экстраполяции, интерполяции и т.д.) путем статистической обработки результатов наблюдений. Будем рассматривать задачи фильтрации состояния систем, моделями которых могут служить стохастические дифференциальные уравнения с винеровскими и пуассоновскими шумами.

Часто стохастические дифференциальные уравнения модели изучаемой системы могут иметь неизвестные параметры и, как правило, всегда содержат параметры, известные с ограниченной точностью. Поэтому возникает задача непрерывного оценивания неизвестных параметров системы (точнее, ее модели) по результатам непрерывных наблюдений. Предположим, что правые части уравнений зависят от конечного множества неизвестных параметров, которые будем рассматривать как компоненты век-

* Работа выполнена при поддержке РФФИ (проект 15-07-02244).

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

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

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

тора параметров θ . Одним из возможных подходов в таких случаях является следующий прием: неизвестный векторный параметр θ считают стохастическим процессом (СтП) $\Theta = \Theta_t$, который определяется дифференциальным уравнением $\dot{\Theta}_t = 0$, и включают компоненты этого векторного процесса в вектор состояния системы («расширяют» вектор состояния путем включения в него неизвестных параметров в качестве дополнительных компонент). Таким образом, задача непрерывного оценивания неизвестных параметров модели системы сводится к задаче непрерывного оценивания состояния системы с расширенным вектором состояния. От неизвестных параметров могут зависеть и уравнения наблюдения. Эти параметры следует включить в вектор θ и, следовательно, в расширенный вектор состояния.

Пусть векторный СтП $[X_t^T Y_t^T]^T$ определяется системой векторных стохастических дифференциальных уравнений Ито:

$$dX_t = \varphi(X_t, Y_t, \Theta, t) dt + \psi'(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''(X_t, Y_t, \Theta, t, v) P^0(dt, dv),$$

$$X(t_0) = X_0; \quad (1)$$

$$dY_t = \varphi_1(X_t, Y_t, \Theta, t) dt + \psi'_1(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''_1(X_t, Y_t, \Theta, t, v) P^0(dt, dv),$$

$$Y(t_0) = Y_0. \quad (2)$$

Здесь $Y_t = Y(t)$ — n_y -мерный наблюдаемый СтП, $Y_t \in \Delta^y$ (Δ^y — гладкое многообразие наблюдений); $X_t = X(t)$ — n_x -мерный ненаблюдаемый СтП (вектор состояния), $X_t \in \Delta^x$ (Δ^x — гладкое многообразие состояний); $W_0 = W_0(t)$ — n_w -мерный винеровский СтП ($n_w \geq n_y$) интенсивности $\nu_0 = \nu_0(\Theta, t)$; $P^0(\Delta, A) = P(\Delta, A) - \mu_P(\Delta, A)$, $P(\Delta, A)$ представляет собой для любого множества A простой пуассоновский СтП, а $\mu_P(\Delta, A)$ — его математическое ожидание, причем

$$\mu_P(\Delta, A) = MP(\Delta, A) = \int_{\Delta} \nu_P(\tau, A) d\tau;$$

$\nu_P(\Delta, A)$ — интенсивность соответствующего пуассоновского потока событий, $\Delta = (t_1, t_2]$; интегрирование по v распространяется на все пространство R^q с выколотым началом координат; Θ — вектор случайных параметров размерности n_Θ ; $\varphi = \varphi(X_t, Y_t, \Theta, t)$, $\varphi_1 = \varphi_1(X_t, Y_t, \Theta, t)$, $\psi' = \psi'(X_t, Y_t, \Theta, t)$ и $\psi'_1 = \psi'_1(X_t, Y_t, \Theta, t)$ — известные функции, отображающие $R^{n_x} \times R^{n_y} \times R$ со-

ответственно в R^{n_x} , R^{n_y} , $R^{n_x n_w}$ и $R^{n_y n_w}$; $\psi'' = \psi''(X_t, Y_t, \Theta, t, v)$ и $\psi''_1(X_t, Y_t, \Theta, t, v)$ — известные функции, отображающие $R^{n_x} \times R^{n_y} \times R^q$ в R^{n_x} и R^{n_y} . Требуется найти оценку \hat{X}_t СтП X_t в каждый момент времени t по результатам наблюдения СтП $Y(\tau)$ до момента t , $Y_{t_0}^t = \{Y(\tau) : t_0 \leq \tau < t\}$.

Следуя [8], предположим, что

- уравнение состояния имеет вид (1);
- уравнение наблюдения (2), во-первых, не содержит пуассоновского шума ($\psi''_1 \equiv 0$), а во-вторых, коэффициент при винеровском шуме ψ'_1 в уравнениях наблюдения не зависит от состояния ($\psi'_1(X_t, Y_t, \Theta, t) = \psi'_1(Y_t, \Theta, t)$).

В этом случае уравнения задачи нелинейной фильтрации имеют следующий вид:

$$dX_t = \varphi(X_t, Y_t, \Theta, t) dt + \psi'(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''(X_t, Y_t, \Theta, t, v) P^0(dt, dv), \quad X(t_0) = X_0; \quad (3)$$

$$dY_t = \varphi_1(X_t, Y_t, \Theta, t) dt + \psi_1(Y_t, \Theta, t) dW_0, \quad Y(t_0) = Y_0. \quad (4)$$

Будем считать, что выполнены условия существования и единственности СтП $[X_t^T Y_t^T]^T$, определяемого (3) и (4) при соответствующих начальных условиях.

В дальнейшем для стохастического уравнения

$$dZ = a dt + b dW_0 + \int_{R_0^q} c P^0(dt, dv) \quad (5)$$

потребуется обобщенная формула Ито [7] для дифференциала нелинейной функции $U = U(Z, t)$:

$$dU = \left\{ U_t + U_z^T a + \frac{1}{2} \text{tr} [U_{zz} b v b^T] \right\} dt + \int_{R_0^q} [U(Z + c, t)^T - U(Z, t)^T - U_z^T c] \mu_P(dt, dv) + U_z^T b dW_0 + \int_{R_0^q} [U(Z + c, t) - U(Z, t)] P^0(dt, dv). \quad (6)$$

Здесь a , b и c — известные функции Z и t .

Как известно [6, 7], для любых СтП X_t и Y_t оптимальная оценка \hat{X}_t , минимизирующая средний квадрат ошибки в каждый момент времени t , представляет собой апостериорное математическое ожидание СтП X_t : $\hat{X}_t = M [X_t | Y_{t_0}^t]$. Чтобы найти

это условное математическое ожидание, необходимо знать $p_t = p_t(x)$ — апостериорное одномерное распределение СтП X_t .

В основе уравнений оптимальной (в смысле минимума средней квадратической ошибки) фильтрации для уравнений (3) и (4) в силу (6) лежит следующая формула для стохастического дифференциала апостериорного математического ожидания скалярной функции $f = f(X, t)$ вектора состояния [8]:

$$\begin{aligned} d\hat{f} = & dM_{\Delta x}^{p_t} [f_t(X, t) + f_x(X, t)^T \varphi(X, Y, t) + \\ & + \frac{1}{2} \text{tr} \{ f_{xx}(X, t) (\psi' \nu_0 \psi'^T)(X, Y, t) \} + \\ & + \int_{R_0^q} \{ f(X + \psi'', t) - f(X, t) - \\ & - f_x(X, t)^T \psi''(X, Y, t) \} \nu_P(t, dv) | Y_{t_0}^t] dt + \\ & + M_{\Delta x}^{p_t} \{ f(X, t) [\varphi_1(X, Y, t)^T - \hat{\varphi}_1^T] + \\ & + f_x(X, t)^T (\psi \nu_0 \psi_1^T)(X, Y, t) | Y_{t_0}^t \} \times \\ & \times \psi_1 \nu_0 \psi_1^T)^{-1}(Y, t) (dY - \hat{\varphi}_1 dt). \quad (7) \end{aligned}$$

Здесь для краткости аргумент Θ опущен; $X = X_t$, $Y = Y_t$, $\nu = \nu_0$ и ν_P — интенсивности W_0 и P^0 ; $\hat{\varphi}_1$ — апостериорное математическое ожидание φ_1 при заданной условной плотности $p_t = p_t(x, \Theta)$:

$$\hat{\varphi}_1 = M_{\Delta x}^{p_t} [\varphi_1(X, Y, t)].$$

Полагая в (5) $f(X, t) \equiv g_t(\lambda, \Theta) = M_{\Delta x}^{p_t} [\exp(i\lambda^T X)]$, получим точное нелинейное фильтрационное уравнение для характеристической функции $g_t(\lambda, \Theta)$:

$$\begin{aligned} dg_t(\lambda, \Theta) = & M_{\Delta x}^{p_t} \left[\left\{ i\lambda^T \varphi(X_t, Y_t, \Theta, t) - \frac{1}{2} \lambda^T \times \right. \right. \\ & \times (\psi \nu_0 \psi^T)(X_t, Y_t, \Theta, t) \lambda + \gamma(\lambda, X_t, Y_t, \Theta, t) \left. \right\} \times \\ & \times e^{i\lambda^T X_t} | Y_{t_0}^t] dt + M_{\Delta x}^{p_t} \left[\left\{ \varphi_1(X_t, Y_t, \Theta, t)^T - \hat{\varphi}_1^T + \right. \right. \\ & + i\lambda^T (\psi \nu_0 \psi_1^T)(X_t, Y_t, \Theta, t) e^{i\lambda^T X_t} | Y_{t_0}^t \left. \right\} \times \\ & \times (\psi_1 \nu_0 \psi_1^T)^{-1}(Y_t, \Theta, t) (dY_t - \hat{\varphi}_1 dt) \left. \right], \quad (8) \end{aligned}$$

где

$$\begin{aligned} \gamma = \gamma(\lambda, X_t, Y_t, \Theta, t) = & \int_{R_0^q} \left[e^{i\lambda^T \psi''(X_t, Y_t, \Theta, t, v)} - \right. \\ & \left. - 1 - i\lambda^T \psi''(X_t, Y_t, \Theta, t, v) \right] \nu_P(\Theta, t, v) dv. \quad (9) \end{aligned}$$

Функции $g_t(\lambda, \Theta)$ и $p_t(x, \Theta)$ связаны между собой преобразованием Фурье [8].

Отсюда для гауссовской МСтС (3), (4) ($\psi'' \equiv 0$) уравнение (7) при $\gamma = 0$ упрощается и приобретает вид:

$$\begin{aligned} dg_t(\lambda, \Theta) = & M_{\Delta x}^{p_t} \left[\left\{ i\lambda^T \varphi(X_t, Y_t, \Theta, t) - \right. \right. \\ & - \frac{1}{2} \lambda^T (\psi \nu_0 \psi^T)(X_t, Y_t, \Theta, t) \lambda \left. \right\} e^{i\lambda^T X_t} | Y_{t_0}^t] dt + \\ & + M_{\Delta x}^{p_t} \left[\left\{ \varphi_1(X_t, Y_t, \Theta, t)^T - \hat{\varphi}_1^T + \right. \right. \\ & + i\lambda^T (\psi \nu_0 \psi_1^T)(X_t, Y_t, \Theta, t) e^{i\lambda^T X_t} | Y_{t_0}^t \left. \right\} \times \\ & \times (\psi_1 \nu_0 \psi_1^T)^{-1}(Y_t, \Theta, t) (dY_t - \hat{\varphi}_1 dt) \left. \right]. \quad (10) \end{aligned}$$

Если функция ψ'' в (3) допускает представление [8, 9]

$$\psi'' = \psi' \omega(\Theta, v), \quad (11)$$

где $P^0(\Delta, A) = P^0((0, t], dv)$, то уравнения (3), (4) примут следующий вид:

$$\begin{aligned} \dot{X}_t = & \varphi(X_t, Y_t, \Theta, t) + \psi'(X_t, Y_t, \Theta, t) V(\Theta, t), \\ X(t_0) = & X_0; \quad (12) \end{aligned}$$

$$\begin{aligned} \dot{Y}_t = & \varphi(X_t, Y_t, \Theta, t) + \psi_1(Y_t, \Theta, t) V_0(\Theta, t), \\ Y(t_0) = & Y_0. \quad (13) \end{aligned}$$

Здесь $V_0(\Theta, t) = \dot{W}_0(\Theta, t)$; $V(\Theta, t) = \dot{W}(\Theta, t)$,

$$\bar{W}(\Theta, t) = W_0(\Theta, t) + \int_{R_0^q} \omega(\Theta, v) P^0((0, t], dv),$$

где $\nu_P(\Theta, t, v) dv = [\partial \mu(\Theta, t, v) / \partial t] dv$ — интенсивность пуассоновского потока скачков, равных $\omega(\Theta, t)$. При этом логарифмические производные от одномерных характеристических функций определяются известными формулами:

$$\begin{aligned} \chi^{W_0}(\rho; t) = & -\frac{1}{2} \rho^T \nu_0(\Theta, t) \rho; \\ \chi^{\bar{W}}(\rho; t) = & -\frac{1}{2} \rho^T(\Theta, t) \rho^T + \\ & + \int_{R_0^q} \left[e^{i\rho^T \omega(\Theta, v)} - 1 - i\rho^T \omega(\Theta, v) \right] \nu_P(\Theta, t, v) dv. \quad (14) \end{aligned}$$

В таком случае уравнение для апостериорной одномерной характеристической функции имеет вид (8), где функция (9) допускает следующую запись:

$$\begin{aligned} \gamma = & \int_{R_0^q} \left[e^{i\lambda^T \psi'(X_t, Y_t, \Theta, t) \omega(\Theta, v)} - 1 - \right. \\ & \left. - i\lambda^T \psi'(X_t, Y_t, \Theta, t) \omega(\Theta, v) \right] \nu_P(\Theta, t, v) dv. \end{aligned}$$

Таким образом, можно сформулировать следующие утверждения разд. 2.

Теорема 2.1. Пусть для МСтС (3), (4) выполнены условия существования и единственности решения, а матрица $\sigma_1 = \psi_1 \nu_0 \psi_1^T$ не вырождена. Тогда при условии ограниченности соответствующих математических ожиданий точное фильтрационное уравнение для условной одномерной нормированной характеристической функции имеет вид (8).

Теорема 2.2. В условиях теоремы 2.1 при отсутствии пуассоновских шумов точное фильтрационное уравнение для условной одномерной нормированной характеристической функции имеет вид (10).

Теорема 2.3 [8, 9]. Пусть для МСтС (12), (13) выполнены условия существования и единственности решения, имеет место представление (11), а матрица $\sigma_1 = \psi_1 \nu_0 \psi_1^T$ не вырождена. Тогда при условии ограниченности соответствующих математических ожиданий точное фильтрационное уравнение имеет вид (8) при условии (9).

Как известно [8], точное решение фильтрационных уравнений возможно только в случаях, когда уравнения гауссовской дифференциальной МСтС линейны или линейны лишь относительно вектора состояния X_t при независимой от состояния функции ψ . Эти уравнения дают точное решение задачи оптимальной нелинейной фильтрации. Однако это решение не может быть реализовано практически. Для нахождения оптимальной оценки вектора состояния необходимо решить фильтрационное уравнение для апостериорной характеристической функции (или фильтрационное уравнение для апостериорной плотности вектора состояния X_t) после получения результатов наблюдений, затем вычислить оптимальную оценку вектора X_t . Но методов точного решения этих уравнений в общем случае пока еще не существует.

В задачах реального времени численное решение фильтрационных уравнений (или онлайн-оценивания) тоже невозможно, так как для этого требуется много времени, а решать их необходимо каждый раз после получения результатов наблюдений. Кроме того, практическое применение точной теории оптимальной нелинейной фильтрации имеет смысл только в тех случаях, когда оценки можно вычислять в реальном масштабе времени по мере получения результатов наблюдений. Точная теория дает оптимальные оценки в каждый момент t по результатам наблюдений, полученным к этому моменту, без использования последующих результатов наблюдений. Если эти оценки не могут быть вычислены в тот же момент t или хотя бы с фиксированным приемлемым запаздыванием и их вычисление приходится откладывать на будущее, то

нет никакого смысла отказываться от использования наблюдений, получаемых после момента t , для оценивания состояния системы в момент t . Поэтому для статистической обработки результатов после окончания наблюдений, т. е. для офлайн-оценивания, целесообразно применять известные из математической статистики методы постобработки информации [8].

Необходимость обработки результатов наблюдений в реальном масштабе времени непосредственно в процессе эксперимента привела к появлению ряда приближенных методов оптимальной нелинейной фильтрации, называемых обычно методами субоптимальной фильтрации [8]. Одни приближенные методы основаны на приближенном решении фильтрационных уравнений, а другие — на превращении формул для стохастических дифференциалов оптимальной оценки \hat{X}_t и апостериорной ковариационной матрицы ошибки R_t в стохастические дифференциальные уравнения для \hat{X}_t и R_t путем разложения функций φ , φ_1 , ψ_1 или φ , φ_1 , ψ' , ψ'' , ψ , ψ_1 в степенные ряды и отбрасывания остаточных членов.

Для приближенного решения уравнения для апостериорной одномерной характеристической функции $g_1(\lambda, \Theta)$ вектора X_t можно использовать методы аналитического моделирования, основанные на параметризации одномерных распределений СтП, определяемого стохастическим дифференциальным уравнением [8]. Эти методы позволяют изучить стохастические дифференциальные уравнения для параметров апостериорного распределения. Простейшим таким методом является МНА апостериорного распределения. Исключительно важное практическое значение имеют квазилинейные фильтры, получаемые с помощью методов эквивалентной линеаризации [8].

3 Субоптимальные фильтры на основе метода ортогональных разложений

При аппроксимации апостериорной одномерной плотности отрезком ее ортогонального разложения [1, 2]

$$p_t(x, \Theta) \approx p^*(x; \Theta, \vartheta) = w(x; \Theta) \left[1 + \sum_{l=3}^N \sum_{|\nu|=l} c_{\nu} p_{\nu}(x) \right] \quad (15)$$

естественно принять за параметры, образующие вектор ϑ , апостериорные математическое ожидание \hat{X}_t , ковариационную матрицу R_t вектора X_t ,

а также коэффициенты ортогонального разложения (КОР) c_κ ($|\nu| = 3, \dots, N$). Здесь КОР определяется формулой:

$$c_\kappa = \left[q_\kappa \left(\frac{\partial}{i\partial\lambda} \right) g_t(\lambda, \Theta) \right]_{\lambda=0}. \quad (16)$$

Заметим, что полином q_κ зависит от \hat{X}_t и R_t .

На основе (7) и (8) для гауссовской МСТС (3), (4) при $\psi'' = 0$ получим, что ортогональный СОФ (ОСОФ) определяется следующими уравнениями:

$$d\hat{X}_t = f dt + h (dY_t - f^{(1)} dt); \quad (17)$$

$$dR_t = \left(f^{(2)} - h\psi_1\nu_0\psi_1^T h^T \right) dt + \sum_{r=1}^{n_y} \rho_r (dY_r - f_r^{(1)} dt). \quad (18)$$

Здесь введены обозначения:

$$\left. \begin{aligned} f &= f(Y_t, \vartheta, \Theta, t) = M_{\Delta x}^* [\varphi(Y_t, X, \Theta, t)]; \\ f^{(1)} &= \{f_r^{(1)}\} = f^{(1)}(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* [\varphi_1(Y_t, X, \Theta, t)]; \\ f^{(2)} &= f^{(2)}(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* \left[(X - \hat{X}_t) \varphi(Y_t, X, \Theta, t)^T + \right. \\ &\quad \left. + \varphi(Y_t, X, \Theta, t) (X^T - \hat{X}_t^T) + \right. \\ &\quad \left. + (\psi\nu_0\psi^T)(Y_t, X, \Theta, t) \right]; \\ h &= h(Y_t, \vartheta, \Theta, t) = \\ &= \left\{ M_{\Delta x}^* [X\varphi_1(Y_t, X, \Theta, t)^T + \right. \\ &\quad \left. + (\psi\nu_0\psi_1^T)(Y_t, X, \Theta, t)] - \right. \\ &\quad \left. - \hat{X}_t f^{(1)T} \right\} (\psi_1\nu_0\psi_1^T)^{-1}(Y_t, \Theta, t); \\ \rho_r &= \rho_r(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* \left[(X - \hat{X}_t) (X^T - \hat{X}_t^T) \times \right. \\ &\quad \left. + a_r(Y_t, X, \Theta, t) + (X - \hat{X}_t) b_r(Y_t, X, \Theta, t)^T + \right. \\ &\quad \left. + b_r(Y_t, X, \Theta, t) (X^T - \hat{X}_t^T) \right] \quad (r = 1, \dots, n_y). \end{aligned} \right\} \quad (19)$$

Далее перепишем (17), (18) покоординатно:

$$d\hat{X}_s = f_s dt + h_s (dY_t - f^{(1)} dt) = A^{\hat{X}_s} dt + B^{\hat{X}_s} dY_t \quad (s = 1, \dots, n_x); \quad (20)$$

$$dR_{sq} = \left(f_{sq}^{(2)} - h_s\psi_1\nu_0\psi_1^T h_q^T \right) dt + \eta_{sq} (dY_t - f^{(1)} dt) = A^{R_{sq}} dt + B^{R_{sq}} dY_t, \quad (21)$$

где $\hat{X}_s(t_0) = X_{s0}$; $R_{sq}(t_0) = R_{sq0}$; $s, q = 1, \dots, n_x$; η_{sq} — матрица-строка, элементами которой служат соответствующие элементы матрицы $\rho_1, \dots, \rho_{n_1}$:

$$\eta_{sq} = \eta_{e_s + e_q} = [\rho_{1sq} \dots \rho_{msq}] \quad (s, q, = 1, \dots, n_x).$$

Здесь и далее для краткости индекс t сохраним только у Y_t . По формуле дифференцирования Ито для винеровского СтП, учитывая (20) и (21), находим в силу (16) стохастический дифференциал:

$$\begin{aligned} dc_\kappa &= \left[d \left\{ q_\kappa \left(\frac{\partial}{i\partial\lambda} \right) g_t(\lambda, \Theta) \right\} \right]_{\lambda=0} = \\ &= \sum_{s=1}^{n_x} \left[\partial q_\kappa \left(\frac{\partial}{i\partial\lambda} \right) \partial \hat{X}_s g_t(\lambda, \Theta) \right]_{\lambda=0} d\hat{X}_s + \\ &+ \sum_{s,u=1}^{n_x} \left[\partial q_\kappa \left(\frac{\partial}{i\partial\lambda} \right) \partial R_{su} g_t(\lambda, \Theta) \right]_{\lambda=0} dR_{su} + \\ &+ \left[q_\kappa \left(\frac{\partial}{i\partial\lambda} \right) dg_t(\lambda, \Theta) \right]_{\lambda=0} + \\ &+ \left\{ \frac{1}{2} \sum_{s,u=1}^{n_x} \left[\frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) g_t(\lambda, \Theta)}{\partial \hat{X}_s \partial \hat{X}_u} \right]_{\lambda=0} h_s \psi_1 \nu_0 \psi_1^T h_u^T + \right. \\ &+ \frac{1}{2} \sum_{s,u,k,l=1}^{n_x} \left[\frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) g_t(\lambda, \Theta)}{\partial R_{su} \partial R_{kl}} \right]_{\lambda=0} \times \\ &\quad \left. \times \eta_{su} \psi_1 \nu_0 \psi_1^T \eta_{kl}^T + \right. \\ &+ \left. \sum_{s,k,l=1}^{n_x} \left[\frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) g_t(\lambda, \Theta)}{\partial \hat{X}_s \partial R_{kl}} \right]_{\lambda=0} h_s \psi_1 \nu_0 \psi_1^T \eta_{kl}^T \right\} dt. \end{aligned}$$

Подставив сюда выражения (20), (21) и (7) дифференциалов $d\hat{X}_s$, dR_{sq} и $dg_t(\lambda, \Theta)$ и вспомнив, что для любого полинома $P(x)$ $P[(\partial/(i\partial\lambda))g_t(\lambda)]_{\lambda=0} = P(\alpha)$, получаем стохастические дифференциальные уравнения:

$$\begin{aligned} dc_\kappa &= \left\{ F_\kappa + \sum_{s=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s} f_s + \right. \\ &+ \sum_{s,u=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial R_{su}} (f_{su}^{(2)} - h_s\psi_1\nu_0\psi_1^T h_u^T) + \\ &+ \frac{1}{2} \sum_{s,u=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial \hat{X}_u} h_s \psi_1 \nu_0 \psi_1^T h_u^T + \\ &+ \frac{1}{2} \sum_{s,u,k,l=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial R_{su} \partial R_{kl}} \eta_{su} \psi_1 \nu_0 \psi_1^T \eta_{kl}^T + \\ &+ \left. \sum_{s,k,l=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{kl}} h_s \psi_1 \nu_0 \psi_1^T \eta_{kl}^T \right\} dt + \\ &+ \left\{ H_\kappa + \sum_{s=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s} h_s + \sum_{s,u=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial R_{su}} \eta_{su} \right\} \times \end{aligned}$$

$$\begin{aligned} \times \left(dY_t - f^{(1)} dt \right) &= A^{c_\kappa} dt + B^{c_\kappa} dY_t, \\ c_\kappa(t_0) &= c_{\kappa 0} \quad (|\kappa| = 3, \dots, N). \end{aligned} \quad (22)$$

Здесь в дополнение к прежним обозначениям принято:

$$\left. \begin{aligned} F_\kappa &= F_\kappa(Y_t, \Theta, \vartheta, t) = \\ &= \sum_{s=1}^{n_x} M_{\Delta x}^{p_s^*} \left[\varphi_s(Y_t, X, \Theta, t) \frac{\partial q_\kappa(X)}{\partial X_s} \right] + \\ &+ \frac{1}{2} \sum_{s,u=1}^{n_x} M_{\Delta x}^{p_{su}^*} \left[\sigma_{su}(Y_t, X, \Theta, t) \frac{\partial^2 q_\kappa(X)}{\partial X_s \partial X_u} \right]; \\ H_\kappa &= H_\kappa(Y_t, \vartheta, \Theta, t) = \\ &= \left\{ M_{\Delta x}^{p_s^*} \left[\varphi_1(Y_t, X, \Theta, t)^T q_\kappa(X) \right] + \right. \\ &+ \sum_{s=1}^{n_x} M_{\Delta x}^{p_s^*} \left[(\psi \nu_0 \psi_1^T)_s(Y_t, X, \Theta, t) \frac{\partial q_\kappa(X)}{\partial X_s} \right] - \\ &\left. - c_\kappa f^{(1)T} \right\} (\psi_1 \nu_0 \psi_1^T)^{-1}(Y_t, \Theta, t), \end{aligned} \right\} (23)$$

где через $(\psi \nu_0 \psi_1^T)_s$ обозначена s -я строка матрицы $\psi \nu_0 \psi_1^T$; $\sigma = \psi \nu_0 \psi_1^T = \{\sigma_{su}\}$.

Функции $f_s, f^{(1)}, f_{su}^{(2)}, h_s, \eta_{su}, F_\kappa$ и H_κ в уравнениях (20), (21) и (22) представляют собой линейные комбинации величин c_ν ($|\nu| = 3, \dots, N$) с коэффициентами, зависящими от \hat{X}_t и R_t . Величины $\partial q_\kappa(\alpha)/\partial \hat{X}_s, \partial q_\kappa(\alpha)/\partial R_{su}, \partial^2 q_\kappa(\alpha)/(\partial \hat{X}_s \partial \hat{X}_u), \partial^2 q_\kappa(\alpha)/(\partial R_{su} \partial R_{kl})$ и $\partial^2 q_\kappa(\alpha)/(\partial \hat{X}_s \partial R_{kl})$ после замены моментов их выражениями через c_ν тоже будут линейными комбинациями величин c_ν с коэффициентами, зависящими от \hat{X}_t и R_t .

Таким образом, имеем следующие утверждения.

Теорема 3.1. Пусть МСтС (3), (4) — гауссовская ($\psi'' = 0$), выполнены условия существования и единственности решения, а матрица $\sigma_1 = \psi_1 \nu_0 \psi_1^T$ не вырождена. Тогда в основе алгоритма ОСОФ по МОР лежат уравнения (15), (20)–(22) при условии ограниченности функций (23).

Теорема 3.2. Пусть для МСтС (3), (4) выполнены условия существования и единственности решения, а матрица $\sigma_1 = \psi_1 \nu_0 \psi_1^T$ не вырождена. Тогда алгоритм ОСОФ согласно МОР задается уравнениями (15), (20)–(22) при условии ограниченности функций $f, f^{(1)}, \bar{f}^{(2)}, h, \rho_r, F_\kappa$ и H_κ , определяемых (19) и (23).

Применяя методы теории чувствительности [9, 10] для приближенного анализа фильтрационных уравнений и учитывая случайность параметров Θ , приходим к следующим уравнениям для функций чувствительности первого порядка:

$$\begin{aligned} d\nabla^\Theta \hat{X}_s &= \nabla^\Theta A^{\hat{X}_s} dt + \nabla^\Theta B^{\hat{X}_s} dY_t, & \nabla^\Theta B^{\hat{X}_s}(t_0) &= 0; \\ d\nabla^\Theta R_{sq} &= \nabla^\Theta A^{R_{sq}} dt + \nabla^\Theta B^{R_{sq}} dY_t, & \nabla^\Theta R_{sq}(t_0) &= 0; \\ d\nabla^\Theta c_\kappa &= \nabla^\Theta A^{c_\kappa} dt + \nabla^\Theta B^{c_\kappa} dY_t, & \nabla^\Theta c_\kappa(t_0) &= 0. \end{aligned}$$

Здесь процедура взятия производных осуществляется по всем входящим переменным, а коэффициенты чувствительности вычисляются при $\Theta = m^\Theta$. При этом предполагается малость дисперсий по сравнению с их математическими ожиданиями. Очевидно, что при дифференцировании по Θ ($\nabla^\Theta = \partial/\partial\Theta$) порядок уравнений возрастает пропорционально числу производных.

Аналогично составляются уравнения для элементов матриц вторых функций чувствительности.

Для оценки качества ОСОФ, следуя [1, 2], при гауссовских Θ с математическим ожиданием m^Θ и ковариационной матрицей K^Θ введем условную функцию потерь, допускающую квадратичную аппроксимацию:

$$\begin{aligned} \rho^{\hat{X}_s} &= \rho^{\hat{X}_s}(\Theta) = \rho(m^\Theta) + \sum_{ii=1}^{n^\Theta} \rho'_i(m^\Theta) \Theta_s^0 + \\ &+ \sum_{i,j=1}^{n^\Theta} \sum_{i,j=1}^{n^\Theta} \rho''_{ij}(m^\Theta) \Theta_i^0 \Theta_j^0, \end{aligned}$$

а также показатель ε :

$$\varepsilon = \varepsilon_2^{1/4}.$$

Здесь введены обозначения:

$$\begin{aligned} \varepsilon_2 &= M^N [\rho(\Theta)^2] - \rho(m^\Theta)^2; \\ M^N [\rho(\Theta)^2] &= \rho(m^\Theta)^2 + \rho'(m^\Theta)^T K^\Theta \rho'(m^\Theta) + \\ &+ 2\rho(m^\Theta) \text{tr} [\rho''(m^\Theta) K^\Theta] + \\ &+ \{ \text{tr} [\rho''(m^\Theta) K^\Theta] \}^2 + 2\text{tr} [\rho''(m^\Theta) K^\Theta]^2, \end{aligned}$$

а функции ρ' и ρ'' по известным формулам [9, 10] определяются на основе первых и вторых функций чувствительности.

Изложенные выше методы синтеза ОСОФ дают принципиальную возможность получить фильтр, близкий к оптимальному по оценке с любой степенью точности. Чем выше максимальный порядок учитываемых моментов, КОР, тем выше будет точность приближения к оптимальной оценке. Однако число уравнений, определяющих параметры апостериорного одномерного распределения, быстро растет с увеличением числа учитываемых параметров. Соответствующие оценки можно найти в [4–6].

4 Эллипсоидальная аппроксимация распределений

Для конечномерных МСтС часто оказывается полезной структурная аппроксимация распределений посредством эллипсоидальных распределений. Следуя [4, 11], для структурной аппроксимации плотностей вероятности случайных векторов будем использовать плотности, имеющие эллипсоидальную структуру, т. е. плотности, у которых поверхностями уровней равной вероятности являются подобные концентрические эллипсоиды (эллипсы для двумерных векторов, эллипсоиды для трехмерных векторов, гиперэллипсоиды для векторов размерности больше трех). В частности, эллипсоидальную структуру имеет нормальное распределение в любом конечномерном пространстве. Характерная особенность таких распределений состоит в том, что их плотности вероятности являются функциями положительно определенной квадратичной формы $u = u(y) = (y^T - m^T)C(y - m)$, где m — математическое ожидание случайного вектора Y ; C — некоторая положительно определенная матрица.

Для нахождения ЭА плотности вероятности r -мерного случайного вектора будем пользоваться конечным отрезком разложения по биортонормальной системе полиномов $\{p_{r,\nu}(u(y)), q_{r,\nu}(u(y))\}$, которые зависят только от квадратичной формы $u = u(y)$ и функцией веса для которых служит некоторая плотность вероятности эллипсоидальной структуры $w(u(y))$:

$$M_{\Delta y}^w [w(u(Y))p_{r,\nu}(Y)q_{r,\mu}(u(Y))] = \delta_{\nu\mu}. \quad (24)$$

Индексы ν и μ у полиномов означают их степень относительно переменной u . Конкретный вид и свойства полиномов определены ниже. Однако без потери общности можно принять, что $q_{r,0}(u) = p_{r,0}(u) = 1$. Тогда плотность вероятности вектора Y может быть приближенно представлена в виде:

$$f(y) \approx f^*(u) = w(u) \sum_{\nu=0}^N c_{r,\nu} p_{r,\nu}(u), \quad (25)$$

где коэффициенты $c_{r,\nu}$ определяются по формуле:

$$c_{r,\nu} = M_{\Delta y}^{\text{ЭА}} [q_{r,\nu}(U)] \quad (\nu = 1, \dots, N). \quad (26)$$

Учитывая, что $p_{r,0}(u)$ и $q_{r,0}(u)$ — взаимно обратные постоянные (полиномы нулевой степени), то всегда $c_{r,0}p_{r,0} = 1$. Поэтому из формулы (25) следует, что

$$f(y) \approx f^*(u) = w(u) \left[1 + \sum_{\nu=2}^N c_{r,\nu} p_{r,\nu}(u) \right].$$

Для приложений большое значение имеет случай, когда за распределение $w(u)$ выбирается нормальное (гауссовское) распределение:

$$w(u) = w(y^T C y) = \frac{1}{\sqrt{(2\pi)^r |K|}} \exp \left(-y^T K^{-1} \frac{y}{2} \right).$$

Учитывая, что $C = K^{-1}$, приведем условие биортонормальности (24) к виду:

$$\frac{1}{2^{r/2} \Gamma(r/2)} \int_0^\infty p_{r,\nu}(u) q_{r,\mu}(u) u^{r/2-1} e^{-u/2} du = \delta_{\nu\mu}.$$

Задача выбора системы полиномов $\{p_{r,\nu}(u), q_{r,\mu}(u)\}$, используемой при ЭА (25) и (26), сводится к нахождению биортонормальной системы полиномов, для которой весом служит χ^2 -распределение с r степенями свободы, при этом используются следующие формулы:

$$\begin{aligned} p_{r,\nu}(u) &= S_{r,\nu}(u), \quad q_{r,\nu}(u) = \\ &= \frac{(r-2)!!}{(r+2\nu-2)!!(2\nu)!!} S_{r,\nu}(u), \quad r \geq 2; \end{aligned}$$

$$\begin{aligned} S_{r,\nu}(u) &= S_\nu^{r/2-1}(u) = \\ &= \sum_{\mu=0}^{\nu} (-1)^{\nu+\mu} C_\nu^\mu \frac{(r+2\nu-2)!!}{(r+2\mu-2)!!} u^\mu. \end{aligned}$$

При разложении по полиномам $S_{r,\nu}(u)$ плотности вероятности случайного вектора Y и всех его возможных проекций согласованы.

5 Субоптимальные фильтры на основе метода эллипсоидальной аппроксимации (линеаризации)

Применим МЭА для приближенного решения задачи оптимальной нелинейной фильтрации в дифференциальной гауссовской СтС. Для этого аппроксимируем апостериорную плотность $p_t(x)$ формулой:

$$p_t(x) = p^*(x; \Theta) = w_1(u_t) \left[1 + \sum_{\nu=2}^N c_\nu p_\nu(u_t) \right], \quad (27)$$

где

$$u_t = (x^T - \hat{X}_t^T) C_t (x - \hat{X}_t);$$

C_t — матрица, обратная по отношению к ковариационной матрице ошибки фильтрации R_t , $C_t = R_t^{-1}$;

$$c_\kappa = M_{\Delta^x}^{f_1} [q_\kappa(U_t)] = \left[q_\kappa \left(\frac{\partial^T}{i\partial\lambda} - m_t^T \right) \times \right. \\ \left. \times C_t \left(\frac{\partial^T}{i\partial\lambda} - m_t \right) g_1(\lambda; t) \right]_{\lambda=0}. \quad (28)$$

Чтобы найти стохастический дифференциал c_κ , применим формулу Ито [7], учитывая, что c_κ представляет собой функцию трех случайных процессов \hat{X}_t , R_t и $g_t(\lambda)$, стохастические дифференциалы Ито которых определяются формулами (20) и (21). В результате получим уравнение (22), которое на основании равенства $q_\kappa(\alpha) = M_{\Delta^x}^{f_1} q_\kappa(U_t)$ имеет вид:

$$dc_\kappa = \left\{ F_\kappa + \sum_{s=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial q_\kappa(U_t)}{\partial \hat{X}_s} \right] + \right. \\ \left. + \sum_{s=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial q_\kappa(U_t)}{\partial R_{su}} \left(f_{su}^{(2)} - h_s \psi_1 \nu \psi_1^T h_u^T \right) \right] + \right. \\ \left. + \frac{1}{2} \sum_{s,u=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial \hat{X}_s \partial \hat{X}_u} h_s \psi_1 \nu \psi_1^T h_u^T \right] + \right. \\ \left. + \frac{1}{2} \sum_{s,u,k,l=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial R_{su} \partial R_{kl}} \eta_{su} \psi_1 \nu \psi_1^T \eta_{kl}^T \right] + \right. \\ \left. + \sum_{s,k,l=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial \hat{X}_s \partial R_{kl}} h_s \psi_1 \nu \psi_1^T \eta_{kl}^T \right] \right\} dt + \\ + \left\{ H_\kappa + \sum_{s=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial q_\kappa(U_t)}{\partial \hat{X}_s} h_s \right] + \right. \\ \left. + \sum_{s,u=1}^n M_{\Delta^x}^{f_1} \left[\frac{\partial q_\kappa(U_t)}{\partial R_{su}} \eta_{su} \right] \right\} (dY_t - f^{(1)} dt). \quad (29)$$

Вычислим входящие в (29) математические ожидания производных полинома $q_\kappa(U_t)$. Имеем:

$$M_{\Delta^x}^{f_1} \left[\frac{\partial q_\kappa(U_t)}{\partial \hat{X}_s} \right] = M_{\Delta^x}^{f_1} \left[q'_\kappa \frac{\partial U}{\partial \hat{X}_s} \right] = \\ = M_{\Delta^y}^{f_1} \left[q'_\kappa(U_t) \left(-2 \sum_{j=1}^n c_{sj} (X_j - \hat{X}_j) \right) \right] = 0; \quad (30) \\ M_{\Delta^x}^{f_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial \hat{X}_s \partial \hat{X}_u} \right] = \\ = M_{\Delta^x}^{f_1} \left[4q''_\kappa(U_t) \left(\sum_{j=1}^n c_{uj} (X_j - \hat{X}_j) \right) \times \right. \\ \left. \times \left(\sum_{j=1}^n c_{sj} (X_j - \hat{X}_j) \right) + 2q'_\kappa(U_t) c_{su} \right]; \quad (31)$$

$$M_{\Delta^y}^{f_1} \left[q''_\kappa(U_t) C_t (X_t - \hat{X}_t) (X_t^T - \hat{X}_t^T) C_t \right] = \\ = \left\{ \frac{a}{n} M_{\Delta^U}^{w_1} \left[q''_\kappa(U) U^{n/2} \right] + \right. \\ \left. + \sum_{\nu=2}^N c_\nu \frac{a}{n} M_{\Delta^U}^{w_1 p_\nu} \left[q''_\kappa(U) U^{n/2} \right] \right\} C_t,$$

где нормирующий множитель a определяется соотношением:

$$a^{-1} = M_{\Delta^U}^{w_1 p_\nu} \left[U^{n/2-1} \right].$$

Введя обозначения

$$\xi_{\kappa 0} = \frac{a}{n} M_{\Delta^U}^{w_1} \left[q''_\kappa(U) U^{n/2} \right];$$

$$\xi_{\kappa \nu} = \frac{a}{n} M_{\Delta^U}^{w_1 p_\nu} \left[q''_\kappa(U) U^{n/2} \right]$$

и заметив, что вследствие ортогональности $p_\nu(u)$ ко всем функциям u^λ при $\lambda < \nu'$ величина $\xi_{\kappa \nu}$ обращается в нуль при $\nu > \kappa - 1$, получим:

$$M_{\Delta^U}^{w_1} \left[q''_\kappa(U_t) C_t (X_t - \hat{X}_t) (X_t^T - \hat{X}_t^T) C_t \right] = \\ = \left(\xi_{\kappa 0} + \sum_{\nu=2}^{\kappa-1} c_\nu \xi_{\kappa \nu} \right) C_t. \quad (32)$$

На основании (32) имеем:

$$4M_{\Delta^U}^{w_1} \left[q''_\kappa(U_t) \left(\sum_{j=1}^n c_{uj} (X_j - \hat{X}_j) \right) \times \right. \\ \left. \times \left(\sum_{j=1}^n c_{sj} (X_j - \hat{X}_j) \right) \right] = \\ = 4 \left(\xi_{\kappa 0} + \sum_{\nu=2}^{\kappa-1} c_\nu \xi_{\kappa \nu} \right) c_{su}. \quad (33)$$

Математическое ожидание во втором слагаемом в (31) определяется формулой:

$$M_{\Delta^U}^{w_1} [q'_\kappa(U_t)] = M_{\Delta^U}^{w_1} \left[q'_\kappa(U) U^{n/2-1} \right] + \\ + \sum_{\nu=2}^N c_\nu M_{\Delta^U}^{w_1 p_\nu} [q'_\kappa(U)].$$

Введя обозначения

$$\zeta'_{\kappa 0} = a M_{\Delta^U}^{w_1} \left[q'_\kappa(U) U^{n/2-1} \right];$$

$$\zeta'_{\kappa \nu} = a M_{\Delta^U}^{w_1 p_\nu} \left[q'_\kappa(U) U^{n/2-1} \right]$$

и заметив, что вследствие ортогональности $p_\nu(u)$ ко всем функциям u^λ при $\lambda < \nu$ величина $\zeta_{\kappa\nu}$ обращается в нуль при $\nu > \kappa - 1$, получим:

$$M_{\Delta U}^{w_1} [q'_\kappa(U_t)] = \zeta'_{\kappa 0} + \sum_{\nu=2}^{\kappa-1} c_\nu \zeta'_{\kappa\nu}. \quad (34)$$

На основании (33) и (34) формула (31) принимает вид:

$$M_{\Delta U}^{w_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial \hat{X}_s \partial \hat{X}_u} \right] = 2 \left(\zeta_{\kappa 0} + \sum_{\nu=2}^{\kappa-1} c_\nu \zeta_{\kappa\nu} \right) c_{su},$$

где для краткости положено:

$$\zeta_{\kappa 0} = 2\xi_{\kappa 0} + \zeta'_{\kappa 0}; \quad \zeta_{\kappa\nu} = 2\xi_{\kappa\nu} + \zeta'_{\kappa\nu}.$$

Вычислив производные $\partial q_\kappa(U_t)/\partial R_{su}$, получим:

$$M_{\Delta U}^{w_1} \left[\frac{\partial q_\kappa(U_t)}{\partial R_{ss}} \right] = -c_{ss} \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right);$$

$$M_{\Delta U}^{w_1} \left[\frac{\partial q_\kappa(U_t)}{\partial R_{su}} \right] = -2c_{su} \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right), \quad s \neq u,$$

где $\gamma_{\kappa 0}, \gamma_{\kappa 2}, \dots, \gamma_{\kappa\kappa}$ определяются формулами:

$$\gamma_{\kappa 0} = \frac{a}{n} M_{\Delta U}^{w_1} [q'_\kappa(U)U^{n/2}];$$

$$\gamma_{\kappa\nu} = \frac{a}{n} M_{\Delta U}^{w_1 p_\nu} [q'_\kappa(U)U^{n/2}].$$

Дифференцируя эти формулы по компонентам вектора \hat{X}_t и элементам матрицы R_t и имея в виду, что

$$\frac{\partial c_{ij}}{\partial R_{rr}} = -c_{ri}c_{rj}; \quad \frac{\partial c_{ij}}{\partial R_{rs}} = -(c_{ri}c_{sj} + c_{si}c_{rj}), \quad (35)$$

получаем:

$$M_{\Delta U}^{w_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial \hat{X}_k \partial R_{su}} \right] = 0 \quad (k, s, u = 1, \dots, n); \quad (36)$$

$$M_{\Delta U}^{w_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial R_{ss} \partial R_{rr}} \right] = c_{rs}^2 \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right); \quad (37)$$

$$M_{\Delta U}^{w_1} \left[\frac{\partial^2 q_\kappa(U_t)}{\partial R_{ss} \partial R_{kl}} \right] = 2c_{ls}c_{ks} \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right), \quad k \neq l; \quad (38)$$

$$M_{\Delta U}^{w_1} \left[\frac{\partial q_\kappa(U_t)}{\partial R_{su} \partial R_{kl}} \right] = 2(c_{ks}c_{lu} + c_{ls}c_{ku}) \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right), \quad s \neq u, \quad k \neq l. \quad (39)$$

Подставив выражения (30), (35)–(39) в уравнение для стохастического дифференциала величины c_κ , приведем его к виду:

$$dc_\kappa = \left\{ F_\kappa - \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right) \times \right. \\ \times \text{tr} \left[\bar{C}_t \left(f^{(2)} - h\psi_1\nu\psi_1^T h^T \right) \right] + \\ \left. + \left(\zeta_{\kappa 0} + \sum_{\nu=2}^{\kappa-1} c_\nu \zeta_{\kappa\nu} \right) \text{tr} \left[C_t h\psi_1\nu\psi_1^T h^T \right] + \right. \\ \left. + \frac{1}{2} \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right) \sum_{s,u,k,l=1}^n A_{sukl} \eta_{su} \psi_1 \nu \psi_1^T \eta_{kl}^T \right\} dt + \\ + \left\{ H_\kappa - \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right) \sum_{s,u=1}^n \bar{c}_{su} \eta_{su} \right\} \times \\ \times (dY_t - f^{(1)} dt) \quad (\kappa = 2, \dots, N), \quad (40)$$

где

$$A_{ssrr} = c_{sr}^2, \quad A_{sskl} = 2c_{ks}c_{ls}, \quad k \neq l;$$

$$A_{sukl} = 2(c_{ks}c_{lu} + c_{ls}c_{ku}), \quad s \neq u, \quad k \neq l;$$

\bar{c}_{su} — элементы матрицы \bar{C}_t , определяемой формулой:

$$M_{\Delta U}^{w_1} [q_\kappa^K(U_t)] = - \left(\gamma_{\kappa 0} + \sum_{\nu=2}^{\kappa} c_\nu \gamma_{\kappa\nu} \right) C_t.$$

Формулы (19) для $f, f^{(1)}, f^{(2)}, h$ и ρ_r при аппроксимации (27) апостериорной плотности принимают вид:

$$f = f(Y_t, \Theta, t) = M_{\Delta^x}^{w_1} [\varphi(Y_t, X, \Theta, t)] + \\ + \sum_{\nu=2}^N c_\nu M_{\Delta^x}^{w_1 p_\nu} [\varphi(Y_t, X, \Theta, t)]; \quad (41)$$

$$f^{(1)} = f^{(1)}(Y_t, \Theta, t) = M_{\Delta^x}^{w_1} [\varphi_1(Y_t, X, \Theta, t)] + \\ + \sum_{\nu=2}^N c_\nu M_{\Delta^x}^{w_1 p_\nu} [\varphi_1(Y_t, X, \Theta, t)]; \quad (42)$$

$$f^{(2)} = f^{(2)}(Y_t, \Theta, t) = \\ = M_{\Delta^x}^{w_1} \left[(X - \hat{X}) \varphi(Y_t, X, \Theta, t)^T + \right. \\ \left. + \varphi(Y_t, X, \Theta, t) (X^T - \hat{X}_t^T) + \right. \\ \left. + (\psi\nu\psi^T)(Y_t, X, \Theta, t) \right] + \\ + \sum_{\nu=2}^N c_\nu M_{\Delta^x}^{w_1 p_\nu} \left[(X - \hat{X}) \varphi(Y_t, X, \Theta, t)^T + \right.$$

$$+ \varphi(Y_t, X, \Theta, t) \left(X - \hat{X} \right)^T + \left. \left. + (\psi \nu \psi^T) (Y_t, X, \Theta, t) \right] ; \quad (43)$$

$$h = h(Y_t, \Theta, t) = \mathbf{M}_{\Delta^x}^{w_1} \left[X \varphi_1(Y_t, X, \Theta, t)^T + (\psi \nu \psi^T) (Y_t, X, \Theta, t) \right] + \sum_{\nu=2}^N c_\nu \mathbf{M}_{\Delta^x}^{w_{1p\nu}} \left[X \varphi(Y_t, X, \Theta, t)^T + (\psi \nu \psi^T) (Y_t, X, \Theta, t) \right] - \hat{X}_t f^{(1)T} (\psi_1 \nu \psi_1^T)^{-1} (Y_t, X, \Theta, t) ; \quad (44)$$

$$\rho_r = \rho_r(Y_t, \Theta, t) = \mathbf{M}_{\Delta^x}^{w_1} \left[\left(X - \hat{X}_t \right) \left(X^T - \hat{X}_t^T \right) a_r(Y_t, X, \Theta, t) + \left(X - \hat{X}_t \right) b_r(Y_t, X, \Theta, t)^T + b_r(Y_t, X, \Theta, t) \left(X^T - \hat{X}_t^T \right) \right] + \sum_{\nu=2}^N c_\nu \mathbf{M}_{\Delta^x}^{w_{1p\nu}} \left[\left(X - \hat{X}_t \right) \left(X^T - \hat{X}_t^T \right) \times a_r(Y_t, X, \Theta, t) + \left(X - \hat{X}_t \right) b_r(Y_t, X, \Theta, t)^T + b_r(Y_t, X, \Theta, t) \left(X^T - \hat{X}_t^T \right) \right]. \quad (45)$$

Формулы (23) для F_κ и H_κ преобразуются к виду:

$$F_\kappa = F_\kappa(Y_t, \Theta, t) = \mathbf{M}_{\Delta^x}^{w_1} \left[q'_\kappa(U) 2\varphi(Y_t, X, \Theta, t)^T C_t (X - \hat{X}_t) + \text{tr} [C_t \sigma(Y_t, X, \Theta, t)] \right] + \sum_{\nu=2}^N c_\nu \mathbf{M}_{\Delta^x}^{w_{1p\nu}} \left[2\varphi(Y_t, X, \Theta, t)^T C_t (X - \hat{X}_t) + \text{tr} [C_t \sigma(Y_t, X, \Theta, t)] \right]; \quad (46)$$

$$H_\kappa = H_\kappa(Y_t, \Theta, t) = \left\{ \mathbf{M}_{\Delta^x}^{w_1} \left[\varphi_1(Y_t, X, \Theta, t)^T q_\kappa(U) \right] + \sum_{\nu=2}^N c_\nu \mathbf{M}_{\Delta^x}^{w_{1p\nu}} \left[\varphi_1(Y_t, X, \Theta, t)^T q_\kappa(U) \right] + 2\mathbf{M}_{\Delta^x}^{w_1} \left[q'(U) \left(X^T - \hat{X}_t^T \right) C_t (\psi \nu \psi_1^T) (Y_t, X, \Theta, t) \right] + 2 \sum_{\nu=2}^N c_\nu \mathbf{M}_{\Delta^x}^{w_{1p\nu}} \times \left[q'(U) \left(X^T - \hat{X}_t^T \right) C_t (\psi \nu \psi_1^T) (Y_t, X, \Theta, t) \right] - f^{(1)T} c_\kappa \right\} (\psi_1 \nu \psi_1^T)^{-1} (Y_t, \Theta, t). \quad (47)$$

Таким образом, получен следующий результат.

Теорема 5.1. Пусть уравнения дифференциальной гауссовской СтС (3), (4) допускают применение МЭА. Тогда фильтрационные уравнения ЭСОФ имеют вид (28), (29), (40) при условиях (41)–(47).

Аналогично выводятся уравнения ЭСОФ для дифференциальной СтС с винеровскими и пуассоновскими шумами вида (3) и (4).

Далее, следуя [4, 5], построим квазилинейный СОФ на основе МЭЛ для МСтС (1), (2) при $\psi' = \psi'(\Theta, t)$, $\psi'' = \psi''(\Theta, t, v)$, $\psi'_1 = \psi'_1(\Theta, t)$ и $\psi''_1 = \psi''_1(\Theta, t, v)$ (т.е. с аддитивными винеровскими и пуассоновскими шумами). Уравнения НСОФ проще получаются, если нелинейные функции φ и φ_1 на основе эллипсоидального распределения с известным c_ν заменить на статистически линеаризованные [6–8]:

$$\left. \begin{aligned} \varphi &= \varphi(X_t, Y_t, \Theta, t) \approx \\ &\approx \varphi_0^3 + k_x^{\varphi} (X_t - m_t^x) + k_y^{\varphi} (Y_t - m_t^y) ; \\ \varphi_1 &= \varphi_1(X_t, Y_t, \Theta, t) \approx \\ &\approx \varphi_{10}^3 + k_x^{\varphi_1} (X_t - m_t^x) + k_y^{\varphi_1} (Y_t - m_t^y) , \end{aligned} \right\} \quad (48)$$

а затем использовать уравнения линейной фильтрации [6–8]. Входящие в (48) коэффициенты с МЭЛ зависят от математических ожиданий, дисперсий и ковариаций:

$$Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} ; m_t^z = \begin{bmatrix} m_t^x \\ m_t^y \end{bmatrix} ; K_t^z = \begin{bmatrix} K_t^{xx} & K_t^{xy} \\ K_t^{xy} & K_t^{yy} \end{bmatrix} .$$

Они определяются из уравнений:

$$\begin{aligned} \dot{Z}_t &= A^z Z_t + A_0^z + B_0^z V, & V &= \dot{W} ; \\ \dot{m}_t^z &= A^z m_t^z + A_0^z, & m_{t_0}^z &= m_0^z ; \\ \dot{K}_t^z &= B^z K_t^z + K_t^z (B^z)^T + B_0^z \nu^m (B_0^z)^T, & K_{t_0}^z &= K_0^z . \end{aligned}$$

Здесь введены следующие обозначения:

$$\left. \begin{aligned} A_0^z &= \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} ; A^z = \begin{bmatrix} a_1 & a \\ b_1 & b \end{bmatrix} ; B_0^z = \begin{bmatrix} \bar{\psi} \\ \bar{\psi}_1 \end{bmatrix} ; \\ a &= k_y^{\varphi} ; a_1 = k_x^{\varphi} ; a_0 = \varphi_0^3 - k_x^{\varphi} m_t^x - k_y^{\varphi} m_t^y ; \\ b &= k_y^{\varphi_1} ; b_1 = k_x^{\varphi_1} ; b_0 = \varphi_0^3 - k_x^{\varphi_1} m_t^x - k_y^{\varphi_1} m_t^y ; \\ \psi dW_0 &+ \int_{R_0^q} \psi'' P^0(dt, dv) = \bar{\psi} dW ; \\ \psi'_1 dW_0 &+ \int_{R_0^q} \psi''_1 P^0(dt, dv) = \bar{\psi}_1 dW , \end{aligned} \right\} \quad (49)$$

где ν^W — интенсивность СтП с независимыми вращениями, состоящего из винеровской и пуассоновской частей (49). Тогда уравнения квазилинейного НСОФ будут иметь вид:

$$\dot{\hat{X}}_t = aY_t + a_1\hat{X}_t + a_0 + \beta_t \left[Z_t - (bY_t + b_1\hat{X}_t + b_0) \right]; \quad (50)$$

$$\beta_t = (R_t b_1^T + \bar{\psi}\nu^W \bar{\psi}_1^T) (\bar{\psi}_1\nu^W \bar{\psi}_1^T)^{-1}; \quad (51)$$

$$\begin{aligned} \dot{R}_t = & a_1 R_t + R_t a_1^T + \bar{\psi}\nu^W \bar{\psi}^T - \\ & - (R_t b_1^T + \bar{\psi}\nu^W \bar{\psi}_1^T) (\bar{\psi}_1\nu^W \bar{\psi}_1^T)^{-1} \times \\ & \times (b_1 R_t + \bar{\psi}_1\nu^W \bar{\psi}^T). \quad (52) \end{aligned}$$

Теорема 5.2. Пусть МСтС (1), (2) содержит только аддитивные винеровские и пуассоновские шумы и допускает замену статистически линеаризованной, а матрица $\sigma_1 = \bar{\psi}_1\nu^W \bar{\psi}_1^T$ не вырождена. Тогда в основе алгоритма квазилинейного НСОФ лежат уравнения (50)–(52) при соответствующих начальных условиях.

Аналогично разд. 3 (в условиях теорем 5.1 и 5.2) составляются уравнения для оценки точности и чувствительности к параметрам Θ .

6 Заключение

Разработана теория аналитического синтеза ЭСОФ для нелинейных дифференциальных МСтС). Рассмотрены случаи гауссовских и негауссовских СтС. Алгоритмы ЭСОФ положены в основу модуля экспериментального программного обеспечения StS-Filter (version 2016).

Результаты допускают развитие на случай дискретных СтС.

Теоретический и практический интерес представляет теория ЭСОФ на основе ненормированных апостериорных распределений [10].

Литература

1. Сеницын И. Н. Аналитическое моделирование распределений на основе ортогональных разложений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 17–24.
2. Сеницын И. Н. Применение ортогональных разложений для аналитического моделирования многомерных распределений в нелинейных стохастических системах на многообразиях // Системы и средства информатики, 2015. Т. 25. № 3. С. 3–22.
3. Сеницын И. Н. Ортогональные субоптимальные фильтры для нелинейных стохастических систем на многообразиях // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 34–44.
4. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы / Пер. с англ. — М.: Наука, 1986. 448 с. (Watanabe S, Ikeda N. Stochastic differential equations and diffusion processes. — Amsterdam — Oxford — New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd., 1981. 476 p.)
5. Справочник по теории вероятностей и математической статистике / Под ред. В. С. Королюка, Н. И. Портенко, А. В. Скорохода, А. Ф. Турбина. — М.: Наука, 1985. 640 с.
6. Пугачев В. С., Сеницын И. Н. Пугачёв В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с.
7. Пугачёв В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
8. Сеницын И. Н. Фильтры Калмана и Пугачева. — 2-е изд. — М.: Логос, 2007. 776 с.
9. Wonham W. M. Some applications of stochastic differential equations to optimal nonlinear filtering // J. Soc. Ind. Appl. Math. A, 1964. Vol. 2. No. 3. P. 347–369.
10. Zakai M. On the optimal filtering of diffusion processes // Z. Wahrscheinlichkeit., 1969. Bd. 11. S. 230–243.
11. Сеницын И. Н., Сеницын В. И. Лекции по теории нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.

Поступила в редакцию 29.02.16

ELLIPSOIDAL SUBOPTIMAL FILTERS FOR NONLINEAR STOCHASTIC SYSTEMS ON MANIFOLDS

I. N. Sinitsyn, V. I. Sinitsyn, and E. R. Korepanov

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

Abstract: For nonlinear differential stochastic systems on manifolds (MStS) with Wiener and Poisson noises, the synthesis theory of ellipsoidal suboptimal filters based on ellipsoidal approximation and ellipsoidal linearization

methods is developed. Special attention is paid to MStS with additive non-Gaussian noises based on ellipsoidal linearization method. The algorithms are the basis of the experimental software tool «StS-Filter» (version 2016). Accuracy and sensitivity equations are presented. Some generalizations are mentioned.

Keywords: ellipsoidal approximation method (EAM); ellipsoidal linearization method (ELM); orthogonal expansions method (OEM); Poisson noise; stochastic system on manifolds (MStS); suboptimal filter (SOF); Wiener noise

DOI: 10.14357/19922264160203

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 15-07-002244).

References

1. Sinitsyn, I. N. 2015. Analiticheskoe modelirovanie raspredeleniy na osnove ortogonal'nykh razlozheniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Analytical modeling in stochastic systems on manifolds based on orthogonal expansions]. *Informatika i ee Primeneniya — Inform Appl.* 9(2):17–24.
2. Sinitsyn, I. N. 2015. Primenenie ortogonal'nykh razlozheniy dlya analiticheskogo modelirovaniya mnogomernykh raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Applications of orthogonal expansions for analytical modeling of multidimensional distributions in stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):3–22.
3. Sinitsyn, I. N. 2016. Ortogonal'nye suboptimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraznykh [Orthogonal suboptimal filters for nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform Appl.* 10(1):34–44.
4. Watanabe, S., and N. Ikeda. 1981. *Stochastic differential equations and diffusion processes*. Amsterdam — Oxford — New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd. 476 p.
5. Korolyuk, V. S., N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, eds. 1985. *Spravochnik po teorii veroyatnosti i matematicheskoy statistike* [Handbook: Probability theory and mathematical statistics]. Moscow: Nauka. 640 p.
6. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester — New York, NY: John Wiley. 549 p.
7. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
8. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd ed. Moscow: Logos. 776 p.
9. Wonham, W. M. 1964. Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Ind. Appl. Math. A* 2(3):347–369.
10. Zakai, M. 1969. On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeit.* 11:230–243.
11. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po teorii 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.

Received February 29, 2016

Contributors

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

Sinitsyn Vladimir I. (b. 1968) — Doctor of Science in physics and mathematics, associate professor, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vsinitsin@ipiran.ru

Korepanov Eduard R. (b. 1966) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ekorepanov@ipiran.ru

МЕТРИЧЕСКАЯ КЛАССИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ СО ВЗВЕШЕННЫМ ВЫРАВНИВАНИЕМ ОТНОСИТЕЛЬНО ЦЕНТРОИДОВ КЛАССОВ*

А. В. Гончаров¹, В. В. Стрижов²

Аннотация: Рассматривается задача метрического анализа и классификации временных рядов. Метрические методы используют матрицу попарных расстояний, строящуюся при помощи фиксированной функции расстояния. Вычислительная сложность алгоритмов, использующих такую матрицу, по меньшей мере квадратична относительно числа временных рядов. Проблема снижения вычислительной сложности решается путем предварительного выделения эталонных объектов, центроидов классов и последующего их использования для описания классов. В качестве базовой модели классификации выбрана модель, использующая динамическое выравнивание временных рядов для построения центроида. В работе предлагается ввести функцию весов центроида, влияющую на вычисление расстояния между объектами. Для анализа алгоритма построения центроида использованы как временные ряды элементарных функций, так и временные ряды физической активности человека с акселерометра мобильного телефона. Свойства построенной модели исследуются и сравниваются со свойствами модели, выбранной в качестве базовой.

Keywords: взвешенное динамическое выравнивание; классификация временных рядов; центроид; функция расстояния

DOI: 10.14357/19922264160204

1 Введение

Рассматривается задача анализа и классификации временных рядов. Существуют различные способы ее решения: построение признакового пространства, использование нейронных сетей, аппроксимация параметрическими функциями. Так, в [1] исследованы методы построения признакового описания временных рядов, в частности метод экспертного построения признаков и метод построения признакового описания на основе гипотезы порождения данных. Результаты [1] показывают, что построенное признаковое пространство адекватно описывает зависимую переменную. В [2] для решения задачи классификации использованы нейронные сети с небольшим числом связей между нейронами, обладающие свойством устойчивости к возмущениям данных. В [3] для классификации предложен алгоритм разбиения исходных временных рядов на периоды и их очистки от шумов. Предложены модификации алгоритма k ближайших соседей и нейронной сети для решения поставленной задачи. В вычислительном эксперименте, проведенном на реальных данных, оценена эффективность, а также проведено сравнение дан-

ных алгоритмов между собой. При этом показан высокий процент правильной классификации.

Построение матрицы попарных расстояний между всеми объектами в задаче метрической классификации является вычислительно трудоемкой задачей. Для снижения размерности задачи и вычислительных затрат предлагается решать задачу с предварительным выделением эталонных объектов, или же центроидов классов, и последующим их использованием для описания множества временных рядов.

Метрические методы используют различные функции расстояния для построения матрицы попарных расстояний: евклидово расстояние [4], метод динамического выравнивания временных рядов [5, 6], метод, основанный на нахождении наибольшей общей последовательности [7], Edit Distance with Real Penalty [8], Edit Distance on Real sequence [9], DISSIM [10], Sequence Weighted Alignment model [11], Spatial Assembling Distance [12] и др. В качестве базового метода для построения функции расстояния в настоящей работе предлагается использовать динамическое выравнивание временных рядов (Dynamic Time Warping) [13]. Как показано в [14], этот метод находит наилучшее со-

* Работа выполнена при финансовой поддержке РФФИ (проект 16-07-01163) и в рамках инициативы «МИТ-Сколтех».

¹Московский физико-технический институт, alex.goncharov@phystech.edu

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, strijov@ccas.ru

ответствие между двумя временными рядами, если они нелинейно деформированы друг относительно друга — растянуты, сжаты или смещены вдоль оси времени.

Базовой моделью классификации и анализа временных рядов принята модель, описанная в [15]. Там в качестве центроида выбирается объект выборки, являющийся ближайшим ко всем остальным объектам. Применяться же будет метод точного его вычисления. Это метод DBA (Dynamic Bandwidth Allocation), решающий задачу оптимизации для нахождения центроида. Алгоритм применения и доказательство корректности приведены в [16].

В настоящей работе вводятся понятия вектора весов и матрицы весов центроида и описываются методы vwDTW и mwDTW (векторно- и матрично-взвешенный DTW) вычисления функции расстояния, основывающиеся на следующем предположении о форме временных рядов: в одном и том же классе находятся временные ряды, имеющие схожую форму с точностью до линейной или нелинейной деформации, локальных или глобальных сдвигов по оси времени. Предполагается, что в центроиде присутствуют характерные для всего класса участки, которым соответствуют большие веса вектора весов этого центроида. А функция расстояния, основанная на vwDTW и использующая вектор весов центроида, точнее объединит объекты одного класса и разделит объекты разных классов, чем основанная на DTW. Поэтому в предлагаемой модели используется метод vwDTW как для вычисления центроида по методу DBA, так и для построения матрицы попарных расстояний. Также исследуются свойства, вид матрицы весов центроида и эффективность применения расстояния, вычисленного с ее помощью — mwDTW, в прикладных задачах.

Для дальнейшей классификации рядов по полученной матрице расстояний сниженной размерности применяется метод k ближайших соседей, как и в базовой модели. Процедура классификации выполняется в три шага:

- (1) отбор эталонных объектов каждого класса;
- (2) построение матрицы попарных расстояний сниженной размерности между временными рядами и эталонными объектами каждого класса;
- (3) классификация временных рядов методом k ближайших соседей с помощью матрицы попарных расстояний.

Для проверки работоспособности такой модели проведен вычислительный эксперимент на реальных и синтетических данных. Эксперимент включает в себя анализ и классификацию данных

при помощи построенной модели. Полученные результаты сравниваются с результатами применения базовой модели к тем же исходным данным.

2 Постановка задачи

Задана выборка $\mathcal{D} = \{(\mathbf{s}_i, y_i)\}_{i=1}^m$, состоящая из пар объект–ответ. Объектами служат временные ряды $\mathbf{s}_i \in \mathbb{R}^n$, а ответами являются метки класса — $y_i \in Y = \{1, \dots, E\}$, где $E \ll m$. Выборка разбита на обучающую \mathcal{D}_l и контрольную \mathcal{D}_t .

Определение 1. Модель классификации f — параметрическая функция объектов выборки, приближающая целевую зависимость y_i . В данной работе параметрами модели примем множество центроидов $\mathbf{C} = \{\mathbf{c}_e\}_{e=1}^E$ и множество векторов весов центроидов $\mathbf{W} = \{\mathbf{w}_e\}_{e=1}^E$ или же матриц весов центроидов $\hat{\mathbf{W}} = \{\mathbf{W}_e\}_{e=1}^E$.

Определение 2. Функцией ошибки S модели f для задачи классификации будем считать

$$S(f, \mathcal{D}_t) = \frac{1}{|\mathcal{D}_t|} \sum_{i=1}^{|\mathcal{D}_t|} [f(\mathbf{s}_i) \neq y_i].$$

Требуется построить модель классификации $f: \mathbb{R}^n \rightarrow Y$, минимизирующую функцию ошибки S на контрольной выборке:

$$f_{\mathbf{C}, \hat{\mathbf{W}}} = \underset{\mathbf{C}, \hat{\mathbf{W}}}{\operatorname{argmin}} (S(f, \mathcal{D}_t)).$$

3 Вычисление значения функции расстояния

3.1 Общие понятия

В данной работе в качестве метрического расстояния между объектами предлагается использовать стоимость *взвешенного пути наименьшей стоимости* между этими объектами.

Даны два временных ряда: \mathbf{s}_1 и \mathbf{s}_2 . Будем считать, что $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$. Пусть $\Omega^{n \times n}$ — это матрица, такая что ее элемент Ω_{ij} равен квадрату разности между i -м и j -м элементами последовательностей \mathbf{s}_1 и \mathbf{s}_2 :

$$\Omega_{ij} = (s_{1i} - s_{2j})^2.$$

Определение 3. Путем π между последовательностями \mathbf{s}_1 и \mathbf{s}_2 назовем упорядоченное множество пар индексов элементов матрицы Ω :

$$\pi = \{\pi_r\} = \{(i_r, j_r)\}, \quad r = 1, \dots, R, \\ i, j \in \{1, \dots, n\},$$

где R — длина пути, зависящая от выбора пути. Он должен удовлетворять следующим условиям.

Граничные условия. $\pi_1 = (1, 1)$ и $\pi_R = (n, n)$, т. е. начало и конец π находятся на диагонали в противоположных углах Ω .

Непрерывность. Пусть $\pi_r = (p_1, p_2)$ и $\pi_{r-1} = (q_1, q_2)$, $r = 2, \dots, R$. Тогда

$$p_1 - q_1 \leq 1, \quad p_2 - q_2 \leq 1.$$

Это ограничение нужно, чтобы в шаге пути π участвовали только соседние элементы матрицы (включая соседние по диагонали).

Монотонность. Пусть $\pi_r = (p_1, p_2)$ и $\pi_{r-1} = (q_1, q_2)$, $r = 2, \dots, R$. Тогда выполняется хотя бы одно из условий

$$p_1 - q_1 \geq 1; \quad p_2 - q_2 \geq 1.$$

Это ограничение обусловлено природой рассматриваемых последовательностей и предназначено для монотонности функции выравнивания времени.

Физические ограничения. Как уже говорилось во введении, предполагается, что временные ряды одного класса имеют схожую форму и являются линейно или нелинейно деформированными или же смещены друг относительно друга. При этом считается, что подобного рода деформации и смещения являются малыми, локальными. В этом предположении выравнивающий путь в матрице слабо отклоняется от диагонали, т. е.

$$\text{для каждого } \{i_r, j_r\} \in \pi \quad i_r - k \leq j_r \leq i_r + k,$$

где k определяется типом задачи и ее физическими ограничениями.

3.2 Векторно-взвешенный путь наименьшей стоимости

Дадим определение векторно-взвешенного пути наименьшей стоимости, vwDTW. Дан вектор весов $\mathbf{w} \in \mathbb{R}^n$.

Определение 4. Стоимостью $\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}, \pi)$ векторно-взвешенного пути π между последовательностями \mathbf{s}_1 и \mathbf{s}_2 с весом \mathbf{w} назовем

$$\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}, \pi) = \sum_{(i,j) \in \pi} w_j \Omega_{ij}.$$

Определение 5. Векторно-взвешенным путем наименьшей стоимости (векторно-взвешенным выравнивающим путем) $\hat{\pi}$ между последовательностями \mathbf{s}_1 и \mathbf{s}_2 назовем взвешенный путь, имеющий наименьшую стоимость среди всех возможных век-

торно-взвешенных путей между последовательностями \mathbf{s}_1 и \mathbf{s}_2 :

$$\hat{\pi} = \underset{\pi}{\operatorname{argmin}} \text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}, \pi). \quad (1)$$

Обозначим стоимость векторно-взвешенного выравнивающего пути между последовательностями \mathbf{s}_1 и \mathbf{s}_2 через $\rho(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}) = \text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}, \hat{\pi})$.

Для вычисления стоимости такого пути в данной работе используется модифицированный метод DTW — vwDTW (vector-weighted DTW). Согласно этому методу необходимо построить новую матрицу Γ , элементы которой определяются следующим образом:

$$\begin{aligned} \Gamma_{1j} &= w_j \Omega_{1j}, \quad \Gamma_{i1} = w_1 \Omega_{i1}; \quad i, j = 1, \dots, n; \\ \Gamma_{ij} &= w_j \Omega_{ij} + \min(\Gamma_{i,j-1}, \Gamma_{i-1,j}, \Gamma_{i-1,j-1}), \\ & \quad i, j = 2, \dots, n. \end{aligned}$$

Элемент Γ_{ij} матрицы Γ равен стоимости векторно-взвешенного выравнивающего пути между последовательностями $\{s_{1a}\}_{a=1}^i$ и $\{s_{2a}\}_{a=1}^j$.

В качестве значения функции расстояния между двумя объектами выберем стоимость векторно-взвешенного выравнивающего пути между ними (1):

$$\rho(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}) = \Gamma_{nn}. \quad (2)$$

Заметим, что при единичном векторе весов vwDTW эквивалентен обычному DTW, описание которого приведено в [15].

3.3 Матрично-взвешенный путь наименьшей стоимости

Дадим определение матрично-взвешенного пути наименьшей стоимости, mwDTW. Дана матрица весов $\mathbf{W} \in \mathbb{R}^{n \times n}$.

Определение 6. Стоимостью $\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{W}, \pi)$ матрично-взвешенного пути π между последовательностями \mathbf{s}_1 и \mathbf{s}_2 с весом \mathbf{W} назовем

$$\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{W}, \pi) = \sum_{(i,j) \in \pi} W_{ij} \Omega_{ij}.$$

Определение 7. Матрично-взвешенным путем наименьшей стоимости (матрично-взвешенным выравнивающим путем) $\hat{\pi}$ между последовательностями \mathbf{s}_1 и \mathbf{s}_2 назовем взвешенный путь, имеющий наименьшую стоимость среди всех возможных матрично-взвешенных путей между последовательностями \mathbf{s}_1 и \mathbf{s}_2 :

$$\hat{\pi} = \underset{\pi}{\operatorname{argmin}} \text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{W}, \pi). \quad (3)$$

Обозначим стоимость матрично-взвешенного выравнивающего пути между последовательностями \mathbf{s}_1 и \mathbf{s}_2 через $\rho(\mathbf{s}_1, \mathbf{s}_2, \mathbf{W}) = \text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \mathbf{W}, \hat{\pi})$.

Вычисление его стоимости происходит с помощью еще одного модифицированного метода DTW — mwDTW (matrix-weighted DTW). Согласно этому методу необходимо построить новую матрицу Γ , элементы которой определяются следующим образом:

$$\begin{aligned} \Gamma_{1j} &= W_{1j}\Omega_{1j}, \quad \Gamma_{i1} = W_{i1}\Omega_{i1}, \quad i, j = 1, \dots, n, \\ \Gamma_{ij} &= W_{ij}\Omega_{ij} + \min(\Gamma_{i,j-1}, \Gamma_{i-1,j}, \Gamma_{i-1,j-1}), \\ & \quad i, j = 2, \dots, n. \end{aligned}$$

Элемент Γ_{ij} матрицы Γ равен стоимости матрично-взвешенного выравнивающего пути между последовательностями $\{s_{1a}\}_{a=1}^i$ и $\{s_{2a}\}_{a=1}^j$.

В качестве значения функции расстояния между двумя объектами выберем стоимость матрично-взвешенного выравнивающего пути между ними (3):

$$\rho(s_1, s_2, \mathbf{W}) = \Gamma_{nn}.$$

Заметим, что при использовании матрицы, состоящей из одних единиц, mwDTW переходит в обычный DTW, описание которого приведено в [15].

4 Вычисление параметров модели классификатора

4.1 Построение центроида

Пусть множество весов $\hat{\mathbf{W}}$ фиксировано. Построим множество центроидов \mathbf{C} .

4.1.1 Постановка задачи построения центроида

Определение 8. Пусть \mathcal{D}_e — множество элементов из \mathcal{D} , принадлежащих одному классу e из Y . Центроидом множества векторов $\mathcal{D}_e = \{s_i | y_i = e\}_{i=1}^m$ по расстоянию ρ назовем вектор $\mathbf{c}_e \in \mathbb{R}^n$ такой, что

$$\mathbf{c}_e = \operatorname{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{s_i \in \mathcal{D}_e} \rho(s_i, \mathbf{c}),$$

где ρ — стоимость векторно-взвешенного (матрично-взвешенного) пути наименьшей стоимости vwDTW (mvDTW).

Центроид найдем как решение оптимизационной задачи для vwDTW:

$$\mathbf{c}_e = \operatorname{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{s_i \in \mathcal{D}_e} \sum_{(t,t') \in \hat{\pi}_i} \mathbf{w}_e(t) (s_i(t') - \mathbf{c}(t))^2 \quad (4)$$

или же для mwDTW:

$$\mathbf{c}_e = \operatorname{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{s_i \in \mathcal{D}_e} \sum_{(t,t') \in \hat{\pi}_i} \mathbf{W}_e(t, t') (s_i(t') - \mathbf{c}(t))^2,$$

где $\hat{\pi}_i$ — векторно-взвешенный (матрично-взвешенный) выравнивающий путь между временными рядами s_i и \mathbf{c} .

4.1.2 Решение задачи нахождения центроида методом DBA

Теорема 1 [16]. Пусть дано множество векторов $\mathcal{D}_e = \{s_i | y_i = e\}_{i=1}^m$ одного класса, начальное приближение центроида \mathbf{c}_e и множество выравнивающих путей между каждым рядом и начальным приближением центроида $\{\tilde{\pi}_i\}_{i=1}^m$. Тогда локальный минимум задачи оптимизации (4) при единичном векторе весов \mathbf{w} в (2) (функция расстояния DTW) достигается при

$$\mathbf{c}_e(t) = \frac{1}{N} \sum_{s_i \in \mathcal{D}_e} \sum_{t': (t,t') \in \tilde{\pi}_i} s_i(t'),$$

где

$$N = \sum_{s_i \in \mathcal{D}_e} \sum_{t': (t,t') \in \tilde{\pi}_i} 1.$$

Доказательство. Для поиска центроида и решения задачи оптимизации (4) воспользуемся необходимым условием экстремума. Запишем частные производные функционала по $\mathbf{c}_e(t)$, $t = 1, \dots, T$, и приравняем их к 0:

$$\frac{\partial F(\mathbf{c}_e, \mathcal{D}_e)}{\partial \mathbf{c}_e(t)} = \sum_{s_i \in \mathcal{D}_e} \sum_{t': (t,t') \in \tilde{\pi}_i} 2(\mathbf{c}_e(t) - s_i(t')) = 0.$$

Откуда и находим значение $\mathbf{c}_e(t)$, $t = 1, \dots, n$:

$$\mathbf{c}_e(t) = \frac{1}{N} \sum_{s_i \in \mathcal{D}_e} \sum_{t': (t,t') \in \tilde{\pi}_i} s_i(t').$$

Данный метод вычисления центроида приведен в [16] и называется методом DBA. Там же приведено и его доказательство. При нахождении нового центроида множество выравнивающих рядов меняется, данную процедуру нужно проводить несколько раз, пока центроид не стабилизируется. При замене единичного вектора весов в функции расстояния vwDTW на произвольный справедливо следующее

Следствие 1. При использовании произвольного вектора весов центроида \mathbf{w} (замене DTW на vwDTW с вектором весов \mathbf{w}) в задаче оптимизации (4) алгоритм DBA вычисления центроида находит локальный минимум при замене множества путей наименьшей стоимости $\{\tilde{\pi}\}_{i=1}^m$ на множество взвешенных путей наименьшей стоимости $\{\hat{\pi}_i\}_{i=1}^m$.

Доказательство данного следствия повторяет доказательство теоремы 1 при замене множества путей наименьшей стоимости $\{\hat{\pi}\}_{i=1}^m$ на множество взвешенных путей наименьшей стоимости $\{\pi_i\}_{i=1}^m$.

Для функции расстояния mwDTW следствие сохраняет свою формулировку и доказательство при замене вектора весов центроида на его матрицу весов \mathbf{W} .

4.2 Оптимизация и ограничения вектора весов

Положим теперь множество центроидов \mathbf{C} фиксированным. Каждому центроиду \mathbf{c}_e из множества \mathbf{C} поставлен в соответствие вектор неотрицательных весов \mathbf{w}_e , принадлежащий множеству $\hat{\mathbf{W}}$. Значения данного вектора весов выделяют наиболее типичные для класса участки центроида, сопоставляя им большие веса. Вычислим этот вектор, решая задачу оптимизации:

$$\mathbf{w}_e = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathcal{D}_e} \sum_{(t,t') \in \pi_i} \mathbf{w}(t) (\mathbf{s}_i(t') - \mathbf{c}_e(t))^2. \quad (5)$$

При отсутствии ограничений на веса \mathbf{w}_e минимум (5) достигается при $\mathbf{w}_e = \mathbf{0}$. Чтобы избежать такого тривиального решения, введем ограничения на сумму элементов вектора весов:

$$\sum_{t=1}^T \mathbf{w}_e(t) = T.$$

Предположим, что при решении задачи (5) нашлось t , для которого выполняется

$$\sum_{\mathbf{s}_i \in \mathcal{D}} \sum_{t':(t,t') \in \pi_i} (\mathbf{s}_i(t') - \mathbf{c}_e(t))^2 = 0.$$

Для таких t элемент решения задачи оптимизации (5) примет большие значения, которые обеспечат выполнение ограничений на сумму элементов. Это приведет к локальному скоплению больших значений вектора весов, что сильно ухудшит дальнейшую интерпретацию вектора весов, а также сделает метод чувствительным к малым изменениям входных данных. Поэтому введем ограничения на элементы вектора весов сверху:

$$\mathbf{w}_e(t) \leq \text{const}, \quad t \in \{1, \dots, T\}.$$

Таким образом, исходная задача (5) примет следующий вид:

$$\left. \begin{aligned} \mathbf{w}_e = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathcal{D}_e} \sum_{(t,t') \in \pi_i} \mathbf{w}_e(t) (\mathbf{s}_i(t') - \mathbf{c}_e(t))^2; \\ \sum_{t=1}^T \mathbf{w}_e(t) = T, \quad 0 \leq \mathbf{w}_e(t) \leq \text{const}, \\ t \in \{1, \dots, T\}, \end{aligned} \right\} \quad (6)$$

где const — некоторая заданная константа.

4.3 Оптимизация матрицы весов

4.3.1 Ограничения матрицы весов

Как и при оптимизации вектора весов, положим множество центроидов \mathbf{C} фиксированным. каждо-

му центроиду \mathbf{c}_e некоторого класса e из множества \mathbf{C} сопоставлена матрица неотрицательных весов \mathbf{W}_e , принадлежащая множеству $\hat{\mathbf{W}}$.

Используя те же соображения, что и для случая использования расстояния vwDTW , определим задачу нахождения матрицы весов центроида как задачу оптимизации с ограничениями при использовании расстояния mwDTW :

$$\left. \begin{aligned} \mathbf{W}_e = \operatorname{argmin}_{\mathbf{W} \in \mathbb{R}^{n \times n}} \sum_{\mathbf{s}_i \in \mathcal{D}_e} \sum_{(t,t') \in \pi_i} \mathbf{W}_e(t, t') (\mathbf{s}_i(t') - \mathbf{c}_e(t))^2; \\ \sum_{t=1}^T \sum_{t'=1}^T \mathbf{W}_e(t, t') = T^2, \\ 0 \leq \mathbf{W}_e(t) \leq \text{const}, \quad t \in \{1, \dots, T\}, \end{aligned} \right\} \quad (7)$$

где const — некоторая заданная константа.

4.3.2 Сглаживание полученной матрицы

Полученная матрица не является устойчивой к изменению входных данных: при использовании других временных рядов выравнивающие пути будут иметь другой вид, а значит, и решение задачи оптимизации будет другое. Более устойчивой матрица получится после процедуры сглаживания.

Будем говорить, что элемент матрицы $\mathbf{W}_e(t, t')$ содержится во множестве Φ , если существует временной ряд $\mathbf{s}_i \in \mathcal{D}_e$ такой, что путь наименьшей матрично-взвешенной стоимости проходит через элемент $\Omega(t, t')$ в матрице Ω , построенной для временного ряда \mathbf{s}_i и центроида.

При решении задачи оптимизации элементы $\{\mathbf{W}_e(t, t')\} \in \Phi$ будут изменяться в меньшую сторону, при этом $\mathbf{W}_e(t, t') \notin \Phi$ достигнет своей верхней границы для выполнения ограничений, накладываемых на сумму элементов матрицы.

Выберем произвольный элемент матрицы весов $\mathbf{W}_e(t, t') \notin \Phi$. Вероятность того, что при добавлении нового временного ряда (например, из тестовой выборки) выполнится $\mathbf{W}_e(t, t') \in \Phi$, выше, если среди ближайших к $\mathbf{W}_e(t, t')$ элементов в строке матрицы весов многие содержатся в множестве Φ . Тогда значение такого элемента должно быть похожим на значения соседних. Добьемся этого, выполнив сглаживание матрицы весов центроида:

$$\widetilde{\mathbf{W}}_e(t, t') = \frac{1}{2\delta} \sum_{k=-\delta}^{\delta} \mathbf{W}_e(t, t' + k),$$

где δ — величина окна сглаживания, а $\widetilde{\mathbf{W}}_e$ — искомая матрица весов центроида.

4.4 Задача оптимизации параметров модели

Задача оптимизации параметров модели сведена к комбинации задач оптимизации (4) и (6), (7) для vwDTW:

$$\mathbf{w}_e, \mathbf{c}_e = \operatorname{argmin}_{\mathbf{c}, \mathbf{w} \in \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathcal{D}_e} \sum_{(t, t') \in \pi_i} \left(\mathbf{w}(t) (\mathbf{s}_i(t') - \mathbf{c}(t))^2 \right);$$

$$e = 1, \dots, E,$$

$$\sum_{t=1}^T \mathbf{w}_e(t) = T, \quad 0 \leq \mathbf{w}_e(t) \leq \text{const}, \quad t \in \{1, \dots, T\},$$

или же для mwDTW:

$$\mathbf{W}_e, \mathbf{c}_e = \operatorname{argmin}_{\mathbf{W} \in \mathbb{R}^n \times \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathcal{D}_e} \sum_{(t, t') \in \pi_i} \left(\mathbf{W}_e(t, t') (\mathbf{s}_i(t') - \mathbf{c}_e(t))^2 \right), \quad e = 1, \dots, E;$$

$$\sum_{t=1}^T \sum_{t'=1}^T \mathbf{W}_e(t, t') = T^2, \quad 0 \leq \mathbf{W}_e(t) \leq \text{const},$$

$$t \in \{1, \dots, T\}.$$

Эту задачу будем решать, вычисляя сначала множество центроидов \mathbf{C} при фиксированном начальном приближении множества весов центроидов $\hat{\mathbf{W}}$, а затем вычисляя множество весов центроидов $\hat{\mathbf{W}}$ при фиксированном множестве центроидов \mathbf{C} . Таким образом, алгоритм вычисления параметров модели будет иметь следующий вид.

Шаг 1. Начальное приближение вектора весов центроида:

$$\mathbf{w}_e = \mathbf{1}, \quad e = 1, \dots, E,$$

или же начальное приближение матрицы весов центроида:

$$\mathbf{W}_e = \mathbf{1}, \quad e = 1, \dots, E.$$

Шаг 2. Начальное приближение центроида класса — произвольный элемент класса:

$$\mathbf{c}_e = \mathbf{s}_j \in \mathcal{D}_e, \quad e = 1, \dots, E.$$

Шаг 3. Вычисление центроида при фиксированном векторе (матрице) весов центроида как решение задачи оптимизации (4).

Шаг 4. Вычисление вектора (матрицы) весов центроида при фиксированном центроиде как решение задачи оптимизации (6), (7).

5 Вычислительный эксперимент

Для проверки свойств введенной функции расстояния, а также выбранной модели был проведен

вычислительный эксперимент на реальных и синтетических данных. Свойства функций расстояния vwDTW и mwDTW для наглядности продемонстрированы на синтетических данных, представляющих собой смещенные и линейно деформированные временные ряды аналитических функций: $\sin x$, \sqrt{x} , x^2 — длиной 100 точек. В выборке находилось 100 временных рядов каждого класса: 50 в обучающей и 50 в контрольной. Обозначим классы как 1 — $\sin x$, 2 — \sqrt{x} и 3 — x^2 соответственно.

Пример такой выборки и результаты нахождения параметров модели по обучающей выборке показаны на рис. 1. По оси абсцисс отложены значения времени, а по оси ординат — значения временного ряда. На рис. 1, *a* приведены примеры аналитических функций, используемых в создании синтетической выборки временных рядов. На рис. 1, *б* изображены центроид (нижний временной ряд) и вектор весов (верхний временной ряд) для класса 1. Аналогично на рис. 1, *в* и 1, *г* показаны результаты для классов 2 и 3.

Векторы весов описывают наиболее информативные участки центроида. Так, для центроида класса 1 ($\sin x$) наиболее информативными оказались минимумы и максимумы, в отличие от точек перегиба.

Для сравнения свойств полученной функции расстояния с функцией расстояния DTW были посчитаны расстояния до центроидов для всех временных рядов контрольной выборки с помощью DTW и vwDTW, после чего производилась классификация. Каждому временному ряду контрольной выборки ставилась в соответствие метка того класса, расстояние до центроида которого было минимальным. Результат классификации с помощью функции расстояния vwDTW — 97%, а для функции расстояния DTW — 84%, что на 15% меньше. Построенная в работе функция расстояния лучше разделила временные ряды разных классов, сгруппировала их вокруг соответствующих центроидов.

Также были построены матрицы весов центроида на синтетических данных. Вследствие хорошей интерпретируемости в работе приведено визуальное отображение матрицы весов центроида только для класса 1. На рис. 2, *a* изображены примеры временных рядов. На рис. 2, *б* — множество выравнивающих путей по матрице Ω . На рис. 2, *в* — матрица весов центроида до сглаживания. На рис. 2, *г* — после сглаживания. Темные цвета соответствуют маленьким значениям элементов матрицы, серые — большим, более светлые — промежуточным. Хорошо просматривается периодичность в матрице, напоминающая периодичность вектора

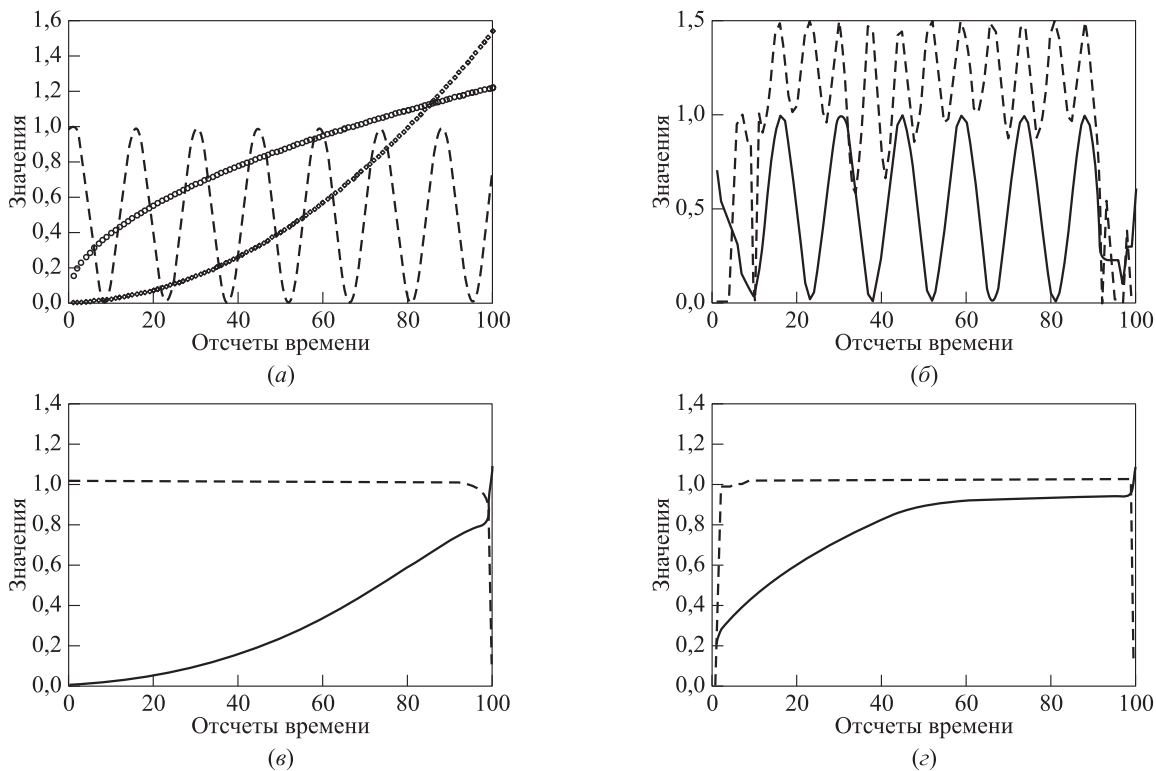


Рис. 1 Примеры синтетических временных рядов аналитических функций (выборка) (а) и результаты построения центраида и вектора весов для синтетических данных: (б) $\sin x$; (в) x^2 ; (г) \sqrt{x}

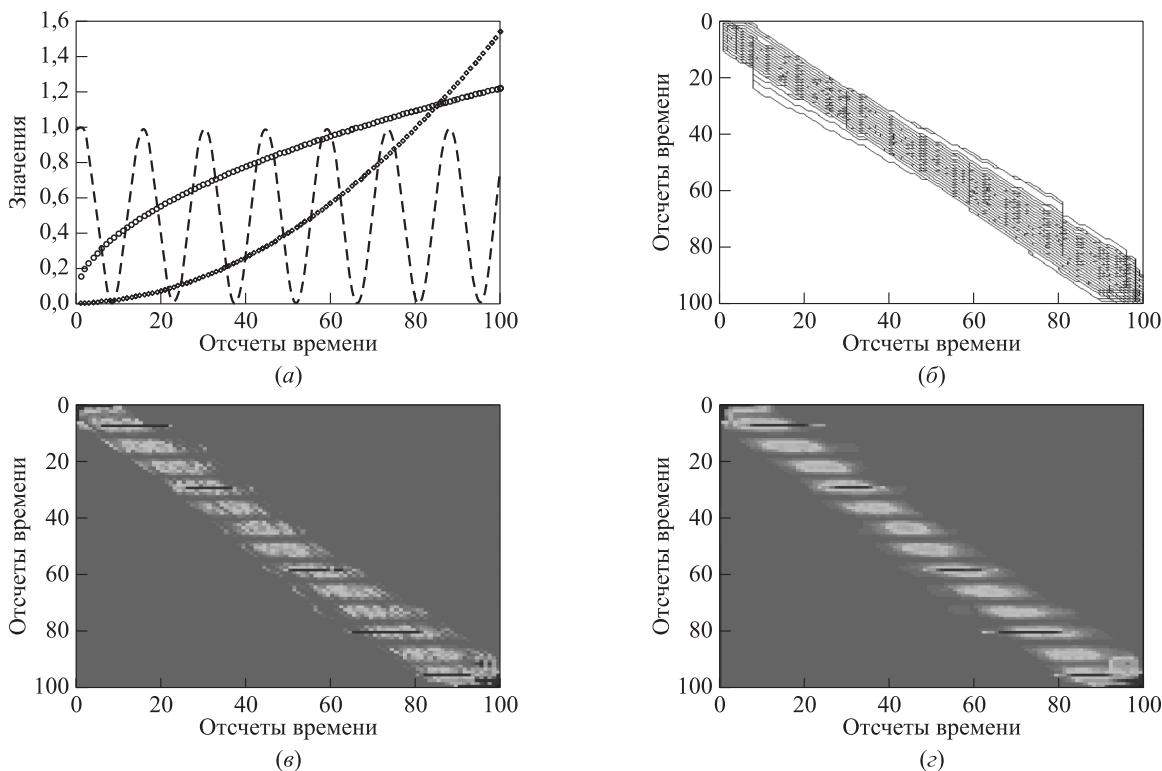


Рис. 2 Примеры синтетических временных рядов аналитических функций (выборка) (а) и результаты построения центраида класса $\sin x$: (б) множество выравнивающих путей; (в) начальная матрица весов; (г) сглаженная матрица весов

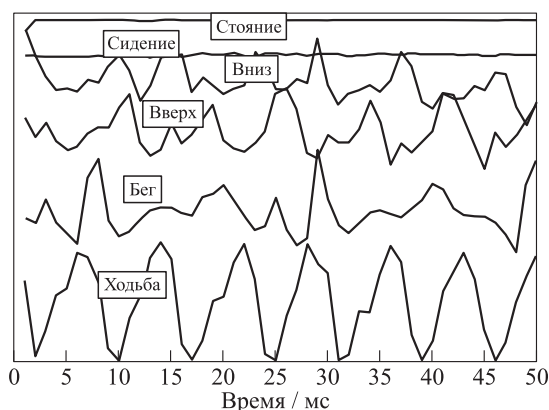


Рис. 3 Примеры временных рядов измерений акселерометра для разных видов физической активности

весов. Предполагается, что функция $m\text{wDTW}$ будет лучше разделять классы, которые сильно различаются между собой, так как матрица весов центроида учитывает и среднее отклонение выравнивающего пути от диагонали в матрице Ω . При отклонениях пути сильнее типичного для данного класса элементам пути будут приписываться большие веса, что видно из структуры матрицы весов центроида.

Использование функции $m\text{wDTW}$ для классификации синтетических временных рядов улучшает классификацию по сравнению с DTW . Этот результат аналогичен случаю использования vwDTW — 97%.

Для демонстрации работы предложенной модели на реальных данных и ее сравнения с базовой моделью классификации были использованы данные акселерометра мобильного телефона. Вследствие большой вычислительной сложности они сравнивались на данных, представляющих собой 600 временных рядов длиной 200 точек, каждый из которых представляет собой абсолютные значения ускорения мобильного телефона, объединяя три временных ряда: временной ряд ускорения по оси X (200 измерений), оси Y (200 измерений) и оси Z (200 измерений). Выделено шесть типов физической активности: ходьба, бег, сидение, стояние, подъем, спуск. Временные ряды записывались акселерометром, который находился в кармане у человека, выполняющего один из

типов физической активности, после чего разделялись на 10-секундные сегменты. Примеры таких временных рядов приведены на рис. 3.

Данные разделялись на обучающую и контрольную выборку. В обучающую выборку входило по 70 временных рядов каждого вида физической активности, а в контрольную — по 30 временных рядов. Производилась классификация методом k ближайших соседей, использующим построенную матрицу попарных расстояний. Осуществлялся контроль качества при помощи кросс-валидации.

В табл. 1 приведены результаты классификации при применении новой модели, использующей функцию vwDTW , и базовой модели классификации.

Качество классификации базовой модели ухудшилось в сравнении с вычислительным экспериментом в работе [15], так как теперь используются абсолютные значения ускорения, а не последовательно соединенные временные ряды ускорения вдоль трех координат. Использование функции расстояния vwDTW улучшает классификацию для всех классов физической активности, повышая общий уровень классификации на 9%.

Полученные параметры модели для реальных данных изображены на рис. 4. Верхние линии соответствуют центроидам, а нижние — их весам. Для класса бег, например, хорошо просматривается периодичность временного ряда как для центроида, так и для вектора весов центроида.

Для сравнения модели, использующей $m\text{wDTW}$, с базовой моделью классификации были использованы данные, представляющие собой 600 временных рядов длиной 100 точек. Временные ряды такой длины выбраны для разумного ограничения времени работы методов численной оптимизации. В остальном этот эксперимент повторяет тот, что выполнен для сравнения базовой модели и модели, использующей vwDTW . Результаты классификации приведены в табл. 2. Видно, что длина временных рядов, а следовательно, и количество периодов временного ряда, существенно влияют на качество классификации. Качество базовой модели сильно упало по сравнению с временными рядами длиной 200 точек. При этом использование $m\text{wDTW}$ улучшает классификацию на 13%.

Таблица 1 Сравнение эффективности предложенной (vwDTW) и базовой модели классификации на данных [17]

Модель	Точность по критерию скользящего контроля, %						Общее
	Бег	Ходьба	Подъем	Спуск	Сидение	Стояние	
vwDTW	97	95	79	75	95	95	89
DTW [15]	95	92	60	60	85	90	80

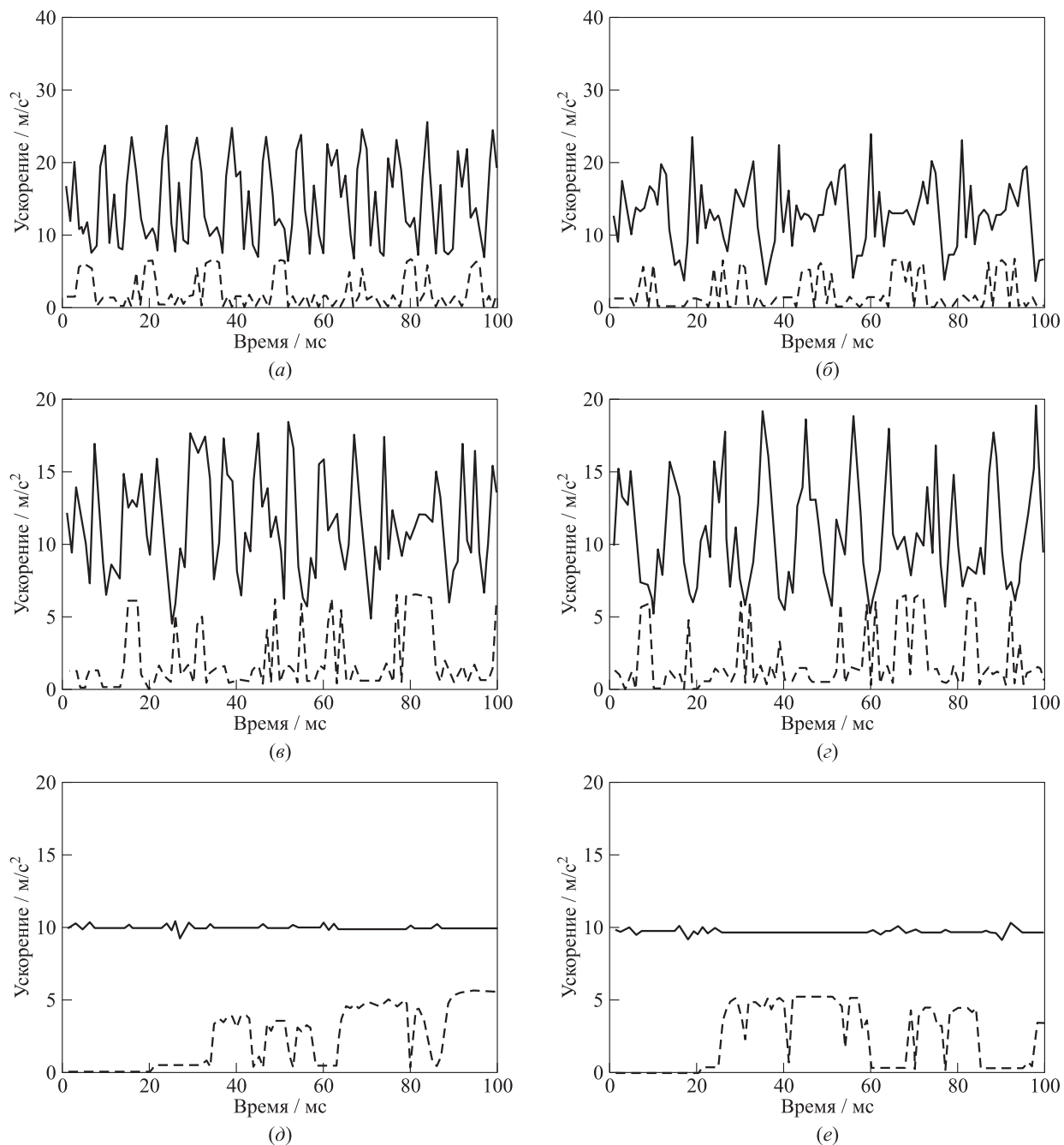


Рис. 4 Примеры временных рядов измерений акселерометра для разных видов физической активности: (а) ходьба; (б) бег; (в) подъем; (г) спуск; (д) сидение; (е) стояние

Таблица 2 Сравнение эффективности предложенной (mwDTW) и базовой модели классификации и алгоритма разделяющей классификации на данных [17]

Модель	Точность по критерию скользящего контроля, %						Общее
	Бег	Ходьба	Подъем	Спуск	Сидение	Стояние	
mwDTW	95	95	78	76	90	90	87
DTW [15]	88	83	55	60	82	80	74

6 Заключение

В работе описан новый подход к работе с центроидами временных рядов, построена модель, использующая веса и матрицы центроидов, и показаны ее преимущества перед моделью, описанной в работе [15]. Продемонстрировано влияние структуры, длины и физического смысла временных рядов на результаты классификации. В последующих работах данный подход будет совершенствоваться. Предполагается ускорить данный алгоритм за счет эффективного нахождения пути наименьшей стоимости, его аппроксимации. Такой подход применим и ко всем типам временных рядов, где возможна нелинейная деформация как по оси времени, так и по оси значений временного ряда.

Литература

1. Кузнецов М. П., Ивкин Н. П. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию // Машинное обучение и анализ данных, 2015. Т. 1. Вып. 13. С. 1471–1483.
2. Попова М. С., Стрижов В. В. Выбор оптимальной модели классификации физической активности по измерениям акселерометра // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 79–89.
3. Ignatov A. D., Strijov V. V. Human activity types recognition using quasiperiodic sets of time series collected from a single tri-axial accelerometer // Multimedia tools and applications. — Springer, 2015. P. 1–14.
4. Faloutsos C., Ranganathan M., Manolopoulos Y. Fast subsequence matching in time-series databases // SIGMOD Conference (International) on Management of Data Proceedings. — Minneapolis, MN, USA: ACM, 1994. P. 419–429.
5. Berndt D. J., Clifford J. Using dynamic time warping to find patterns in time series // Workshop on Knowledge Discovery in Databases at the 12th Conference (International) on Artificial Intelligence Proceedings. — Seattle, WA, 1994. P. 359–370.
6. Keogh E. J., Ratanamahatana C. A. Exact indexing of dynamic time warping // Knowl. Inf. Syst., 2005. Vol. 7. No. 3. P. 358–386.
7. Vlachos M., Gunopulos D., Kollios G. Discovering similar multidimensional trajectories // 18th Conference (International) on Data Engineering (ICDE'02) Proceedings. — San Jose, CA, USA: IEEE Computer Society, 2002. P. 673–684.
8. Chen L., Ng R. T. On the marriage of lp-norms and edit distance // 30th Conference (International) on Very Large Data Bases Proceedings. — Toronto: Morgan Kaufmann, 2004. P. 792–803.
9. Chen L., Özsu M. T., Oria V. Robust and fast similarity search for moving object trajectories // 24th ACM Conference (International) on Management of Data Proceedings. — Baltimore, MD, USA: ACM, 2005. P. 491–502.
10. Frentzos E., Gratsias K., Theodoridis Y. Index-based most similar trajectory search // 23rd Conference (International) on Data Engineering Proceedings. — Istanbul: IEEE Computer Society, 2007. P. 816–825.
11. Morse M. D., Patel J. M. An efficient and accurate method for evaluating time series similarity // ACM Conference (International) on Management of Data Proceedings. — Beijing: ACM, 2007. P. 569–580.
12. Chen Y., Nascimento M. A., Ooi B. C., Tung A. K. H. SPADe: On Shape-based pattern detection in streaming time series // 23rd Conference (International) on Data Engineering Proceedings. — Istanbul: IEEE Computer Society, 2007. P. 786–795.
13. Keogh E. J., Pazzani M. J. Scaling up dynamic time warping to massive datasets // Principles of Data Mining and Knowledge Discovery: 3rd European Conference Proceedings. — Berlin–Heidelberg: Springer, 1999. P. 1–11.
14. Salvador S., Chan P. Fastdtw: Toward accurate dynamic time warping in linear time and space // 3rd Workshop on Mining Temporal and Sequential Data Proceedings. — Seattle, WA, USA, 2004. P. 70–80.
15. Гончаров А. В., Попова М. С., Стрижов В. В. Метрическая классификация временных рядов с выравниванием относительно центроидов классов // Системы и средства информатики, 2015. Т. 25. Вып. 4. С. 52–64.
16. Petitjean F., Forestier G., Webb G. I., Nicholson A. E., Chen Y., Keogh E. Dynamic time warping averaging of time series allows faster and more accurate classification // 30th IEEE Conference (International) on Data Engineering Proceedings. — Chicago, IL, USA: IEEE Computer Society, 2014. P. 470–479.
17. Data from accelerometer. https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group274/Goncharov2015MetricClassification/data/preprocessed_large.csv.

Поступила в редакцию 31.12.15

METRIC TIME SERIES CLASSIFICATION USING WEIGHTED DYNAMIC WARPING RELATIVE TO CENTROIDS OF CLASSES

A. V. Goncharov¹ and V. V. Strijov²

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

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

Abstract: The paper discusses the problem of metric time series analysis and classification. The proposed classification model uses a matrix of distances between time series which is built with a fixed distance function. The dimension of this distance matrix is very high and all related calculations are time-consuming. The problem of reducing computational complexity is solved by selecting reference objects and using them for describing classes. The model that uses dynamic time warping for building reference objects or centroids is chosen as the basic model. This paper introduces a function of weights for each centroid that influences calculation of the distance measure. Time series of different analytic functions and time series of human activity from an accelerometer of a mobile phone are used as the objects for classification. The properties and the classification result of this model are investigated and compared with the properties of the basic model.

Keywords: metric classification; weighted dynamic time warping; time series classification; centroid; distance function

DOI: 10.14357/19922264160204

Acknowledgments

The work was financially supported by the Russian Foundation for Basic Research (project 16-07-01163) and fulfilled in the frames of the initiative “MIT-Skoltech.”

References

1. Kuznecov, M. P., and N. P. Ivkin. 2015. Algoritm klasifikatsii vremennykh ryadov akselerometra po kombinirovannomu priznakovomu opisaniyu [Time series classification algorithm using combined feature description]. *Mashinnoe Obuchenie i Analiz Danykh* [Machine Learning and Data Analysis] 1(13):1471–1483.
2. Popova, M. S., and V. V. Strijov. 2015. Vybór optimal'noy modeli klassifikatsii fizicheskoy aktivnosti po izmereniyam akselerometra [Selection of optimal physical activity classification model using measurements of accelerometer]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):79–89.
3. Ignatov, A. D. and V. V. Strijov. 2015. Human activity types recognition using quasiperiodic sets of time series. *Multimedia tools and applications*. Springer. 1–14.
4. Faloutsos, C., M. Ranganathan, and Y. Manolopoulos. 1994. Fast subsequence matching in time-series databases. *ACM Conference on Management of Data (SIGMOD) International Proceedings*. Minneapolis, MN: ACM. 419–429.
5. Berndt, D. J., and J. Clifford. 1994. Using dynamic time warping to find patterns in time series. *Workshop on Knowledge Discovery in Databases at the 12th Conference (International) on Artificial Intelligence Proceedings*. Seattle, WA. 359–370.
6. Keogh, E. J., and C. A. Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* 7(3):358–386.
7. Vlachos, M., D. Gunopulos, and G. Kollios. 2002. Discovering similar multidimensional trajectories. *IEEE Conference (International) on Data Engineering Proceedings*. San Jose, CA: IEEE Computer Society. 673–684.
8. Chen, L., and R. T. Ng. 2004. On the marriage of lp-norms and edit distance. *30th Conference (International) on Very Large Data Bases Proceedings*. Toronto: Morgan Kaufmann. 792–803.
9. Chen, L., M. T. Özsu, and V. Oria. 2005. Robust and fast similarity search for moving object trajectories. *24th ACM Conference (International) on Management of Data Proceedings*. Baltimore, MD: ACM. 491–502.
10. Frentzos, E., K. Gratsias, and Y. Theodoridis. 2007. Index-based most similar trajectory search. *23rd IEEE Conference (International) on Data Engineering Proceedings*. Istanbul: IEEE Computer Society. 816–825.
11. Morse, M. D., and J. M. Patel. 2007. An efficient and accurate method for evaluating time series similarity. *ACM Conference (International) on Management of Data Proceedings*. Beijing: ACM. 569–580.
12. Chen, Y., M. A. Nascimento, B. C. Ooi, and A. K. H. Tung. 2007. SpADe: On shape-based pattern detection in streaming time series. *23rd Conference (Inter-*

- national) on Data Engineering Proceedings*. Istanbul: IEEE Computer Society. 786–795.
13. Keogh, E. J., and M. J. Pazzani. 1999. Scaling up dynamic time warping to massive datasets. *Principles of Data Mining and Knowledge Discovery: 3rd European Conference Proceedings*. Berlin–Heidelberg: Springer. 1–11.
 14. Salvador, S., and P. Chan. 2004. Fastdtw: Toward accurate dynamic time warping in linear time and space. *3rd Workshop on Mining Temporal and Sequential Data Proceedings*. Seattle, WA. 70–80.
 15. Goncharov, A. V., M. S. Popova, and V. V. Strijov. 2015. Metricheskaya klassifikatsiya vremennykh ryadov s vyravnivaniem otnositel'no tsentroidov klassov [Metric time series classification using dynamic warping relative to centroids of classes]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(4):52–64.
 16. Petitjean, F., G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh. 2014. Dynamic time warping averaging of time series allows faster and more accurate classification. *30th IEEE Conference (International) on Data Engineering Proceedings*. Chicago, IL: IEEE Computer Society. 470–479.
 17. Data from accelerometer. Available at: https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group274/Goncharov2015MetricClassification/data/preprocessed_large.csv (accessed April 28, 2016).

Received December 31, 2015

Contributors

Goncharov Alexey V. (b. 1995) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; alex.goncharov@phystech.edu

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

МЕТРИЧЕСКОЕ ОБУЧЕНИЕ В ЗАДАЧАХ МУЛЬТИКЛАССОВОЙ КЛАССИФИКАЦИИ ВРЕМЕННЫХ РЯДОВ*

Р. В. Исаченко¹, В. В. Стрижов²

Аннотация: Работа посвящена построению модели многоклассовой классификации временных рядов. Предлагается выравнять временные ряды относительно центроидов классов. Процедура нахождения центроидов и выравнивания временных рядов осуществляется с помощью алгоритма динамической трансформации времени. Для повышения качества классификации в данной работе используются методы метрического обучения. Метрическое обучение позволяет модифицировать расстояния между временными рядами, сближая временные ряды из одного класса и отдаляя временные ряды из разных классов. Расстояние между временными рядами измеряется с помощью метрики Махаланобиса. Процедура метрического обучения состоит в определении оптимальной матрицы трансформаций в метрике Махаланобиса. Для анализа качества построенного алгоритма проведен вычислительный эксперимент на синтетических и реальных данных показаний с акселерометра мобильного телефона.

Keywords: классификация временных рядов; выравнивание; метрическое обучение; алгоритм LMNN

DOI: 10.14357/19922264160205

1 Введение

Решается задача мультиклассовой классификации временных рядов [1, 2]. Для решения этой задачи ранее использовались: метод опорных векторов [3, 4], нейронные сети [5], байесовский подход [6]. В данной работе для классификации временных рядов используется идея ближайших соседей [7].

Для повышения качества классификации предлагается использовать методы метрического обучения [8–10]. Метрическое обучение позволяет модифицировать расстояния между временными рядами с помощью линейного преобразования признакового пространства объектов. В результате преобразования временные ряды одного класса оказываются ближе друг к другу по выбранной метрике, а временные ряды, принадлежащие разным классам, отдаляются друг от друга. Методы метрического обучения применяются при ранжировании поисковой выдачи [11], идентификации лиц [12], распознавании рукописных цифр [13]. В данной работе в качестве алгоритма метрического обучения был выбран алгоритм LMNN (Large Margin Nearest Neighbor) [14]. Данный алгоритм основан на идее метода k ближайших соседей. Алгоритм для каждого объекта минимизирует расстояния до k ближайших соседей, принадлежащих тому же классу, и штрафует объекты из других классов, попавшие

на расстояние порядка расстояния до k -го ближайшего соседа.

Алгоритм LMNN позволяет произвести отбор признаков. С помощью линейного преобразования алгоритм помещает объекты в новое пространство. Если размерность нового пространства меньше размерности исходного пространства, то происходит снижение размерности, т. е. отбор признаков.

Для вычисления расстояний между временными рядами в данной работе производится их выравнивание относительно центроидов классов методом динамической трансформации времени DTW (Dynamic Time Warping) [15]. Задача поиска оптимального центроида класса решается с помощью метода выравненного взвешенного усреднения DBA (DTW Bayescenter Averaging) [16]. Классификация, основанная на идее ближайших соседей, чувствительна к изменению масштабов признаков. Для повышения устойчивости классификации выравненные временные ряды были отнормированы.

Таким образом, полученная модель классификации представляет собой суперпозицию алгоритмов построения центроидов, выравнивания временных рядов относительно центроидов классов, метрического обучения и классификации.

В данной работе вычислительный эксперимент проводился на синтетических временных рядах, представляющих аналитические функции, и реаль-

* Работа выполнена при финансовой поддержке РФФИ (проект 16-07-01158).

¹Московский физико-технический институт, isa-ro@yandex.ru

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, strijov@ccas.ru

ных данных показаний акселерометра мобильного телефона. Цель эксперимента — определить вид активности человека по форме сигнала акселерометра. Получена оценка качества работы построенного алгоритма и проведен анализ его свойств.

2 Постановка задачи

Пусть объект $\mathbf{x}_i \in \mathbb{R}^n$ — временной ряд, последовательность измерений некоторой исследуемой величины в различные моменты времени. Пусть \mathbf{X} — множество всех временных рядов фиксированной длины n ; $Y = \{1, \dots, K\}$ — множество меток классов. Пусть задана выборка $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{\ell}$ — множество объектов с известными метками классов $y_i \in Y$.

Требуется построить точную, простую, устойчивую модель классификации

$$a : \mathbf{X} \rightarrow Y.$$

Данную модель представим в виде суперпозиции

$$a(\mathbf{x}) = b \circ \mathbf{f} \circ G(\mathbf{x}, \{\mathbf{c}_e\}_{e=1}^K),$$

где G — процедура выравнивания временных рядов относительно центроидов классов $\{\mathbf{c}_e\}_{e=1}^K$; \mathbf{f} — алгоритм метрического обучения; b — алгоритм многоклассовой классификации.

2.1 Выравнивание временных рядов

Для повышения качества и устойчивости алгоритма классификации предлагается провести выравнивание временных рядов каждого класса относительно центроида.

Пусть \mathbf{X}_e — множество объектов обучающей выборки \mathcal{D} , принадлежащих одному классу $e \in \{1, \dots, K\}$. Центроидом множества объектов $\mathbf{X}_e = \{\mathbf{x}_i | y_i = e\}_{i=1}^{\ell}$ по расстоянию ρ назовем вектор $\mathbf{c}_e \in \mathbb{R}^n$ такой, что

$$\mathbf{c}_e = \operatorname{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{\mathbf{x}_i \in \mathbf{X}_e} \rho(\mathbf{x}_i, \mathbf{c}). \quad (1)$$

Для нахождения центроида предлагается в качестве расстояния между временными рядами использовать путь наименьшей стоимости [17], найденный методом динамической трансформации времени. Псевдокод решения оптимизационной задачи (1) приведен в алгоритме 1.

Общая процедура выравнивания имеет следующий вид:

- (1) построить множество центроидов классов $\{\mathbf{c}_e\}_{e=1}^K$;
- (2) по множеству центроидов найти пути наименьшей стоимости между каждым временным рядом \mathbf{x}_i и центроидом его класса \mathbf{c}_{y_i} ;

Алгоритм 1 Нахождение центроида DBA(\mathbf{X}_e, n_iter)

Вход: \mathbf{X}_e — множество временных рядов, принадлежащих одному и тому же классу, n_iter — количество итераций алгоритма.

Выход: \mathbf{c} — центроид множества \mathbf{X}_e .

1: задать начальное приближение приближение центроида \mathbf{c} ;

2: для $i = 1, \dots, n_iter$

3: для $\mathbf{x} \in \mathbf{X}_e$

4: вычислить выравнивающий путь между \mathbf{c} и \mathbf{x}
 $\text{alignment}(\mathbf{x}) := \text{DTWalignment}(\mathbf{c}, \mathbf{x});$

5: объединить поэлементно множества индексов для каждого отсчета времени
 $\text{alignment} := \bigcup_{\mathbf{x} \in \mathbf{X}_e} \text{alignment}(\mathbf{x});$

6: $\mathbf{c} = \text{mean}(\text{alignment})$

ПРОЦЕДУРА $\text{DTWalignment}(\mathbf{c}, \mathbf{x})$

Вход: \mathbf{c}, \mathbf{x} — временные ряды.

Выход: alignment — выравнивающий путь. // каждый индекс временного ряда \mathbf{x} поставлен в однозначное соответствие индексу временного ряда \mathbf{c}

1: построить $n \times n$ -матрицу деформаций DTW
 $\text{cost} := \text{DTW}(\mathbf{c}, \mathbf{x});$

2: вычислить выравнивающий путь по матрице деформаций
 $\text{alignment} := \text{DTWpath}(\text{cost});$

- (3) по каждому пути восстановить выравненный временной ряд;
- (4) привести множества выравненных временных рядов к нулевому среднему и нормировать на дисперсию.

Результатом выравнивания должно стать множество выравненных временных рядов.

2.2 Метрическое обучение

Введем на множестве выравненных временных рядов расстояние Махаланобиса

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{\top} \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)},$$

где матрица трансформаций $\mathbf{A} \in \mathbb{R}^{n \times n}$ является симметричной и неотрицательно определенной ($\mathbf{A}^{\top} = \mathbf{A}$, $\mathbf{A} \succeq 0$). Представим матрицу \mathbf{A} в виде разложения $\mathbf{A} = \mathbf{L}^{\top} \mathbf{L}$. Матрица $\mathbf{L} \in \mathbb{R}^{p \times n}$ — матрица линейного преобразования, где p задает размерность преобразованного пространства. Если параметр $p < n$, то происходит снижение размерности признакового пространства.

Расстояние $d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j)$ есть евклидово расстояние между $\mathbf{L}\mathbf{x}_i$ и $\mathbf{L}\mathbf{x}_j$:

$$\begin{aligned} d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{\top} \mathbf{L}^{\top} \mathbf{L} (\mathbf{x}_i - \mathbf{x}_j)} = \\ &= \sqrt{(\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j))^{\top} (\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j))} = \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2. \end{aligned}$$

В качестве алгоритма метрического обучения в данной работе был выбран алгоритм LMNN. Данный алгоритм сочетает в себе идеи метода k ближайших соседей. Первая идея заключается в минимизации расстояний между k ближайшими объектами, находящимися в одном классе. Запишем функционал качества в виде

$$Q_1(\mathbf{L}) = \sum_{j \rightsquigarrow i} \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 \rightarrow \min_{\mathbf{L}},$$

где $j \rightsquigarrow i$ означает, что \mathbf{x}_j является одним из k ближайших соседей для \mathbf{x}_i . Вторая идея состоит в максимизации расстояния между каждым объектом и его объектами-нарушителями. Объектом-нарушителем для \mathbf{x}_i назовем объект \mathbf{x}_l такой, что

$$\|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2 \leq \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 + 1, \text{ где } j \rightsquigarrow i. \quad (2)$$

Таким образом, необходимо минимизировать следующий функционал:

$$\begin{aligned} Q_2(\mathbf{L}) &= \sum_{j \rightsquigarrow i} \sum_l (1 - y_{il}) [1 + \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 - \\ &\quad - \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2]_+ \rightarrow \min_{\mathbf{L}}, \end{aligned}$$

где $y_{il} = 1$, если $y_i = y_l$, и $y_{il} = 0$ в противном случае. Положительная срезка позволяет штрафовать только те объекты, которые удовлетворяют условию (2).

Задача метрического обучения состоит в нахождении линейного преобразования $f(\mathbf{x}) = \mathbf{L}\mathbf{x}$, т.е. нахождении матрицы \mathbf{L} в виде решения оптимизационной задачи

$$Q(\mathbf{L}) = \mu Q_1(\mathbf{L}) + (1 - \mu) Q_2(\mathbf{L}) \rightarrow \min_{\mathbf{L}}, \quad (3)$$

где $\mu \in (0, 1)$ — весовой параметр, определяющий вклад каждого из функционалов. Задача (3) представляет собой задачу полуопределенного программирования [18] и может быть решена существующими оптимизационными пакетами.

2.3 Классификация временных рядов

Пусть $\mathbf{x} \in \mathbf{X}$ — неразмеченный временной ряд. Выравниваем временной ряд \mathbf{x} относительно всех центроидов классов $\hat{\mathbf{x}}_e = G(\mathbf{x}, \mathbf{c}_e)$, где $e = \{1, \dots, K\}$.

Отнесем временной ряд к классу, для которого минимально расстояние до соответствующего центроида. В качестве расстояния используем обученную метрику Махаланобиса с фиксированной матрицей \mathbf{A} :

$$\hat{y} = \operatorname{argmin}_{e \in \{1, \dots, K\}} d_{\mathbf{A}}(\hat{\mathbf{x}}_e, \mathbf{c}_e).$$

После нахождения оптимальных центроидов классов и нахождения оптимальной матрицы трансформаций процедура классификации заключается в измерении расстояния между найденными центроидами и новыми неразмеченными объектами.

Для оценки качества работы алгоритма будем вычислять ошибку классификации как долю неправильно классифицированных объектов тестовой выборки \mathcal{U} :

$$\text{error} = \frac{1}{|\mathcal{U}|} \sum_{i=1}^{|\mathcal{U}|} [a(\mathbf{x}_i) \neq y_i].$$

3 Вычислительный эксперимент

Цель вычислительного эксперимента — проверить работоспособность предложенного подхода. Предполагается, что построенный алгоритм мультиклассовой классификации способен определить тип активности человека по форме сигнала акселерометра мобильного телефона.

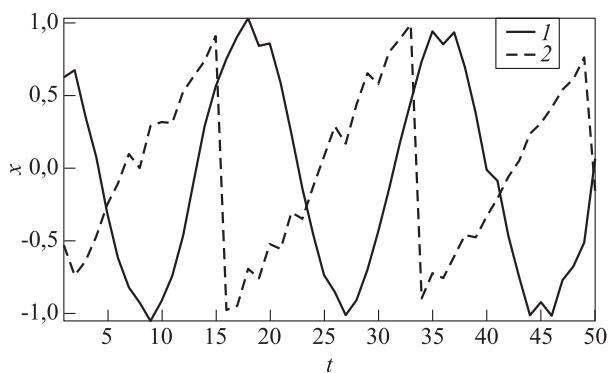


Рис. 1 Центроиды синтетических временных рядов двух классов: 1 — $\sin(x + b)$; 2 — пилообразные функции с различными сдвигами по временной шкале

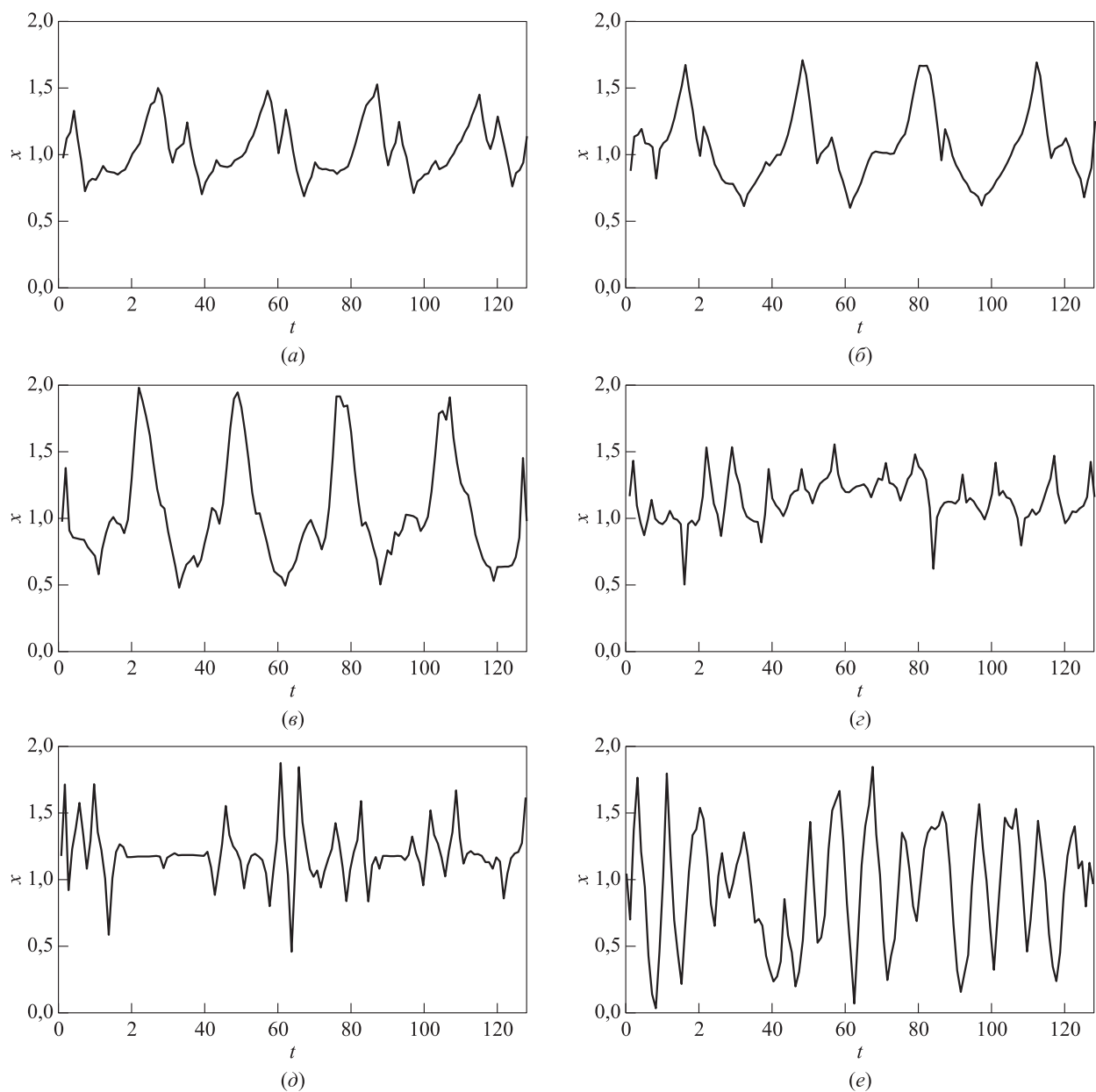


Рис. 2 Центроиды временных рядов акселерометра: (а) ходьба; (б) подъем; (в) спуск; (г) сидение; (д) стояние; (е) лежание

Для проведения базового вычислительного эксперимента были подготовлены синтетические временные ряды, принадлежащие двум классам [19].

Первый класс — синусы вида $\sin(x+b)$, где параметр b определяет сдвиг каждого временного ряда.

Второй класс — пилообразные функции с различными сдвигами по временной шкале. На каждый временной ряд был наложен нормальный шум. Число временных рядов каждого класса = 60. Длина каждого временного ряда $n = 50$. Построенные центры классов проиллюстрированы на рис. 1.

Видно, что процедура корректно определяет сдвиги временных рядов.

Чтобы убедиться в целесообразности применения метрического обучения, данные временные ряды классифицировались в пространстве с евклидовой метрикой и в пространстве с метрикой Махаланобиса. Число ближайших соседей $k = 5$; размерность преобразованного пространства $p = 40$. Полученные ошибки классификации составили:

- евклидова метрика — 27%;
- метрика Махаланобиса — 6%.

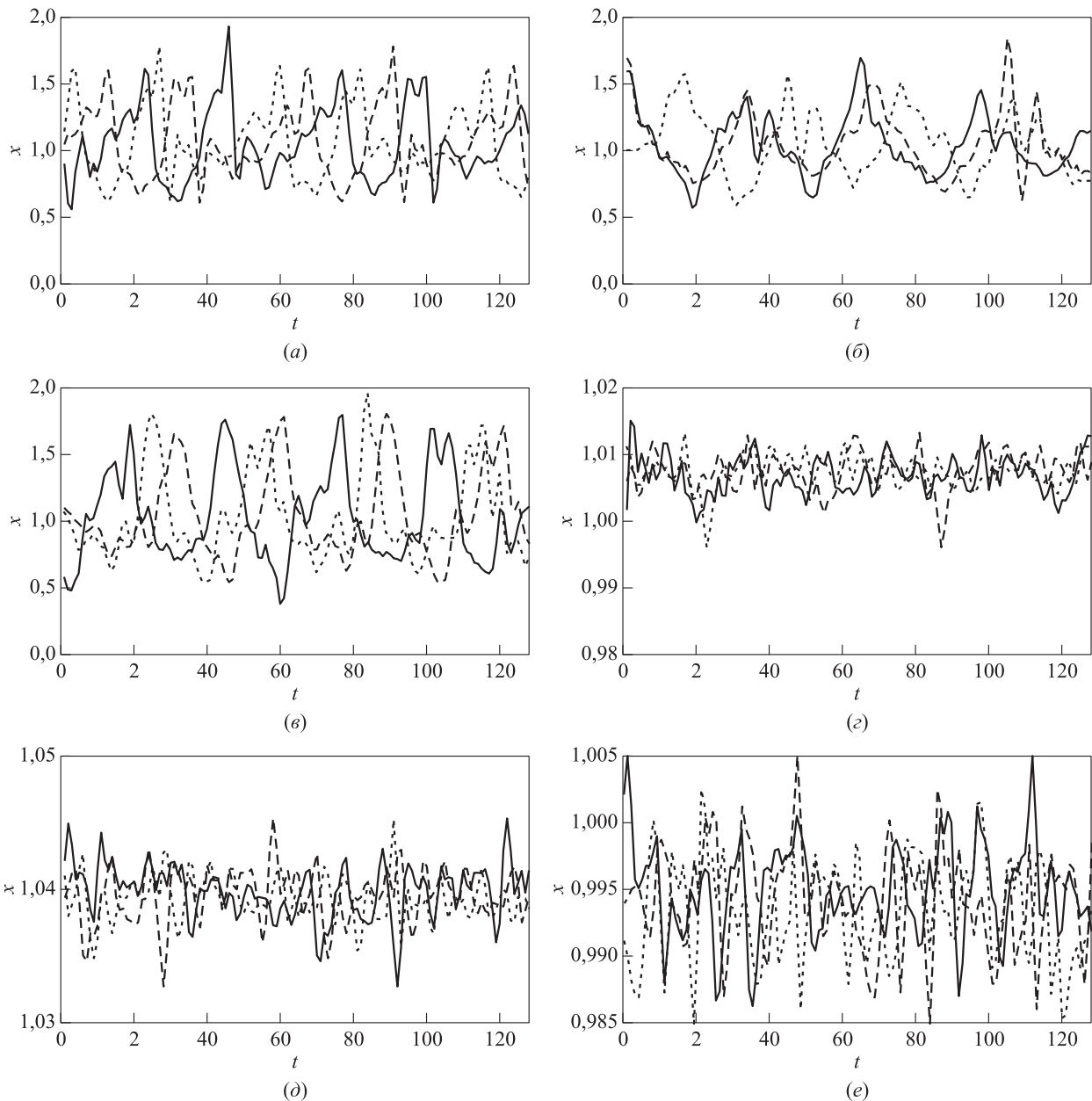


Рис. 3 Временные ряды акселерометра: (а) ходьба; (б) подъем; (в) спуск; (г) сидение; (д) стояние; (е) лежание

Реальные данные [20] представляли собой временные ряды акселерометра мобильного телефона. Каждый из шести классов соответствовал определенной физической активности испытуемых. Для проведения вычислительного эксперимента было выбрано по 200 объектов каждого класса. Длина каждого временного ряда равнялась $n = 128$ отсчетам времени.

Построенные центры классов изображены на рис. 2. Найденные центры обладают периодичностью, свойственной временным рядам по-

казаний активности человека. На рис. 3 показаны примеры временных рядов каждого класса. Эти же временные ряды после процедуры выравнивания относительно построенных центров изображены на рис. 4.

Ошибка классификации без использования метрического обучения составила 37,5%. Алгоритм LMNN позволяет настроить параметры: число ближайших соседей k , размерность преобразованного евклидова пространства p . Для выбора оптимальных параметров воспользуемся процедурой кросс-

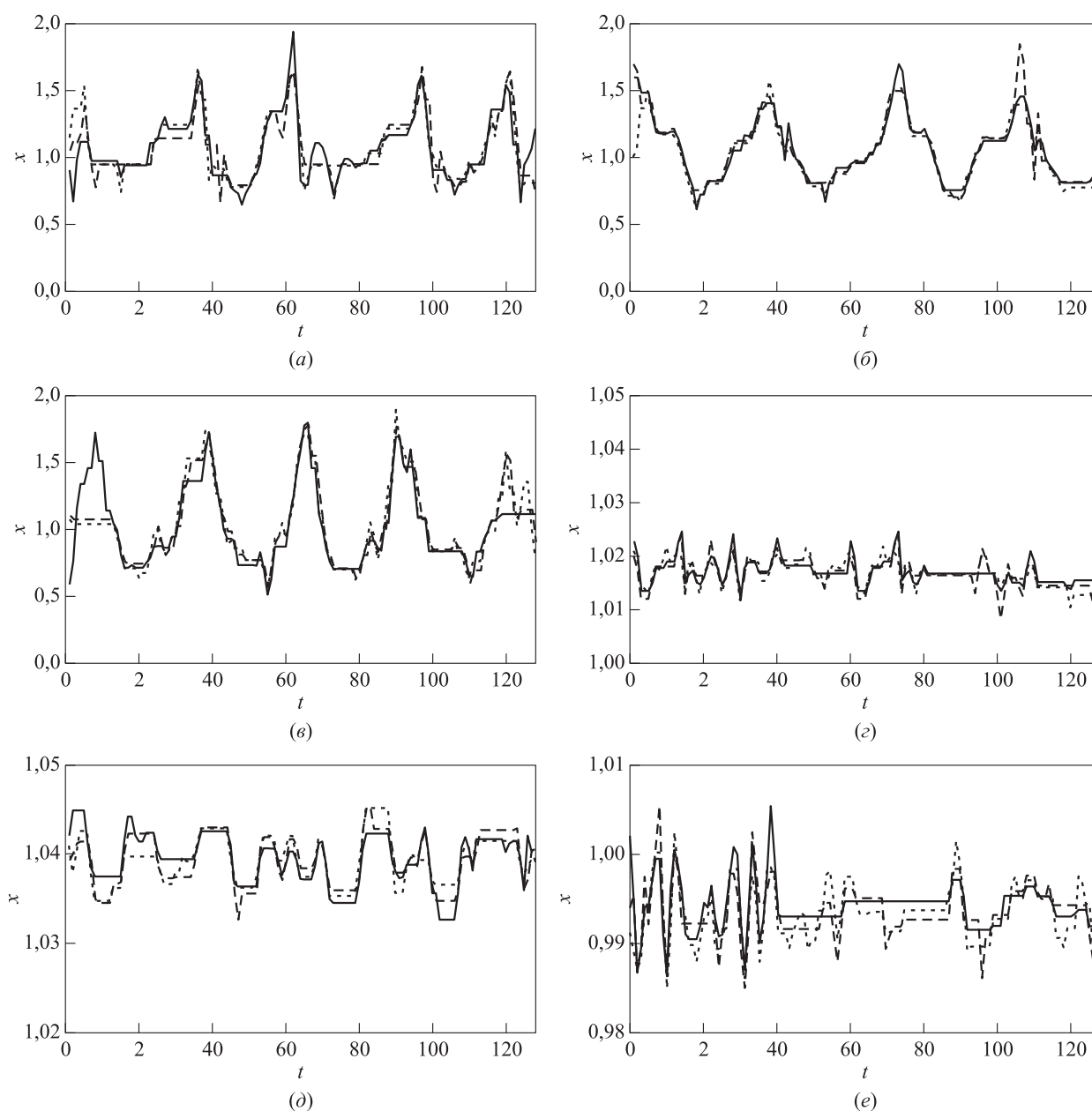


Рис. 4 Выравненные временные ряды акселерометра: (а) ходьба; (б) подъем; (в) спуск; (г) сидение; (д) стояние; (е) лежание

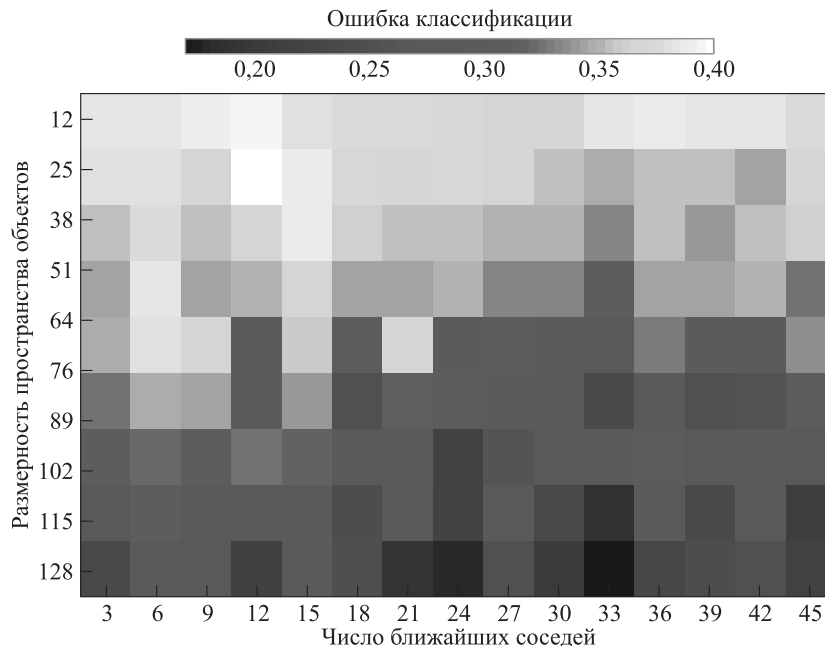


Рис. 5 Ошибка классификации в зависимости от параметров

проверки. На рис. 5 показана ошибка классификации алгоритма в зависимости от его параметров. На данной выборке алгоритм LMNN оказывается слабо чувствителен к числу ближайших соседей, и при уменьшении размерности пространства объектов ошибка классификации растет.

Настроим алгоритм LMNN со следующими параметрами: число ближайших соседей $k = 30$, размерность выходного пространства $p = 128$. Ошибка классификации составила 17,25%, что вдвое меньше ошибки классификации с использованием евклидовой метрики.

В табл. 1 представлены матрицы несоответствий результатов классификации при использовании евклидовой метрики и метрики Махаланобиса. Столбцы соответствуют истинным меткам классов объектов, строки — предсказанным меткам. Диагональное преобладание матрицы несоответствий указывает на высокую предсказательную способность алгоритма.

В табл. 2 продемонстрировано увеличение точности классификации при использовании в качестве меры расстояния метрики Махаланобиса. Пересечение i -го столбца и j -й строки отвечает изменению доли объектов класса i , отнесенных к классу j . Положительное суммарное значение диагональных элементов таблицы соответствует увеличению качества классификации. Значительное улучшение предсказания происходит при классификации первых трех классов. Данные клас-

Таблица 1 Матрицы несоответствий

Предсказанные метки	Истинные метки классов					
	1	2	3	4	5	6
Евклидова метрика						
1	80	0	5	0	0	0
2	4	56	33	0	0	0
3	5	5	86	0	0	0
4	7	8	5	168	4	21
5	51	61	57	12	192	11
6	53	70	14	20	2	168
Метрика Махаланобиса						
1	151	12	13	0	0	0
2	10	142	14	0	0	0
3	9	10	171	0	0	0
4	10	7	0	173	9	21
5	2	11	0	12	186	9
6	18	18	2	15	5	170

сы соответствуют следующим видам физической активности: ходьба, подъем, спуск.

4 Заключение

В данной работе предложен новый подход к решению задачи многоклассовой классификации временных рядов. Сравнивались результаты классификации множества временных рядов, основанных на измерении расстояний с помощью евклидовой метрики и обученной метрики Махаланобиса. Проведен вычислительный эксперимент на реальных данных показаний акселерометра мобильного

Таблица 2 Увеличение точности классификации при использовании адекватной оценки матрицы трансформаций

Предсказанные метки	Истинные метки классов					
	1	2	3	4	5	6
1	0,355	0,06	0,04	0	0	0
2	0,03	0,43	-0,095	0	0	0
3	0,02	0,025	0,425	0	0	0
4	0,015	-0,005	-0,025	0,025	0,025	0
5	-0,245	-0,25	-0,28	0	-0,03	-0,01
6	-0,175	-0,26	-0,06	-0,025	0,005	-0,01

телефона. Построенная модель классификации показала высокое качество распознавания активности человека по форме сигнала акселерометра.

Литература

1. Попова М. С., Стрижов В. В. Выбор оптимальной модели классификации физической активности по измерениям акселерометра // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 79–89.
2. Ignatov A. D., Strijov V. V. Human activity recognition using quasiperiodic time series collected from a single triaxial accelerometer // *Multimed. Tools Appl.*, 2015. 14 p. doi: 10.1007/s11042-015-26430.
3. Güler I., Übeyli E. D. Multiclass support vector machines for eeg-signals classification // *IEEE Trans. Inf. Technol. Biomedicine*, 2007. Vol. 11. No. 2. P. 117–126.
4. Übeyli E. D. Ecg beats classification using multiclass support vector machines with error correcting output codes // *Digit. Signal Process.*, 2007. Vol. 17. No. 3. P. 675–684.
5. Anand R., Mehrotra K., Mohan C. K., Ranka S. Efficient classification for multiclass problems using modular neural networks // *IEEE Trans. Neural Networ.*, 1995. Vol. 6. No. 1. P. 117–124.
6. Kafai M., Bhanu B. Dynamic bayesian networks for vehicle classification in video // *IEEE Trans. Ind. Inform.*, 2012. Vol. 8. No. 1. P. 100–109.
7. Chaovalitwongse W. A., Fan Y.-J., Sachdeo R. C. On the time series k -nearest neighbor classification of abnormal brain activity // *IEEE Trans. Syst. Man Cy. A*, 2007. Vol. 37. No. 6. P. 1005–1016.
8. Yang L., Jin R. Distance metric learning: A comprehensive survey. — Michigan State University, 2006. Vol. 2. 51 p.
9. Bellet A., Habrard A., Sebban M. A survey on metric learning for feature vectors and structured data. arXiv preprint arXiv:1306.6709, 2013.
10. Wang F., Sun J. Survey on distance metric learning and dimensionality reduction in data mining // *Data Min. Knowl. Disc.*, 2015. Vol. 29. No. 2. P. 534–564.
11. McFee B., Lanckriet G. R. Metric learning to rank // 27th Conference (International) on Machine Learning Proceedings, 2010. P. 775–782.
12. Guillaumin M., Verbeek J., Schmid C. Is that you? Metric learning approaches for face identification // IEEE 12th Conference (International) on Computer Vision Proceedings. — IEEE, 2009. P. 498–505.
13. Weinberger K. Q., Saul L. K. Fast solvers and efficient implementations for distance metric learning // 25th Conference (International) on Machine Learning Proceedings, 2008. P. 1160–1167.
14. Weinberger K. Q., Blitzer J., Saul L. K. Distance metric learning for large margin nearest neighbor classification // *Advances in Neural Information Processing Systems*. — Cambridge, MA, USA: MIT Press, 2006. P. 1473–1480.
15. Berndt D. J., Clifford J. Using dynamic time warping to find patterns in time series // *KDD Workshop*, 1994. Vol. 10. No. 16. P. 359–370.
16. Petitjean F., Ketterlin A., Gançarski P. A global averaging method for dynamic time warping, with applications to clustering // *Pattern Recogn.*, 2011. Vol. 44. No. 3. P. 678–693.
17. Гончаров А. В., Попова М. С., Стрижов В. В. Метрическая классификация временных рядов с выравниванием относительно центроидов классов // *Системы и средства информатики*, 2015. Т. 25. № 4. С. 52–64.
18. Vandenberghe L., Boyd S. Semidefinite programming // *SIAM Rev.*, 1996. Vol. 38. No. 1. P. 49–95.
19. Исаченко П. В. Реализация алгоритма классификации временных рядов // Sourceforge.net, 2015. <http://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group274/Isachenko2015TimeSeries/code>.
20. UCI repository. Human activity recognition using smartphones dataset. <https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>.

Поступила в редакцию 18.03.16

METRIC LEARNING IN MULTICLASS TIME SERIES CLASSIFICATION PROBLEM

R. V. Isachenko¹ and V. V. Strijov²

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

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

Abstract: This paper is devoted to the problem of multiclass time series classification. It is proposed to align time series in relation to class centroids. Building of centroids and alignment of time series is carried out by the dynamic time warping algorithm. The accuracy of classification depends significantly on the metric used to compute distances between time series. The distance metric learning approach is used to improve classification accuracy. The metric learning procedure modifies distances between objects to make objects from the same cluster closer and from the different clusters more distant. The distance between time series is measured by the Mahalanobis metric. The distance metric learning procedure finds the optimal transformation matrix for the Mahalanobis metric. To calculate quality of classification, a computational experiment on synthetic data and real data of human activity recognition was carried out.

Keywords: time series classification; time series alignment; distance metric learning; LMNN algorithm

DOI: 10.14357/19922264160205

Acknowledgments

The work was financially supported by the Russian Foundation for Basic Research (project 16-07-01158).

References

1. Popova, M. S., and V. V. Strijov. 2015. Vybor optimal'noy modeli klassifikatsii fizicheskoy aktivnosti po izmereniyam akselerometra [Selection of optimal physical activity classification model using measurements of accelerometer]. *Informatika i ee primeneniya — Inform. Appl.* 9(1):79–89.
2. Ignatov, A. D., and V. V. Strijov. 2015. Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. *Multimed. Tools Appl.* 14 p. doi: 10.1007/s11042-015-26430.
3. Güler, I., and E. D. Übeyli. 2007. Multiclass support vector machines for eeg-signals classification. *IEEE Trans. Inf. Technol. Biomedicine* 11(2):117–126.
4. Übeyli, E. D. 2007. Ecg beats classification using multiclass support vector machines with error correcting output codes. *Digit. Signal Process.* 17(3):675–684.
5. Anand, R., K. Mehrotra, C. K. Mohan, and S. Ranka. 1995. Efficient classification for multiclass problems using modular neural networks. *IEEE Trans. Neural Networ.* 6(1):117–124.
6. Kafai, M., and B. Bhanu. 2012. Dynamic bayesian networks for vehicle classification in video. *IEEE Trans. Ind. Inform.* 8(1):100–109.
7. Chaovalitwongse, W. A., Y.-J. Fan, and R. C. Sachdeo. 2007. On the time series k -nearest neighbor classification of abnormal brain activity. *IEEE Trans. Syst. Man Cy. A* 37(6):1005–1016.
8. Yang, L., and R. Jin. 2006. *Distance metric learning: A comprehensive survey*. Michigan State University. Vol. 2. 51 p.
9. Bellet, A., A. Habrard, and M. Sebban. 2013. A survey on metric learning for feature vectors and structured data. arXiv preprint arXiv:1306.6709.
10. Wang, F., and J. Sun. 2015. Survey on distance metric learning and dimensionality reduction in data mining. *Data Min. Knowl. Disc.* 29(2):534–564.
11. McFee, B., and G. R. Lanckriet. 2010. Metric learning to rank. *27th Conference (International) on Machine Learning Proceedings.* 775–782.
12. Guillaumin, M., J. Verbeek, and C. Schmid. 2009. Is that you? Metric learning approaches for face identification. *IEEE 12th Conference (International) on Computer Vision.* 498–505.
13. Weinberger, K. Q., and L. K. Saul. 2008. Fast solvers and efficient implementations for distance metric learning. *25th Conference (International) on Machine Learning Proceedings.* 1160–1167.
14. Weinberger, K. Q., J. Blitzer, and L. K. Saul. 2005. Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems*. Cambridge, MA: MIT Press. 1473–1480.
15. Berndt, D. J., and J. Clifford. 1994. Using dynamic time warping to find patterns in time series. *KDD Workshop.* 10(16):359–370.

16. Petitjean, F., A. Ketterlin, and P. Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn.* 44(3):678–693.
17. Goncharov, A. B., M. S. Popova, and V. V. Strijov. 2015. Metricheskaya klassifikatsiya vremennykh ryadov s vyravnivaniem otnositel'no tsentroidov klassov [Metric time series classification using dynamic warping relative to centroids of classes]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(4):52–64.
18. Vandenberghe, L., and S. Boyd. 1996. Semidefinite programming. *SIAM Rev.* 38(1):49–95.
19. Isachenko, R. V. 2015. Project code. *Sourceforge.net*. Available at: <https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group274/Isachenko2015TimeSeries/code/> (accessed March 18, 2016).
20. UCI repository. Human activity recognition using smartphones dataset. Available at: <https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones> (accessed March 18, 2016).

Received March 18, 2016

Contributors

Isachenko Roman V. (b. 1994) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; isa-ro@yandex.ru

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

РАСПОЗНАВАНИЕ ЗАВИСИМОСТЕЙ НА ОСНОВЕ ОБРАТНОГО ОТОБРАЖЕНИЯ*

А. Н. Тырсин¹, С. М. Серебрянский²

Аннотация: Описан метод распознавания зависимостей, основанный на использовании обратного отображения. Из заданного конечного множества моделей выбирают ту, которая в наибольшей степени соответствует выборке данных. Для каждой модели по выборке определяют соответствующую ей выборочную зависимость. Для одномерного случая с помощью обратного отображения каждой выборочной зависимости ставится в соответствие одна и та же эталонная модель в виде уравнения прямой. Для каждой модели выборочные данные отображаются с некоторыми ошибками на одно и то же уравнение прямой. В качестве критерия адекватности построенной модели выборки данных предложено использовать минимум дисперсии ошибок. В случае многомерных зависимостей предложен эвристический прием, согласно которому для каждой модели рассматривают совокупность обратных функций для каждой из независимых переменных. Проведена апробация метода с помощью статистического моделирования методом Монте Карло.

Ключевые слова: распознавание; функциональная зависимость; модель; обратная функция; выборка; дисперсия; аппроксимация

DOI: 10.14357/19922264160206

1 Введение

Одной из актуальных проблем во многих приложениях является установление типа модели по выборке экспериментальных данных. Действительно, за исключением относительно малого числа задач, приходится исследовать плохо организованные системы, для которых установить адекватную математическую модель очень сложно или невозможно [1–4]. В этом случае истинный вид модели неизвестен.

В настоящее время предложен ряд методов восстановления зависимостей по экспериментальным данным. Данная задача не может быть строго решена, так как между конечной выборкой и бесконечным количеством функциональных моделей невозможно построить однозначное соответствие. Поэтому на практике упрощают постановку задачи, используя ту или иную априорную информацию. Кроме того, данный подход требует наличия экспериментальных данных сравнительно большого объема, что часто бывает затруднительно обеспечить практически.

Часто может оказаться эффективным подход, не требующий восстанавливать по конечной выборке неизвестную зависимость, а выбрать модель, адекватно ее описывающую. Это означает, что по экспериментальным данным из заданного множества

моделей выбирают ту, которая в наибольшей степени соответствует выборке. Одна и та же выборка может принадлежать с различной вероятностью каждой из рассматриваемых зависимостей. В качестве искомой зависимости выбираем из конечного множества моделей наиболее вероятную модель для данной выборки. Очевидно, без гарантии, что выбрана действительно «наилучшая модель». Однако включение во множество возможных моделей таких, которые достаточно адекватно описывают исследуемую зависимость, позволит получить приемлемое решение.

Один из таких методов распознавания равномерно дискретизированных зависимостей основан на сопоставлении каждой модели соответствующей структурной разностной схемы [5, 6]. Как отмечено в [7, с. 67–73], метод показал свою эффективность для распознавания. Его основным недостатком является ограниченность применения, поскольку не все зависимости можно представить в виде разностных схем.

Задача этой статьи — расширение области применения метода распознавания на основе структурных разностных схем. Это расширение включает в себя следующие случаи: (1) распознавание зависимостей во временных рядах; (2) распознавание одномерных зависимостей; (3) распознавание многомерных зависимостей.

* Работа выполнена при поддержке РФФИ (проект 16-06-00048).

¹ Научно-инженерный центр «Надежность и ресурс больших систем и машин» УрО РАН, Екатеринбург, at2001@yandex.ru

² Троицкий филиал Челябинского государственного университета, tf_chelgu@mail.ru

2 Метод распознавания зависимостей по экспериментальным данным на основе обратного отображения

2.1 Распознавание одномерных зависимостей

Рассмотрим некоторую строго монотонную и непрерывную в заданной области D функцию $y = F_0(x)$. Также зададим конечное множество строго монотонных и непрерывных в заданной области D модельных функций $W = \{F_1(x), \dots, F_J(x)\}$.

Так как функция $F_0(x)$ является строго монотонной, то она имеет обратную функцию. Зададим произвольную монотонную и непрерывную в области D функцию $F(x)$. Для функций $F(x)$ и $F_0(x)$ введем отображение $G(F, F_0): F(x) \rightarrow f(x)$ как

$$\{F^{-1}[F_0(x)]\} = f(x), \quad (1)$$

где F^{-1} — функция, обратная к функции F .

Из свойств обратной функции следует, что если $F_0(x) = F(x)$, то согласно (1) $F_0(x) \xrightarrow{G(F_0, F_0)} x$, или $\{F_0^{-1}[F_0(x)]\} = x$, т. е. в этом случае отображение (1) даст прямую линию $y = x$. Очевидно, что $\forall F(x)$, такой что $F_0(x) \neq F(x)$, получим $\{F^{-1}[F_0(x)]\} \neq x$. Поскольку обратное преобразование является непрерывным, то в результате удастся установить отношение порядка на любом конечном множестве

модельных функций W . В идеале, если одна из функций множества W совпадет с функцией $F_0(x)$, отображение (1) полностью совпадет с линейной функцией $y = x$. В противном случае получим множество функций $\{F_j^{-1}[F_0(x)]\} = f_j(x) \neq x, j = 1, 2, \dots, J$.

В табл. 1 приведены примеры нескольких распространенных моделей функции и обратных к ним. Не всегда обратную функцию можно выразить в явном виде. Например, для функции $y = xe^x$ обратная функция задается неявно как $x \ln x = y$.

Пусть имеется экспериментальная выборка данных $(x_1, y_1), \dots, (x_n, y_n)$, состоящая из измеренных значений независимой и зависимой переменной. Величины y_i — это измеренные со случайными погрешностями ε_i оценки значений z_i , которые описываются некоторой строго монотонной и непрерывной в заданной области D функциональной зависимостью $z = F_0(x)$. Считаем, что дисперсия случайных погрешностей конечна. Зададим конечное множество строго монотонных и непрерывных в заданной области D модельных функций $W = \{F_1(x), \dots, F_J(x)\}$. Также потребуем, чтобы количество параметров модельных функций было значительно меньше числа наблюдений.

Необходимо определить среди множества функций W наиболее вероятную модель для данной выборки.

Для выборки данных $(x_1, y_1), \dots, (x_n, y_n)$ в результате аппроксимации, например с помощью метода наименьших квадратов [8], вместо функции $F_0(x)$ получим функцию $\hat{F}_0(x)$ того же типа, но ее

Таблица 1 Примеры обратных функций

j	Модель $y = F_j(x)$	Обратное преобразование $x = F_j^{-1}(y)$
1	$y = A + Bx^C$	$x = \left(\frac{y - A}{B}\right)^{1/C}$
2	$y = A + B \ln x$	$x = \exp\left(\frac{y - A}{B}\right)$
3	$y = A + Be^{Cx}$	$x = \frac{1}{C} \ln\left(\frac{y - A}{B}\right)$
4	$y = \frac{A}{1 + Be^{Cx}} + D$	$x = \frac{1}{C} \ln\left(\frac{A - (y - D)}{B(y - D)}\right)$
5	$y = A \exp\{Be^{Cx}\} + D, A > 0, B < 0, C < 0$	$x = \frac{1}{C} \ln\left(\frac{1}{B} \ln \frac{y - D}{A}\right)$
6	$y = \frac{A + Bx}{C + x}$	$x = \frac{A - Cy}{y - B}$
7	$y = \frac{Ax^2 + B}{C + x}$	$x = \frac{y - D \pm \sqrt{(y - D)^2 + 4A(C(y - D) - B)}}{2A}$

параметры будут отличаться от параметров функции $F_0(x)$. В результате получим множество значений $\hat{F}_0(x_i)$, $i = 1, 2, \dots, n$. Отметим, что в случае конечной дисперсии случайных погрешностей ε_i метод наименьших квадратов дает состоятельные оценки параметров модели, т. е. имеет место сходимость по вероятности $\hat{F}_0(x) \xrightarrow{P} F_0(x)$. Если плотность вероятности случайных погрешностей имеет более вытянутые хвосты по сравнению с нормальным законом, то оценку параметров модели можно выполнить иначе, например с помощью метода наименьших модулей [9].

Суть распознавания состоит в следующем. Для каждого типа модели $F_j(x)$, $j = 1, 2, \dots, J$, строим соответствующую оценку $\hat{F}_j(x)$.

Далее всем значениям y_i , $i = 1, 2, \dots, n$, с помощью (1) ставим в соответствие некоторые величины:

$$\hat{x}_i^{(j)} = \hat{F}_j^{-1}(y_i), \quad j = 1, 2, \dots, J, \quad (2)$$

где $\hat{F}_j^{-1}(\cdot)$ — функция, обратная функции $\hat{F}_j(\cdot)$.

Очевидно, что преобразование (2) не даст прямой линии $y = x$ и имеет ошибки:

$$e_i^{(j)} = \hat{x}_i^{(j)} - x_i, \quad i = 1, 2, \dots, n.$$

Ошибки $e_i^{(j)}$ отображения $\hat{F}_j^{-1}(y_i)$ обусловлены случайной погрешностью, а также различием между функцией $\hat{F}_j(x)$ и истинной зависимостью $F_0(x)$.

Пусть $\hat{F}_j(x_i^{(j)*}) = y_i$, где $x_i^{(j)*}$ — решение уравнения:

$$\hat{F}_j(x_i^{(j)*}) = y_i. \quad (3)$$

Тогда

$$\begin{aligned} e_i^{(j)} &= \hat{x}_i^{(j)} - x_i = \hat{F}_j^{-1}(y_i) - x_i = \\ &= \hat{F}_j^{-1}(\hat{F}_j(x_i^{(j)*})) - x_i = x_i^{(j)*} - x_i. \end{aligned} \quad (4)$$

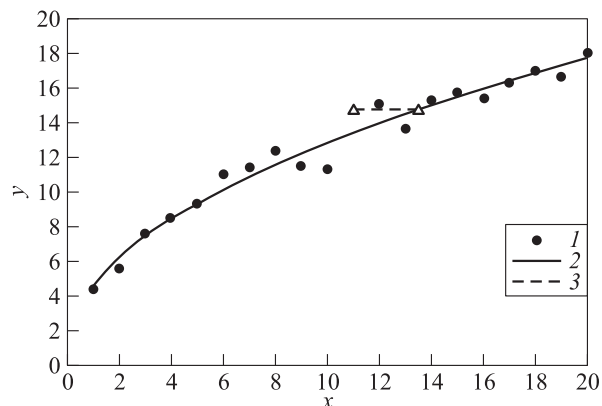


Рис. 1 Построение зависимости $\hat{F}(x)$ по выборке данных: 1 — y ; 2 — $\hat{F}(x)$; 3 — e_{11}

Отсюда (2) с учетом (3) и (4) примет вид:

$$\hat{x}_i^{(j)} = \hat{F}_j^{-1}(y_i) = x_i^{(j)*}. \quad (5)$$

Соотношение (5) позволяет строить обратные отображения (2) без непосредственного вычисления обратной функции. Вместо решения уравнения (5) можно рассматривать задачу минимизации:

$$\hat{x}_i^{(j)} = \arg \min_x \left(\hat{F}_j(x) - y_i \right)^2. \quad (6)$$

Использование задачи (6) значительно упрощает реализацию метода, так как нахождение обратной функции в ряде случаев может оказаться трудоемкой процедурой, например, если обратная функция может быть задана только неявно. Кроме того, можно расширить область применения и рассматривать и немонотонные функции. При этом согласно (6) в качестве оценки $\hat{x}_i^{(j)}$ следует использовать ближайшее к x_i решение уравнения (6).

Отметим, что если уравнения (2) или (6) для некоторого значения y_i не имеют решения, то это говорит о том, что для этого значения не существует обратного отображения $\hat{F}_j^{-1}(y_i)$, т. е. модель $F_j(x)$ не соответствует экспериментальным данным и должна быть исключена из рассмотрения.

Пример 1. Проиллюстрируем преобразование (2) с помощью рис. 1 и 2.

На рис. 1 показаны:

- исходное множество наблюдений (x_i, y_i) (1), $i = 1, 2, \dots, 20$, для наглядности задано $\forall i \ x_i = i$;
- график полученной в результате аппроксимации моделью $F(x) = b_0 + b_1\sqrt{x}$ зависимости $\hat{F}(x) = 0,945 + 3,758\sqrt{x}$ (2);

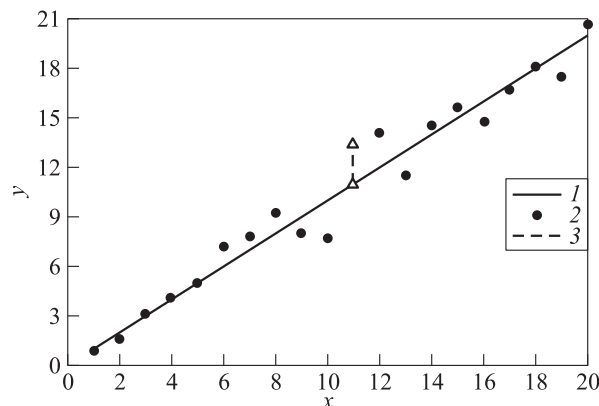


Рис. 2 Построение обратного отображения $\hat{x}_i = \hat{F}^{-1}(y_i)$ по выборке данных: 1 — $y = x$; 2 — \hat{x}_i ; 3 — e_{11}

– точки (x_{11}, y_{11}) и (x_{11}^*, y_{11}) , обозначенные треугольниками, где x_{11}^* — решение уравнения $\hat{F}(x_{11}^*) = y_{11}$;

– ошибка для 11-го наблюдения $e_{11} = x_{11}^* - x_{11}$, выделенная пунктиром (3).

На рис. 2 изображены:

– график прямой $y = x$ (1), соответствующей идеальному случаю, когда в (1) $F_0(x) = F(x)$;

– множество точек (x_i, \hat{x}_i) (2), $i = 1, 2, \dots, 20$, где $\hat{x}_i = \hat{F}^{-1}(y_i) = ((y_i - 0,945)/3,758)^2$;

– точки (x_{11}, x_{11}) и (x_{11}, \hat{x}_{11}) , обозначенные треугольниками;

– ошибка для 11-го наблюдения $e_{11} = \hat{x}_{11} - x_{11}$, выделенная пунктиром (3).

Далее из множества моделей W на основании рассмотрения множеств наборов

$$\left\{ (x_1, \hat{x}_1^{(j)}), \dots, (x_n, \hat{x}_n^{(j)}) \right\}, \quad j = 1, 2, \dots, J, \quad (7)$$

выбираем тот, который наиболее соответствует линейной зависимости $y = x$. В качестве критерия соответствия считаем выбор наиболее вероятной модели.

Возможны два случая.

Случай 1. Данные имеют вид временного ряда $y_k = y(t_k) = y(k\Delta)$, где Δ — интервал дискретизации, т. е. вместо независимой переменной имеем номер измерения. При этом вместо множеств наборов (7) имеем множество временных рядов $\left\{ \hat{x}_1^{(j)}, \dots, \hat{x}_n^{(j)} \right\}$, $j = 1, 2, \dots, J$.

Утверждение. Пусть задана числовая последовательность $\{f(k)\}$, где $f(k)$ — некоторая непрерывная монотонная функция. Зададим произвольное множество последовательностей $W = \{f_j(k)\}$, $f_j(k)$ — непрерывная монотонная функция, $j \in \Omega$, Ω — множество индексов. Тогда множество W упорядочено по отношению порядка в смысле близости $\{f_j(k)\}$ к $\{f(k)\}$.

Доказательство. Пусть $\{u_k\}$ — некоторая числовая последовательность. Поставим ей в соответствие вектор $\mathbf{a} = (a_1, a_2)$ коэффициентов разностной схемы с помощью отображения $\Gamma(\{u_k\})$: $\{u_k\} \rightarrow \mathbf{a}$ как

$$\begin{aligned} \mathbf{a} &= \\ &= \arg \min_{\mathbf{a} \in \mathbf{R}} \left[\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{k=1}^{\infty} (u_k - a_1 u_{k-1} - a_2 u_{k-2})^2 \right]. \quad (8) \end{aligned}$$

Из определения обратного отображения следует $\{f^{-1}f(k)\} = \{k\}$, где f^{-1} — отображение, обрат-

ное к f . Последовательности $\{k\}$ согласно (8) будет соответствовать вектор $\mathbf{a}^* = (a_1^*, a_2^*) = (2, -1)$ [5]. Таким образом, $\{k\} \xrightarrow{\Gamma(\{k\})} \mathbf{a}^*$.

Очевидно, что для каждой функции f_j существует последовательность $\{u_k^j\}$, элементы которой $u_k^j = f_j^{-1}[f(k)]$, где f_j^{-1} — отображение, обратное к f_j . Если $f_j \neq f$, то $\{u_k^j\} \xrightarrow{\Gamma(\{u_k^j\})} (a_1^j, a_2^j) = \mathbf{a}^j \neq \mathbf{a}^*$.

Поставим в соответствие каждой последовательности $\{f_j(k)\}$ функционал $d_j(\{f(k)\}) = \sqrt{(\mathbf{a}^j, \mathbf{a}^*)}$. Из непрерывности отображения (8) [10] вытекает упорядоченность множества W по отношению порядка: если $d_j(\{f(k)\}) \leq d_l(\{f(k)\})$, то $f_j \preceq f_l$.

На основании утверждения реализуем алгоритм распознавания. Для всех J последовательностей $\{\hat{x}_k^{(j)}\}$, $j = 1, 2, \dots, J$, определяем векторы $\hat{\mathbf{a}}^{(j)}$, решая задачи минимизации:

$$\begin{aligned} \hat{\mathbf{a}}^{(j)} &= \arg \min_{\mathbf{a} \in \mathbf{R}} \sum_{k=3}^n (u_k - a_1 u_{k-1} - a_2 u_{k-2})^2, \\ & \quad j = 1, 2, \dots, J. \end{aligned}$$

В качестве наилучшей модели выбираем ту, для которой

$$d_j(\{\hat{x}_k^{(j)}\}) = \sqrt{(\hat{\mathbf{a}}^{(j)}, \mathbf{a}^*)}$$

имеет минимальное значение.

Случай 2. Исходные данные $(x_1, y_1), \dots, (x_n, y_n)$, а значит, и множество наборов (7) имеют общий вид.

Если модель $F_0(x)$ входит в состав множества W , то с учетом сходимости по вероятности $\hat{F}_0(x) \xrightarrow{P} F_0(x)$ получим, что $\hat{x}_i^{(0)} \xrightarrow{P} x_i$.

Преобразование (2) независимо от функциональной формы моделей из множества W отображает их оценки $\hat{F}_j(x_i)$ на одну и ту же прямую $y = x$. Чем более адекватна прямая $y = x$ множеству точек $\left\{ (x_1, \hat{x}_1^{(j)}), \dots, (x_n, \hat{x}_n^{(j)}) \right\}$, а значит, по вероятности и функции $F_0(x)$, тем ближе будут расположены точки $\hat{x}_i^{(k)}$ к значениям x_i . Известно [8, 11], что для однотипных моделей критерием качества (степени адекватности построенной модели выборочной совокупности) является дисперсия ошибок:

$$s_j^2 = \frac{1}{n - l_j} \sum_{i=1}^n (e_i^{(j)})^2, \quad (9)$$

где l_j — число параметров в j -й модели.

Согласно этому критерию в качестве наиболее достоверной модели для исходной выборки выбираем модельную функцию $F_j(x)$, которая обеспечит минимум дисперсии ошибок (9).

2.2 Распознавание многомерных зависимостей

Пусть имеем выборку данных $(y_i, x_{i1}, x_{i2}, \dots, x_{im}), i = 1, 2, \dots, n$. Требуется подобрать из заданного множества моделей $F_j(x_1, x_2, \dots, x_m), j = 1, 2, \dots, J$, наилучшую для этой выборки данных.

Для функций нескольких переменных непосредственно задать обратную функцию нельзя. Поэтому воспользуемся следующим приемом. Для каждой функции $F_j(x_1, x_2, \dots, x_m)$ будем рассматривать совокупность m обратных функций одной переменной $F_{jk}^{-1}(y)$, область значений каждой из которых принадлежит оси $Ox_k, k = 1, 2, \dots, m$.

Для всех моделей строим по выборке данных оценки $\hat{F}_j(x_1, x_2, \dots, x_m), j = 1, 2, \dots, J$.

Далее для каждой зависимости $\hat{F}_j(x_1, x_2, \dots, x_m)$ решаем задачу Лагранжа:

$$\left. \begin{aligned} & (x_{i1}^{(j)*}, x_{i2}^{(j)*}, \dots, x_{im}^{(j)*}) = \\ & = \arg \min_{(x_1, \dots, x_m) \in \mathbf{R}^m} \sum_{k=1}^m (x_i - x_{ik})^2; \\ & \hat{F}_j(x_1, x_2, \dots, x_m) = y_i, \quad i = 1, 2, \dots, n. \end{aligned} \right\} \quad (10)$$

По аналогии с (4) суммарные ошибки совокупности отображений $F_{jk}^{-1}(y), k = 1, 2, \dots, m$, будут равны:

$$e_i^{(j)} = \sqrt{\sum_{k=1}^m (x_i^{*(j)} - x_{ik})^2}.$$

В качестве наиболее вероятной модели выбираем ту, у которой дисперсия ошибок (9) будет минимальной.

Таким образом, в многомерном случае вместо уравнения (5) или задачи безусловной минимизации (6) приходится решать задачу Лагранжа (10) с одним ограничением в форме равенства, согласно которому поиск оптимального вектора $\mathbf{x}_i^{(j)*}$ ищут на поверхности уровня $\hat{F}_j(x_1, x_2, \dots, x_m) = y_i$.

3 Результаты вычислительных экспериментов

Воспользуемся методом статистических испытаний Монте Карло [12].

Пример 2. Рассмотрим множество W из семи моделей F_j , приведенных в табл. 1. Проведем численный эксперимент. Для каждой из моделей $F_j, j = 1, 2, \dots, 7$, смоделируем выборку $(x_1, y_1), \dots, (x_n, y_n)$, где x_i — случайное число, равномерно распределенное от 1 до 20; $y_i^{(j)} = F_j(x_i, \mathbf{b}^{(j)}) + \varepsilon_i$, где ε_i — случайные погрешности

измерений, имеющие нормальное распределение $N(0, \sigma_\varepsilon^2); n = 100; \mathbf{b}^{(j)}$ — векторы параметров каждой из моделей, число повторений эксперимента зададим $M = 100$. Для определенности выберем векторы $\mathbf{b}^{(j)}$ так, чтобы конкретные функции $F_j(x)$ были достаточно похожи на отрезке, а именно:

- для 1-й модели $A = 0,2, B = 4, C = 0,5$;
- для 2-й модели $A = 1, B = 5$;
- для 3-й модели $A = -11, B = 11, C = 0,05$;
- для 4-й модели $A = 20, B = 60, C = -0,4, D = 0$;
- для 5-й модели $A = 20, B = -3, C = -0,18, D = 0$;
- для 6-й модели $A = -8, B = 21, C = 6$;
- для 7-й модели $A = 2, B = 15, C = 20, D = 0$.

В каждом эксперименте для всех моделей $F_j, j = 1, 2, \dots, 7$, выполним по 7 преобразований (5) и (4), т. е. определим

$$\hat{x}_i^{(j,k)} = \hat{F}_k^{-1}(\hat{y}_i^{(j)}), e_i^{(j,k)} = \hat{x}_i^{(j,k)} - x_i, \quad k = 1, 2, \dots, 7,$$

а затем по формуле (9) вычислим дисперсии ошибок как

$$s_{j,k}^2 = \frac{1}{n - l_j} \sum_{i=1}^n (e_i^{(j,k)})^2. \quad (11)$$

Очевидно, что в идеале минимальными значения дисперсий (11) должны быть при совпадении моделей, т. е. когда $k = j$.

Результаты статистического моделирования показали, что при среднем квадратическом отклонении случайных погрешностей измерений $\sigma_\varepsilon < 0,14$ практически всегда минимальное значение оценки (11) достигалось при $k = j$. В качестве иллюстрации в табл. 2 приведены усредненные по 100 испытаниям значения оценок дисперсий (11) для случая $\sigma_\varepsilon = 0,1$.

В табл. 3 приведены относительные частоты правильного выбора модели ($k = j$) в процентах от общего числа испытаний $M = 100$ для нескольких значений среднего квадратического отклонения случайных погрешностей измерений σ_ε .

Анализируя полученные результаты, можно сказать, что метод для рассматриваемых моделей показал хорошие результаты. Фактическая модель для приемлемых на практике значений σ_ε в большинстве случаев правильно распознается.

Таблица 2 Усредненные значения оценок дисперсий $s_{j,k}^2$ для случая $\sigma_\varepsilon = 0,1$

j	k						
	1	2	3	4	5	6	7
1	0,0293	2,3709	0,8086	0,0888	0,0816	0,0486	0,0374
2	0,1022	0,0661	4,0069	1,1407	1,3204	0,1823	0,0933
3	0,0473	29,7966	0,0135	3,0720	0,0167	2,7542	0,1455
4	0,9851	7,8095	1,1530	0,0566	0,4344	3,3365	0,9882
5	0,3657	1,6192	1,0961	0,3881	0,0319	0,2322	0,4834
6	0,6450	0,1349	4,1431	1,5220	0,4890	0,0851	1,4468
7	0,0291	32,1426	0,0793	32,9608	0,0222	3,1160	0,0189

Таблица 3 Относительные частоты выбора фактической модели в качестве истинной, %

σ_ε	1	2	3	4	5	6	7
0	100	100	100	100	100	100	100
0,1	100	97	98	100	100	99	97
0,15	96	91	93	100	100	94	90
0,2	84	77	80	100	100	82	75
0,25	65	55	61	100	100	62	53
0,3	44	33	37	100	100	40	30

4 Заключение

1. Рассмотрен метод распознавания зависимостей на основе обратного отображения. Метод позволяет выбрать среди заданного множества зависимостей наиболее вероятную модель на основе минимума дисперсии ошибок обратного отображения для соответствующей модели. При этом экспериментальные данные отображаются на одну и ту же линейную зависимость $y = x$. Дисперсия ошибок для однотипных зависимостей является показателем адекватности моделей экспериментальным данным.
2. Множество рассматриваемых зависимостей формируется из приемлемых для конкретных экспериментальных данных моделей. Число типовых зависимостей может быть любым.
3. Примеры реализации метода на тестовых данных показали его работоспособность.

Литература

1. *Налимов В. В.* Теория эксперимента. — М.: Наука, 1971. 208 с.

2. *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Исследование зависимостей. — М.: Финансы и статистика, 1985. 487 с.

3. *Клейнер Г. Б., Смоляк С. А.* Эконометрические зависимости: принципы и методы построения. — М.: Наука, 2000. 104 с.

4. *Орлов А. И.* Прикладная статистика. — 2-е изд., перераб. и доп. — М.: Экзамен, 2007. 671 с.

5. *Тырсин А. Н.* Идентификация зависимостей на основе моделей авторегрессии // Автометрия, 2005. Т. 41. № 1. С. 43–49.

6. *Тырсин А. Н., Серебрянский С. М.* Распознавание зависимостей во временных рядах на основе структурных разностных схем // Автометрия, 2015. Т. 51. № 2. С. 54–60.

7. *Букреев В. Г., Колесникова С. И., Янковская А. Е.* Выявление закономерностей во временных рядах в задачах распознавания состояний динамических объектов. — 2-е изд., испр. и доп. — Томск: ТПУ, 2011. 254 с.

8. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс. — 6-е изд., перераб. и доп. — М.: Дело, 2004. 576 с.

9. *Мудров В. И., Кушко В. Л.* Методы обработки измерений: Квазиправдоподобные оценки. — 2-е изд., перераб. и доп. — М.: Радио и связь, 1983. 304 с.

10. *Тырсин А. Н.* Модель авторегрессии как отображение функциональной зависимости временного ряда // Системы управления и информационные технологии, 2005. № 1(18). С. 27–29.

11. *Greene W. H.* Econometric analysis. — 7th ed. — Prentice Hall, 2011. 1230 p.

12. *Михайлов Г. А., Войтишек А. В.* Численное статистическое моделирование. Методы Монте-Карло. — М.: Академия, 2006. 368 с.

Поступила в редакцию 16.02.16

RECOGNITION OF DEPENDENCES ON THE BASIS OF INVERSE MAPPING

A. N. Tyrsin¹ and S. M. Serebryanskii²

¹Science and Engineering Center “Reliability and Resource of Large Systems and Machines,” Ural Branch of the Russian Academy of Sciences; 54a Studencheskaya Str., Yekaterinburg 620049, Russian Federation

²Troitsk Branch of Chelyabinsk State University, 9 S. Rasin Str., Troitsk 457100, Russian Federation

Abstract: The article describes the method of recognition of dependences based on the use of inverse mapping. From a given finite set of models, one chooses the model that best fits the sample data. For each model, the selective dependence corresponding to it is determined by the sample. For the one-dimensional case, each selective dependence is mapped to the same reference model in the form of the straight line equation by means of inverse mapping. For each model, sample data are mapped to the same equation of the straight line with some mistakes. It is suggested to use the minimum of variance of mistakes as the criterion of adequacy of the constructed model of sample of data. In the case of multidimensional dependences, a heuristic method is suggested according to which a set of inverse functions for each of explanatory variables is considered for each model. Approbation of the method by means of statistical modeling by the Monte-Carlo method is carried out.

Keywords: recognition; functional dependence; model; inverse function; sample; variance; approximation

DOI: 10.14357/19922264160206

Acknowledgments

The work is supported by the Russian Foundation for Basic Research (grant No. 16-06-00048).

References

1. Nalimov, V. V. 1971. *Teoriya eksperimenta* [The theory of the experiment]. Moscow: Nauka. 208 p.
2. Ajvazyan, S. A., I. S. Enjukov, and L. D. Meshalkin. 1985. *Prikladnaya statistika: Issledovanie zavisimostey* [Applied statistics: Research of dependences]. Moscow: Finance and Statistics. 487 p.
3. Kleyner, G. B., and S. A. Smolyak. 2000. *Ekonometricheskie zavisimosti: Printsipy i metody postroeniya* [Econometric dependences: Principles and methods of construction]. Moscow: Nauka. 104 p.
4. Orlov, A. I. 2007. *Prikladnaya statistika* [Applied statistics]. 2nd ed. Moscow: Examen. 671 p.
5. Tyrsin, A. N. 2005. Identifikatsiya zavisimostey na osnove modeley avtoregressii [Identification of dependences on the basis of autoregressive models]. *Avtometriya* 41(1):43–49.
6. Tyrsin, A. N., and S. M. Serebryanskii. 2015. Raspoznavanie zavisimostey vo vremennykh ryadakh na osnove strukturnykh raznostnykh skhem [Dependence identification in a time series on the basis of structural difference schemes]. *Avtometriya* 51(2):54–60.
7. Bukreev, V. G., S. I. Kolesnikova, and A. E. Jankovskaya. 2011. *Vyyavlenie zakonmernostey vo vremennykh ryadakh v zadachakh raspoznavaniya sostoyaniy dinamicheskikh ob"ektov* [To identify patterns in time series in the tasks of recognition of dynamic objects]. 2nd ed. Tomsk: Tomsk Polytechnic University Publ. 254 p.
8. Magnus, Ja. R., P. K. Katyshev, and A. A. Pereseckiy. 2004. *Ekonometrika. Nachal'nyy kurs* [Econometrics. Basic course]. 6th ed. Moscow: Delo. 576 p.
9. Mudrov, V. I., and V. L. Kushko. 1983. *Metody obrabotki izmereniy: Kvazipravdopodobnye otsenki* [Treatment methods of measurement: Quasi-plausible estimations]. 2nd ed. Moscow: Radio and Communication. 304 p.
10. Tyrsin, A. N. 2005. Model' avtoregressii kak otobrazhenie funktsional'noy zavisimosti vremennogo ryada [Autoregressive model as a mapping of the functional dependence of time series]. *Sistemy Upravleniya i Informatzionnye Tekhnologii* [Control Systems and Information Technology] 1(18):27–29.
11. Greene, W. H. 2011. *Econometric analysis*. 7th ed. Prentice Hall. 1230 p.
12. Mikhaylov, G. A., and A. V. Voytishchek. 2006. *Chislennoe statisticheskoe modelirovanie. Metody Monte-Karlo* [Numerical statistical modeling. Monte-Carlo methods]. Moscow: Academy Publ. 368 p.

Received February 16, 2016

Contributors

Tyrsin Alexander N. (b. 1961) — Doctor of Science in technology, leading researcher, Science and Engineering Center “Reliability and Resource of Large Systems and Machines,” Ural Branch of the Russian Academy of Sciences; 54a Studencheskaya Str., Yekaterinburg 620049, Russian Federation; at2001@yandex.ru

Serebryanskii Sergey M. (b. 1983) — senior lecturer, Troitsk Branch of Chelyabinsk State University, 9 S. Rasin Str., Troitsk 457100, Russian Federation; tf_chelgu@mail.ru

СТАТИСТИЧЕСКИЕ СВОЙСТВА МЕТОДА ПОДАВЛЕНИЯ ШУМА, ОСНОВАННОГО НА СТАБИЛИЗИРОВАННОЙ ЖЕСТКОЙ ПОРОГОВОЙ ОБРАБОТКЕ*

О. В. Шестаков¹

Аннотация: Методы пороговой обработки коэффициентов вейвлет-разложения функций сигналов и изображений стали популярным аппаратом для подавления шума благодаря своей простоте, вычислительной эффективности и возможности адаптации к функциям, имеющим на разных участках различную степень регулярности. Рассматривается предложенный недавно стабилизированный метод жесткой пороговой обработки, в котором устранены основные недостатки мягкой и жесткой пороговой обработки, и исследуются статистические свойства этого метода. В модели данных с аддитивным гауссовским шумом проводится анализ несмещенной оценки среднеквадратичного риска и показывается, что при определенных условиях данная оценка является сильно состоятельной и асимптотически нормальной. Данные свойства позволяют строить асимптотические доверительные интервалы для теоретического среднеквадратичного риска метода.

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

DOI: 10.14357/19922264160207

1 Введение

При применении методов подавления шума в сигналах или изображениях важным условием является «экономное» представление функций сигналов в некотором специальном базисе. Данное условие позволяет применять популярные процедуры мягкой и жесткой пороговой обработки [1] к коэффициентам разложения функций сигналов. Эти процедуры просты, не затратны и часто позволяют достичь удовлетворительных результатов. Однако они имеют свои недостатки. Жесткая пороговая обработка использует разрывную пороговую функцию, что приводит к появлению дополнительных артефактов, отсутствию устойчивости при выборе порога [2] и невозможности построения несмещенной оценки среднеквадратичной погрешности [3]. При мягкой пороговой обработке все коэффициенты подвергаются изменению, вследствие чего в оценке функции сигнала появляется дополнительное смещение.

В работе [4] предложен стабилизированный вариант жесткой пороговой обработки, позволяющий обойти указанные недостатки. Для выбора параметров данного метода и анализа погрешности можно использовать несмещенную оценку среднеквадратичного риска [5]. В данной работе исследуются

статистические свойства этой оценки при разложении функции сигнала по вейвлет-базису и выборе «универсального» порога. Для метода мягкой пороговой обработки подобные исследования проводились в работах [6–9].

2 Метод пороговой обработки

Для функции сигнала $f \in L^2(\mathbb{R})$ разложение по вейвлет-базису имеет вид:

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

где $\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k)$, а $\psi(t)$ — некоторая вейвлет-функция (семейство $\{\psi_{jk}\}_{j,k \in \mathbb{Z}}$ образует ортонормированный базис в $L^2(\mathbb{R})$). Индекс j в (1) называется масштабом, а индекс k — сдвигом. В данной работе рассматриваются функции сигнала на конечном отрезке $[a, b]$, равномерно регулярные по Липшицу с некоторым показателем $\gamma > 0$. Кроме того, предполагается, что вейвлет-функция ψ имеет M непрерывных производных ($M \geq \gamma$), M нулевых моментов и удовлетворяет условию:

$$\int_{-\infty}^{\infty} |t^\gamma \psi(t)| dt < \infty.$$

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 16-07-00736).

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

В этом случае найдется такая константа $C_f > 0$, что [10]

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

Модель данных, рассматриваемая в данной работе, предполагает, что функция сигнала задана в дискретных отсчетах и наблюдения содержат шум:

$$X_i = f_i + \epsilon_i, \quad i = 1, \dots, 2^J,$$

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

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

где ϵ_{jk}^W независимы и имеют такое же распределение, как и ϵ_i , а $\mu_{jk} \approx 2^{J/2} \langle f, \psi_{jk} \rangle$ [10].

Для построения оценки функции сигнала к коэффициентам Y_{jk} обычно применяется функция жесткой пороговой обработки $\rho_H(y, T) = x \mathbf{1}(|y| > T)$ или функция мягкой пороговой обработки $\rho_S(y, T) = \text{sgn}(x) (|y| - T)_+$ с порогом T .

Как уже отмечалось, функция ρ_H разрывна, что приводит к отсутствию устойчивости, а функция ρ_S приводит к появлению дополнительного смещения в оценке функции сигнала. В работе [4] предложен новый метод пороговой обработки, являющийся сглаженным аналогом жесткой пороговой обработки. В этом методе оценки μ_{jk} вычисляются по формулам:

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

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

$$\begin{aligned} \hat{\mu}_{jk} = Y_{jk} & \left[\Phi \left(-\frac{T + Y_{jk}}{\lambda} \right) + 1 - \Phi \left(\frac{T - Y_{jk}}{\lambda} \right) \right] + \\ & + \lambda \left[\phi \left(\frac{T - Y_{jk}}{\lambda} \right) - \phi \left(\frac{T + Y_{jk}}{\lambda} \right) \right]. \end{aligned}$$

Достоинством такого метода является бесконечная дифференцируемость $\hat{\mu}_{jk}$ по Y_{jk} , что приводит к более устойчивым оценкам [4]. Заметим также, что при $\lambda \rightarrow 0$ получается обычный метод жесткой пороговой обработки.

3 Несмещенная оценка среднеквадратичного риска и ее свойства

Среднеквадратичная погрешность (риск) описанного выше метода определяется по формуле:

$$R_J(T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} (\hat{\mu}_{jk} - \mu_{jk})^2. \quad (3)$$

В [4] показано, что

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

где

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

Таким образом, величина

$$\begin{aligned} \hat{R}_J(T) & = \\ & = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left[(Y_{jk} - \hat{\mu}_{jk})^2 + 2\sigma^2 \frac{\partial}{\partial Y_{jk}} \hat{\mu}_{jk} - \sigma^2 \right] \end{aligned} \quad (4)$$

является несмещенной оценкой $R_J(T)$, не зависящей от ненаблюдаемых «чистых» значений μ_{jk} .

В данной работе параметр λ будем полагать фиксированным, а в качестве T рассмотрим так называемый «универсальный» порог $T_U = \sigma \sqrt{2 \ln 2^J}$, который позволяет достичь хороших результатов при подавлении шума и обеспечивает близость среднеквадратичного риска к минимальному при жесткой и мягкой пороговых обработках [10].

Покажем, что оценка (4) является асимптотически нормальной. Это свойство позволяет строить асимптотические доверительные интервалы для риска (3).

Теорема 1. Пусть f задана на конечном отрезке и является равномерно регулярной по Липшицу с параметром $\gamma > 0$. Тогда

$$\mathbb{P} \left(\frac{\hat{R}_J(T_U) - R_J(T_U)}{\sigma^2 \sqrt{2^{J+1}}} < x \right) \Rightarrow \Phi(x). \quad (5)$$

Доказательство. Выберем p такое, что $(2\gamma + 1)^{-1} < p < 1$. Обозначим слагаемые в (4) через $F_{jk}(T)$ и запишем разность $\widehat{R}_J(T_U) - R_J(T_U)$ в виде:

$$\begin{aligned} \widehat{R}_J(T_U) - R_J(T_U) &= \sum_{j=0}^{[pJ]} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)] + \\ &+ \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)]. \quad (6) \end{aligned}$$

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

$$|F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)| \leq C_F T_U \text{ п.в.} \quad (7)$$

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

$$\begin{aligned} \mathbb{P} \left(\frac{\left| \sum_{j=0}^{[pJ]} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)] \right|}{\sigma^2 \sqrt{2^{J+1}}} > \delta \right) &\leq \\ &\leq \exp \left\{ -C_\delta \frac{2^{J-pJ}}{J} \right\}, \end{aligned}$$

т. е.

$$\frac{\sum_{j=0}^{[pJ]} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)]}{\sigma^2 \sqrt{2^{J+1}}} \xrightarrow{\mathbb{P}} 0 \text{ при } J \rightarrow \infty. \quad (8)$$

Для всех слагаемых во второй сумме в (6) в силу (2) выполнено $\mu_{jk} \rightarrow 0$ при $J \rightarrow \infty$. Рассмотрим дисперсии слагаемых в этой сумме. Имеем $D F_{jk}(T_U) = D[Y_{jk}^2 - \sigma^2 + U_{jk}(T_U)]$, где

$$\begin{aligned} U_{jk}(T_U) &= F_{jk}(T_U) - Y_{jk}^2 + \sigma^2 = \\ &= (\widehat{\mu}_{jk} - 2Y_{jk})\widehat{\mu}_{jk} + 2\sigma^2 \frac{\partial}{\partial Y_{jk}} \widehat{\mu}_{jk}. \end{aligned}$$

Пусть $A_J = \sigma \sqrt{A \ln 2^J}$, где $0 < A < 2$. Тогда

$$\begin{aligned} D U_{jk}(T_U) &\leq \mathbb{E} U_{jk}^2(T_U) = \\ &= \mathbb{E} [\mathbf{1}(|Y_{jk}| > A_J) U_{jk}^2(T_U)] + \\ &+ \mathbb{E} [\mathbf{1}(|Y_{jk}| \leq A_J) U_{jk}^2(T_U)]. \end{aligned}$$

Поскольку $T_U - A_J \rightarrow \infty$ при $J \rightarrow \infty$, учитывая определение $\widehat{\mu}_{jk}$ и $(\partial/\partial Y_{jk})\widehat{\mu}_{jk}$, заключаем, что найдется такое $\varepsilon_J > 0$, что $|\mathbf{1}(|Y_{jk}| \leq A_J) U_{jk}^2(T_U)| \leq \varepsilon_J$ п.в. и $\varepsilon_J \rightarrow 0$ при $J \rightarrow \infty$. Следовательно, $\mathbb{E}[\mathbf{1}(|Y_{jk}| \leq A_J) U_{jk}^2(T_U)] \rightarrow 0$ при $J \rightarrow \infty$.

Далее найдется такая константа $C > 0$, что $\mathbb{E}[\mathbf{1}(|Y_{jk}| > A_J) U_{jk}^2(T_U)] \leq C \mathbb{E}[\mathbf{1}(|Y_{jk}| > A_J) Y_{jk}^4] \rightarrow 0$ при $J \rightarrow \infty$. Таким образом, учитывая очевидное соотношение для дисперсии суммы случайных величин $X + Y$: $(\sqrt{DX} - \sqrt{DY})^2 \leq D(X + Y) \leq (\sqrt{DX} + \sqrt{DY})^2$, получаем:

$$\lim_{J \rightarrow \infty} \frac{D \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)]}{D \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} Y_{jk}^2} = 1. \quad (9)$$

Кроме того, поскольку Y_{jk} независимы, $D Y_{jk}^2 = 2\sigma^4 + 4\sigma^2 \mu_{jk}^2$ и в (9) $\mu_{jk} \rightarrow 0$ при $J \rightarrow \infty$, получаем:

$$\lim_{J \rightarrow \infty} \frac{D \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} Y_{jk}^2}{\sigma^4 2^{J+1}} = 1. \quad (10)$$

Наконец, выполнено условие Линдберга: для любого $\epsilon > 0$ при $J \rightarrow \infty$

$$\begin{aligned} \frac{1}{D^2} \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} \left[(F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U))^2 \mathbf{1}(|F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)| > \epsilon D_J) \right] &\rightarrow \\ &\rightarrow 0, \quad (11) \end{aligned}$$

где

$$D_J^2 = D \sum_{j=[pJ]+1}^{J-1} \sum_{k=0}^{2^j-1} [F_{jk}(T_U) - \mathbb{E}F_{jk}(T_U)].$$

Действительно, в силу (7), (9) и (10) начиная с некоторого J все индикаторы в (11) обращаются в ноль. Объединяя (8)–(11), получаем (5). Теорема доказана.

Докажем теперь свойство сильной состоятельности оценки (4), справедливое при более слабых ограничениях на функцию сигнала.

Теорема 2. Пусть $f \in L^2(\mathbb{R})$ и задана на конечном отрезке, тогда при любом $\alpha > 1/2$ имеет место сходимость:

$$\frac{\widehat{R}_J(T_U) - R_J(T_U)}{2^{\alpha J}} \rightarrow 0 \text{ п.в. при } J \rightarrow \infty. \quad (12)$$

Доказательство. Используя неравенство Хефдинга с учетом (7) и вида T_U , получаем, что для любого $\delta > 0$ найдется константа $C_\delta > 0$ такая, что

$$\begin{aligned} p_J &= \mathbb{P} \left(\frac{|\widehat{R}_J(T_U) - R_J(T_U)|}{2^{\alpha J}} > \delta \right) \leq \\ &\leq \exp \left\{ -C_\delta \frac{2^{2\alpha J - J}}{J} \right\}. \quad (13) \end{aligned}$$

Далее

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

и в силу леммы Бореля–Кантелли осуществляется лишь конечное число событий под вероятностью в (13). Следовательно, выполнено (12). Теорема доказана.

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Breiman L. Heuristics of instability and stabilization in model selection // *Ann. Statist.*, 1996. Vol. 24. No. 6. P. 2350–2383.
3. Jansen M. Noise reduction by wavelet thresholding. — Lecture notes in statistics ser. — Springer Verlag, 2001. Vol. 161. 218 p.
4. Huang H.-C., Lee T. C. M. Stabilized thresholding with generalized sure for image denoising // *IEEE 17th Confer-*
- ence (International) on Image Processing Proceedings. — IEEE, 2010. P. 1881–1884.
5. Stein C. Estimation of the mean of a multivariate normal distribution // *Ann. Stat.*, 1981. Vol. 9. No. 6. P. 1135–1151.
6. Маркин А. В. Предельное распределение оценки риска при пороговой обработке вейвлет-коэффициентов // *Информатика и её применения*, 2009. Т. 3. Вып. 4. С. 57–63.
7. Маркин А. В., Шестаков О. В. О состоятельности оценки риска при пороговой обработке вейвлет-коэффициентов // *Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн.*, 2010. № 1. С. 26–34.
8. Шестаков О. В. Асимптотическая нормальность оценки риска пороговой обработки вейвлет-коэффициентов при выборе адаптивного порога // *Докл. РАН*, 2012. Т. 445. № 5. С. 513–515.
9. Шестаков О. В. О точности приближения распределения оценки риска пороговой обработки вейвлет-коэффициентов сигнала нормальным законом при неизвестном уровне шума // *Системы и средства информатики*, 2012. Т. 22. № 1. С. 142–152.
10. Mallat S. A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.

Поступила в редакцию 22.01.16

STATISTICAL PROPERTIES OF THE DENOISING METHOD BASED ON THE STABILIZED HARD THRESHOLDING

O. V. Shestakov^{1,2}

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

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

The thresholding techniques for the wavelet coefficients of the signal and image functions have become a popular denoising tool because of their simplicity, computational efficiency, and possibility to adapt to the functions with different amounts of smoothness in different locations. The paper considers the recently proposed stabilized hard thresholding method which avoids the main disadvantages of the popular soft and hard thresholding techniques. The statistical properties of this method are studied. The unbiased risk estimate is analyzed in the model with an additive Gaussian noise. Wavelet thresholding risk analysis is an important practical task, because it allows determining the quality of the techniques themselves and the equipment which is being used. The paper proves that under certain conditions, the unbiased risk estimate is strongly consistent and asymptotically normal. These properties allow constructing the asymptotic confidence intervals for the theoretical mean squared risk of the method.

Keywords: wavelets; thresholding; unbiased risk estimate; asymptotic normality; strong consistency

DOI: 10.14357/19922264160207

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 16-07-00736).

References

1. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455.
2. Breiman, L. 1996. Heuristics of instability and stabilization in model selection. *Ann. Stat.* 24(6):2350–2383.
3. Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. Springer Verlag. Vol. 161. 218 p.
4. Huang, H.-C., and T. C. M. Lee. 2010. Stabilized thresholding with generalized sure for image denoising. *IEEE 17th Conference (International) on Image Processing Proceedings*. IEEE. 1881–1884.
5. Stein, C. 1981. Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* 9(6):1135–1151.
6. Markin, A. V. 2009. Predel'noe raspredelenie otsenki riska pri porogovoy obrabotke veyvlet-koeffitsientov [Limit distribution of risk estimate of wavelet coefficient thresholding]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):57–63.
7. Markin, A. V., and O. V. Shestakov. 2010. Consistency of risk estimation with thresholding of wavelet coefficients. *Moscow Univ. Comput. Math. Cybern.* 34(1):22–30.
8. Shestakov, O. V. 2012. Asymptotic normality of adaptive wavelet thresholding risk estimation. *Dokl. Math.* 86(1):556–558.
9. Shestakov, O. V. 2012. O tochnosti priblizheniya raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov signala normal'nym zakonom pri neizvestnom urovne shuma [On the accuracy of normal approximation for risk estimate distribution when thresholding signal wavelet coefficients in case of unknown noise level]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):142–152.
10. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.

Received January 22, 2016

Contributor

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

ОБ ОПТИМАЛЬНОМ ПОРОГОВОМ ЗНАЧЕНИИ ДЛИНЫ ОЧЕРЕДИ В ОДНОЙ ЗАДАЧЕ МАКСИМИЗАЦИИ ДОХОДА СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ ТИПА $M/G/1$ *

Я. М. Агаларов¹, М. Я. Агаларов², В. С. Шоргин³

Аннотация: Рассматривается задача максимизации среднего дохода системы $M/G/1$ в единицу времени на множестве стационарных пороговых стратегий ограничения доступа с одной «точкой переключения». Доход определяется следующими параметрами, измеряемыми в стоимостных единицах: плата, получаемая за обслуживание заявок; затраты на техническое обслуживание прибора; вычет из дохода за задержку заявок в очереди; штраф за необслуженные заявки; штраф за простой системы. Получены условия существования конечного оптимального порогового значения, предложены метод и алгоритм расчета оценок снизу для оптимального порога и соответствующего значения максимального дохода в единицу времени. Решена вспомогательная задача максимизации дохода системы, усредненного по числу обслуженных заявок, на множестве рассматриваемых пороговых стратегий. Получены необходимые и достаточные условия существования решения вспомогательной задачи, предложен метод и алгоритм ее решения.

Ключевые слова: система массового обслуживания; пороговая стратегия; оптимизация

DOI: 10.14357/19922264160208

1 Введение

Для повышения эффективности работы современных систем и сетей передачи данных используются алгоритмы управления потоками (ограничения нагрузки), наиболее применяемыми из которых стали различные модификации пороговых алгоритмов (пороговых стратегий) [1]. Одним из основных методов исследования эффективности пороговых стратегий является математическое моделирование с использованием аппарата теории массового обслуживания, предметом исследования которой выступают системы массового обслуживания (СМО) различного типа.

Большинство работ, в которых рассмотрены СМО с пороговой стратегией управления потоками, посвящено методам расчета характеристик системы (средней длины очереди, среднего времени пребывания, вероятности отклонения заявки, времени простоя приборов и т. д.) при заданной пороговой стратегии (краткий обзор некоторых из них проведен в [2]). В ряде работ, посвященных данной тематике, ставится задача оптимизации пороговой стратегии в смысле максимизации дохода системы, представленного в виде стоимостного функционала (см., например, [3–7]). Хотя практический интерес

к такой постановке задачи в смысле повышения эффективности систем, как представляется, выше, чем к задачам расчета характеристик систем при фиксированной пороговой стратегии, вопрос существования эффективных методов и алгоритмов поиска оптимальных пороговых стратегий для СМО остается открытым, за исключением самых простых СМО ($M/M/1$, $M/M/n$ [4]) и простых целевых функций (допустимой средней задержки заявок в системе, допустимой интенсивности потерь).

Для более сложных СМО (например, $G/M/1$, $M/G/1$) с более сложными целевыми функциями результаты исследований ограничиваются математической постановкой задач и эвристическими алгоритмами их решения. Отметим работу [5], где для системы $G/M/n$ сформулирована математическая постановка максимизации дохода системы на множестве пороговых стратегий с одним переключением (одной гистерезисной петлей), фиксированными платой за своевременное обслуживание и штрафом за невыполнение этого условия для допущенной в систему заявки. В работе предложен эвристический алгоритм поиска оптимальной стратегии и выдвинута гипотеза о том, что для систем $G/G/n$ целевая функция унимодальна. Аналогичная задача рассмотрена в работе [6] для системы

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 15-07-03406).

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

² ПАО Промсвязьбанк, murad-agalarov@yandex.ru

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

$M/D/1$, где также сформулирована математическая постановка задачи и приведена нижняя оценка для оптимального порогового значения. В работе [3] рассматривается задача управления запасами, для решения которой в качестве модели используется управляемая на множестве пороговых стратегий СМО типа $M/G/1$. Постановка задачи, наиболее близкая к рассматриваемой в настоящей статье, представлена в работе [7], где решение задачи максимизации среднего дохода, получаемого системой $M/G/1$ в единицу времени, ищется на множестве смешанных пороговых стратегий. Решение ищется при условии, что поступающая на вход нагрузка меньше единицы. Доход системы складывается из следующих составляющих (констант):

- плата, получаемая системой за единицу времени обслуживания одной заявки прибором;
- вычет из дохода системы в единицу времени технического обслуживания прибора;
- вычет из дохода системы в единицу времени ожидания одной заявки в очереди;
- штраф за потерю одной заявки;
- вычет из дохода в единицу времени простоя системы.

В работе [7] доказано, что если решение задачи существует, то оно принадлежит множеству чистых стратегий. Для поиска решения задачи предлагается использовать численный метод.

В данной статье рассматривается аналогичная задача максимизации среднего дохода системы $M/G/1$ на множестве чистых пороговых стратегий для двух вариантов формирования дохода: в одном варианте плату за обслуживание система получает в момент приема заявки и плата не зависит от длительности обслуживания прибором, в другом плату за обслуживание система получает в момент окончания обслуживания заявки и величина платы прямо пропорциональна длительности обслуживания этой заявки прибором (случай, рассмотренный в [7]). Проведены исследования, касающиеся вопросов существования решения задачи максимизации дохода и метода поиска оптимальной стратегии.

2 Постановка задачи

Рассматривается СМО типа $M/G/1$ с накопителем бесконечной емкости и одним прибором обслуживания, на которую поступает пуассоновский поток заявок с интенсивностью $\lambda > 0$ и время обслуживания каждой заявки распределено по произвольному закону $H(t)$. Поступившая заявка допускается в накопитель системы (занимает любое

свободное место в накопителе), если в момент ее поступления число занятых мест в накопителе меньше k , $k > 0$ — некоторое заданное значение (тривиальный случай $k = 0$ не рассматривается). Такую процедуру доступа заявок в систему называют пороговой стратегией управления доступом (далее для краткости — стратегией). Обозначим стратегию соответствующим пороговым значением k . Если заявка допущена в накопитель, она занимает любое свободное место в накопителе и обслуживается на приборе в порядке поступления. Заявка покидает систему только при завершении обслуживания, освободив одновременно прибор и накопитель, а на освободившийся прибор поступает очередная заявка из накопителя (если таковая есть). Система получает доход, который определяется следующими составляющими:

- $C_0 \geq 0$ — плата, получаемая системой, если поступившая заявка будет обслужена системой (допущена в накопитель);
- $C_1 \geq 0$ — величина штрафа, который платит система, если поступившая заявка отклонена;
- $C_2 \geq 0$ — вычет из дохода системы за единицу времени ожидания заявки в системе;
- $C_3 \geq 0$ — вычет из дохода системы за единицу времени простоя прибора (за время отсутствия заявок в системе);
- $C_4 \geq 0$ — затраты системы в единицу времени на техническое обслуживание системы.

Всюду ниже под доходом системы будем понимать суммарный доход с учетом всех указанных выше составляющих.

Отметим, что процесс обслуживания заявок в данной системе описывается цепью Маркова, где переходы цепи определяются моментами окончания обслуживания и состояние системы есть число заявок, остающихся в ней в момент ухода с прибора обслуженной заявки (см., например, [8, 9]). Отметим также, что при заданной стратегии k указанная цепь Маркова имеет один положительный возвратный класс состояний $i = 0, \dots, k - 1$.

Введем обозначения:

- $\pi_i^k, 0 \leq i \leq k - 1$, — стационарное распределение вероятностей цепи при стратегии k (π_i^k — вероятность того, что цепь находится в состоянии i);
- g^k — среднее значение суммарного предельного дохода, усредненного по числу обслуженных заявок;
- Q^k — средний предельный доход, получаемый системой в единицу времени при стратегии k ;
- q_i^k — средний доход, получаемый системой в состоянии i при стратегии $k, i \geq 0$;

$\bar{v} = \int_0^\infty t dH(t)$ — среднее время пребывания системы в состоянии i , $0 < \bar{v} < \infty$;

$B_{v,m}$ — событие, состоящее в том, что за время v поступило ровно m заявок, $m \geq 0$.

Среднее значение предельного дохода, усредненного по числу обслуженных заявок, при стратегии k равно

$$g^k = \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}} \frac{d_n^k}{N_{\text{вых}}(T)},$$

где d_n^k — доход, полученный системой при стратегии k за обслуживание n -й заявки, $N_{\text{вх}}(T)$ — число поступивших заявок за отрезок времени $[0, T]$; $N_{\text{вых}}(T)$ — число обслуженных заявок за отрезок времени $[0, T]$. Из определения вложенной цепи Маркова следует:

$$g^k = \sum_{i=0}^{k-1} \pi_i^k q_i^k. \quad (1)$$

Среднее значение предельного дохода системы в единицу времени равно:

$$\begin{aligned} Q^k &= \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^k}{T} = \\ &= \lim_{T \rightarrow \infty} \frac{N_{\text{вых}}(T)}{T} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^k}{N_{\text{вых}}(T)} = \\ &= \lim_{T \rightarrow \infty} \frac{N_{\text{вых}}(T)}{T} \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^k}{N_{\text{вых}}(T)} = \\ &= \lambda (1 - \theta_k^k) g^k, \quad (2) \end{aligned}$$

где θ_k^k — вероятность того, что поступившая заявка будет допущена в систему.

Ставится задача: найти оптимальные стационарные стратегии $k^0 > 0$ и $k^* > 0$ такие, что

$$\max_{k>0} g^k = g^{k^0}; \quad \max_{k>0} Q^k = Q^{k^*}. \quad (3)$$

3 Метод решения

Перейдем к рассмотрению параметров q_i^k и π_j^k , $j = 0, \dots, k-1$, $k \geq 1$.

Для вероятностей переходов вложенной цепи Маркова p_{ij}^k справедливы формулы:

$$p_{ij}^k = \begin{cases} r_{j-i+1}, & i \leq j \leq k-2; \\ 1 - \sum_{l=0}^{k-i-1} r_l, & j = k-1, \end{cases}$$

если $1 \leq i \leq k-1$, $p_{0j}^k = p_{1j}^k, p_{ij}^k = 0$ при $i > k-1$ или $j < i-1$, где

$$r_l = \int_0^\infty \frac{(\lambda v)^l}{l!} e^{-\lambda v} dH(v) \quad \text{при } 0 \leq l \leq k-2. \quad (4)$$

Для рассматриваемой цепи Маркова при стратегии k стационарное распределение вероятностей является единственным решением системы уравнений [9]:

$$\left. \begin{aligned} \pi_j^k &= \sum_{i=0}^{j+1} \pi_i^k p_{ij}^k, \quad j = 0, \dots, k-2; \\ \pi_{k-1}^k &= \sum_{i=0}^{k-1} \pi_i^k p_{ik-1}^k; \\ \sum_{j=0}^{k-1} \pi_j^k &= 1. \end{aligned} \right\} \quad (5)$$

Из (4) и (5) следуют рекуррентные формулы для расчета стационарного распределения при стратегии k :

$$\pi_j^k = \pi_0^k R_j, \quad j = 1, \dots, k-1, \quad (6)$$

где

$$\begin{aligned} \pi_0^k &= \left(\sum_{i=0}^{k-1} R_i \right)^{-1}, \quad R_0 = 1, \quad R_1 = \frac{1-r_0}{r_0}; \\ R_{i+1} &= \frac{1}{r_0} \left(R_i - r_i - \sum_{j=0}^i R_j r_{i-j+1} \right), \\ & \quad i = 1, \dots, k-2. \end{aligned}$$

Величина платы, получаемой системой при пороге $k > 1$ за время нахождения в состоянии $1 \leq i \leq k-1$ (т. е. за один шаг соответствующей цепи Маркова), вычисляется как сумма плат, получаемых за обслуживание допущенных в систему за это время новых заявок, и выражается формулой:

$$\begin{aligned} & \int_0^\infty \sum_{m=1}^{k-i} \frac{(\lambda v)^m}{m!} e^{-\lambda v} C_0 m dH(v) + \\ & + \int_0^\infty \sum_{m=k-i+1}^\infty \frac{(\lambda v)^m}{m!} e^{-\lambda v} C_0 (k-i) dH(v) = \\ & = C_0 \left[\sum_{m=1}^{k-i} m r_m + (k-i) \sum_{m=k-i+1}^\infty r_m \right]. \quad (7) \end{aligned}$$

Величина штрафа, которую система платит за не допущенные в состоянии $1 \leq i \leq k-1$ заявки, равна:

$$C_1 \sum_{m=k-i+1}^\infty (m-k+i) r_m. \quad (8)$$

Величина вычета из дохода системы за задержку заявок в системе, накопленная за время пребывания системы в состоянии $1 \leq i \leq k-1$, равна

$$\frac{C_2}{\lambda} \left[\frac{1}{2} \sum_{m=2}^{k-i+1} (m-1)mr_m + (k-i) \sum_{m=k-i+2}^{\infty} mr_m - \frac{1}{2}(k-i)(k-i+1) \sum_{m=k-i+2}^{\infty} r_m \right] + C_2(i-1)\bar{v}. \quad (9)$$

Формула (9) получается из следующих рассуждений. Фиксируем состояние системы i , $1 \leq i \leq k-1$. Пусть v — длительность нахождения системы в состоянии i ; T_l — последовательные моменты поступления заявок в промежутке времени $[0, v]$; m — число поступивших за время v заявок; $U_l^i(m, v)$ — длительность времени из интервала $[0, v]$, в течение которого заявка, поступившая в момент T_l , $l+i \leq k$, ожидала в системе.

Так как входящий поток пуассоновский, то совместное распределение величин T_l совпадает с распределением порядковых статистик из выборки m , взятой из равномерного распределения на $(0, v]$ (см., например, [8]) и маргинальным распределением случайной величины T_l/v является бета-распределение с плотностью

$$f(x) = \begin{cases} \frac{m!}{(l-1)!(m-l)!} x^{l-1}(1-x)^{m-l}, & 0 < x < 1; \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда поскольку $U_l^i(m, v) = v - T_l$, $l+i \leq k$, то условное среднее значение $M[U_l^i(m, v)] = v - lv/(m+1)$ (при условии, что длительность нахождения системы в состоянии i равна v и за это время поступило ровно m заявок). При $m \leq k-1$ получаем, что среднее значение $M[\sum_{l=1}^m (v - T_l)] = mv - mv/2 = mv/2$, а при $m > k-i$, так как поступившие в моменты T_l , $l > k-i$, отклоняются, условное среднее значение суммарного времени ожидания в системе поступивших в состоянии i заявок равно:

$$M \left[\sum_{l=1}^{k-i} (v - T_l) \right] = (k-i)v - v \sum_{l=1}^{k-i} \frac{l}{m+1} = (k-i)v - \frac{(k-i)(k-i+1)v}{2(m+1)}.$$

Суммарное время ожидания в системе заявок, находящихся в системе в момент перехода в состояние i , равно $(i-1)v$. Отсюда получаем, что на интервале времени $(0, v]$ (во время пребывания

системы в состоянии i) при условии, что в этом интервале поступило m заявок, суммарное время ожидания находящихся в системе заявок равно:

$$\frac{mv}{2} \chi(k-i-m) + \frac{(k-i)(k-i+1)v}{2(m+1)} [1 - \chi(k-i-m)] + (i-1)v,$$

где $\chi(k-i-m)$ — функция Хевисайда. Используя формулу полной вероятности и взяв среднее по распределению $H(v)$ (аналогично формуле (7)), из последнего выражения получим формулу (9).

Используя формулы (7)–(9) и то, что величина среднего значения вычета из дохода за техническое обслуживание системы в состоянии $1 \leq i \leq k-1$ равно $C_4\bar{v}$, получим:

$$q_i^k = C_0 \left[\sum_{m=1}^{k-i} mr_m + (k-i) \sum_{m=k-i+1}^{\infty} r_m \right] - C_1 \sum_{m=k-i+1}^{\infty} (m-k+i)r_m - \frac{C_2}{\lambda} \left[\frac{1}{2} \sum_{m=1}^{k-i+1} (m-1)mr_m + (k-i) \sum_{m=k-i+2}^{\infty} mr_m - \frac{1}{2}(k-i)(k-i+1) \sum_{m=k-i+2}^{\infty} r_m \right] - C_2(i-1)\bar{v} - C_4\bar{v}, \quad 1 \leq i \leq k-1. \quad (10)$$

Состояние $i=0$ отличается от состояния $i=1$ в смысле дохода только тем, что в состоянии $i=0$ возможен простой длительности $1/\lambda$, из-за которого доход уменьшается на величину $(C_3 + C_4)/\lambda$, и завершает обслуживание в состоянии $i=0$ заявка, которая поступила, когда система находилась в этом же состоянии. Поэтому справедливо равенство:

$$q_0^k = q_1^k + C_0 - \frac{C_3 + C_4}{\lambda}.$$

Докажем следующее вспомогательное утверждение.

Лемма 1. *Справедливы равенства*

$$q_j^{k+1} = q_j^k + (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \frac{C_2}{\lambda} \left[\sum_{m=k-i+1}^{\infty} mr_m - (k-i+1) \sum_{m=k-i+1}^{\infty} r_m \right]; \quad (11)$$

$$\pi_j^{k+1} = (1 - \pi_k^{k+1}) \pi_j^k \quad j = 0, \dots, k-1. \quad (12)$$

Доказательство. Из (9) имеем:

$$\begin{aligned}
 q_i^{k+1} - q_i^k &= C_0 \left[\sum_{m=1}^{k-i} mr_m + (k-i+1)r_{k-i+1} + \right. \\
 &+ \left. \sum_{m=k-i+2}^{\infty} r_m + (k-i) \sum_{m=k-i+2}^{\infty} r_m \right] - \\
 &- C_1 \sum_{m=k-i+2}^{\infty} (m-k+i)r_m + C_1 \sum_{m=k-i+2}^{\infty} r_m - \\
 &- \frac{C_2}{\lambda} \left[\frac{1}{2} \sum_{m=2}^{k-i+1} (m-1)mr_m + \right. \\
 &+ \frac{1}{2} (k-i+1)(k-i+2)r_{k-i+2} + \\
 &+ (k+1-i) \sum_{m=k-i+3}^{\infty} mr_m - \\
 &- \left. \frac{1}{2} (k-i+1)(k-i+2) \sum_{m=k-i+3}^{\infty} r_m \right] - \\
 &- C_0 \left[\sum_{m=1}^{k-i} mr_m + (k-i)r_{k-i+1} + (k-i) \sum_{m=k-i+2}^{\infty} r_m \right] + \\
 &+ C_1 r_{k-i+1} + C_1 \sum_{m=k-i+2}^{\infty} (m-k+i)r_m + \\
 &+ \frac{C_2}{\lambda} \left[\frac{1}{2} \sum_{m=2}^{k-i+1} (m-1)mr_m + \right. \\
 &+ (k-i)(k-i+2)r_{k-i+2} + (k-i) \sum_{m=k-i+3}^{\infty} mr_m - \\
 &- \left. \frac{1}{2} (k-i)(k-i+1)r_{k-i+2} + \right. \\
 &+ \left. \frac{1}{2} (k-i)(k-i+1) \sum_{m=k-i+3}^{\infty} r_m \right].
 \end{aligned}$$

Проведя преобразования, получим:

$$\begin{aligned}
 q_i^{k+1} - q_i^k &= (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \\
 &- \frac{C_2}{\lambda} \left[\sum_{m=k-i+3}^{\infty} mr_m - (k-i+1) \sum_{m=k-i+3}^{\infty} r_m + r_{k-i+2} \right] = \\
 &= (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \\
 &- \frac{C_2}{\lambda} \left[\sum_{m=k-i+1}^{\infty} mr_m - (k-i+1) \sum_{m=k-i+1}^{\infty} r_m \right].
 \end{aligned}$$

Второе равенство утверждения следует непосредственно из (6):

$$\begin{aligned}
 \pi_j^k - \pi_j^{k+1} &= \frac{R_j}{R_0 + \dots + R_{k-1}} - \frac{R_j}{R_0 + \dots + R_k} = \\
 &= \frac{R_j R_k}{(R_0 + \dots + R_{k-1})(R_0 + \dots + R_k)} = \pi_j^k \pi_k^{k+1}.
 \end{aligned}$$

Получим вспомогательное выражение для разности $g^k - g^{k+1}$, $0 < k < \infty$. Из формулы (1) следует:

$$\begin{aligned}
 g^k - g^{k+1} &= \sum_{j=0}^{k-1} \pi_j^k q_j^k - \sum_{j=0}^k \pi_j^{k+1} q_j^{k+1} = \\
 &= \sum_{j=0}^{k-1} \pi_j^k q_j^k - \sum_{j=0}^{k-1} \pi_j^{k+1} q_j^{k+1} - \pi_k^{k+1} q_k^{k+1}.
 \end{aligned}$$

Подставив вместо q_i^{k+1} правую часть равенства (11), получим:

$$\begin{aligned}
 g^k - g^{k+1} &= \sum_{j=0}^{k-1} \pi_j^k q_j^k - \sum_{j=0}^{k-1} \pi_j^{k+1} q_j^k - \\
 &- \sum_{j=0}^{k-1} \pi_j^{k+1} \left\{ (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \right. \\
 &- \left. \frac{C_2}{\lambda} \sum_{m=k-i+1}^{\infty} [m - (k-i+1)]r_m \right\} - \pi_k^{k+1} q_k^{k+1}.
 \end{aligned}$$

Заменив здесь π_j^{k+1} на правую часть равенства (12), имеем:

$$\begin{aligned}
 g^k - g^{k+1} &= \sum_{j=0}^{k-1} \pi_j^k q_j^k - (1 - \pi_k^{k+1}) \sum_{j=0}^{k-1} \pi_j^k q_j^k - \\
 &- (1 - \pi_k^{k+1}) \sum_{j=0}^{k-1} \pi_j^k \left\{ (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \right. \\
 &- \left. \frac{C_2}{\lambda} \sum_{m=k-i+1}^{\infty} [m - (k-i+1)]r_m \right\} - \pi_k^{k+1} q_k^{k+1} = \\
 &= \pi_k^{k+1} \left\{ g^k - \frac{1 - \pi_k^{k+1}}{\pi_k^{k+1}} \sum_{j=0}^{k-1} \pi_j^k \left\{ (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \right. \right. \\
 &- \left. \left. \frac{C_2}{\lambda} \sum_{m=k-i+1}^{\infty} [m - (k-i+1)]r_m \right\} - q_k^{k+1} \right\}. \quad (13)
 \end{aligned}$$

Заменив в правой части последнего неравенства π_k^{k+1} и π_j^k , $j = 0, \dots, k$, на их выражения в (6), перепишем (13) в виде:

$$\begin{aligned}
 g^k - g^{k+1} &= \\
 &= \pi_k^{k+1} \left\{ g^k - \frac{1}{R_k} \sum_{j=0}^{k-1} R_j \left\{ (C_0 + C_1) \sum_{m=k-i+1}^{\infty} r_m - \right. \right. \\
 &\left. \left. - \frac{C_2}{\lambda} \sum_{m=k-i+1}^{\infty} [m - (k - i + 1)] r_m \right\} - q_k^{k+1} \right\}. \quad (14)
 \end{aligned}$$

Из предпоследнего уравнения в (5), подставив (4) и (6), получим:

$$R_k = \frac{\sum_{j=0}^{k-1} R_j \sum_{i=k-j+1}^{\infty} r_i}{r_0}. \quad (15)$$

Заменив в (14) R_k на правую часть (15) и q_{i+1}^{k+1} на $q_i^k - C_2 \bar{v}$ (см. формулу (10)), получим:

$$\begin{aligned}
 g^k - g^{k+1} &= \pi_k^{k+1} \left\{ g^k - (C_0 + C_1) r_0 + \right. \\
 &+ \frac{C_2}{\lambda} r_0 \frac{\sum_{j=0}^{k-1} R_j \sum_{m=k-j+1}^{\infty} [m - (k - j)] r_m}{\sum_{j=0}^{k-1} R_j \sum_{i=k-j+1}^{\infty} r_i} - \\
 &\left. - \frac{C_2}{\lambda} r_0 - q_1^2 + C_2(k - 1) \bar{v} \right\}. \quad (16)
 \end{aligned}$$

Обозначим

$$\begin{aligned}
 F(k) &= \\
 &= \frac{\sum_{j=0}^{k-1} R_j \sum_{m=k-j+1}^{\infty} [m - (k - j)] r_m}{\sum_{j=0}^{k-1} R_j \sum_{i=k-j+1}^{\infty} r_i}. \quad (17)
 \end{aligned}$$

Заметим, что $F(k)$ — среднее число заявок, отклоненных при стратегии k в стационарном режиме работы системы за время обслуживания одной заявки, при условии, что поступило больше заявок, чем число свободных мест в накопителе.

Используя обозначение (17), перепишем (16) в виде

$$\begin{aligned}
 g^k - g^{k+1} &= \pi_k^{k+1} \left\{ g^k - \left(C_0 + C_1 + \frac{C_2}{\lambda} \right) r_0 + \right. \\
 &\left. + \frac{C_2}{\lambda} r_0 F(k) - q_1^2 + C_2(k - 1) \bar{v} \right\}. \quad (18)
 \end{aligned}$$

Как видно из (18), неравенство $g^k - g^{k+1} < 0$ эквивалентно неравенству:

$$\begin{aligned}
 g^k &< \left(C_0 + C_1 + \frac{C_2}{\lambda} \right) r_0 - \frac{C_2}{\lambda} r_0 F(k) + \\
 &+ q_1^2 - C_2(k - 1) \bar{v}. \quad (19)
 \end{aligned}$$

Правую часть (19) обозначим через $G(k)$:

$$\begin{aligned}
 G(k) &= \left(C_0 + C_1 + \frac{C_2}{\lambda} \right) r_0 - \frac{C_2}{\lambda} r_0 F(k) + \\
 &+ q_1^2 - C_2(k - 1) \bar{v}. \quad (20)
 \end{aligned}$$

Докажем следующее утверждение.

Теорема 1. Пусть $F(k) - F(k + 1) < \lambda \bar{v}$, $0 < \lambda < \infty$, $0 < \bar{v} < \infty$, $k > 0$. Тогда при любых значениях параметров $0 \leq C_i < \infty$, $i = 0, \dots, 4$, $0 < C_2 < \infty$ существует стационарная стратегия $0 < k^0 < \infty$. При этом если $g^1 \geq G(1)$, то $k^0 = 1$, а если $C_2 = 0$ и $g^1 < G(1)$, то $k^0 = \infty$.

Доказательство. Из (20) и условия теоремы следует:

$$\begin{aligned}
 G(k) - G(k+1) &= -\frac{C_2}{\lambda} r_0 [F(k) - F(k+1)] + C_2 \bar{v} > \\
 &> C_2 \bar{v} (1 - r_0) > 0. \quad (21)
 \end{aligned}$$

Рассмотрим поведение функции g^k при последовательном увеличении значения $k > 0$. Как следует из (18), неравенства $g^k \leq g^{k+1}$ и $g^k \leq G(k)$ эквивалентны. Отсюда вытекает, что если при текущем значении $k > 0$ выполняется неравенство $g^k < G(k)$, то при следующем значении $k + 1$ выполняется неравенство $g^{k+1} > g^k$, и наоборот: если при текущем значении $k > 0$ выполняется неравенство $g^k > G(k)$, то при следующем значении $k + 1$ выполняется неравенство $g^{k+1} < g^k$, а при $g^k = G(k)$ выполняется равенство $g^{k+1} = g^k$.

Рассмотрим сначала случай $C_2 > 0$. Пусть k_1 — произвольное число, $k_1 \geq 1$. Рассмотрим два альтернативных случая: (1) $g^{k_1} \geq G(k_1)$ и (2) $g^{k_1} < G(k_1)$. Всюду ниже под k^0 будем понимать одну из стратегий, удовлетворяющих условию: $\max_{k > 0} g^k = g^{k^0}$ (см. (3)).

Пусть $g^{k_1} \geq G(k_1)$. Тогда если $g^{k_1} > G(k_1)$, то согласно (18) (так как $0 < \pi_k^{k+1} < 1$) выполняется $g^{k_1+1} > G(k_1)$, а так как $G(k)$ убывает по $k > 0$ (следует из (21)), выполняется $g^{k_1+1} > G(k_1 + 1)$. Следовательно, $g^k > G(k)$, $k \geq k_1$, и согласно (18) и (19) последовательность $\{g^k, k = k_1, k_1 + 1, \dots\}$ является убывающей, т. е. $g^k < g^{k_1}$, $k > k_1$, откуда следует, что существует $k^0 \leq k_1$. Если $g^{k_1} = G(k_1)$, то из (13) и неравенства (21) следует, что $g^{k_1+1} = g^{k_1}$, $g^{k_1+1} > G(k_1 + 1)$. Тогда, как уже показано, существует $k^0 \leq k_1 + 1$. Так как $g^{k_1+1} = g^{k_1}$, то существует также $k^0 \leq k_1$. При этом если $k_1 = 1$, то получаем $k^0 = 1$.

Пусть $g^{k_1} < G(k_1)$. Тогда согласно неравенству (19) из $g^{k_1} < G(k_1)$ следует $g^{k_1+1} > g^{k_1}$ и неравенство $g^{k_1+1} < G(k_1)$, так как $g^{k_1+1} - g^{k_1} < G(k_1) - g^{k_1}$ (это следует из равенства (18) и неравенства

$0 < \pi_k^{k+1} < 1$). Следовательно, пока выполняется неравенство $G(k) > g^k$, $k = k_1, k_1 + 1, \dots$, будет выполняться неравенство $G(k) > g^{k+1} > g^k$. Так как согласно (21) $G(k)$ убывает по $k > 0$ и $G(k) \xrightarrow[k]{\infty} -\infty$, то существует k_2 , $k_1 < k_2 < \infty$, такое, что выполняется система неравенств $g^{k_2+1} \geq G(k_2 + 1)$, $G(k_1) > G(k) > g^{k+1} > g^k$, $k_1 + 1 \leq k \leq k_2$. Тогда поскольку при $k_2 + 1$ наступает случай 1, то, как уже доказано, существует $k^0 \leq k_2 + 1$, т. е. утверждение теоремы 1 для случая 2 верно.

Пусть теперь $C_2 = 0$. В этом случае из (20) следует $G(k) = const$, $k > 0$. Рассуждая точно так же, как и в случае $C_2 > 0$, получаем последовательность $\{g^k, k \geq 1\}$, которая удовлетворяет неравенствам $g^1 \geq g^2 \geq \dots \geq g^n \geq \dots \geq const$ при условии $g^1 \geq const$ и неравенствам $g^1 < g^2 < \dots < g^n < \dots < const$ при условии $g^1 < const$. Как видим, $k = 1$ при $g^1 \geq const$ и $k = \infty$ при $g^1 < const$ удовлетворяют определению k^0 в условии теоремы.

Следствие 1. Пусть $F(k)$ удовлетворяет условию теоремы. Стратегия k^0 удовлетворяет условию $\max_{k>0} g^k = g^{k^0}$ тогда и только тогда, когда k^0 удовлетворяет одному из трех условий:

- (1) $G(1) \leq g^1, k^0 = 1$;
- (2) $G(1) > g^1, k^0 = \min\{k : G(k) \leq g^k\}$;
- (3) $g^{k^0-1} < g^{k^0}, g^{k^0+1} > g^{k^0}, 1 < k^0 < \infty$.

Доказательство. Доказательство следствия 1 следует непосредственно из доказательства теоремы 1.

Следствие 2. Пусть $F(k)$ удовлетворяет условию теоремы и $g^{k^0} > 0$. Тогда решение (одно из решений) k^* задачи (3) удовлетворяет условиям $k^* \geq k^0$, $\lambda(1 - \theta_{k^0}^{k^0})g^{k^0} \leq Q^{k^*} \leq \lambda g^{k^0}$.

Доказательство. Так как θ_k^k убывает по k , для любой стратегии $0 < k \leq k^0$ выполняются условия $Q^k \leq \lambda(1 - \theta_{k^0}^{k^0})g^{k^0} = Q^{k^0} \leq Q^{k^*}$, если $g^{k^0} \geq 0$ (см. (2)). Кроме того, $0 < Q^{k^*} = \lambda(1 - \theta_{k^*}^{k^*})g^{k^*} < \lambda g^{k^*} < \lambda g^{k^0}$.

На основании следствия 1 стратегию k^0 можно найти с помощью следующего простого алгоритма.

1. Положить $k = 1$.
2. Если $G(k) > g^k$, увеличить k на единицу, иначе перейти к п. 5.
3. Вычислить $G(k)$ и g^k .
4. Если выполняется неравенство $G(k) > g^k$, перейти к п. 2, иначе перейти к п. 5.
5. Положить $k^0 = k$.
6. Конец алгоритма.

4 Пример

В качестве примера рассмотрим СМО $M/H_n/1$ с функцией распределения времени обслуживания $H_n(t) = \sum_{i=1}^n f_i(1 - e^{-\mu_i t})$. Справедливы соотношения:

$$r_{k-j+1} = \sum_{i=1}^n f_i \frac{\mu_i}{\lambda + \mu_i} \left(\frac{\lambda}{\lambda + \mu_i} \right)^{k-j+1};$$

$$\sum_{m=k-j+1}^{\infty} r_m = \sum_{m=k-j+1}^{\infty} \sum_{i=1}^n f_i \frac{\mu_i}{\lambda + \mu_i} \left(\frac{\lambda}{\lambda + \mu_i} \right)^m = \sum_{i=1}^n f_i \left(\frac{\lambda}{\lambda + \mu_i} \right)^{k-j+1};$$

$$\sum_{m=k-j+1}^{\infty} m r_m = \sum_{i=1}^n f_i \left(\frac{\lambda}{\lambda + \mu_i} \right)^{k-j+1} \left(\frac{\lambda + \mu_i}{\mu_i} + k - j \right).$$

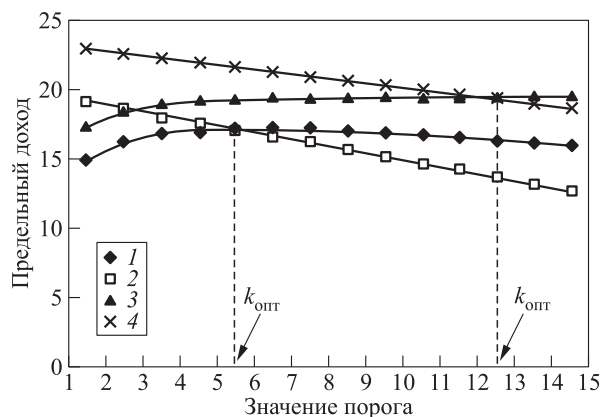
Функции $F(k)$ (см. (17)) и $G(k)$ принимают вид:

$$F(k) = \frac{\sum_{j=0}^{k-1} R_j \sum_{m=k-j+1}^{\infty} [m - (k - j + 1)] r_m}{\sum_{j=0}^{k-1} R_j \sum_{i=k-j+1}^{\infty} r_i} + 1 = \frac{\sum_{i=1}^n f_i (1/\mu_i) \sum_{j=0}^{k-1} R_j (\lambda/(\lambda + \mu_i))^{k-j+1}}{\sum_{i=1}^n f_i \sum_{j=0}^{k-1} R_j (\lambda/(\lambda + \mu_i))^{k-j+1}} + 1;$$

$$G(k) = r_o (C_0 + C_1) - C_2 r_0 \frac{\sum_{i=1}^n f_i (1/\mu_i) \sum_{j=0}^{k-1} R_j (\lambda/(\lambda + \mu_i))^{k-j+1}}{\sum_{i=1}^n f_i \sum_{j=0}^{k-1} R_j (\lambda/(\lambda + \mu_i))^{k-j+1}} + q_1^2 - C_2 \sum_{i=1}^n \frac{f_i}{\mu_i} (k - 1).$$

На рисунке приведены зависимости функций g^k и $G(k)$ от порогового значения и проиллюстрирована зависимость выполнения условия $g^{k+1} > g^k$ от условия $g^k < G(k)$. Графики 1 и 2 на рисунке соответствуют g^k и $G(k)$ при значениях параметров $C_0 = 20, C_1 = 10, C_2 = 0,5, C_3 = 0,01, C_4 = 0,01, \lambda = 1, n = 2, f_1 = 0,3, f_2 = 0,7$ и $\mu_1 = \mu_2 = 1$; графики 3 и 4 изображают функции g^k и $G(k)$ при $\mu_1 = 2$ и $\mu_2 = 1$ и тех же значениях остальных параметров. Оптимальный порог на рисунке обозначен через $k_{\text{опт}}$. В случае $M/M/1$ имеем:

$$r_{k-j+1} = \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^{k-j+1};$$



Зависимость предельного дохода и функции $G(k)$ от порогового значения

$$\sum_{m=k-j+1}^{\infty} r_m = \sum_{m=k-j+1}^{\infty} \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^m = \left(\frac{\lambda}{\lambda + \mu} \right)^{k-j+1};$$

$$\sum_{m=k-j+1}^{\infty} mr_m = \left(\frac{\lambda}{\lambda + \mu} \right)^{k-j+1} \left(\frac{\lambda + \mu}{\mu} + k - j \right);$$

$$F(k) = \lambda \bar{v} + 1;$$

$$G(k) = r_0 \left(C_0 + C_1 - \frac{C_2}{\mu} \right) + q_1^2 - \frac{C_2}{\mu} (k - 1).$$

Как видим, $F(k) - F(k + 1) = 0 < \lambda \bar{v}$ и $G(k)$ убывает линейно по $k > 0$; следовательно, согласно следствию 1 предложенный выше алгоритм дает оптимальный порог.

5 Заключение

В отличие от постановки задачи, рассмотренной в [7], выше предполагалось, что плата за обслуживание система получает в момент приема задачи в накопитель системы и величина платы не зависит от длительности обслуживания на приборе. Покажем, что и в случае, рассмотренном в [7], приведенные выше утверждения также справедливы.

Если плату за обслуживание система получает в момент окончания обслуживания заявки и величина платы прямо пропорциональна длительности обслуживания на приборе, средняя величина платы, получаемой системой за обслуживание заявки во время нахождения в данном состоянии, равна $C_0 \bar{v}$ (так как в каждом состоянии на приборе обслуживается одна заявка). Заметим, что эта величина не зависит от стратегии $k > 0$ и остальные составляющие дохода системы в этом случае

и в случае, рассмотренном выше, совпадают. После замены в (10) правой части формулы (7) на $C_0 \bar{v}$ формула (11) принимает вид:

$$q_j^{k+1} = q_j^k + C_1 \sum_{m=k-i+1}^{\infty} r_m - \frac{C_2}{\lambda} \left[\sum_{m=k-i+1}^{\infty} mr_m - (k - i + 1) \sum_{m=k-i+1}^{\infty} r_m \right].$$

Как видим, в формуле (11) произошла только замена константы $(C_0 + C_1)$ на константу C_1 , и, следовательно, все рассуждения, проведенные выше относительно задачи (3), остаются в силе. Отметим, что выше нигде не требовалось выполнение условия $\lambda \bar{v} < 1$ (в отличие от [7]).

В системах и сетях передачи данных буферная память узлов связи имеет ограниченную емкость, линии связи — ограниченную пропускную способность. Поэтому в качестве модели узла связи (или канала связи) больше подходит (в смысле адекватности) СМО типа $M/G/1/r$ с числом мест хранения $r < \infty$, чем $M/G/1$ с бесконечным числом мест ожидания. Вложенная цепь Маркова, которая рассматривается как модель процесса работы $M/G/1/r$, при любой стационарной смешанной пороговой стратегии имеет конечное число состояний и один эргодический класс. Тогда то, что задача (3) имеет решение и оно принадлежит множеству чистых стратегий, следует из [10]. Поэтому с практической точки зрения в рассматриваемой выше задаче максимизации (3) (как и в аналогичных других задачах) достаточно оптимизировать только чистые пороговые стратегии.

Полученные в данной работе результаты могут быть использованы для поиска оптимальных пороговых стратегий для систем, моделируемых с помощью СМО типа $M/G/1$ ($M/G/1/r$).

Дальнейшим развитием результатов данной статьи могут стать алгоритмы поиска оптимальных пороговых стратегий для СМО других типов, построенные с использованием аналогичных математических приемов, например путем построения соотношений, аналогичных (17).

Литература

1. *Welzl M.* Network congestion control. — New York, NY, USA: Wiley, 2005. 282 p.
2. *Печинкин А. В., Разумчик Р. В.* Время пребывания в различных режимах системы обслуживания с неординарными пуассоновскими входящими потоками, рекуррентным обслуживанием и гистерезисной политикой // Информационные процессы, 2015. Т. 15. № 3. С. 324–336.

3. Nino-Mora J. Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock $M/G/1$ queues // *Math. Oper. Res.*, 2006. Vol. 31. No. 1. P. 50–84.
4. Жерновы́й Ю. В. Решение задач оптимального синтеза для некоторых марковских моделей обслуживания // *Информационные процессы*, 2010. Т. 10. № 3. С. 257–274.
5. Коновалов М. Г. Об одной задаче оптимального управления нагрузкой на сервер // *Информатика и её применения*, 2013. Т. 7. Вып. 4. С. 34–43.
6. Агаларов Я. М. Пороговая стратегия ограничения доступа к ресурсам в СМО $M/D/1$ с функцией штрафов за несвоевременное обслуживание заявок // *Информатика и её применения*, 2015. Т. 9. Вып. 3. С. 56–65.
7. Гришунина Ю. Б. Оптимальное управление очередью в системе $M/G/1/\infty$ с возможностью ограничения приема заявок // *Автоматика и телемеханика*, 2015. № 3. С. 79–93.
8. Карлин С. Основы теории случайных процессов / Пер. с англ. — М.: Мир, 1971. 536 с. (*Karlin S. A first course in stochastic processes.* — New York, NY, USA — London, U.K.: Academic Press, 1968. 502 p.)
9. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
10. Майн Х., Осаки С. Марковские процессы принятия решений. — М.: Наука, 1977. 176 с.

Поступила в редакцию 18.02.16

ABOUT THE OPTIMAL THRESHOLD OF QUEUE LENGTH IN A PARTICULAR PROBLEM OF PROFIT MAXIMIZATION IN THE $M/G/1$ QUEUING SYSTEM

Ya. M. Agalarov¹, M. Ya. Agalarov², and V. S. Shorgin¹

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

²PromsvyazBank OJSC, 10 Smirnovskaya Str., Moscow 109052, Russian Federation

Abstract: The paper considers the problem of maximizing the average profit per time in the $M/G/1$ system on the set of access restriction stationary threshold strategies with one “switch point.” Profit in the described model is defined as the following measures: service fee, hardware maintenance fee, fine for service delay, fine for unhandled requests, and fine for system idle. The conditions of existence of optimal and finite threshold values are obtained. The method and the algorithm for calculating the lower bound for the optimal threshold and corresponding value of maximal profit per time are proposed. The auxiliary problem of maximizing the system profit, averaged by number of handled requests on the set of the considered threshold strategies, is solved. The necessary and sufficient conditions of existence of solution of the auxiliary problem are found. The method and algorithm for its solution are proposed.

Keywords: queuing system; threshold strategy; optimization

DOI: 10.14357/19922264160208

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 15-07-03406).

References

1. Welzl, M. 2005. *Network congestion control*. New York, NY: Wiley. 282 p.
2. Pechinkin, A. V., and R. V. Razumchik. 2015. Vremya prebyvaniya v razlichnykh rezhimakh sistemy obsluzhivaniya s neordinarnymi puassonovskimi vkhodyashchimi potokami, rekurrentnym obsluzhivaniem i gistereziy noy politikoy [First passage times between modes in the queueing system with batch Poisson arrivals, general service, and hysteresis policy]. *Informatsionnye Protsessy* [Inform. Proc.] 15(3):324–336.
3. Nino-Mora, J. 2006. Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock $M/G/1$ queues. *Math. Oper. Res.* 31(1):50–84.
4. Zhernovyy, Yu. V. 2010. Reshenie zadach optimal'nogo sinteza dlya nekotorykh markovskikh modeley obsluzhivaniya [Solution of optimum synthesis problem for

- some Markov models of service]. *Informatsionnye Protsesty* [Inform. Proc.] 10(3):257–274.
5. Konovalov, M. G. 2013. Ob odnoy zadache optimal'nogo upravleniya nagruzkoy na server [About one task of overload control]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):34–43.
 6. Agalarov, Ya. M. 2015. Porogovaya strategiya ogranicheniya dostupa k resursam v SMO $M/D/1$ s funktsiyey shtrafov za nesvoevremennoe obsluzhivanie zayavok [The threshold strategy for restricting access in the $M/D/1$ queuing system with penalty function for late service]. *Informatika i ee Primeneniya — Inform. Appl.* 9(3): 56–65.
 7. Grishunina, Yu. B. 2015. Optimal'noe upravlenie ochered'yu v sisteme $M/G/1/\infty$ s vozmozhnost'yu ogranicheniya priema zayavok [Optimal control of queue in the $M/G/1/\infty$ system with possibility of customer admission restriction]. *Avtomatika i Telemekhanika* [Automation Remote Control] 3:79–93.
 8. Karlin, S. 1968. *A first course in stochastic processes*. New York, N.Y. – London: Academic Press. 502 p.
 9. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
 10. Mine, H., and S. Osaki. 1970. *Markovian decision processes*. New York, NY: American Elsevier Publishing Co. 142 p.

Received February 18, 2016

Contributors

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agglar@yandex.ru

Agalarov Murad Ya. (b. 1987) — Head of System Analysis Department, PromsvyazBank OJSC, 10 Smirnovskaya Str., Moscow 109052, Russian Federation; murad-agalarov@yandex.ru

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VShorgin@ipiran.ru

БАЙЕСОВСКАЯ РЕКУРРЕНТНАЯ МОДЕЛЬ РОСТА НАДЕЖНОСТИ: ПАРАБОЛИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ПАРАМЕТРОВ

А. А. Кудрявцев¹, С. И. Палионная²

Аннотация: Данная работа посвящена рассмотрению случая параболического распределения параметров в байесовской рекуррентной модели роста надежности сложных модифицируемых информационных систем. Надежность системы зависит от соотношения параметров, интерпретируемых в теории надежности как показатели «дефективности» и «эффективности» средства, исправляющего ошибки в системе. При использовании байесовских моделей применительно к задачам теории надежности предполагается, что основные параметры системы не являются заданными, а известны только их априорные распределения. Приводятся формулы для вычисления средней предельной надежности системы для параболического априорного распределения параметров. Также приводятся численные результаты для модельных примеров.

Ключевые слова: модифицируемые информационные системы; теория надежности; байесовский подход; параболическое распределение

DOI: 10.14357/19922264160209

1 Введение

Как правило, любая впервые созданная сложная информационная система не обладает необходимой надежностью. Вследствие этого она подвергается различным модификациям, целью которых является устранение дефектов, препятствующих правильному функционированию системы. Изменения в систему могут вноситься на стадиях разработки, испытаний и опытной эксплуатации, а также во время штатного функционирования.

Для формализации понятия надежности системы будем характеризовать ее в каждый момент времени t параметром $p(t)$, при этом время считаем непрерывным. Так как система подвергается модификациям в случайные моменты времени $0 = Y_0 \leq \leq Y_1 \leq Y_2 \leq \dots$, параметр $p(t)$ изменяется, принимая соответственно значения $p(t) = p(Y_j) = p_j$ при $Y_j \leq t < Y_{j+1}$ (предполагается, что траектории процесса $p(t)$ непрерывны справа, а модификации происходят мгновенно). Причем при модификациях параметр надежности может как увеличиваться, так и уменьшаться из-за некачественных модификаций.

К числу математических моделей, описывающих изменение надежности модифицируемых информационных систем, относятся рекуррентные модели роста надежности. Рассмотрим, в частно-

сти, дискретную экспоненциальную модель, которая определяется следующим образом:

$$p_{j+1} = \eta_{j+1}p_j + \theta_{j+1}(1 - p_j).$$

Здесь $\{(\theta_j, \eta_j)\}$, $j \geq 1$, — последовательность независимых одинаково распределенных двумерных случайных векторов таких, что $0 < \eta_1 < 1$; $0 < \theta_1 < 1$ почти наверное. Начальная надежность p_0 считается заданной, случайные величины η_j (параметры «дефективности») и θ_j (параметры «эффективности») описывают соответственно возможное уменьшение и увеличение надежности.

Обозначим $\lambda = 1 - E\theta_j$, $\mu = E\eta_j$. В книге [1] доказано, что при условии $\lambda + \mu \neq 1$

$$p = \lim_{j \rightarrow \infty} E p_j = \frac{\mu}{\lambda + \mu}.$$

Величина p характеризует асимптотическое значение надежности системы в рамках некоторой рекуррентной модели, задаваемой набором $\{(\theta_j, \eta_j)\}$.

В рамках байесовского подхода к постановке задач теории надежности можно рассмотреть более сложную ситуацию, где основные параметры системы λ и μ предполагаются случайными. В таком случае наиболее естественной и удобной для изучения характеристикой является усредненное значение предельной надежности, т. е.

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

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, sofapalonnaya@gmail.com

$$p_{\text{сред}} = \mathbb{E}p = \mathbb{E} \frac{\mu}{\lambda + \mu},$$

где усреднение ведется по совместному распределению случайных величин λ и μ .

Так как случайные величины η_1 и θ_1 удовлетворяют ограничениям $0 < \eta_1 < 1, 0 < \theta_1 < 1$, средние значения λ и μ величин $1 - \mathbb{E}\theta_j$ и $\mathbb{E}\eta_j$ соответственно также находятся на отрезке $[0, 1]$. Поэтому в качестве априорных распределений параметров λ и μ следует выбирать только распределения, сосредоточенные на $[0, 1]$.

Далее будут рассмотрены независимые случайные параметры λ и μ , имеющие параболические распределения на некоторых (вообще говоря, разных) отрезках, являющихся подмножествами отрезка $[0, 1]$.

2 Основные результаты

Пусть средний параметр «эффективности» λ и средний параметр «дефективности» μ независимы и имеют параболическое распределение $P(a_\lambda, b_\lambda)$, $0 \leq a_\lambda < b_\lambda \leq 1$, и $P(a_\mu, b_\mu)$, $0 \leq a_\mu < b_\mu \leq 1$, соответственно.

Для плотности $f_\xi(x)$ некоторой случайной величины ξ , имеющей параболическое распределение с параметрами (a_ξ, b_ξ) , справедливо

$$f_\xi(x) = \frac{6(x - a_\xi)(b_\xi - x)}{(b_\xi - a_\xi)^3}, \quad x \in [a_\xi, b_\xi]. \quad (1)$$

Заметим, что плотность параболического распределения может быть представлена в виде полинома:

$$f_\xi(x) = \sum_{i=0}^2 c_{\xi,i} x^i \cdot \mathbf{1}(x \in [a_\xi, b_\xi]),$$

где

$$c_{\xi,0} = -\frac{6a_\xi b_\xi}{(b_\xi - a_\xi)^3}; \quad c_{\xi,1} = \frac{6(a_\xi + b_\xi)}{(b_\xi - a_\xi)^3};$$

$$c_{\xi,2} = -\frac{6}{(b_\xi - a_\xi)^3}.$$

Очевидно, что при байесовском подходе вычисление вероятностных характеристик надежности $p = \mu/(\lambda + \mu)$ удобно производить, базируясь на известном распределении величины $\rho = \lambda/\mu$. В работе [2] были получены формулы для плотности распределения случайной величины ρ в предположении, что плотности параметров λ и μ имеют полиномиальный вид. Основываясь на характеристиках распределения параметра ρ и учитывая, что

$p = 1/(1 + \rho)$, имеем для функции распределения и плотности p следующие соотношения:

$$F_p(x) = 1 - F_\rho \left(\frac{1-x}{x} \right); \quad f_p = \frac{1}{x^2} f_\rho \left(\frac{1-x}{x} \right).$$

Введем обозначение:

$$L(a, b, x) = \frac{1}{x^2} \int_a^b y f_\lambda \left(\frac{1-x}{x} y \right) f_\mu(y) dy =$$

$$= \sum_{i=0}^2 \sum_{j=0}^2 c_{\lambda,i} c_{\mu,j} \frac{b^{i+j+2} - a^{i+j+2}}{i+j+2} \frac{(1-x)^i}{x^{i+2}}, \quad (2)$$

где a и b одновременно принадлежат отрезкам $[a_\mu, b_\mu]$ и $[a_\lambda/x, b_\lambda/x]$.

Рассмотрев всевозможные комбинации взаимного расположения точек $a_\mu, b_\mu, a_\lambda/x$ и b_λ/x на отрезке $[0, 1]$, убеждаемся в справедливости следующего утверждения.

Теорема 1. Пусть независимые случайные величины λ и μ имеют параболическое распределение, а их плотности $f_\lambda(x)$ и $f_\mu(x)$ определяются соотношением (1) с соответствующими параметрами. Тогда случайная величина $p = \mu/(\lambda + \mu)$ имеет плотность:

$$f_p(x) =$$

$$= \mathbf{1} \left(\frac{a_\mu}{b_\lambda + a_\mu} < x \leq \min \left\{ \frac{a_\mu}{a_\lambda + a_\mu}, \frac{b_\mu}{b_\lambda + b_\mu} \right\} \right) \times$$

$$\times L \left(a_\mu, \frac{b_\lambda x}{1-x}, x \right) +$$

$$+ \mathbf{1} \left(\frac{b_\mu}{b_\lambda + b_\mu} < x \leq \frac{a_\mu}{a_\lambda + a_\mu} \right) L(a_\mu, b_\mu, x) +$$

$$+ \mathbf{1} \left(\frac{a_\mu}{a_\lambda + a_\mu} < x \leq \frac{b_\mu}{b_\lambda + b_\mu} \right) L \left(\frac{a_\lambda x}{1-x}, \frac{b_\lambda x}{1-x}, x \right) +$$

$$+ \mathbf{1} \left(\max \left\{ \frac{a_\mu}{a_\lambda + a_\mu}, \frac{b_\mu}{b_\lambda + b_\mu} \right\} < x \leq \frac{b_\mu}{a_\lambda + b_\mu} \right) \times$$

$$\times L \left(\frac{a_\lambda x}{1-x}, b_\mu, x \right),$$

где величины $L(a, b, x)$ определены соотношением (2).

Для $\xi \in \{\lambda, \mu\}$ введем дополнительное обозначение:

$$J_\xi(d, b) = \sum_{i=0}^2 \sum_{j=0}^2 c_{\lambda,i} c_{\mu,j} \frac{d^{i+j+2}}{i+j+2} \times$$

$$\times \sum_{k=0}^{l_\xi} C_{l_\xi}^k (-1)^k \left[\mathbf{1}(k \neq l_\xi) \frac{b^{k-l_\xi}}{k-l_\xi} + \mathbf{1}(k = l_\xi) \ln b \right], \quad (3)$$

где $l_\xi = i \mathbf{1}(\xi = \lambda) + (j+1) \mathbf{1}(\xi = \mu)$.

Таблица 1 Частные значения $p_{\text{сред}} (\lambda \sim P(a_\lambda, b_\lambda), \mu \sim P(a_\mu, b_\mu))$

$a_\lambda; b_\lambda$	$a_\mu; b_\mu$									
	0,0; 1	0,1; 1	0,2; 1	0,3; 1	0,4; 1	0,5; 1	0,6; 1	0,7; 1	0,8; 1	0,9; 1
0,0; 1	0,50	0,53	0,56	0,58	0,60	0,62	0,63	0,65	0,66	0,67
0,1; 1	0,47	0,50	0,53	0,55	0,57	0,59	0,60	0,62	0,63	0,65
0,2; 1	0,44	0,47	0,50	0,52	0,54	0,56	0,58	0,59	0,61	0,62
0,3; 1	0,42	0,45	0,48	0,50	0,52	0,54	0,56	0,57	0,59	0,60
0,4; 1	0,40	0,43	0,46	0,48	0,50	0,52	0,54	0,55	0,57	0,58
0,5; 1	0,38	0,41	0,44	0,46	0,48	0,50	0,52	0,53	0,55	0,56
0,6; 1	0,37	0,40	0,42	0,44	0,46	0,48	0,50	0,52	0,53	0,54
0,7; 1	0,35	0,38	0,41	0,43	0,45	0,47	0,48	0,50	0,51	0,53
0,8; 1	0,34	0,37	0,39	0,41	0,43	0,45	0,47	0,49	0,50	0,51
0,9; 1	0,33	0,35	0,38	0,40	0,42	0,44	0,46	0,47	0,49	0,50

Таблица 2 Частные значения $p_{\text{сред}} (\lambda \sim P(a_\lambda, b_\lambda), \mu \sim P(a_\mu, b_\mu))$

$a_\lambda; b_\lambda$	$a_\mu; b_\mu$									
	0; 0,1	0; 0,2	0; 0,3	0; 0,4	0; 0,5	0; 0,6	0; 0,7	0; 0,8	0; 0,9	0; 1,0
0; 0,1	0,50	0,65	0,72	0,77	0,81	0,83	0,85	0,86	0,88	0,89
0; 0,2	0,35	0,50	0,59	0,65	0,69	0,72	0,75	0,77	0,79	0,81
0; 0,3	0,28	0,41	0,50	0,56	0,61	0,65	0,68	0,70	0,72	0,74
0; 0,4	0,23	0,35	0,44	0,50	0,55	0,59	0,62	0,65	0,67	0,69
0; 0,5	0,19	0,31	0,39	0,45	0,50	0,54	0,57	0,60	0,63	0,65
0; 0,6	0,17	0,28	0,35	0,41	0,46	0,50	0,53	0,56	0,59	0,61
0; 0,7	0,15	0,25	0,32	0,38	0,43	0,47	0,50	0,53	0,55	0,58
0; 0,8	0,14	0,23	0,30	0,35	0,40	0,44	0,47	0,50	0,53	0,55
0; 0,9	0,12	0,21	0,28	0,33	0,37	0,41	0,45	0,47	0,50	0,52
0; 1,0	0,11	0,19	0,26	0,31	0,35	0,39	0,42	0,45	0,48	0,50

Воспользовавшись теоремой 1 для вычисления средней надежности системы

$$p_{\text{сред}} = E p = \int x f_p(x) dx,$$

убеждаемся в справедливости следующего утверждения.

Теорема 2. Пусть средний параметр «эффективности» λ и средний параметр «дефективности» μ удовлетворяют условиям теоремы 1. Тогда средняя предельная надежность системы имеет вид:

$$p_{\text{сред}} = J_\mu \left(b_\lambda, \frac{b_\lambda}{b_\lambda + a_\mu} \right) + J_\mu \left(a_\lambda, \frac{a_\lambda}{a_\lambda + b_\mu} \right) + J_\lambda \left(a_\mu, \frac{a_\mu}{b_\lambda + a_\mu} \right) + J_\lambda \left(b_\mu, \frac{b_\mu}{a_\lambda + b_\mu} \right) - J_\lambda \left(a_\mu, \frac{a_\mu}{a_\lambda + a_\mu} \right) - J_\lambda \left(b_\mu, \frac{b_\mu}{b_\lambda + b_\mu} \right) - J_\mu \left(b_\lambda, \frac{b_\lambda}{b_\lambda + b_\mu} \right) - J_\mu \left(a_\lambda, \frac{a_\lambda}{a_\lambda + a_\mu} \right), \quad (4)$$

где величины $J_\lambda(d, b)$ и $J_\mu(d, b)$ определены соотношением (3).

3 Численные результаты

Основываясь на формуле (4), приведем табл. 1 и 2, содержащие частные значения (с точностью до сотых) средней предельной надежности системы для некоторых наборов параметров.

Литература

1. Королев В. Ю., Соколов И. А. Основы математической теории надежности модифицируемых систем. — М.: ИПИ РАН, 2006. 102 с.
2. Кудрявцев А. А. Байесовские модели массового обслуживания и надежности: априорные распределения с компактным носителем // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 67–71.

Поступила в редакцию 29.03.16

BAYESIAN RECURRENT MODEL OF RELIABILITY GROWTH: PARABOLIC DISTRIBUTION OF PARAMETERS

A. A. Kudryavtsev and S. I. Palionnaia

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

Abstract: This work is devoted to the study of the parabolic distribution of parameters in the Bayesian recurrent model of reliability growth of complex modifiable information systems. In the reliability theory, the reliability of the system depends on the ratio of parameters which are interpreted as indexes of “defectiveness” and “efficiency” of the tool correcting the deficiencies in the system. In the framework of Bayesian models, it is assumed that only the information about the *a priori* distributions of the system’s parameters is given. In this work, the average marginal system reliability is calculated for the *a priori* parabolic distribution of the parameters. The numerical results for the model examples are obtained.

Keywords: modifiable information system; reliability theory; Bayesian approach; parabolic distribution

DOI: 10.14357/19922264160209

References

1. Korolev, V. Y., and I. A. Sokolov. 2006. *Osnovy matematicheskoy teorii nadezhnosti modifitsiruemykh sistem* [The fundamentals of mathematical reliability theory of modifiable systems]. Moscow: IPI RAN. 102 p.
2. Kudryavtsev, A. A. 2016. Bayesovskie modeli v teorii massovogo obsluzhivaniya i nadezhnosti: Apriornie raspredeleniya s kompaktnym nositelem [Bayesian queueing and reliability models: *A priori* distributions with compact support]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):67–71.

Received March 29, 2016

Contributors

Kudryavtsev Alexey A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

Palionnaia Sofia I. (b. 1995) — student, 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; sofiaPalionnaya@gmail.com

РАЗЛОЖЕНИЯ ТИПА КОРНИША–ФИШЕРА ДЛЯ РАСПРЕДЕЛЕНИЙ СТАТИСТИК, ПОСТРОЕННЫХ ПО ВЫБОРКАМ СЛУЧАЙНОГО РАЗМЕРА*

А. С. Марков¹, М. М. Монахов², В. В. Ульянов³

Аннотация: Для квантилей выборочного среднего по выборке случайного объема построены обобщенные разложения Корниша–Фишера на базе квантилей распределений Лапласа и Стьюдента. В последние годы интерес к разложениям Корниша–Фишера значительно вырос в связи с исследованиями по управлению рисками. Широко распространенная мера риска Value at Risk (VaR) является квантилью функции потерь. Используется общая теорема переноса, позволяющая получать асимптотические разложения для функций распределения статистик по выборкам случайного объема из асимптотических разложений для функции распределения случайного объема выборки и асимптотических разложений для функций распределения статистик по выборкам неслучайного объема. Проведен вычислительный эксперимент, иллюстрирующий полученные разложения Корниша–Фишера.

Ключевые слова: обобщенные разложения Корниша–Фишера; выборка случайного объема; распределение Лапласа; распределение Стьюдента

DOI: 10.14357/19922264160210

1 Введение

В классических задачах математической статистики объем выборки традиционно считается детерминированным и играет роль *известного* параметра, как правило, неограниченно возрастающего. Однако на практике очень часто возникают ситуации, при которых размер выборки не определен заранее и может считаться случайным. Обычно это возникает ввиду того, что данные накапливаются в течение некоторого промежутка времени, длительность которого по разным причинам нельзя считать фиксированной. В этом случае встает вопрос об аппроксимирующих разложениях для различных характеристик статистик, например функций распределения или квантилей, основанных на выборках случайного объема.

Предельные распределения для случайно индексированных последовательностей и их применения см., например, в монографии [1]. Ранее Б. В. Гнеденко в работе [2] продемонстрировал, что в математической статистике при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут радикально изменяться.

В [3] доказана общая теорема переноса, позволяющая получать асимптотические разложения для функций распределения статистик, основанных на выборках случайного объема, из асимптотических разложений для функции распределения случайного объема выборки и асимптотических разложений для функций распределения статистик, построенных по выборкам неслучайного объема.

Настоящая работа развивает результаты работы [3]: на основе разложения функции распределения статистики, основанной на выборке случайного объема специального вида, получены обобщенные разложения Корниша–Фишера на базе квантилей распределений Лапласа и Стьюдента.

Классические разложения Корниша–Фишера на базе квантилей нормального распределения введены в [4], их обобщение было предложено в [5]. В последние годы интерес к разложениям Корниша–Фишера значительно вырос в связи с исследованиями по управлению рисками. Широко распространенная мера риска VaR является, по существу, квантилью функции потерь, связанной с описанием инвестиционного портфеля из финансовых инструментов (см., например, [6]).

* Исследование выполнено при поддержке Российского научного фонда (проект 14-11-00364).

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, markov.ctc@yandex.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, mih_monah@mail.ru

³Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, vulyanov@cs.msu.ru

Рассмотрим случайные величины (с.в.) N_1, N_2, \dots и X_1, X_2, \dots , заданные на одном и том же вероятностном пространстве $(\Omega, \mathbb{A}, \mathbb{P})$. В статистике с.в. X_1, X_2, \dots, X_n имеют смысл наблюдений, n — неслучайный объем выборки, а с.в. N_n — случайный объем выборки, зависящий от натурального параметра $n \in \mathbb{N}$. Предположим, что при каждом $n \geq 1$ с.в. N_n принимает только натуральные значения (т.е. $N_n \in \mathbb{N}$) и независима от последовательности с.в. X_1, X_2, \dots .

Обозначим через $T_n \equiv T_n(X_1, \dots, X_n)$ некоторую статистику. Для каждого $n \geq 1$ определим с.в. T_{N_n} , полагая

$$T_{N_n}(\omega) \equiv T_{N_n(\omega)}(X_1(\omega), \dots, X_{N_n(\omega)}(\omega)), \omega \in \Omega,$$

т.е. T_{N_n} — это статистика, построенная на основе статистики T_n по выборке случайного объема N_n .

Определение 1. Квантилью порядка α (α -квантилью) случайной величины X с функцией распределения $F(x) = \mathbb{P}(X < x)$ называется число x_α , такое что

$$x_\alpha = \inf \{x; F(x) > \alpha\}.$$

2 Основные результаты

Всюду ниже X_1, X_2, \dots — независимые одинаково распределенные с.в. с $\mathbb{E}(X_1) = \mu$, $0 < < \mathbb{D}(X_1) = \sigma^2$, $\mathbb{E}|X_1|^{3+2\delta} < \infty$ для $\delta \in (0, 1/2)$ и $\mathbb{E}(X_1 - \mu)^3 = \mu_3$. Для натурального n обозначим $T_n = (X_1 + \dots + X_n)/n$.

В дальнейшем предполагаем, что с.в. X_1 удовлетворяет условию Крамера:

$$\limsup_{|t| \rightarrow \infty} |\mathbb{E}e^{itX_1}| < 1.$$

2.1 Распределение Стьюдента как предельное

Предположим, что с.в. N_n имеет отрицательное биномиальное распределение с параметрами $p = = 1/n$ и $r > 0$, т.е. для $k \in \mathbb{N}$

$$\mathbb{P}(N_n = k) = \frac{(k+r-2) \cdots r}{(k-1)!} \frac{1}{n^r} \left(1 - \frac{1}{n}\right)^{k-1}.$$

Пусть $G_f(x)$ — функция распределения Стьюдента с параметром f , соответствующая плотности вида:

$$g_f(x) = \frac{\Gamma((f+1)/2)}{\sqrt{\pi f} \Gamma(f/2)} \left(1 + \frac{x^2}{f}\right)^{-(f+1)/2}, \quad x \in \mathbb{R},$$

где $\Gamma(\cdot)$ — эйлерова гамма-функция, а $f > 0$ — параметр формы. Если параметр f натурален, то он называется числом степеней свободы.

Теорема 1. Пусть $x = x_\alpha$ — α -квантиль нормированной статистики $\sigma\sqrt{r(n-1)+1}(T_{N_n} - \mu)$, $u = = u_\alpha$ — α -квантиль распределения Стьюдента с параметром $2r$. Тогда в обозначениях, введенных выше, справедливо следующее асимптотическое разложение:

$$x = u - \frac{\mu_3 \sigma^3 \Gamma(r)}{6\sqrt{n}} \frac{1+ru^2}{r-1/2} \left[\frac{1+u^2/2r}{1+u^2/2}\right]^{r+1/2} + R,$$

где

$$R = \begin{cases} \mathcal{O}\left(\frac{1}{n}\right), & r \geq \frac{2}{1+2\delta}; \\ o\left(\frac{1}{\sqrt{n}}\right), & \frac{1}{1+2\delta} < r < \frac{2}{1+2\delta}. \end{cases}$$

Замечание 1. При $r = 1$ объем выборки N_n имеет геометрическое распределение. Асимптотическое разложение для квантилей в этом случае принимает вид:

$$x = u - \frac{\mu_3 \sigma^3}{3} (1+u^2) \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

2.2 Распределение Лапласа как предельное

Предположим, что с.в. $N_n = N_n(s)$, где $s \in \in \mathbb{N}$ — фиксированный параметр, имеет распределение вида:

$$\mathbb{P}(N_n(s) = k) = \binom{k}{s+k}^n - \binom{k-1}{s+k-1}^n, \quad k \in \mathbb{N}.$$

Пусть $\Lambda_\theta(x)$ — функция распределения Лапласа с параметром $\theta > 0$, соответствующая плотности вида:

$$\lambda_\theta(x) = \frac{1}{\theta\sqrt{2}} e^{-\sqrt{2}|x|/\theta}, \quad x \in \mathbb{R}.$$

Теорема 2. Пусть $x = x_\alpha$ есть α -квантиль нормированной статистики $\sigma\sqrt{n}(T_{N_n(s)} - \mu)$, $u = u_\alpha$ — α -квантиль распределения Лапласа с параметром $1/s$. Тогда в обозначениях, введенных выше, справедливо следующее асимптотическое разложение:

$$x = u - \frac{\mu_3 \sigma^3}{6} \left(\frac{|u|}{\sqrt{2s}} + \frac{1}{2s} - u^2\right) \frac{\lambda_{1/\sqrt{s}}(u)}{\lambda_{1/s}(u)} \frac{1}{\sqrt{n}} + + o\left(\frac{1}{\sqrt{n}}\right).$$

Замечание 2. При $s = 1$ основной результат значительно упрощается:

$$x = u - \frac{\mu_3 \sigma^3}{6\sqrt{n}} \left(\frac{|u|}{\sqrt{2}} + \frac{1}{2} - u^2 \right) + o\left(\frac{1}{\sqrt{n}}\right).$$

Замечание 3. В теоремах 1 и 2 построены приближения квантилей x нормированных статистик некоторыми функциями от квантилей u распределений Стьюдента и Лапласа соответственно. При этом точность приближения дана в виде порядка по n , в частности как $o(1/\sqrt{n})$. В [7] в общей ситуации показано, что для ошибок приближения квантилей можно получать более информативные вычислимые оценки при наличии таковых для ошибок приближения распределений нормированных статистик.

3 Вывод основных результатов

3.1 Вспомогательная лемма

Пусть $F_n(x)$ есть последовательность функций распределения, каждая из которых допускает разложение типа Эджворда–Чебышёва по степеням $\varepsilon = n^{-1/2}$ или $\varepsilon = n^{-1}$ (см., например, [8]):

$$\begin{aligned} F_n(x) &= G_{k,n}(x) + O(\varepsilon^k); \\ G_{k,n}(x) &= G(x) + \{\varepsilon a_1(x) + \dots + \varepsilon^{k-1} a_{k-1}(x)\} g(x), \end{aligned}$$

где $g(x)$ — плотность распределения $G(x)$.

В частности, для $k = 2$ имеем:

$$F_n(x) = G(x) + \varepsilon a_1(x)g(x) + O(\varepsilon^2). \quad (1)$$

Лемма 1 (см., например, [8]). *В сформулированных выше обозначениях имеет место следующее разложение:*

$$x(u) = u + \varepsilon b_1(u) + O(\varepsilon^2), \quad (2)$$

где $x(u)$ и u суть квантили соответственно распределений F_n и G одинакового порядка, т. е. $F_n(x(u)) = G(u)$ и $b_1(u) = -a_1(u)$.

3.2 Доказательство теоремы

для распределения Стьюдента как предельного

Для вывода необходимого асимптотического разложения квантилей распределения статистики, построенной по выборке случайного объема, воспользуемся результатом, полученным в работе [3].

Введем следующее обозначение:

$$f_r(x) \equiv \int_0^\infty \phi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} dH_r(y),$$

где $H_r(x)$ — функция гамма-распределения с параметром $r > 0$:

$$H_r(x) = \frac{r^r}{\Gamma(r)} \int_0^x e^{-ry} y^{r-1} dy, x \geq 0.$$

В работе [3] показано, что в условиях теоремы для функции распределения нормированной статистики T_{N_n} справедлив следующий результат:

$$\sup_x \left| \mathbb{P}(\sigma\sqrt{r(n-1)+1}(T_{N_n} - \mu) < x) - G_{2r}(x) - \frac{\mu_3 \sigma^3}{6\sqrt{r(n-1)+1}} f_r(x) \right| = R,$$

где

$$R = \begin{cases} O\left(\left(\frac{\log n}{n}\right)^{1/2+\delta}\right), & r = 1; \\ O\left(\frac{1}{n^{\min(1, r(1/2+\delta))}}\right), & r > 1; \\ O\left(\frac{1}{n^{r(1+2\delta)}}\right), & \frac{1}{1+2\delta} < r < 1. \end{cases}$$

Отсюда при любом $x \in \mathbb{R}$ имеем асимптотическое разложение вида:

$$\begin{aligned} P\left(\sigma\sqrt{r(n-1)+1}(T_{N_n} - \mu) < x\right) &= \\ &= G_{2r}(x) - \frac{\mu_3 \sigma^3}{6\sqrt{r(n-1)+1}} f_r(x) + R_1, \quad (3) \end{aligned}$$

где

$$R_1 = \begin{cases} O\left(\frac{1}{n}\right), & r \geq \frac{2}{1+2\delta}; \\ o\left(\frac{1}{\sqrt{n}}\right), & \frac{1}{1+2\delta} < r < \frac{2}{1+2\delta}. \end{cases}$$

Введем дополнительно следующие обозначения:

$$\left. \begin{aligned} F_n(x) &\equiv \mathbb{P}(\sigma\sqrt{r(n-1)+1}(T_{N_n} - \mu) < x); \\ G(x) &\equiv G_{2r}(x); \\ \varepsilon &\equiv n^{-1/2}; \\ \varepsilon a_1(x)g(x) &\equiv \frac{\mu_3 \sigma^3}{6\sqrt{r(n-1)+1}} f_r(x), \end{aligned} \right\} \quad (4)$$

где $g(x)$ есть плотность распределения G . Заметим теперь, что в новых обозначениях разложение (3) является схожим с разложением типа Эджворта–Чебышёва из леммы 1. Поэтому, подставляя (2) в разложение (3) согласно обозначениям (4), находим выражение для b_1 и приходим к искомому разложению типа Корниша–Фишера.

Для упрощения вывода введем дополнительное обозначение, полагая $\theta \equiv \mu_3 \sigma^3 / 6$. Имеем (ср. (1)):

$$F_n(x) = G_{2r}(u) + g_{2r}(u) \frac{b_1}{\sqrt{n}} + \frac{\theta}{\sqrt{r(n-1)+1}} + f_r(u) + \frac{\theta}{\sqrt{r(n-1)+1}} \frac{1}{\sqrt{n}} f_r'(u) b_1(u) + \mathcal{O}\left(\frac{1}{n}\right).$$

Заметим, что последнее слагаемое этого равенства является функцией порядка $\mathcal{O}(1/n)$, а следовательно, и $o(1/\sqrt{n})$.

Таким образом, учитывая условие $F_n(x) = G_{2r}(u)$, т.е. x и u суть квантили одного порядка, получаем следующее выражение для коэффициента $b_1(u)$:

$$b_1(u) = -\frac{\theta \sqrt{n}}{\sqrt{r(n-1)+1}} \frac{f_r(u)}{g_{2r}(u)} = -\frac{\theta}{\sqrt{r}} \frac{f_r(u)}{g_{2r}(u)} + \mathcal{O}\left(\frac{1}{n}\right).$$

Для упрощения получившегося выражения воспользуемся следующей леммой.

Лемма 2. Для функции $f_r(x)$, введенной выше, имеем:

$$f_r(x) = \frac{1}{\sqrt{2\pi}} \Gamma\left(r - \frac{1}{2}\right) [1 + rx^2] \left(1 + \frac{x^2}{2}\right)^{-(r+1/2)}.$$

Для доказательства леммы подставим функцию гамма-распределения $H_r(y)$ и плотность стандартного нормального распределения $\phi(x\sqrt{y})$ в выражение для функции $f_r(x)$:

$$\begin{aligned} f_r(x) &= \int_0^\infty \phi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} dH_r(y) = \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(x^2y)/2} \frac{1-x^2y}{\sqrt{y}} y^{r-1} e^{-y} dy = \\ &= \frac{1}{\sqrt{2\pi}} \left(\int_0^\infty e^{-y(1+x^2/2)} y^{(r-1/2)-1} dy + \right. \\ &\quad \left. + x^2 \int_0^\infty e^{-y(1+x^2/2)} y^{(r+1/2)-1} dy \right). \end{aligned}$$

Рассмотрим вспомогательный интеграл:

$$h(\alpha, p) \equiv \int_0^\infty t^{\alpha-1} e^{-pt} dt = \Gamma(\alpha) p^{-\alpha}.$$

Тогда $f_r(x)$ запишется в виде:

$$\begin{aligned} f_r(x) &= \frac{1}{\sqrt{2\pi}} \left(h\left(r - \frac{1}{2}, 1 + \frac{x^2}{2}\right) + \right. \\ &\quad \left. + x^2 h\left(r + \frac{1}{2}, 1 + \frac{x^2}{2}\right) \right) = \\ &= \frac{1}{\sqrt{2\pi}} \left(1 + \frac{x^2}{2}\right)^{-(r+1/2)} \left(\Gamma\left(r - \frac{1}{2}\right) \left(1 + \frac{x^2}{2}\right) + \right. \\ &\quad \left. + x^2 \Gamma\left(r + \frac{1}{2}\right) \right) = \\ &= \frac{1}{\sqrt{2\pi}} \left(1 + \frac{x^2}{2}\right)^{-(r+1/2)} \Gamma\left(r - \frac{1}{2}\right) (1 + rx^2). \end{aligned}$$

Воспользуемся результатом леммы 2 и получим утверждение теоремы.

Рассмотрим случай $r = 1$. Тогда объем выборки N_n имеет геометрическое распределение. Имеем: $\delta \in (0, 1/2)$; значит, $2/(1+2\delta) \in (1, 2)$. Следовательно, остаточный член в случае $r = 1$ имеет порядок $o(1/\sqrt{n})$. Найдем асимптотическое разложение для квантилей в этом случае, подставив значение параметра в итоговое выражение:

$$\begin{aligned} x_{r=1}(u) &= \\ &= u - \frac{\mu_3 \sigma^3 \Gamma(1)}{6\sqrt{n}} \frac{1+u^2}{1-1/2} \left[\frac{1+u^2/2}{1+u^2/2} \right]^{3/2} + \\ &+ o\left(\frac{1}{\sqrt{n}}\right) = u - \frac{\mu_3 \sigma^3}{3} (1+u^2) \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

3.3 Доказательство теоремы для распределения Лапласа как предельного

В работе [3] показано, что в условиях данной теоремы для функции распределения нормированной статистики $T_{N_n(s)}$ справедливо асимптотическое разложение вида:

$$\begin{aligned} \max_x \left| \mathbb{P}(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x) - \Lambda_{1/s}(x) - \right. \\ \left. - \frac{\mu_3 \sigma^3 l_s(x)}{6\sqrt{n}} \right| = O\left(\frac{1}{n^{1/2+\delta}}\right), \quad n \rightarrow \infty, \quad (5) \end{aligned}$$

где функции $\Lambda_{1/s}(x)$ и $l_s(x)$ определены следующим образом:

$$\Lambda_{1/s}(x) = \int_0^\infty \Phi(x\sqrt{y}) de^{-s/y},$$

$$l_s(x) = \int_0^\infty \phi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} de^{-s/y},$$

а $\Phi(x)$ и $\phi(x)$ — функция распределения и плотность стандартного нормального распределения соответственно.

Из формулы (5) вытекает следующее соотношение:

$$\mathbb{P}(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x) = \Lambda_{1/s}(x) + \frac{\mu_3\sigma^3 l_s(x)}{6\sqrt{n}} + O\left(\frac{1}{n^{1/2+\delta}}\right), \quad n \rightarrow \infty. \quad (6)$$

Формула (6) похожа на разложение типа Эджворда (формула (1)), однако не соответствует ему в точности.

Введем следующие обозначения:

$$\left. \begin{aligned} F_n(x) &\equiv \mathbb{P}(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x); \\ G(x) &\equiv \Lambda_{1/s}(x); \\ \gamma &\equiv \frac{\mu_3\sigma^3 l_s(x)}{6}. \end{aligned} \right\} \quad (7)$$

Используем представление из [9]

$$\int_0^\infty x^{-n-1/2} e^{-px-q/x} dx = (-1)^n \sqrt{\frac{\pi}{p}} \frac{\partial^n}{\partial q^n} e^{-2\sqrt{pq}},$$

$$p > 0, \quad q > 0, \quad (8)$$

и рассмотрим $l_s(x)$:

$$l_s(x) = \int_0^\infty \phi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} de^{-s/y} =$$

$$= \frac{s}{\sqrt{2\pi}} \left(\int_0^\infty y^{-5/2} e^{-(x^2y)/2-s/y} dy - x^2 \int_0^\infty y^{-3/2} e^{-(x^2y)/2-s/y} dy \right) = l_s^*(x).$$

Возникают два случая.

1. Случай, когда $x \neq 0$:

$$l_s^*(x) = \frac{2\pi s}{\sqrt{2\pi x^2}} \left(\frac{x^2 e^{-\sqrt{2sx^2}}}{2s} + \frac{x^4 e^{-\sqrt{2sx^2}}}{2\sqrt{2}(sx^2)^{3/2}} - x^2 \frac{x^2 e^{-\sqrt{2sx^2}}}{\sqrt{2sx^2}} \right) = \lambda_{1/\sqrt{s}}(x) \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right).$$

2. Случай, когда $x = 0$:

$$l_s^*(x) = \frac{s}{\sqrt{2\pi}} \int_0^\infty y^{-5/2} e^{-s/y} dy =$$

$$= \frac{s}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2s^{3/2}} = \frac{1}{2\sqrt{2s}}.$$

Поскольку тождество (8) применимо, только когда $x > 0$, проверим на непрерывность полученный результат:

$$\lim_{x \rightarrow 0-0} \lambda_{1/\sqrt{s}}(x) \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) =$$

$$= \frac{1}{2\sqrt{2s}} = l_s(0-0);$$

$$\lim_{x \rightarrow 0+0} \lambda_{1/\sqrt{s}}(x) \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) =$$

$$= \frac{1}{2\sqrt{2s}} = l_s(0+0).$$

Таким образом, в силу (7) исходное приближение принимает следующий вид:

$$F_n(x) = G(x) + \frac{\gamma}{\sqrt{n}} \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) \lambda_{1/\sqrt{s}}(x) + O\left(\frac{1}{n^{1/2+\delta}}\right), \quad n \rightarrow \infty. \quad (9)$$

Поскольку

$$O\left(\frac{1}{n^{1/2+\delta}}\right) = o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty,$$

формула (9) переписывается в виде:

$$F_n(x) = G(x) + \frac{\gamma}{\sqrt{n}} \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) \lambda_{1/\sqrt{s}}(x) + o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.$$

Так как $s \in \mathbb{N}$, то $\lambda_{1/\sqrt{s}}(x) > 0$ для любого $x \in \mathbb{R}$. Тогда представим $\lambda_{1/\sqrt{s}}(x) > 0$ в виде:

$$\lambda_{1/\sqrt{s}}(x) = \frac{\lambda_{1/\sqrt{s}}(x)}{\lambda_{1/s}(x)} \lambda_{1/s}(x)$$

и подставим в предыдущую формулу:

$$F_n(x) = G(x) + \frac{\gamma}{\sqrt{n}} \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) \frac{\lambda_{1/\sqrt{s}}(x)}{\lambda_{1/s}(x)} \times$$

$$\times \lambda_{1/s}(x) + o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.$$

Очевидно, что $O(\varepsilon^2) = o(\varepsilon)$ при $\varepsilon \rightarrow 0$, поэтому полученная формула в точности совпадает с формулой (1), в которой

$$a(x) \equiv \gamma \left(\frac{|x|}{\sqrt{2s}} + \frac{1}{2s} - x^2 \right) \frac{\lambda_{1/\sqrt{s}}(x)}{\lambda_{1/s}(x)};$$

$$g(x) \equiv \lambda_{1/s}(x);$$

$$\varepsilon \equiv \frac{1}{\sqrt{n}}.$$

Следовательно, формула (2) переписывается в виде:

$$x(u) =$$

$$= u - \frac{\mu_3 \sigma^3}{6} \left(\frac{|u|}{\sqrt{2s}} + \frac{1}{2s} - u^2 \right) \frac{\lambda_{1/\sqrt{s}}(u)}{\lambda_{1/s}(u)} \frac{1}{\sqrt{n}} +$$

$$+ o\left(\frac{1}{\sqrt{n}}\right).$$

Теорема доказана.

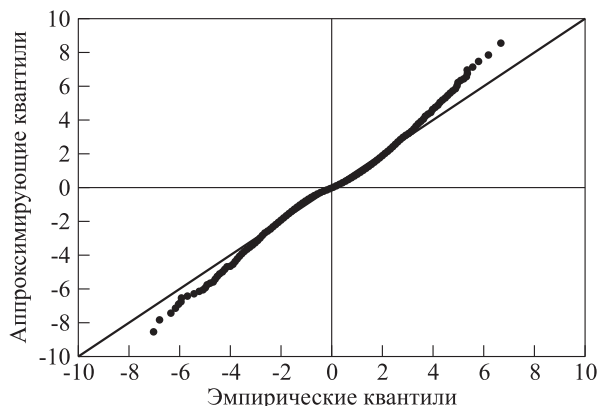
4 Вычислительный эксперимент

План эксперимента.

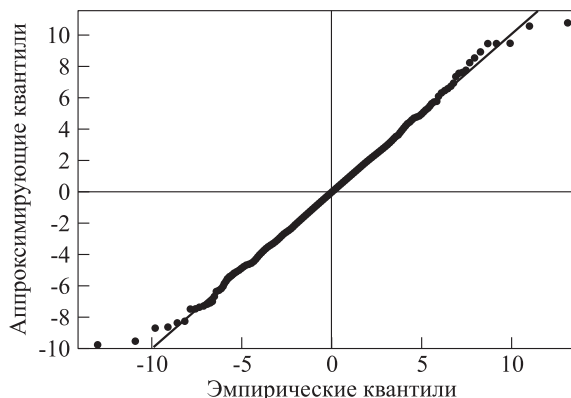
1. Задаются следующие параметры (сначала для случая, когда распределение Стьюдента является предельным, затем в скобках для случая предельного распределения Лапласа):
2. Производится расчет в следующей последовательности: для каждой точки эмпирической функции моделируется случайный объем N_n , далее моделируется вектор случайных величин X_1, \dots, X_{N_n} , затем рассчитывается значение статистики T_{N_n} , после чего она нормируется; отдельно для каждой точки аппроксимирующей функции моделируется значение соответствующей квантили распределения Стьюдента (Лапласа), значение самой функции рассчитывается согласно полученным выражениям в теоремах 1 и 2.
3. Результаты даются в виде таблицы и графиков на рисунке. В таблице используются следующие расстояния между эмпирической и аппроксимирующей функциями:

Значения метрик для приближения Лапласа и Стьюдента на разных интервалах

n	Приближение Лапласа		Приближение Стьюдента	
	$M_{\text{dif}}, 10,000$	$L_1, 10,000$	$M_{\text{dif}}, 1000$	$L_1, 1000$
0%–99,99%	1,93	0,15	3,00	0,06
0%–1%	1,51	0,65	2,42	0,83
0%–5%	1,51	0,28	2,42	0,29
1%–99%	0,26	0,14	1,16	0,05
5%–95%	0,17	0,14	0,21	0,04



(а)



(б)

Графики квантилей для приближений Лапласа (а) и Стьюдента (б)

$$M_{\text{dif}} : \rho(f, g) = \max_{1 \leq i \leq k} |f(x_i) - g(x_i)| ;$$

$$L_1 : \rho(f, g) = \frac{1}{k} \sum_{i=1}^k |f(x_i) - g(x_i)| .$$

Отметим, что для расчета даже одного эмпирического значения статистики необходимо для каждой из $k = 10\,000$ точек моделировать N_n случайных величин X_1, \dots, X_{N_n} , затем выполнить расчет статистики, что занимает несколько минут уже для $n = 1000$.

Таблица иллюстрирует известный факт, что разложения Корниша–Фишера дают хорошее приближение в центральной зоне, т.е. для вероятностей от 0,05 до 0,95. Качество приближения ухудшается для значений вероятности около 0 и 1. О том же говорит и рисунок.

На рисунке приведены графики квантилей для приближения Лапласа с параметрами $n = 10\,000$, $s = 1$ (а) и Стьюдента с параметрами $n = 1000$, $r = 1$ (б). По оси OX откладываются значения эмпирических квантилей, по оси OY — значения, даваемые разложениями Корниша–Фишера.

Литература

1. Королев В. Ю. Предельные распределения для случайно индексированных последовательностей и их применения. — М.: МГУ, 1993. 269 с.
2. Гнеденко Б. В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Тр. Тбилисского мат. ин-та, 1989. Т. 92. С. 146–150.
3. Бенинг В. Е., Королев В. Ю., Галиева Н. К. Асимптотические разложения для функций распределения статистик, построенных по выборкам случайного объема // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 75–83.
4. Cornish E. A., Fisher R. A. Moments and cumulants in the specification of distributions // Rev. Inst. Int. Statist., 1937. Vol. 4. P. 307–320.
5. Hill G. W., Davis A. W. Generalized asymptotic expansions of Cornish–Fisher type // Ann. Math. Stat., 1968. Vol. 39. P. 1264–1273.
6. Jaschke S. The Cornish–Fisher expansion in the context of delta-gamma-normal approximations // J. Risk, 2002. Vol. 4. No. 4. P. 33–52.
7. Ulyanov V. V., Aoshima M., Fujikoshi Y. Non-asymptotic results for Cornish–Fisher expansions. Technical Report Hiroshima Statistical Research Group. No. 16-03. — Hiroshima: Hiroshima University, 2016. 8 p.
8. Ulyanov V. V. Cornish–Fisher expansions // International encyclopedia of statistical science / Ed. M. Lovric. — Berlin: Springer, 2011. P. 312–315.
9. Прудников А. П., Брычков Ю. А., Маричев О. И. Интегралы и ряды. Элементарные функции. — М.: Наука, 1981. 344 с.

Поступила в редакцию 02.12.15

GENERALIZED CORNISH–FISHER EXPANSIONS FOR DISTRIBUTIONS OF STATISTICS BASED ON SAMPLES OF RANDOM SIZE

A. S. Markov, M. M. Monakhov, and V. V. Ulyanov

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

Abstract: Generalized Cornish–Fisher expansions are constructed for quantiles of sample mean for a sample of random size in terms of quantiles for the Laplace distribution and Student’s t -test. In recent years, the interest in Cornish–Fisher expansions grew significantly in the context of research on risk management. The widespread risk measure Value at Risk, or VaR, is, in fact, the quantile of the loss function. The authors use the general transfer theorem that makes it possible to obtain asymptotic expansions for the distribution functions of statistics based on samples of random size by asymptotic expansions for the distribution function of the random sample size and asymptotic expansions for the distribution functions of statistics based on nonrandom samples. A computational experiment was performed to illustrate the obtained Cornish–Fisher expansions.

Keywords: quantiles; generalized Cornish–Fisher expansions; random size sample; Laplace distribution

DOI: 10.14357/19922264160210

Acknowledgments

The work was supported by the Russian Science Foundation (project 14-11-00364).

References

1. Korolev, V. Yu. 1993. *Predel'nye raspredeleniya dlya sluchayno indeksirovannykh posledovatel'nostey i ikh primeneniya* [Limit theorems for randomly indexed sequences and its applications]. Moscow: MSU. 269 p.
2. Gnedenko, B. V. 1989. Ob otsenke neizvestnykh parametrov raspredeleniya pri sluchaynom chisle nezavisimykh nablyudeny [On estimation of unknown parameters of distributions from a random number of independent observations]. *Tr. Tbilisskogo mat. in-ta* [Proceedings of Tbilisi Mathematical Institute] 92:146–150.
3. Bening, V. E., V. Yu. Korolev, and N. K. Galieva. 2013. Asimptoticheskie razlozheniya dlya funktsiy raspredeleniya statistik, postroennykh po vyborkam sluchaynogo ob"ema [Asymptotic expansions for the distribution functions of statistics constructed from samples with random sizes]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):75–83.
4. Cornish, E. A., and R. A. Fisher. 1937. Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Stat.* 4:307–320.
5. Hill, G. W., and A. W. Davis. 1968. Generalized asymptotic expansions of Cornish–Fisher type. *Ann. Math. Stat.* 39:1264–1273.
6. Jaschke, S. 2002. The Cornish–Fisher expansion in the context of delta-gamma-normal approximations. *J. Risk* 4(4):33–52.
7. Ulyanov, V. V., M. Aoshima, and Y. Fujikoshi. 2016. Non-asymptotic results for Cornish–Fisher expansions. Technical Report Hiroshima Statistical Research Group. No. 16-03. Hiroshima: Hiroshima University. 8 p.
8. Ulyanov, V. V. 2011. Cornish–Fisher expansions. *International encyclopedia of statistical science*. Ed. M. Lovric. Berlin: Springer. 312–315.
9. Prudnikov, A. P., Yu. A. Brychkov, and O. I. Marichev. 1981. *Integraly i ryady. Elementarnye funktsii* [Integrals and series. Elementary functions]. Moscow: Nauka. 344 p.

Received December 2, 2015

Contributors

Markov Alexander S. (b. 1993) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; markov.cmc@yandex.ru

Monakhov Mikhail M. (b. 1993) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; mih_monah@mail.ru

Ulyanov Vladimir V. (b. 1953) — Doctor of Science in physics and mathematics, professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; vulyanov@cs.msu.ru

СИСТЕМА ОБСЛУЖИВАНИЯ С ГИПЕРЭКСПОНЕНЦИАЛЬНЫМ ВХОДЯЩИМ ПОТОКОМ И ПРОФИЛАКТИКАМИ ПРИБОРА*

В. Г. Ушаков¹

Аннотация: Изучена одноканальная система массового обслуживания с бесконечным числом мест для ожидания, гиперэкспоненциальным входящим потоком и профилактиками обслуживающего прибора при освобождении системы. Найдено нестационарное распределение числа требований в системе. Профилактики прибора заключаются в том, что в момент освобождения системы от требований прибор на случайное время с заданным распределением становится недоступным для обслуживания. Если за время профилактики поступает хотя бы одно требование, начинается нормальное функционирование системы. Если требования не поступают, то прибор отправляется на новую профилактику. Такие системы хорошо описывают функционирование большого числа реальных вычислительных и информационных систем. В частности, такие модели можно использовать при анализе систем, в которых наряду с основными требованиями имеются второстепенные. Второстепенные требования всегда присутствуют в системе, а их обслуживание может проводиться только тогда, когда нет основных, т. е. в фоновом режиме.

Ключевые слова: гиперэкспоненциальный поток; профилактики обслуживающего прибора; одноканальная система; длина очереди

DOI: 10.14357/19922264160211

1 Введение

Исследованию систем массового обслуживания, в которых при осуществлении определенных событий обслуживающий прибор становится на случайное время полностью или частично недоступным, уделяется в последнее время большое внимание. В работах [1–4] можно найти обзор известных результатов, большое число постановок задач, описание различных приложений и обширную библиографию.

Наиболее изученным и важным для приложений является случай, когда таким событием является освобождение системы от требований. В англоязычной литературе для такой особенности работы системы принят термин *vacation*. В литературе на русском языке общепринятого термина нет. Чаще всего используют один из следующих: каникулы, прогулки, профилактики обслуживающего прибора. Будем использовать последнее название.

Системы обслуживания с профилактиками прибора хорошо описывают функционирование большого числа реальных вычислительных и информационных систем. В частности, такие модели можно использовать при анализе систем, в которых наряду с основными требованиями имеются второстепенные. Второстепенные требования всегда

присутствуют в системе, а их обслуживание может проводиться только тогда, когда нет основных, т. е. в фоновом режиме.

В настоящей работе исследуется длина очереди в нестационарном режиме в однолинейной системе с ожиданием и гиперэкспоненциальным входящим потоком. С пуассоновским потоком аналогичная система обслуживания исследована в [2].

2 Описание модели

В однолинейную систему обслуживания с неограниченным числом мест для ожидания поступает гиперэкспоненциальный поток требований с функцией распределения интервалов между поступлениями вида:

$$A(t) = \sum_{i=1}^k c_i (1 - e^{-a_i t}), \quad t > 0, \quad a_i > 0,$$

$$c_i > 0, \quad \sum_{i=1}^k c_i = 1.$$

Гиперэкспоненциальный поток можно рассматривать как пуассоновский поток со случайной интенсивностью a , которая принимает k различных

* Работа выполнена при финансовой поддержке Российского научного фонда (проект 14-11-00397).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vgushakov@mail.ru

значений a_1, \dots, a_k с вероятностями c_1, \dots, c_k . Текущее значение a разыгрывается в момент поступления требования и не меняется между двумя соседними поступлениями. Введем случайный процесс $j(t)$ такой, что если $a = a_j$ в момент времени t , то $j(t) = j$.

Предполагаем, что длительности обслуживания требований — независимые в совокупности и не зависящие от входящего потока случайные величины с функцией распределения $B(x)$. Если в некоторый момент времени система освободилась от требований, то обслуживающий прибор отправляется на профилактику, которая длится случайное время с функцией распределения $C(x)$. Не ограничивая общности, будем считать, что $B(x) < 1$, $C(x) < 1$ для любого x и существуют плотности распределения $b(x)$ и $c(x)$ (окончательные результаты от этих предположений не зависят, но их доказательства становятся технически более простыми и менее громоздкими). Обозначим:

$$\begin{aligned} \beta(s) &= \int_0^{\infty} e^{-sx} b(x) dx, \\ \gamma(s) &= \int_0^{\infty} e^{-sx} c(x) dx, \\ \psi(s) &= \int_0^{\infty} e^{-sx} dx \int_0^{\infty} \frac{c(u+x)d(u)}{1-C(u)} du. \end{aligned}$$

Пока прибор находится на профилактике, он не доступен для обслуживания. Если за время профилактики поступают требования, после ее завершения начинается их обслуживание. Если ни одного требования не поступает, то прибор отправляется на новую профилактику. Длительности различных профилактик являются независимыми случайными величинами и не зависят от входящего потока и времен обслуживания. Основным объектом изучения будет случайный процесс $L(t)$ — число требований в системе в момент времени t . При его исследовании потребуются свойства ряда функций, определяемых через параметры системы. Рассмотрим многочлен по μ степени k вида:

$$\prod_{i=1}^k (\mu + a_i) - z \sum_{j=1}^k c_j a_j \prod_{i \neq j} (\mu + a_i). \quad (1)$$

Занумеруем его корни $\mu_1(z), \dots, \mu_k(z)$ таким образом, чтобы они были непрерывными функциями z и $\mu_1(1) = 0$. Тогда $\operatorname{Re} \mu_j(z) < 0$, $|z| < 1$, $\mu_i(z) \neq \mu_j(z)$, $i \neq j$, $j = 1, \dots, k$. Обозначим

$$\alpha_m(z) = \prod_{j \neq m} (\mu_m(z) - \mu_j(z)).$$

В дальнейшем понадобится следующая лемма.

Лемма 1. При каждом m , $m = 1, \dots, k$, уравнение

$$z = \beta(s - \mu_m(z))$$

имеет в области $\operatorname{Re} s > 0$ единственное решение $z = z_m(s)$ такое, что $|z_m(s)| < 1$.

3 Распределение длины очереди

Для нахождения распределения случайного процесса $L(t)$ потребуются и другие процессы, связанные с функционированием системы обслуживания. Пусть $\nu(t) = 1$, если в момент времени t прибор занят обслуживанием требования, и $\nu(t) = 0$, если в момент времени t прибор находится на профилактике. Случайный процесс $x(t)$ определим следующим образом. Если $L(t) > 0$ и $\nu(t) = 1$, то $x(t)$ есть время, прошедшее с начала обслуживания требования, находящегося на приборе, до момента t . Если $\nu(t) = 0$, то $x(t)$ есть время, прошедшее с начала профилактики прибора до момента t .

Случайный процесс $(L(t), j(t), x(t), \nu(t))$ является однородным марковским процессом. Положим:

$$\begin{aligned} P_j(n, x, t) &= \\ &= \frac{\partial}{\partial x} \mathbf{P}(L(t) = n, j(t) = j, \nu(t) = 1, x(t) < x), \\ & \quad n > 0, x \geq 0, j = 1, \dots, k; \end{aligned}$$

$$\begin{aligned} Q_j(n, x, t) &= \\ &= \frac{\partial}{\partial x} \mathbf{P}(L(t) = n, j(t) = j, \nu(t) = 0, x(t) < x), \\ & \quad n \geq 0, x \geq 0, j = 1, \dots, k; \end{aligned}$$

$$\begin{aligned} \eta(x) &= \frac{b(x)}{1-B(x)}; \quad \varphi(x) = \frac{c(x)}{1-C(x)}; \\ \delta_{i,j} &= \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases} \end{aligned}$$

Функции $P_j(n, x, t)$ и $Q_j(n, x, t)$ удовлетворяют при $x > 0$ следующим системам дифференциальных уравнений:

$$\begin{aligned} \frac{\partial P_j(n, x, t)}{\partial t} + \frac{\partial P_j(n, x, t)}{\partial x} &= \\ &= -(a_j + \eta(x)) P_j(n, x, t) + \\ &+ (1 - \delta_{n,1}) c_j \sum_{l=1}^k a_l P_l(n-1, x, t); \quad (2) \end{aligned}$$

$$\begin{aligned} \frac{\partial Q_j(n, x, t)}{\partial t} + \frac{\partial Q_j(n, x, t)}{\partial x} = & \\ = -(a_j + \varphi(x)) Q_j(n, x, t) + & \\ + (1 - \delta_{n,0}) c_j \sum_{l=1}^k a_l Q_l(n-1, x, t) & \quad (3) \end{aligned}$$

и краевым условиям при $x = 0$:

$$\begin{aligned} Q_j(n, 0, t) = 0, \quad n > 0, \quad Q_j(0, 0, t) = & \\ = \int_0^\infty Q_j(0, x, t) \varphi(x) dx + \int_0^\infty P_j(1, x, t) \eta(x) dx; & \quad (4) \end{aligned}$$

$$\begin{aligned} P_j(n, 0, t) = \int_0^\infty P_j(n+1, x, t) \eta(x) dx + & \\ + \int_0^\infty Q_j(n, x, t) \varphi(x) dx. & \quad (5) \end{aligned}$$

Будем предполагать, что в начальный момент времени $t = 0$ система свободна от требований, а с начала профилактики прибора прошло случайное время с заданным распределением с плотностью $d(x)$. Таким образом,

$$\begin{aligned} P_j(n, x, 0) = 0, \quad n > 0; & \\ Q_j(n, x, 0) = \delta_{n,0} c_j d(x), \quad j = 1, \dots, k. & \end{aligned}$$

Положим

$$\begin{aligned} p_j(z, x, s) = \sum_{n=1}^\infty z^n \int_0^\infty e^{-st} P_j(n, x, t) dt; & \\ q_j(z, x, s) = \sum_{n=0}^\infty z^n \int_0^\infty e^{-st} Q_j(n, x, t) dt. & \end{aligned}$$

Тогда, учитывая начальные условия, из (2) и (3) получаем:

$$\begin{aligned} \frac{\partial p_j(z, x, s)}{\partial x} = -(s + a_j + \eta(x)) p_j(z, x, s) + & \\ + c_j z \sum_{l=1}^k a_l p_l(z, x, s); & \quad (6) \end{aligned}$$

$$\begin{aligned} \frac{\partial q_j(z, x, s)}{\partial x} = -(s + a_j + \varphi(x)) q_j(z, x, s) + & \\ + c_j z \sum_{l=1}^k a_l q_l(z, x, s) + c_j d(x). & \quad (7) \end{aligned}$$

Решения (6) и (7) имеют вид:

$$\begin{aligned} p_j(z, x, s) = & \\ = (1 - B(x)) c_j \sum_{m=1}^k \frac{\gamma^{(m)}(z, s)}{\mu_m(z) + a_j} e^{-(s - \mu_m(z))x}; & \quad (8) \end{aligned}$$

$$\begin{aligned} q_j(z, x, s) = (1 - C(x)) c_j \sum_{m=1}^k e^{-(s - \mu_m(z))x} \times & \\ \times \left(\delta^{(m)}(z, s) + \alpha_m^{-1}(z) \prod_{l=1}^k (\mu_m(z) + a_l) \times & \right. \\ \times \int_0^x e^{(s - \mu_m(z))u} \frac{d(u)}{1 - C(u)} du \Big) / (\mu_m(z) + a_j), & \quad (9) \end{aligned}$$

где функции $\gamma^{(m)}(z, s)$ и $\delta^{(m)}(z, s)$ являются произвольными функциями указанных переменных и определяются из краевых условий. Переходя в (4) и (5) к преобразованиям Лапласа по t и производящим функциям, получаем:

$$\begin{aligned} p_j(z, 0, s) = z^{-1} \int_0^\infty p_j(z, x, s) \eta(x) dx + & \\ + \int_0^\infty q_j(z, x, s) \varphi(x) dx - q_j(z, 0, s). & \quad (10) \end{aligned}$$

Заметим, что $q_j(z, 0, s)$ не зависит от z , т.е. $q_j(z, 0, s) = q_j(s)$. Подставляя (8) и (9) в (10), получаем:

$$\begin{aligned} \sum_{m=1}^k \left(\frac{\gamma^{(m)}(z, s) (1 - z^{-1} \beta(s - \mu_m(z)))}{\mu_m(z) + a_j} + & \right. \\ + \frac{\delta^{(m)}(z, s) (1 - \gamma(s - \mu_m(z)))}{\mu_m(z) + a_j} \Big) = & \\ = \sum_{m=1}^k \frac{\alpha_m^{-1}(z) \prod_{l=1}^k (\mu_m(z) + a_l) \psi(s - \mu_m(z))}{\mu_m(z) + a_j}, & \\ j = 1, \dots, k. & \end{aligned}$$

Отсюда

$$\begin{aligned} \gamma^{(m)}(z, s) (1 - z^{-1} \beta(s - \mu_m(z))) = & \\ = \delta^{(m)}(z, s) (\gamma(s - \mu_m(z)) - 1) + & \\ + \alpha_m^{-1}(z) \prod_{l=1}^k (\mu_m(z) + a_l) \psi(s - \mu_m(z)). & \quad (11) \end{aligned}$$

В силу леммы 1 левая часть (11) обращается в 0 при $z = z_m(s)$. Следовательно,

$$\begin{aligned} \delta^{(m)}(z_m(s), s) = \frac{\psi(s - \lambda_m(s))}{\alpha_m(z_m(s))(1 - \gamma(s - \lambda_m(s)))} \times & \\ \times \prod_{l=1}^k (\lambda_m(s) + a_l). & \quad (12) \end{aligned}$$

Здесь $\lambda_m(s) = \mu_m(z_m(s))$. Из (9) следует, что

$$q_j(s) = c_j \sum_{m=1}^k \frac{\delta^{(m)}(z, s)}{\mu_m(z) + a_j}, \quad j = 1, \dots, k.$$

Решая эту систему уравнений относительно $\delta^{(m)}(z, s)$, получаем:

$$\delta^{(m)}(z, s) = z \frac{\prod_{j=1}^k (\mu_m(z) + a_j)}{\alpha_m(z)} \sum_{l=1}^k \frac{a_l q_l(s)}{\mu_m(z) + a_l}. \quad (13)$$

Подставляя в (13) $z = z_m(s)$, имеем:

$$\begin{aligned} \delta^{(m)}(z_m(s), s) &= \\ &= z_m(s) \frac{\prod_{j=1}^k (\lambda_m(s) + a_j)}{\alpha_m(z_m(s))} \sum_{l=1}^k \frac{a_l q_l(s)}{\lambda_m(s) + a_l}. \quad (14) \end{aligned}$$

Сравнивая два представления (12) и (14) для $\delta^{(m)}(z_m(s), s)$, получаем систему уравнений для $q_l(s)$:

$$\sum_{l=1}^k \frac{a_l q_l(s)}{\lambda_m(s) + a_l} = \frac{\psi(s - \lambda_m(s))}{z_m(s)(1 - \gamma(s - \lambda_m(s)))}, \quad m = 1, \dots, k,$$

из которой находим:

$$\begin{aligned} q_l(s) &= c_l \prod_{j=1}^k (\lambda_l(s) + a_j) \times \\ &\times \sum_{m=1}^k \left(\frac{\psi(s - \lambda_m(s))}{(1 - \gamma(s - \lambda_m(s))) (\lambda_m(s) + a_l)} \times \right. \\ &\quad \left. \times \frac{1}{\prod_{n \neq m} (\lambda_m(s) - \lambda_n(s))} \right). \quad (15) \end{aligned}$$

Подставляя (15) в (13) и учитывая (1), получаем:

$$\begin{aligned} \delta^{(m)}(z, s) &= \\ &= \frac{z}{\alpha_m(z)} \sum_{j=1}^k \frac{\psi(s - \lambda_j(s)) \prod_{l=1}^k (\lambda_j(s) + a_l)}{z_j(s)(1 - \gamma(s - \lambda_j(s)))} \times \\ &\quad \times \prod_{\nu \neq j} \frac{\mu_m(z) - \lambda_\nu(s)}{\lambda_j(s) - \lambda_\nu(s)}. \quad (16) \end{aligned}$$

Таким образом, все величины, определяющие функции $p_j(z, x, s)$ и $q_j(z, x, s)$, найдены. Заметим, что преобразование Лапласа производящей функции числа требований в системе можно найти по формуле:

$$\begin{aligned} p(z, s) &= \int_0^\infty e^{-st} \mathbf{E} z^{L(t)} dt = \\ &= \sum_{j=1}^k \int_0^\infty (p_j(z, x, s) + q_j(z, x, s)) dx. \end{aligned}$$

Подставляя сюда (8) и (9), имеем:

$$\begin{aligned} p(z, s) &= \sum_{j=1}^k \sum_{m=1}^k \frac{c_j}{(\mu_m(z) + a_j)(s - \mu_m(z))} \times \\ &\times \left(\gamma^{(m)}(z, s)(1 - \beta(s - \mu_m(z))) + \right. \\ &\quad + \delta^{(m)}(z, s)(1 - \gamma(s - \mu_m(z))) + \\ &\quad \left. + \prod_{l=1}^k (\mu_m(z) + a_l) \alpha_m^{-1}(z) (1 - \psi(s - \mu_m(z))) \right). \end{aligned}$$

Далее, из (11) имеем:

$$\begin{aligned} \gamma^{(m)}(z, s) &= \frac{\delta^{(m)}(z, s)(\gamma(s - \mu_m(z)) - 1)}{1 - z^{-1}\beta(s - \mu_m(z))} + \\ &\quad + \frac{\prod_{l=1}^k (\mu_m(z) + a_l) \psi(s - \mu_m(z))}{\alpha_m(z) (1 - z^{-1}\beta(s - \mu_m(z)))}. \end{aligned}$$

Значит,

$$\begin{aligned} p(z, s) &= \sum_{j=1}^k \sum_{m=1}^k \frac{c_j}{(\mu_m(z) + a_j)(s - \mu_m(z))} \times \\ &\times \left(\frac{1 - \beta(s - \mu_m(z))}{1 - z^{-1}\beta(s - \mu_m(z))} \times \right. \\ &\quad \times \left(\delta^{(m)}(z, s)(\gamma(s - \mu_m(z)) - 1) + \right. \\ &\quad \left. \left. + \frac{\prod_{l=1}^k (\mu_m(z) + a_l) \psi(s - \mu_m(z))}{\alpha_m(z)} \right) \right) + \\ &\quad + \delta^{(m)}(z, s)(1 - \gamma(s - \mu_m(z))) + \end{aligned}$$

$$\begin{aligned}
 & + \prod_{l=1}^k (\mu_m(z) + a_l) \alpha_m^{-1}(z) (1 - \psi(s - \mu_m(z))) \Big) = \\
 & = \sum_{j=1}^k \sum_{m=1}^k \frac{c_j}{(\mu_m(z) + a_j)(s - \mu_m(z))} \left(\delta^{(m)}(z, s) \times \right. \\
 & \quad \times \frac{(z-1)(1 - \gamma(s - \mu_m(z)))\beta(s - \mu_m(z))}{z - \beta(s - \mu_m(z))} + \\
 & \quad \quad \quad \frac{\prod_{l=1}^k (\mu_m(z) + a_l)}{\alpha_m(z)} \times \\
 & \quad \left. \times \left(1 + \frac{(1-z)\beta(s - \mu_m(z))\psi(s - \mu_m(z))}{z - \beta(s - \mu_m(z))} \right) \right). \quad (17)
 \end{aligned}$$

Обозначим

$$\tau_m(z) = \sum_{j=1}^k \frac{c_j}{\mu_m(z) + a_j}, \quad m = 1, \dots, k.$$

Подставляя теперь (16) в (17), окончательно получаем:

$$\begin{aligned}
 p(z, s) & = \sum_{m=1}^k \frac{\tau_m(z)}{\alpha_m(z)(s - \mu_m(z))} \times \\
 & \times \left(\frac{z(z-1)(1 - \gamma(s - \mu_m(z)))\beta(s - \mu_m(z))}{z - \beta(s - \mu_m(z))} \times \right.
 \end{aligned}$$

Литература

1. *Doshi B. T.* Queueing systems with vacations — a survey // Queueing Syst., 1986. Vol. 1. P. 29–66.
2. *Takagi H.* Time-dependent analysis of M|G|1 vacation models with exhaustive service // Queueing Syst., 1990. Vol. 6. P. 369–390.
3. *Li J., Tian N., Zhang Z. G., Luh H. P.* Analysis of the M|G|1 queue with exponentially working vacations — a matrix analytic approach // Queueing Syst., 2009. Vol. 61. P. 139–166.
4. *Bouman N., Borst S. C., Boxma O. J., Leeuwaarden J. S. H.* Queues with random back-offs // Queueing Syst., 2014. Vol. 77. P. 33–74.

Поступила в редакцию 05.03.16

QUEUEING SYSTEM WITH WORKING VACATIONS AND HYPEREXPONENTIAL INPUT STREAM

V. G. Ushakov^{1,2}

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

Abstract: The time-dependent process in the single server vacation model with hyperexponential input stream is analyzed. The Laplace transform (with respect to an arbitrary point in time) of the joint distribution of server state, queue size, and elapsed time in that state is obtained. The author restricts themselves to a system with exhaustive service (the queue must be empty when the server starts a vacation). The queueing systems with vacations have been well studied because of their applications in modeling the computer networks, communication, and manufacturing systems. For example, in many digital systems, the processor is multiplexed among a number of jobs and, hence, is not available all the time to handle one job type. Besides such an application, theoretical interest in vacation models has arisen with respect to their relationship with polling models.

Keywords: hyperexponential input stream; working vacations; single server; queue length

DOI: 10.14357/19922264160211

Acknowledgments

This work was supported by the Russian Science Foundation (project 14-11-00397).

References

1. Doshi, B.T. 1986. Queueing systems with vacations — a survey. *Queueing Syst.* 1:29–66.
2. Takagi, H. 1990. Time-dependent analysis of $M|G|1$ vacation models with exhaustive service. *Queueing Syst.* 6:369–390.
3. Li, J., N. Tian, Z. G. Zhang, and H. P. Luh. 2009. Analysis of the $M|G|1$ queue with exponentially working vacations — a matrix analytic approach. *Queueing Syst.* 61:139–166.
4. Bouman, N., S. C. Borst, O. J. Boxma, and J. S. H. Leeuwaarden. 2014. Queues with random back-offs. *Queueing Syst.* 77:33–74.

Received March 5, 2016

Contributor

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

МНОГОМЕРНОЕ ДРОБНОЕ ДВИЖЕНИЕ ЛЕВИ И ЕГО ПРИЛОЖЕНИЯ*

Ю. С. Хохлов¹

Аннотация: С начала 1990-х гг. было проведено большое число эмпирических исследований трафика реальных телекоммуникационных систем. Было обнаружено, что он обладает рядом специфических свойств, отличающих его от обычного голосового трафика, а именно: он обладает свойствами самоподобия и долговременной зависимости и распределение величины нагрузки, поступающей от одного источника, имеет тяжелые хвосты. Были построены новые модели трафика, которые обладали указанными свойствами. Наиболее известные из них — дробное броуновское движение и α -устойчивое движение Леви. Но каждая из этих моделей обладает только частью из перечисленных выше свойств. Были предприняты попытки построить более сложные модели, являющиеся комбинацией этих двух, в частности, предложен некоторый вариант одномерного дробного движения Леви. В настоящей работе рассматривается многомерный аналог дробного движения Леви. Этот процесс представляет собой многомерное дробное броуновское движение со случайной заменой времени, в качестве которой рассматривается α -устойчивое движение Леви с односторонними устойчивыми распределениями. Изучены свойства этого процесса, показано, что он является самоподобным и имеет стационарные приращения. Показано также, что координаты одномерных сечений этого процесса имеют распределения, отличные от устойчивых. Но асимптотика хвостов этих распределений в точности такая же, как и у устойчивых распределений. Далее эта модель использована для анализа неоднородного трафика, и получена нижняя асимптотическая оценка для вероятности переполнения хотя бы одного буфера при условии, что все буферы большие. Возможны и другие приложения.

Ключевые слова: многомерное дробное броуновское движение; α -устойчивый субординатор; самоподобные процессы; вероятность переполнения буфера

DOI: 10.14357/19922264160212

1 Введение

Во многих прикладных задачах часто встречается ситуация, когда процесс риска является принципиально многомерным. Например, нагрузка на сервер может поступать от многих источников по нескольким каналам. При этом нагрузка по каждому каналу формируется таким образом, что эти потоки оказываются зависимыми. Страховые компании ведут свою деятельность по многим направлениям, каждое из которых формирует свой поток рисков, но они взаимосвязаны. При формировании инвестиционного портфеля используются различные типы ценных бумаг, доходности которых меняются со временем, при этом они зависят от некоторых общих факторов, определяющих состояние рынка. Все эти примеры ведут к многомерным процессам риска с зависимыми компонентами. Существует и множество других примеров, возникающих, например, в физических задачах.

Многочисленные эмпирические исследования показали, что процессы риска в описанных вы-

ше областях обладают двумя важными свойствами, а именно: свойством долговременной зависимости и тяжелыми хвостами распределений вероятностей. Обычно при управлении процессами риска используются усредненными характеристиками процессов, которые при сильном агрегировании быстро стабилизируются. Это сильно упрощает задачи управления, так как позволяет работать с практически неслучайными величинами. Но эмпирические исследования последних 20 лет показали, что современные процессы риска в широком диапазоне интервалов агрегирования не стабилизируются, а остаются (после некоторой нормировки) практически такими же. Подобное поведение случайного процесса называют его самоподобием.

Сказанное выше приводит к задаче построения многомерных моделей риска, которые обладают тремя важными свойствами: самоподобием, долговременной зависимостью и тяжелыми хвостами распределений.

В одномерном случае этой проблематике посвящено большое число работ. Обычно рассматривают

* Работа поддержана РФФ (проект 14-11-00364).

¹Московский государственный университет им. М. В. Ломоносова, yskhokhlov@yandex.ru

самоподобные процессы со стационарными приращениями.

Типичными примерами таких процессов служат дробное броуновское движение и устойчивое движение Леви. Первый процесс обладает свойством долговременной зависимости, но, как известно, гауссовское распределение имеет быстро убывающий хвост. Напротив, устойчивое движение Леви имеет независимые приращения, но устойчивые распределения (с показателем устойчивости меньше 2) есть типичный пример распределений с тяжелыми хвостами. В определенном смысле эти два процесса являются антиподами. В то же время и тот и другой класс процессов в качестве частного случая содержит обычное броуновское движение.

Хотелось бы найти такую модель, которая объединяла бы свойства этих двух. Было предложено несколько вариантов решения этой задачи. Одним из них является модель FARIMA, использующая понятие дробного сдвига. Похожий вариант для случая непрерывного времени используется в так называемой модели линейного дробного устойчивого движения. Существует также модель логарифмически-дробного устойчивого движения и некоторые другие. Наиболее близка к настоящему исследованию работа [1], которая использует подход Мандельброта при построении дробного броуновского движения с помощью некоторого стохастического интеграла по обычному броуновскому движению. Построенный авторами процесс является самоподобным и имеет устойчивые распределения. Но его использование в задачах моделирования вызывает некоторые затруднения в силу его сложной структуры. В работе [2] предложен новый вариант одномерного дробного движения Леви, который представляет собой подчиненный процесс, где управляемый процесс есть дробное броуновское движение, а управляющий процесс есть одностороннее устойчивое движение Леви с показателем устойчивости меньше 1. Эта модель была использована для получения нижней оценки вероятности переполнения большого буфера.

В настоящей работе рассматривается аналогичная модель в многомерном случае. В качестве исходного объекта берется многомерное дробное броуновское движение. Подробное исследование этого процесса проведено в работах [3, 4]. Далее рассматривается подчиненный процесс, для которого управляемый процесс есть многомерное дробное броуновское движение, а управляющий процесс (случайное время), как и раньше, есть одностороннее устойчивое движение Леви. Исследованы свойства такого многомерного процесса. Далее эта модель применяется в некоторой задаче из теории телетрафика.

2 Устойчивые распределения и процессы Леви

В предлагаемой модели важную роль играют устойчивые распределения и процессы Леви. Далее приводятся некоторые известные определения и рассматриваются необходимые свойства таких распределений и процессов.

Определение 1. Случайный процесс $Y = (Y(t), t \geq 0)$ со значениями в R^d называется процессом Леви, если

- (1) $Y(0) = 0$ п. н.;
- (2) Y имеет независимые приращения;
- (3) Y имеет стационарные приращения, т. е. для любых $t \geq 0, h > 0$ случайный вектор $Y(t+h) - Y(t)$ имеет распределение, не зависящее от t .

Очень часто из соображений регулярности требуют выполнения следующего свойства: с вероятностью единица все траектории Y должны быть непрерывными справа и иметь конечные пределы слева. Это не является дополнительным ограничением, так как всегда можно построить реализацию процесса Леви с таким свойством.

Хорошо известно, что все конечномерные распределения процесса Y однозначно определяются по распределению случайного вектора $Y(1)$, которое является безгранично делимым.

Одним из наиболее известных одномерных примеров процессов Леви служит процесс броуновского движения (или винеровский процесс).

Определение 2. Процесс Леви $B = (B(t), t \geq 0)$ со значениями в R^1 называется процессом броуновского движения (БМ — Brownian motion), если для любых $t \geq 0, h > 0$ приращение $B(t+h) - B(t)$ имеет гауссовское распределение с нулевым средним и дисперсией $\sigma^2 h$.

Если $\sigma^2 = 1$, то имеет место стандартное броуновское движение. Нетрудно показать, что

$$K(t, s) = \text{Cov}(Y(t), Y(s)) = \sigma^2 \min(t, s).$$

По определению броуновское движение имеет гауссовские распределения. В силу центральной предельной теоремы такие распределения получаются асимптотически для нормированных сумм независимых и одинаково распределенных случайных величин с конечной дисперсией. В случае бесконечных дисперсий приходим к понятию устойчивого распределения.

Определение 3. Говорят, что случайная величина Y имеет α -устойчивое распределение, если ее характеристическая функция имеет следующий вид:

$$\varphi(\omega) := E [e^{i\omega X}] = \exp\{i\mu\omega - \sigma|\omega|^\alpha [1 - i\beta \operatorname{sign}(\omega)\theta(\omega, \alpha)]\},$$

где $0 < \alpha \leq 2$; $\sigma \geq 0$; $-1 \leq \beta \leq 1$; $\mu \in R^1$;

$$\theta(\omega, \alpha) = \begin{cases} \tan\left(\frac{\alpha\pi}{2}\right) & \alpha \neq 1; \\ -\frac{2}{\pi} \ln|\omega| & \alpha = 1. \end{cases}$$

Параметр α называется *характеристическим показателем* и определяет скорость убывания хвостов распределения; σ и μ являются параметрами *масштаба* и *сдвига* соответственно; β называется *параметром асимметрии*. Если $\beta = 0$, то X имеет симметричное относительно μ распределение. Если $0 < \alpha < 1$, $\mu = 0$ и $\beta = 1$, то случайная величина X положительна с вероятностью 1. В дальнейшем будем говорить, что случайная величина Y имеет стандартное α -устойчивое распределение, если $\mu = 0$ и $\sigma = 1$.

α -устойчивое распределение является безгранично делимым. Оно порождает некоторый процесс Леви.

Определение 4. Случайный процесс $L_\alpha = (L_\alpha(t), t \geq 0)$ со значениями в R^1 называется α -устойчивым движением Леви, если это процесс Леви, для которого $L_\alpha(1)$ имеет заданное устойчивое распределение.

Если у распределения $L_\alpha(1)$ $0 < \alpha < 1$, $\beta = 1$, $\mu = 0$, то траектории процесса L_α являются положительными и неубывающими. Такой процесс называется *α -устойчивым субординатором*.

Если $\alpha = 2$, $\mu = 0$, то мы вновь возвращаемся к процессу броуновского движения B .

Между α -устойчивыми движениями Леви с различными α справедливо следующее соотношение.

Теорема 1. Если $(L_{\alpha_1}(t), t \geq 0)$, $0 < \alpha_1 \leq 2$, есть α_1 -устойчивое движение Леви с симметричными распределениями и $(L_{\alpha_2}(t), t \geq 0)$, $0 < \alpha_2 < 1$, есть α_2 -устойчивый субординатор, то случайный процесс $Y = (Y(t) := L_{\alpha_1}(L_{\alpha_2}(t)), t \geq 0)$ есть $\alpha_1\alpha_2$ -устойчивое движение Леви с симметричными распределениями.

Эта теорема есть прямое следствие следующего результата В. М. Золотарева [5, теорема 3.3.1.]:

Теорема 2. Если Y_1 имеет симметричное α_1 -устойчивое распределение, $0 < \alpha_1 \leq 2$, Y_2 имеет одностороннее α_2 -устойчивое распределение, $0 < \alpha_2 < 1$, то случайная величина $Y = Y_1 Y_2^{1/\alpha_1}$ имеет симметричное $\alpha_1\alpha_2$ -устойчивое распределение.

В частности, для $\alpha_1 = 2$ и $0 < \alpha_2 = \alpha/2 < 1$ получается следующий результат.

Теорема 3. Если $B = (B(t), t \geq 0)$ есть броуновское движение, $L_{\alpha/2} = (L_{\alpha/2}(t), t \geq 0)$ есть $\alpha/2$ -устойчивый субординатор, то $L_\alpha = (L_\alpha(t) := B(L_{\alpha/2}(t)), t \geq 0)$, $0 < \alpha < 2$, есть α -устойчивое движение Леви с симметричными распределениями.

Рассмотрим теперь многомерные аналоги приведенных выше определений и результатов.

Определение 5. Процесс Леви $B = (B(t), t \geq 0)$ со значениями в R^d называется многомерным процессом броуновского движения (МВМ — multivariate BM), если для любых $t \geq 0$, $h > 0$ приращение $B(t+h) - B(t)$ имеет гауссовское распределение с нулевым средним и матрицей ковариаций Σh , где Σ — некоторая положительно определенная матрица.

Пусть $Y = (Y_1, \dots, Y_d)$ есть случайный вектор.

Определение 6. Случайный вектор Y имеет многомерное α -устойчивое распределение с параметром $\alpha \in (0, 2]$, если его характеристическая функция имеет следующий вид: для любого $\omega \in R^d$

$$\varphi_Y(\omega) = E(\exp(i(Y, \omega))) = \exp(i(a, \omega) - I(\omega)),$$

где

- (1) если $\alpha = 2$, то $a \in R^d$ — вектор средних, $I(\omega) = (1/2)(\Sigma\omega, \omega)$, Σ — матрица ковариаций случайного вектора Y ;
- (2) если $0 < \alpha < 2$, то $a \in R^d$ и

$$I(\omega) = \int_{S^{d-1}} |(\omega, u)|^\alpha \theta_\alpha(\omega, u) \Gamma(du),$$

Γ — конечная мера на сфере S^{d-1} и

$$\theta_\alpha(\omega, u) = \begin{cases} 1 - i \tan\left(\frac{\alpha\pi}{2}\right) \operatorname{sign}(\omega, u), & \alpha \neq 1, \\ 1 + \frac{2}{\pi} \ln|(\omega, u)| \operatorname{sign}(\omega, u), & \alpha = 1. \end{cases}$$

Такое распределение является безгранично делимым и порождает некоторый процесс Леви $Y = (Y(t), t \geq 0)$ со значениями в R^d . Будем называть его многомерным α -устойчивым движением Леви. Для $\alpha = 2$ и $a = 0$ получаем многомерное броуновское движение.

Важным частным случаем являются так называемые многомерные эллиптически контурированные устойчивые распределения, которые кратко будем называть эллиптическими устойчивыми распределениями. Их характеристическая функция имеет вид:

$$\varphi_Y(\omega) := E(\exp(i(Y, \omega))) = \exp(i(a, \omega) - (\Sigma\omega, \omega)^{\alpha/2}),$$

где $a \in R^d$, Σ — некоторая положительно определенная матрица (см., например, [6]).

Справедлив следующий аналог теоремы 3.

Теорема 4. Если $B = (B(t), t \geq 0)$ есть многомерное броуновское движение с матрицей ковариаций Σ , $L_{\alpha/2} = (L_{\alpha/2}(t), t \geq 0)$ есть $\alpha/2$ -устойчивый субординатор, то $L_\alpha = (L_\alpha(t) := B(L_{\alpha/2}(t)), t \geq 0)$, $0 < \alpha < 2$, есть многомерное α -устойчивое движение Леви с эллиптически контурированными распределениями.

3 Самоподобные процессы

Определение 7. Вещественный случайный процесс $Y = (Y(t), t \geq 0)$ называется самоподобным с параметром Херста $H \geq 0$, если он удовлетворяет следующему условию:

$$Y(t) \stackrel{d}{=} c^{-H} Y(ct), \forall t \geq 0, \forall c > 0,$$

где $\stackrel{d}{=}$ обозначает равенство конечномерных распределений.

Двумя наиболее популярными примерами самоподобных процессов являются дробное броуновское движение (fBM — fractional BM) и α -устойчивое движение Леви.

Определение 8. Дробное броуновское движение с параметром Херста H есть гауссовский процесс $(B_H(t), t \geq 0)$ с нулевыми средними и ковариационной функцией

$$K_H(t, s) = \frac{1}{2} [|t|^{2H} + |s|^{2H} - |t - s|^{2H}].$$

При $H = 1/2$ возвращаемся к обычному броуновскому движению.

Определение α -устойчивого движения Леви было дано выше. Для него $H = 1/\alpha$.

Можно определить многомерный аналог самоподобного процесса.

Определение 9. Случайный процесс $Y = (Y(t) = (Y_1(t), \dots, Y_d(t)), t \in R^1)$ со значениями в R^d называется самоподобным с параметром Херста $H = (H_1, \dots, H_d) \in (0, \infty)^d$, если он удовлетворяет следующему условию:

$$Y(t) \stackrel{d}{=} c^{-H} Y(ct) = (c^{-H_1} Y(ct_1), \dots, c^{-H_d} Y(ct_d)), \forall t \geq 0, \forall c > 0.$$

Возможны и другие более общие определения. Дополнительную информацию об устойчивых и самоподобных процессах можно найти в книгах [7, 8].

4 Многомерное дробное броуновское движение

Основой модели, которая предлагается в данной статье, является многомерное дробное броуновское движение. Далее приводится только его определение. Более подробную информацию можно найти в [3, 4].

Определение 10. Многомерным дробным броуновским движением (MFBM — multivariate fBM) с параметром Херста $H \in (0, 1)^d$ называется d -мерный случайный процесс $Y = (Y_1(t), \dots, Y_d(t), t \in R^1)$, который обладает следующими свойствами:

- (1) Y есть гауссовский процесс;
- (2) Y является самоподобным с параметром Херста H ;
- (3) Y имеет однородные приращения.

В данной работе наиболее интересен случай, когда $1/2 < H_p < 1$ для всех $p : 1 \leq p \leq d$. В этой ситуации при дополнительном ограничении, что процесс Y является обратимым по времени, он имеет нулевые средние и следующие ковариационные функции:

$$E(Y_p(s)Y_q(t)) = \frac{\sigma_{pq}}{2} [|s|^{H_p+H_q} + |t|^{H_p+H_q} - |t - s|^{H_p+H_q}],$$

где $\Sigma = (\sigma_{pq})$ есть некоторая положительно определенная матрица. В частности,

$$D(Y_p(t)) = E(|Y_p(t)|^2) = \sigma_p^2 |t|^{2H_p}.$$

Далее многомерное дробное броуновское движение с параметром H будем обозначать, как и в одномерном случае, B_H .

5 Многомерное дробное движение Леви

В этом разделе предлагается некоторый вариант многомерного дробного движения Леви. Одномерный аналог этого процесса был определен в работе [2].

Пусть $(B_H(t), t \in R^1)$ есть многомерное дробное броуновское движение с параметром Херста H и матрицей ковариаций Σ ; $(L_\alpha^1(t), t \geq 0)$ и $(L_\alpha^2(t), t \geq 0)$ — стандартные α -устойчивые субординаторы, $0 < \alpha < 1$; процессы B_H , L_α^1 и L_α^2 независимы.

Определение 11. Многомерным дробным движением Леви называется случайный процесс $X = (X(t), t \in R^1)$ со значениями в R^d такой, что

$$X(t) := \begin{cases} B_H(L_\alpha^1(t)), & t \geq 0; \\ B_H(L_\alpha^2(-t)), & t < 0, \end{cases}$$

Покажем, что предложенный процесс обладает свойством самоподобия, а именно имеет место следующий результат.

Теорема 5. Построенный выше случайный процесс является самоподобным с параметром Херста H/α .

Доказательство. Процессы $(L_\alpha^k(t), t \geq 0)$, $k = 1, 2$, являются α -устойчивыми и самоподобными с параметром Херста $1/\alpha$. В силу этого для любого $c > 0$ имеем:

$$(L_\alpha^k(ct), t \geq 0) \stackrel{d}{=} (c^{1/\alpha} L_\alpha^k(t), t \geq 0).$$

Тогда

$$\begin{aligned} (X(ct), t \in R^1) &= B_H(\pm L_\alpha^k(c|t|), t \in R^1) \stackrel{d}{=} \\ &\stackrel{d}{=} (B_H(\pm c^{1/\alpha} L_\alpha^k(|t|), t \in R^1). \end{aligned}$$

Используя самоподобие процесса B_H для фиксированного $\tau = L_\alpha^k(|t|)$, имеем для любого $a > 0$

$$(B_H(a\tau), \tau \geq 0) \stackrel{d}{=} (a^H B_H(\tau), \tau \geq 0),$$

или

$$(B_H(\pm c^{1/\alpha} \tau), \tau \geq 0) \stackrel{d}{=} (c^{H/\alpha} B_H(\pm \tau), \tau \geq 0).$$

Применяя формулу полной вероятности, получаем нужный результат.

Используя последнюю теорему, можно получить следующий полезный для дальнейших приложений результат.

Следствие 1. Для любого $t > 0$

$$X(t) \stackrel{d}{=} ((L_\alpha^1(t))^{H_1} Y_1, \dots, (L_\alpha^1(t))^{H_d} Y_d),$$

где случайный вектор $Y = (Y_1, \dots, Y_d)$ имеет многомерное нормальное распределение со средним ноль и матрицей ковариаций Σ , причем $L_\alpha^1(t)$ и Y независимы.

Этот результат легко следует из определения многомерного дробного броуновского движения и независимости B_H и L_α .

Следствие 2. Если B есть многомерное дробное броуновское движение с матрицей ковариаций Σ , то $H_1 = \dots = H_d = 1/2$ и для любого $t > 0$ случайный вектор

$$X(t) \stackrel{d}{=} (L_\alpha^1(t))^{1/2} (Y_1, \dots, Y_d)$$

имеет многомерное α -устойчивое эллиптически контурированное распределение с матрицей ковариаций Σ .

Этот результат есть прямое следствие теоремы 4.

В случае дробного броуновского движения случайный вектор $X(t)$ имеет распределение, отличное от устойчивого, если все $H_p > 1/2$. Для доказательства достаточно рассмотреть одну из компонент этого вектора. Как отмечалось выше, случайная величина $(L_\alpha(t))^{1/2} B(1)$ имеет симметричное устойчивое распределение с характеристическим показателем $\alpha/2$. Но случайные величины $(L_\alpha(t))^{1/2} B(1)$ и $(L_\alpha(t))^{H_1} B(1)$ имеют разные распределения, если $H_1 \neq 1/2$.

Замечание 1. Параметр Херста H/α определенно выше процесса X может быть любым положительным числом. Будем предполагать, что $1/2 < H_k/\alpha < 1$, $k = 1, \dots, d$. В этом случае рассматриваемый процесс имеет конечные математические ожидания и обладает свойством долговременной зависимости.

Теорема 6. Определенный выше случайный процесс X имеет однородные по времени приращения.

Доказательство. Хорошо известно, что дробное броуновское движение B_H имеет однородные приращения. Более того, для любых $0 < t_1 < t_2$

$$B_H(t_2) - B_H(t_1) \stackrel{d}{=} B_H(t_2 - t_1).$$

Тогда для любых $t \geq 0$, $h > 0$ и фиксированных $L_\alpha^k(t+h) = t_2$, $L_\alpha^k(t) = t_1$ имеем:

$$B_H(L_\alpha^k(t+h)) - B_H(L_\alpha^k(t)) \stackrel{d}{=} B_H(L_\alpha^k(t+h) - L_\alpha^k(t)).$$

В силу формулы полной вероятности имеем то же самое и для случайных моментов времени. Процесс $L_\alpha^k(t)$ также имеет однородные приращения. В итоге получаем:

$$\begin{aligned} B_H(L_\alpha^k(t+h)) - B_H(L_\alpha^k(t)) &\stackrel{d}{=} \\ &\stackrel{d}{=} B_H(L_\alpha^k(t+h) - L_\alpha^k(t)) \stackrel{d}{=} B_H(L_\alpha^k(h)). \end{aligned}$$

6 Асимптотика хвостов распределений многомерного дробного движения Леви

В этом разделе изучается поведение хвостов одномерных сечений дробного движения Леви, построенного выше.

В силу свойства самоподобия достаточно рассмотреть только распределение случайного вектора $X(1)$. Рассмотрим случайную величину $Z_\alpha :=$

$:= L_\alpha^1$. Без ограничения общности можно считать, что она имеет стандартное одностороннее α -устойчивое распределение. В силу следствия 2 имеем:

$$X(1) \stackrel{d}{=} ((Z_\alpha)^{H_1} Y_1, \dots, (Z_\alpha)^{H_d} Y_d),$$

где случайный вектор $Y = (Y_1, \dots, Y_d)$ имеет многомерное нормальное распределение со средним ноль и матрицей ковариаций Σ , причем Z_α и Y независимы. Покажем, что этот вектор имеет распределения, отличные от устойчивых. Для этого достаточно проверить, что это верно для отдельно взятой координаты.

Предположим противное: первая координата $V := (Z_\alpha)^{H_1} Y_1$ имеет симметричное устойчивое распределение с показателем $0 < \alpha_1 < 2$. Тогда, как показано выше, существует такая независимая от Y_1 случайная величина S , которая имеет одностороннее устойчивое распределение с показателем $\alpha_1/2$, что

$$V \stackrel{d}{=} S^{1/2} Y_1.$$

Отсюда получаем:

$$(Z_\alpha)^{H_1} Y_1 \stackrel{d}{=} S^{1/2} Y_1.$$

Нормальное распределение обладает свойством идентифицируемости для масштабных смесей. В силу этого получаем

$$(Z_\alpha)^{H_1} \stackrel{d}{=} S^{1/2},$$

или

$$(Z_\alpha)^{2H_1} \stackrel{d}{=} S.$$

Но, как хорошо известно, если случайная величина Z_α имеет устойчивое распределение, то случайная величина S имеет распределение, отличное от устойчивого. Это противоречие доказывает нужное утверждение.

Тем не менее можно показать, что хвосты распределений координат вектора $X(1)$ ведут себя в точности так же, как хвосты устойчивых распределений. Для доказательства этого свойства потребуется результат, известный как теорема Бреймана [9].

Теорема 7. Пусть X и Y есть независимые неотрицательные случайные величины и

$$\bar{F}(x) := P(X > x) = x^{-\alpha} L(x), \quad x \rightarrow \infty,$$

где $\alpha > 0$, L — медленно меняющаяся функция, $E(Y^{\alpha+\varepsilon}) < \infty$ для некоторого $\varepsilon > 0$. Тогда при больших $x > 0$

$$\bar{H}(x) := P(XY > x) \sim E(Y^\alpha) \bar{F}(x).$$

Применим этот результат к распределению k -й координаты вектора $X(1)$. Если Z_α имеет стандартное α -устойчивое распределение, то для больших $x > 0$

$$P(Z_\alpha > x) \sim C(\alpha) x^{-\alpha},$$

где

$$C(\alpha) = \frac{\sin(\pi\alpha)}{\pi} \Gamma(\alpha)$$

(см. [11, теорема 2.4.1]). Тогда, используя теорему Бреймана, для больших $x > 0$ получаем:

$$\begin{aligned} P((Z_\alpha)^{H_k} Y_k > x) &= \frac{1}{2} P((Z_\alpha)^{H_k} |Y_k| > x) = \\ &= \frac{1}{2} P(Z_\alpha |Y_k|^{1/H_k} > x^{1/H_k}) \sim \\ &\sim \frac{1}{2} E(|Y_k|^{\alpha/H_k}) P(Z_\alpha > x^{1/H_k}) \sim \\ &\sim \frac{1}{2} C(\alpha) E(|Y_k|^{\alpha/H_k}) x^{-\alpha/H_k} = \\ &= \frac{1}{2} C(\alpha) E(|Y_k|^{1/H_k}) x^{-1/H_k}. \end{aligned}$$

7 Приложение к моделированию телетрафика

В этом разделе построенный выше процесс применяется для моделирования динамики нагрузки сервера, поступающей по нескольким каналам, и делается нижняя оценка для вероятности переполнения хотя бы одного из буферов при больших размерах всех буферов.

Пусть имеется система массового обслуживания с одним сервером, на которую подается нагрузка по нескольким каналам, которые, вообще говоря, зависимы. Величина нагрузки, поступившей по k -му каналу, определяется по правилу:

$$A_k(t) := m_k t + (\sigma_k m_k)^{1/\beta_k} X_k(t), \quad k = 1, \dots, d.$$

Векторный случайный процесс $X(t) = (X_1(t), \dots, X_d(t))$ есть дробное движение Леви с параметром Херста $H^1 = (H_1^1, \dots, H_d^1) := H/\alpha = (H_1/\alpha, \dots, H_d/\alpha)$ и матрицей $\Sigma = (\sigma_{pq})$, определенное выше, $\sigma_k^2 = 1$, $\beta_k = \alpha/H_k$, $k = 1, \dots, d$. Если обслуживание пакетов осуществляется путем случайного выбора канала с некоторой вероятностью, то естественно предположить, что скорость обслуживания нагрузки из k -го канала постоянна и равна r_k . Чтобы обеспечить устойчивость функционирования системы, предположим, как обычно, что $r_k > m_k$, $k = 1, \dots, d$. Тогда величина

загрузки k -го буфера в момент времени $t \in R^1$ можно записать в виде:

$$Q_k(t, r_k) = \sup_{s \leq t} (A_k(t) - A_k(s) - r_k(t - s))_+.$$

В силу теоремы 6 процессы $Q_k = (Q_k(t, r), t \in R^1)$ являются стационарными. Пусть b_k есть размер k -го буфера.

Обозначим $b = (b_1, \dots, b_d)$. Найдем оценку для следующей вероятности переполнения хотя бы одного из буферов:

$$\begin{aligned} \varepsilon(b) &= P \left(\bigcup_{k=1}^d (Q_k(0, r) > b_k) \right) = \\ &= P \left(\bigcup_{k=1}^d \left(\sup_{\tau \geq 0} (A_k(\tau) - r_k \tau) > b_k \right) \right). \end{aligned}$$

Далее получим нижнюю границу для этой вероятности при больших b_k , используя технику, развитую в работах [1, 10]. Аналогичный результат в одномерном случае был получен ранее в работе [2].

Используя определение входящего потока и свойство самоподобия, получаем:

$$\begin{aligned} \varepsilon(b) &= P \left(\bigcup_{k=1}^d \left(\sup_{\tau \geq 0} (A_k(\tau) - r_k \tau) > b_k \right) \right) \geq \\ &\geq \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d ((A_k(\tau) - r_k \tau) > b_k) \right) = \\ &= \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d (m_k \tau + (\sigma_k m_k)^{1/\beta_k} X_k(\tau) - r_k \tau > b_k) \right) = \\ &= \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d (m_k \tau + (\sigma_k m_k \tau)^{1/\beta_k} X_k(1) - r_k \tau > b_k) \right) = \\ &> b_k) = \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d \left(X_k(1) > \frac{b_k + (r_k - m_k)\tau}{(\sigma_k m_k \tau)^{1/\beta_k}} \right) \right). \end{aligned}$$

Обозначим

$$f_k(\tau) = \frac{b_k + (r_k - m_k)\tau}{(\sigma_k m_k \tau)^{1/\beta_k}}.$$

Используя следствие 1, получаем:

$$\begin{aligned} \varepsilon(b) &\geq \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d ((L_\alpha^1(1))^{H_k} Y_k > f_k(\tau)) \right) \geq \\ &\geq \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d ((L_\alpha^1(1))^{H_k} Y_k > f_k(\tau), \right. \end{aligned}$$

$$\begin{aligned} &\left. \bigcap_{k=1}^d (Y_k > 1) \right) \geq \\ &\geq \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d ((L_\alpha^1(1))^{H_k} > f_k(\tau), \right. \\ &\left. \bigcap_{k=1}^d (Y_k > 1) \right) \geq \\ &\geq \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d ((L_\alpha^1(1))^{H_k} > f_k(\tau)) \right) \times \\ &\times P \left(\bigcap_{k=1}^d (Y_k > 1) \right) = \\ &= C(\Sigma) \sup_{\tau \geq 0} P \left(\bigcup_{k=1}^d (L_\alpha^1(1) > [f_k(\tau)]^{1/H_k}) \right) = \\ &= C(\Sigma) \sup_{\tau \geq 0} P \left(L_\alpha^1(1) > \min_k [f_k(\tau)]^{1/H_k} \right). \end{aligned}$$

Последняя вероятность есть невозрастающая функция от $f_k(\tau)$, которая в точке

$$\tau_k = \frac{b_k H_k^1}{(1 - H_k^1)(r_k - m_k)}$$

принимает наименьшее значение, равное

$$d_k := f_k(\tau_k) \frac{(r_k - m_k)^{H_k^1} (1 - H_k^1)^{-(1-H_k^1)}}{(\sigma_k m_k H_k^1)^{H_k^1}} b_k^{1-H_k^1}.$$

Так как

$$\begin{aligned} \sup_{\tau \geq 0} \min_k [f_k(\tau)]^{1/H_k} &\leq \min_k \sup_{\tau \geq 0} [f_k(\tau)]^{1/H_k} = \\ &= \min_k [d_k]^{1/H_k}, \end{aligned}$$

то получаем

$$\varepsilon(b) \geq C(\Sigma) P \left(L_\alpha^1(1) > \min_k [d_k]^{1/H_k} \right).$$

Если $L_\alpha^1(1)$ имеет стандартное α -устойчивое распределение, то для больших $x > 0$

$$P(L_\alpha^1(1) > x) \sim C(\alpha) x^{-\alpha},$$

где

$$C(\alpha) = \frac{\sin(\pi\alpha)}{\pi} \Gamma(\alpha)$$

(см. [11, теорема 2.4.1].) Отсюда для больших b_k

$$\varepsilon(b) \geq C(\alpha, H, \Sigma) \max_k \left[\sigma_k \frac{m_k}{r_k - m_k} b_k^{-(1-H_k^1)/H_k^1} \right].$$

Сформулируем окончательный результат в виде следующей теоремы.

Теорема 8. *В рамках описанной выше модели асимптотическая нижняя граница для вероятности переполнения хотя бы одного из буферов имеет следующий вид:*

$$\varepsilon(b) \geq C(\alpha, H, \Sigma) \max_k \left[\sigma_k \frac{m_k}{r_k - m_k} b_k^{-(1-H_k^1)/H_k^1} \right],$$

если $b_k \rightarrow \infty, 1 \leq k \leq d$.

Литература

1. *Laskin N., Lambadaris I., Harmantzis F. C., Devetsikiotis M.* Fractional Levy motion and its application to network traffic modeling // *Computer Networks*, 2002. Vol. 40. P. 363–375.
2. *De Nikola C., Khokhlov Yu. S., Pagano M., Sidorova O. I.* Fractional Levy motion with dependent increments and its application to network traffic modeling // *Информатика и её применения*, 2012. Т. 6. Вып. 3. С. 58–63.
3. *Stoev S., Taqqu M.* How rich is the class of multifractional brownian motions? // *Stochastic Processes Their Applications*, 2006. Vol. 116. P. 200–221.
4. *Amblard P. O., Coeurjolly J. F., Lavancier F., Philippe A.* Basic properties of the multivariate fractional Brownian motion // *Bull. Society Mathematique de France, Seminaires et Congres*, 2012. Vol. 28. P. 65–87.
5. *Золотарев В. М.* Одномерные устойчивые распределения. — М.: Наука, 1983. 304 с.
6. *Nolan J. P.* Multivariate elliptically contoured stable distributions: Theory and estimation // *Comput. Stat.*, 2013. Vol. 28. No. 5. P. 2067–2089.
7. *Samorodnitsky G., Taqqu M. S.* Stable non-Gaussian random processes. — London: Chapman & Hall, 1994. 632 p.
8. *Embrechts P., Maejima M.* Selfsimilar process. — Princeton, NJ, USA: Princeton University Press, 2002. 111 p.
9. *Breiman L.* On some limit theorems similar to the arcsin law // *Теория вероятн. и ее примен.*, 1965. Т. 10. Вып. 2. С. 351–359.
10. *Norros I.* A storage model with self-similar input // *Queueing Syst.*, 1994. Vol. 16. P. 387–396.
11. *Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные величины. — М.: Наука, 1965. 524 с.

Поступила в редакцию 01.12.15

MULTIVARIATE FRACTIONAL LEVY MOTION AND ITS APPLICATIONS

Yu. S. Khokhlov

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

Abstract: Since the beginning of the 1990s, many empirical studies of real telecommunication systems traffic have been conducted. It was found that traffic has some specific properties, which are different from common voice traffic, namely, it has the properties of self-similarity and long-range dependence and the distribution of loading size from one source has heavy tails. Some new models have been constructed, where these features were captured. Brownian fractional motion and α -stable Levy motion are the well-known examples. But both of these models do not have all of the above properties. More complicated models have been proposed using some combination of these ones. In particular, the authors have proposed a variant of univariate fractional Levy motion. This paper considers a multivariate analog of fractional Levy motion. This process is multivariate fractional Brownian motion with random change of time, where random change of time is Levy motion with one-sided stable distributions. The properties of this process are investigated and it is proven that it is self-similar and has stationary increments. Next, it is shown that the coordinates of one-dimensional sections of this process have the distributions which are not stable. But asymptotic of tails for these distributions is the same as for the stable ones. This model is applied to analyze heterogeneous traffic and to get a lower asymptotic bound of the probability of overflow of at least one buffer. There are other possible applications.

Keywords: fractional Brownian motion; α -stable subordinator; self-similar processes; buffer overflow probability

DOI: 10.14357/19922264160212

Acknowledgments

The work was supported by the Russian Science Foundation (project 14-11-00364).

References

1. Laskin, N., I. Lambadaris, F. C. Harmantzis, and M. Devetsikiotis. 2002. Fractional Levy motion and its application to network traffic modeling. *Computer Networks* 40:363–375.
2. De Nikola, C., Y.S. Khokhlov, M. Pagano, and O. I. Sidorova. 2012. Fractional Levy motion with dependent increments and its application to network traffic modeling. *Informatika i ee Primeneniya — Inform. Appl.* 6(3):58–63.
3. Stoev, S., and M. Taqqu. 2006. How rich is the class of multifractional brownian motions? *Stochastic Processes Their Applications* 116:200–221.
4. Amblard, P.O., J.F. Coeurjolly, F. Lavancier, and A. Philippe. 2012. Basic properties of the multivariate fractional Brownian motion. *Bull. Society Mathematique de France, Seminaires et Congres.* 28:65–87.
5. Zolotarev, V. M. 1986. *One-dimensional stable distributions*. Translations of Mathematical Monographs. AMS. Vol. 65. 284 p.
6. Nolan, J. P. 2013. Multivariate elliptically contoured stable distributions: Theory and estimation. *Comput. Stat.* 28(5):2067–2089.
7. Samorodnitsky, G., and M. S. Taqqu. 1994. *Stable non-Gaussian random processes*. Chapman & Hall. 632 p.
8. Embrechts, P., and M. Maejima. 2002. *Selfsimilar process*. Princeton University Press. 111 p.
9. Breiman, L. 1965. On some limit theorems similar to the arc-sin law. *Teoriya Veroyatnostei i Primeneniya* [Theory of Probability and its Applications] 10(2):351–359.
10. Norros, I. 1994. A storage model with self-similar input. *Queueing Syst.* 16:387–396.
11. Ibragimov, I. A., and Yu. V. Linnik. 1971. *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen. 443 p.

Received December 1, 2015

Contributor

Khokhlov Yury S. (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, GSP-1, Moscow 119991, Russian Federation; yskhokhlov@yandex.ru

ИНТЕНСИВНОСТЬ ЦИТИРОВАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ В ИЗОБРЕТЕНИЯХ ПО ИНФОРМАЦИОННО-КОМПЬЮТЕРНЫМ ТЕХНОЛОГИЯМ, ПАТЕНТУЕМЫХ В РОССИИ ОТЕЧЕСТВЕННЫМИ И ЗАРУБЕЖНЫМИ ЗАЯВИТЕЛЯМИ*

В. А. Минин¹, И. М. Зацман², В. А. Хавансков³, С. К. Шубников⁴

Аннотация: Рассматриваются информационные взаимосвязи науки и технологий, а также методы индикаторного оценивания процессов переноса (трансфера) знаний из разных областей исследований в сферу технологического развития. Предлагаемые методы предназначены для определения значений индикатора интенсивности цитирования научных работ в описаниях изобретений, патентуемых в России отечественными и зарубежными заявителями. Подобный подход может использоваться для получения косвенных оценок инновационного потенциала направлений научных исследований (НИИ). Значения индикатора интенсивности вычислялись как в целом, так и с распределением по странам заявителей. Представлены результаты определения значений индикатора, для чего в качестве исходной информации использовались полнотекстовые описания изобретений по классу G06 Международной патентной классификации (МПК, англ. *International Patent Classification* — IPC) (Обработка данных; вычисления; счет), опубликованные Роспатентом в 2000–2012 гг. Использование информационных ресурсов Роспатента было обусловлено тем, что они представлены в электронном виде, т. е. доступны для автоматизированной обработки. В результате получены значения индикатора интенсивности цитирования (ИЦ) научных работ с разделением по отечественным, зарубежным и совместным изобретениям, запатентованным в РФ. Такая детализация позволила оценить активность международного технологического сотрудничества и совместного патентования изобретений по информационно-компьютерным технологиям (ИКТ) в России, а также определить тематику сотрудничества в этой области.

Ключевые слова: цитирование научных работ; интенсивность цитирования; взаимосвязи науки и технологий; информационные технологии; Международная патентная классификация; расчет значений индикатора интенсивности цитирования

DOI: 10.14357/19922264160213

1 Введение

Среди различных видов источников знаний, стимулирующих появление новых инженерных идей, научные публикации выделяются тем, что они открыты и доступны для использования [1]. Доля изобретений, в которых цитируются именно научные публикации, зависит от вида технологий. Тематически каждый вид технологий описывается, как правило, в виде списка рубрик МПК. Международная патентная классификация — иерархическая система патентной классификации, которая является средством для рубрицирования патентных до-

кументов (описаний изобретений, промышленных образцов, полезных моделей) единообразно в международном масштабе.

Наиболее часто для такого описания используются списки рубрик МПК из номенклатуры, разработанной Фраунгоферовским институтом системотехники и инновационных исследований (Fraunhofer Gesellschaft-Institute für Systemtechnik und Innovationsforschung — FhG-ISI).

Примеры списков рубрик МПК из номенклатуры FhG-ISI приведены в табл. 1 для пяти видов технологий. Последний столбец содержит значения индикатора интегральной ИЦ (ИИЦ) результатов

* Работа выполнена при финансовой поддержке РФФИ (проект 16-07-00075).

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

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, iz_ipi@i170.ipi.ac.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, havanskov@i170.ipi.ac.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sergeysh50@yandex.ru

Таблица 1 Коды МПК технологий и значения индикатора ИИЦ

Вид технологий	Коды МПК вида технологий	ИИЦ
Биотехнологии	C07G; C12M, N, P, Q, R, S	138,43
Фармацевтические	A61K	83,71
Полупроводниковые	H01L	56,44
Оптические	G02; G03B, C, D, F, G, H; H01S	21,89
Информационные	G06; G11C; G10L	20,39

научных исследований, которые связаны с развитием технологий, указанных в первом столбце. Значения индикатора ИИЦ определены в табл. 1 как число цитируемых научных публикаций на 100 описаний изобретений [2].

Заметим, что индикатор ИИЦ не зависит от того или иного деления всей системы знаний на отрасли науки и научные направления, так как учитываются все цитируемые в изобретениях научные публикации по всем отраслям науки. Одна из наиболее актуальных задач в широком спектре исследований информационных взаимосвязей науки и технологий состоит в вычислении распределения интенсивности цитирования с учетом классификации публикаций по конкретным научным направлениям или дисциплинам.

Иначе говоря, кроме индикатора ИИЦ необходимо вычислить значения индикатора ИЦ для конкретных областей знаний, научных направлений или дисциплин. Значения индикатора ИЦ будут зависеть от выбора конкретной классификации отраслей знаний (Государственный рубрикатор научно-технической информации (ГРНТИ), рубрикатор Российского фонда фундаментальных исследований (РФФИ) и др.).

Исследования по индикаторному оцениванию информационных взаимосвязей науки и технологий проводятся за рубежом с конца прошлого века [3–12]. В России аналогичные работы появились в начале этого века и были выполнены в основном в Институте проблем информатики Российской академии наук (ИПИ РАН) [13–19]. Разработка методов вычисления значений индикатора ИЦ по конкретным научным направлениям выполнялась в более общем контексте проблематики информационного мониторинга в сфере науки [20–26].

В качестве индикатора ИЦ традиционно используется число цитирований научных публикаций, упоминаемых в описаниях изобретений. Данный индикатор характеризует уровень (частоту) использования научных результатов и тем самым степень воздействия фундаментальной науки на развитие технологической сферы.

Результаты оценивания процессов трансфера знаний, полученные в ИПИ РАН, дали возмож-

ность впервые в РФ вычислить значения индикатора ИЦ как косвенного показателя интенсивности взаимосвязи фундаментальной науки и ИКТ, к которым, в частности, относится класс G06 (Обработка данных; вычисления; счет). При этом для описания тематики научных статей, цитируемых в запатентованных изобретениях, использовался ГРНТИ. Вычисленные значения индикатора показали, что наиболее часто в изобретениях по ИКТ цитируются научные статьи по автоматике, вычислительной технике, кибернетике, электронике, радиотехнике, электротехнике и информатике. Таким образом, применение рубрик ГРНТИ отражает в основном прикладной аспект результатов, изложенных в научных публикациях, цитируемых в изобретениях по ИКТ. Было показано также, что для получения многоаспектной картины взаимосвязей отраслей науки с технологиями целесообразно также использовать фундаментальную рубрику отраслей знаний, например классификатор РФФИ [19].

Предметом настоящего исследования является описание и предварительный анализ результатов индикаторного оценивания процессов трансфера знаний из разных областей науки в сферу ИКТ. В качестве исходной информации использовались полнотекстовые описания изобретений по классу G06 МПК, опубликованные Роспатентом в 2000–2012 гг. В результате их автоматизированной обработки получены значения индикатора ИЦ отдельно для отечественных, зарубежных и совместных изобретений, запатентованных в РФ.

Чтобы определить интенсивность цитирования научных работ, были сопоставлены информационные ресурсы двух категорий — патентной и научно-технической:

- массив полнотекстовых описаний изобретений за определенный период времени, которые относятся к ИКТ;
- библиографические описания научных работ, которые цитируются в описаниях изобретений, входящих в первый массив.

Массив полнотекстовых описаний изобретений, на которые выданы патенты, является пуб-

личной информацией, размещенной на сайте Роспатента. Массив цитируемых научных работ был сформирован авторами статьи путем выделения библиографических описаний этих работ из полнотекстовых описаний запатентованных изобретений. В результате обработки и анализа второго массива были определены (по ГРНТИ) рубрики НИИ, к которым относятся научные результаты, излагаемые в публикациях.

Последующее сопоставление рубрик НИИ и индексов МПК позволило определить частотность взаимосвязей между ними, характеризующую интенсивность цитирования в изобретениях по ИКТ публикаций, относящихся к некоторой научной дисциплине, и, как следствие, косвенно оценить интенсивность переноса научных знаний в сферу изобретений и технологий.

Целью статьи является анализ экспериментальных данных, характеризующих интенсивность цитирования научных работ в описаниях изобретений с детализацией по отечественным и зарубежным изобретениям, запатентованным в РФ.

2 Сопоставление патентной активности отечественных и зарубежных заявителей

Использованная в работе патентная информация позволяет сравнить патентную активность в России отечественных и зарубежных заявителей. По данным, которые публикуются в ежегодных отчетах Роспатента, в 2000 г. доля патентов РФ, выданных иностранным заявителям по всем рубрикам МПК, составляла 18% от общего числа выданных патентов (всего было выдано 17 592 патента РФ). К 2014 г. эта доля возросла до 32% (всего было выдано 33 950 патентов) [27, 28]. Индексы МПК опубликованных патентов на изобретения дают возможность определить эту долю по любому виду технологий, патентуемых в России.

В рамках проекта РФФИ № 16-07-00075, первым результатам выполнения которого посвящена

данная статья, анализировались технологии, относящиеся к классу¹ G06 МПК «Обработка данных; вычисление; счет». С этой целью с серверов Роспатента была отобрана информация об изобретениях данного класса МПК, в том числе поля данных о странах заявителей и патентообладателей. Это дало возможность изучить распределение патентов, полученных в РФ, по странам заявителей.

Используемые подходы и методы позволяют кроме индикаторов ИИЦ и ИЦ расширить спектр применяемых индикаторов и вычислять их значения, исходя из потребностей наукометрических исследований в рамках перечня информационных полей, публикуемых Роспатентом, в том числе определять патентную активность в РФ изобретателей различных стран.

Всего в РФ за период с 2000 по 2012 г. Роспатент опубликовал сведения о 6665 патентах РФ на изобретения по классу G06 (учитывались основные и дополнительные индексы МПК, относящиеся к этому классу). Как видно из табл. 2, права на более чем 45% изобретений по G06 принадлежат патентообладателям только из России, более 54% — зарубежным патентообладателям только из одной страны, кроме РФ, и менее 1% — патентообладателям из двух и более стран. Учитывая такое распределение, в дальнейшем будем рассматривать только патенты с правообладателями исключительно из одной страны.

На рис. 1–3 представлено распределение патентов РФ на изобретения класса G06, опубликованных в РФ за период 2000–2012 гг., по годам с указанием прав на изобретение: РФ или другие страны.

Как следует из рис. 1, общее количество патентов РФ по классу G06 выросло за этот период более чем в три раза. Как видно из рис. 2, в последние годы растет не только число патентов класса G06 (см. рис. 1), но и доля их в общем числе патентов.

Как уже отмечалось, доля патентов РФ, выданных иностранным заявителям по всем рубрикам МПК, в 2014 г. составляла 32% [28], причем доля патентов РФ на изобретения по ИКТ, принадлежа-

Таблица 2 Принадлежность прав на изобретения класса G06, опубликованные в РФ за период 2000–2012 гг.

Принадлежность прав на изобретение	Количество патентов	Доля от общего числа патентов
Патентообладатели только из РФ	3012	45,19%
Патентообладатели из РФ и других стран (совместно)	28	0,42%
Патентообладатели только из одной страны (не РФ)	3611	54,18%
Патентообладатели из нескольких стран (не РФ)	14	0,21%

¹ Индексы МПК представляют собой многоуровневую иерархическую структуру и соответственно уровням разделены на разделы, классы, подклассы, группы и подгруппы.

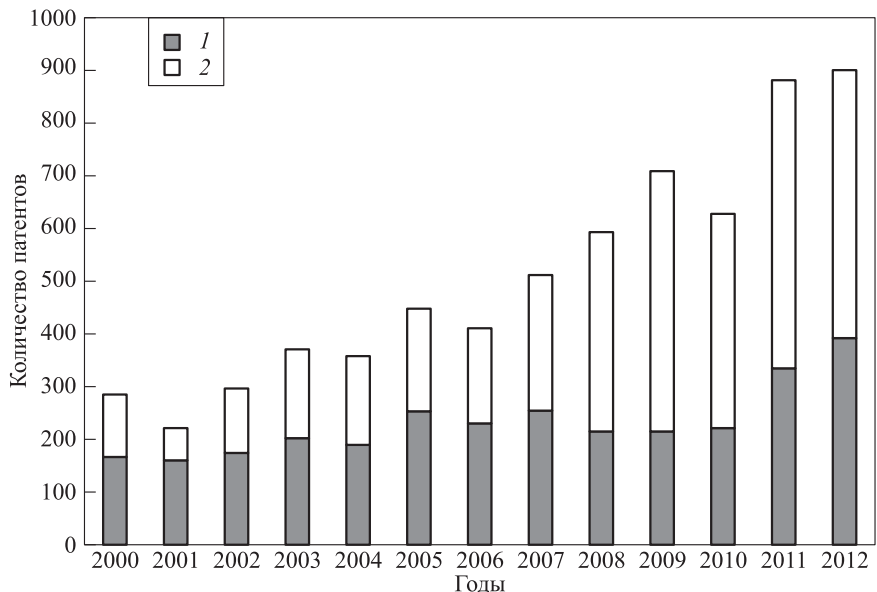


Рис. 1 Распределение по годам патентов РФ на изобретения класса G06, опубликованных за период 2000–2012 гг.: 1 — РФ; 2 — другие страны

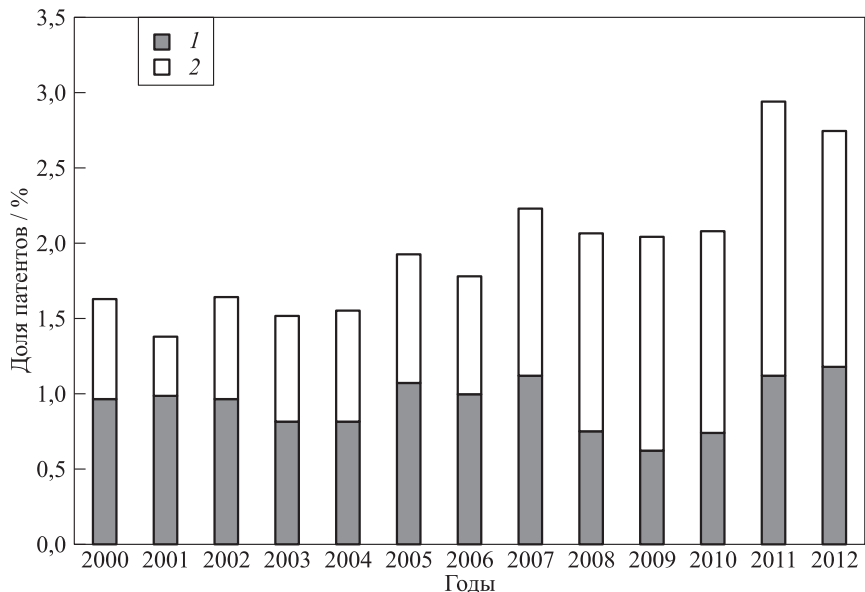


Рис. 2 Доля патентов класса G06 в общем числе опубликованных патентов РФ по годам: 1 — РФ; 2 — другие страны

щих зарубежным патентообладателям, существенно выше.

Как видно из рис. 3, эта доля выросла с 41% в 2000 г. до 57% в 2012 г. (максимум, почти в 70%, приходится на 2009 г.). В целом это говорит о значительном интересе, который проявляют иностранные компании и фирмы, получающие патенты РФ, к российскому рынку ИКТ. Этот интерес сохраняется, несмотря на то что начиная с 2010 г. доля

патентов РФ на изобретения по ИКТ, принадлежащих зарубежным патентообладателям, имеет тенденцию к снижению.

Наибольший интерес к патентованию в РФ проявляют США (1500 патентов, т.е. половина от 3000 патентов российских патентообладателей), Южная Корея, Германия, Япония, Франция и Нидерланды. На рис. 4 представлено распределение патентов РФ на изобретения класса G06 по

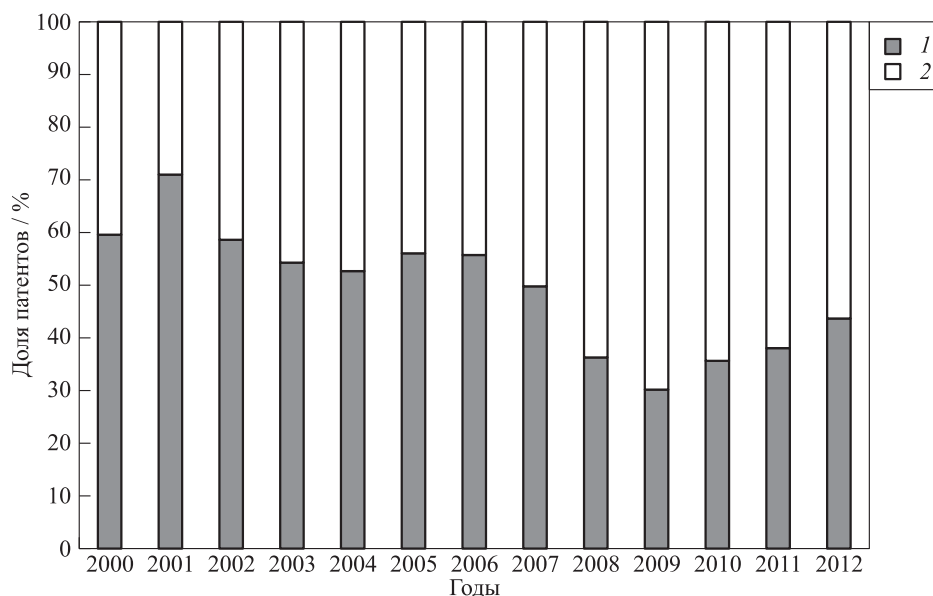


Рис. 3 Распределение патентов РФ на изобретения класса G06 по отечественным и зарубежным патентообладателям, а также по годам: 1 — РФ; 2 — другие страны

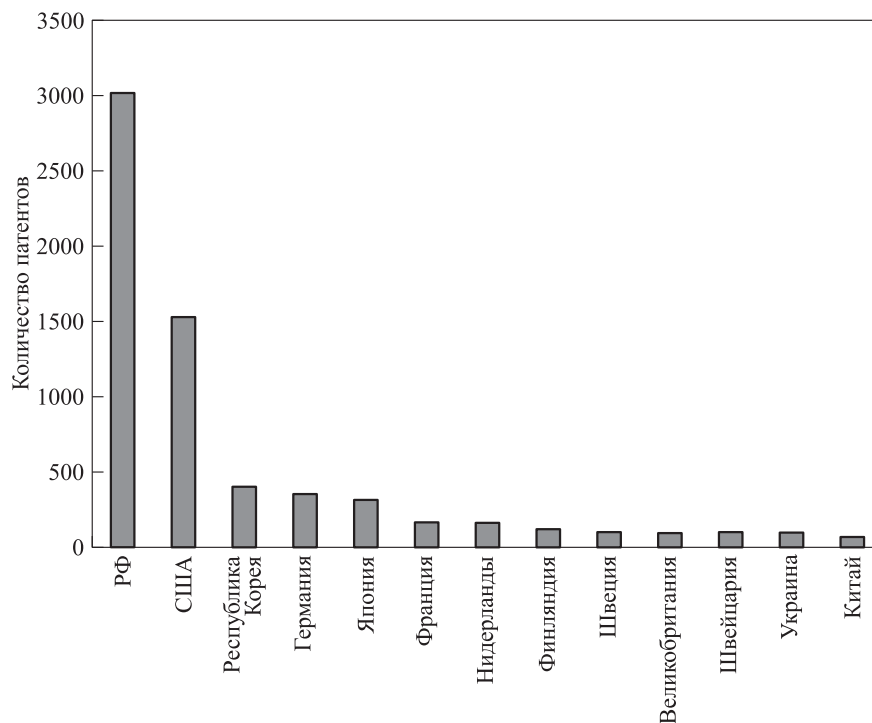


Рис. 4 Распределение патентов РФ на изобретения класса G06 по странам (для стран, у которых число патентов не меньше 50; без патентов с патентообладателями из разных стран)

странам. Для наглядности на рисунке представлены только те страны, которые имеют в РФ не менее 50 патентов.

На рис. 5 представлено распределение по подклассам МПК патентов, в описаниях которых упоминается класс G06.

Подклассы A61B, B42D и H04N, отображенные на рис. 5 (см. их названия в табл. 3) не относятся к классу G06, однако в описаниях обработанных изобретений, где эти три индекса — основные, присутствуют также дополнительные индексы, относящиеся к классу G06.

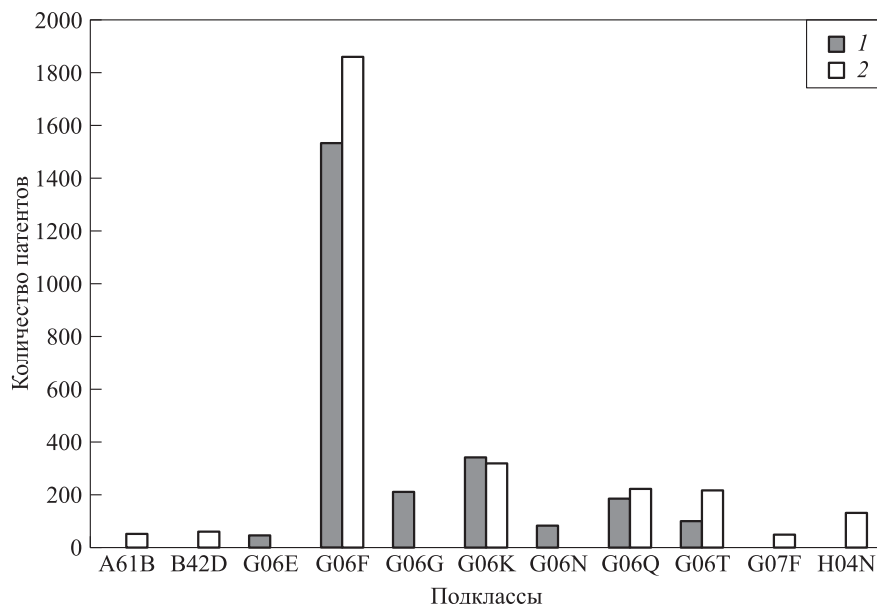


Рис. 5 Распределение патентов РФ на изобретения класса G06 по подклассам основного индекса МПК (расшифровка подклассов дана в табл. 3): 1 — РФ; 2 — другие страны

Интересно отметить, что российские патенто-обладатели указывают для изобретений по ИКТ в качестве основных индексы, относящиеся, как правило, к классу G06, в то время как у зарубежных патентообладателей доля индексов, относящихся к другим классам, выше (согласно проведенным расчетам — 16% и 25% соответственно). Это гово-

рит, скорее всего, о том, что зарубежные заявители чаще обозначают сферы прикладного использования патентуемых ими ИКТ с помощью индексов МПК.

Распределение патентов РФ на изобретения подкласса G06F по группам МПК представлено на рис. 6. Названия групп даны в табл. 4.

Таблица 3 Названия приведенных подклассов МПК

Подкласс МПК	Название
A61B	Диагностика; хирургия; опознание личности
B42D	Книги; книжные обложки; несброшюрованные листы
G06C	Механические цифровые вычислительные машины
G06D	Гидравлические и пневматические цифровые вычислительные устройства
G06E	Оптические вычислительные устройства
G06F	Обработка цифровых данных с помощью электрических устройств
G06G	Аналоговые вычислительные машины
G06J	Гибридные вычислительные устройства
G06K	Распознавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации
G06M	Счетчики; способы и устройства для подсчета предметов, не отнесенные к другим подклассам
G06N	Компьютерные системы, основанные на специфических вычислительных моделях
G06Q	Системы обработки данных или способы, специально предназначенные для административных, коммерческих, финансовых, управленческих, надзорных или прогностических целей
G06T	Обработка или генерация данных изображения
G07F	Монетные или подобные им автоматы
H04N	Передача изображений

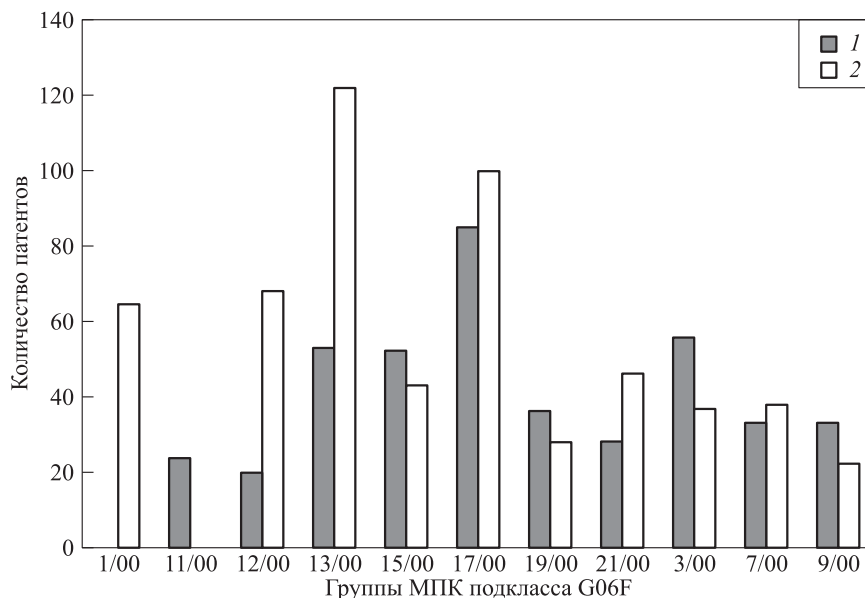


Рис. 6 Распределение патентов РФ на изобретения подкласса G06F по группам МПК (приведены группы подкласса G06F, содержащие более 20 патентов): 1 — РФ; 2 — другие страны

Таблица 4 Названия групп МПК, приведенных на рис. 6

Группа МПК	Название
G06F 1/00	Конструктивные элементы вычислительных машин и устройств для обработки данных
G06F 11/00	Обнаружение ошибок, исправление ошибок; контроль
G06F 12/00	Выборка, адресация или распределение данных в системах или архитектурах памяти
G06F 13/00	Соединение запоминающих устройств, устройств ввода-вывода или устройств центрального процессора
G06F 15/00	Цифровые компьютеры
G06F 17/00	Устройства или методы цифровых вычислений или обработки данных, специально предназначенные для специфических функций
G06F 19/00	Устройства или способы цифровых вычислений или обработки данных для специальных применений
G06F 21/00	Устройства защиты компьютеров или компьютерных систем от несанкционированной деятельности
G06F 3/00	Вводные устройства для передачи данных, подлежащих преобразованию в форму, пригодную для обработки в вычислительной машине; выводные устройства для передачи данных из устройств обработки в устройства вывода
G06F 7/00	Способы и устройства для обработки данных с воздействием на порядок их расположения или на содержание обрабатываемых данных
G06F 9/00	Устройства для программного управления

3 Интенсивность цитирования научных публикаций

В работе [19] были приведены значения индикатора ИЦ (в описаниях изобретений по классу G06) для научных работ, тематика первоисточников которых (журналов и трудов конференций) была описана с помощью рубрик ГРНТИ. Определение значений этого индикатора ранее было выполнено без детализации по странам патентообладателей.

Учет страны позволяет выявить различия в интенсивности цитирования научных работ в российских и иностранных изобретениях по ИКТ, на которые были выданы патенты РФ.

Таблицы 5 и 6 содержат данные о числе процитированных научных публикаций, соответствующих как рубрике ГРНТИ, так и подклассу МПК, что позволяет охарактеризовать интенсивность взаимосвязи конкретного научного направления (по ГРНТИ) и технологии (по подклассу МПК). По

Таблица 5 Число цитирований научных работ для изобретений с патентообладателями из РФ

Подкласс МПК	Код рубрики ГРНТИ							
	13.00.00	20.00.00	27.00.00	28.00.00	45.00.00	47.00.00	50.00.00	76.00.00
G06E	0	0	0	0	0	16	0	0
G06F	2	423	59	758	469	517	915	0
G06G	0	0	0	2	15	27	14	0
G06K	0	539	4	587	550	602	655	2
G06N	0	19	1	14	30	26	19	1
G06Q	0	0	0	8	0	0	1	5
G06T	0	8	51	51	8	17	80	0

Таблица 6 Число цитирований научных работ для изобретений патентообладателей из зарубежных стран

Подкласс МПК	Код рубрики ГРНТИ							
	13.00.00	20.00.00	27.00.00	28.00.00	45.00.00	47.00.00	50.00.00	76.00.00
G06E	0	0	0	0	0	0	0	0
G06F	0	0	45	46	9	9	54	0
G06G	0	0	0	0	0	0	0	0
G06K	0	0	31	13	28	36	65	0
G06N	0	1	1	1	0	0	1	0
G06Q	0	0	0	0	0	0	0	0
G06T	0	0	76	78	32	48	137	2

уровню использования научных результатов, что косвенно отображается числом цитирований научных работ, для технологий подклассов G06F и G06K важны практически все научные направления, указанные в табл. 5 и 6, причем по числу цитирований выделяются результаты по научным направлениям ГРНТИ: «Кибернетика» (28.00.00) «Электроника. Радиотехника» (47.00.00), «Автоматика. Вычислительная техника» (50.00.00). Первые строки этих таблиц содержат коды восьми рубрик ГРНТИ, которые использовались для описания тематики цитируемых научных работ. Названия рубрик даны в табл. 7.

Как видно из табл. 5, самой высокой цитируемостью научных работ характеризуются изобретения российских патентообладателей для технологий подкласса G06F (Обработка цифровых данных с помощью электрических устройств) и G06K (Распознавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации). Показательно, что на долю именно этих технологий приходится наибольшее число выданных патентов РФ, что говорит о более интенсивном развитии данных технологий в РФ.

В то же время, как следует из табл. 6, для изобретений зарубежных патентообладателей по интенсивности цитирования научных работ выделяется технология G06T (Обработка или генерация данных изображения), хотя число патентов в подклас-

се G06T (см. рис. 5) меньше, чем в G06F и G06K (около 10%).

В табл. 7 даны названия ряда рубрик ГРНТИ, для которых приведены данные о числе процитированных научных публикаций для подкласса МПК G06F. Как видно из таблицы, для большинства научных работ, процитированных в описаниях изобретений, указаны рубрики ГРНТИ самого верхнего уровня иерархии, что, конечно, снижает информативность выявленных связей и говорит о необходимости уточнения рубрик публикаций, цитируемых авторами изобретений.

4 Период патентного отклика на научные публикации

Временные аспекты информационных взаимосвязей науки и технологий характеризуются «периодом патентного отклика», т.е. промежутком времени, прошедшим с момента выхода научных публикаций до опубликования патентов, в описании изобретений которых цитируются данные публикации. Значения показателя «период патентного отклика» дают возможность оценить динамику востребованности результатов научных исследований в сфере разработки технологий.

В расчетах, представленных в работе [19], были определены значения этого показателя для научных публикаций, цитируемых в изобретениях по классу

Таблица 7 Рубрики ГРНТИ источников, публикации которых цитируются в изобретениях подкласса МПК G06F

Код ГРНТИ	Название рубрики ГРНТИ	РФ	Другие страны
13.00.00	КУЛЬТУРА. КУЛЬТУРОЛОГИЯ	2	
20.00.00	ИНФОРМАТИКА	423	
27.00.00	МАТЕМАТИКА	44	45
27.03.00	Математическая логика и основания математики	2	
27.35.00	Математические модели естественных наук и технических наук. Уравнения математической физики	2	
27.41.00	Вычислительная математика	11	
28.00.00	КИБЕРНЕТИКА	702	46
28.17.00	Теория моделирования	28	
28.19.00	Теория кибернетических систем управления	2	
28.23.00	Искусственный интеллект	26	
45.00.00	ЭЛЕКТРОТЕХНИКА	468	9
45.53.00	Электротехническое оборудование специального назначения	1	
47.00.00	ЭЛЕКТРОНИКА. РАДИОТЕХНИКА	459	9
47.01.00	Общие вопросы электроники и радиотехники	4	
47.03.00	Теоретические основы электронной техники	14	
47.05.00	Теоретическая радиотехника	2	
47.14.00	Проектирование и конструирование электронных приборов	2	
47.29.00	Электровacuумные и газоразрядные приборы и устройства	2	
47.33.00	Твердотельные приборы	2	
47.37.00	Голография	8	
47.41.00	Радиоэлектронные схемы	2	
47.43.00	Распространение радиоволн	2	
47.45.00	Антенны. Волноводы. Элементы СВЧ-техники	2	
47.47.00	Радиопередающие и радиоприемные устройства	2	
47.49.00	Радиотехнические системы зондирования, локации и навигации	2	
47.51.00	Телевизионная техника	2	
47.53.00	Запись и воспроизведение сигналов	2	
47.55.00	Электроакустика, ультразвуковая и инфразвуковая техника	2	
47.57.00	Инфракрасная техника	2	
47.59.00	Узлы, детали и элементы радиоэлектронной аппаратуры	2	
47.61.00	Приборы для радиотехнических измерений	2	
47.63.00	Системы и устройства отображения информации	2	
50.00.00	АВТОМАТИКА. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА	891	54
50.07.00	Теоретические основы вычислительной техники	4	
50.09.00	Элементы, узлы и устройства автоматики и вычислительной техники	2	
50.11.00	Запоминающие устройства	4	
50.41.00	Программное обеспечение вычислительных машин, комплексов и сетей	2	
50.45.00	Системы телеуправления и телеизмерения	2	
50.51.00	Автоматизация проектирования	4	
50.53.00	Автоматизация научных исследований	6	
59.00.00	ПРИБОРОСТРОЕНИЕ	78	
59.01.00	Общие вопросы приборостроения	2	
59.14.00	Проектирование и конструирование приборов	4	
59.41.00	Приборы для измерения оптических и светотехнических величин и характеристик	6	
59.45.00	Приборы неразрушающего контроля изделий и материалов	2	

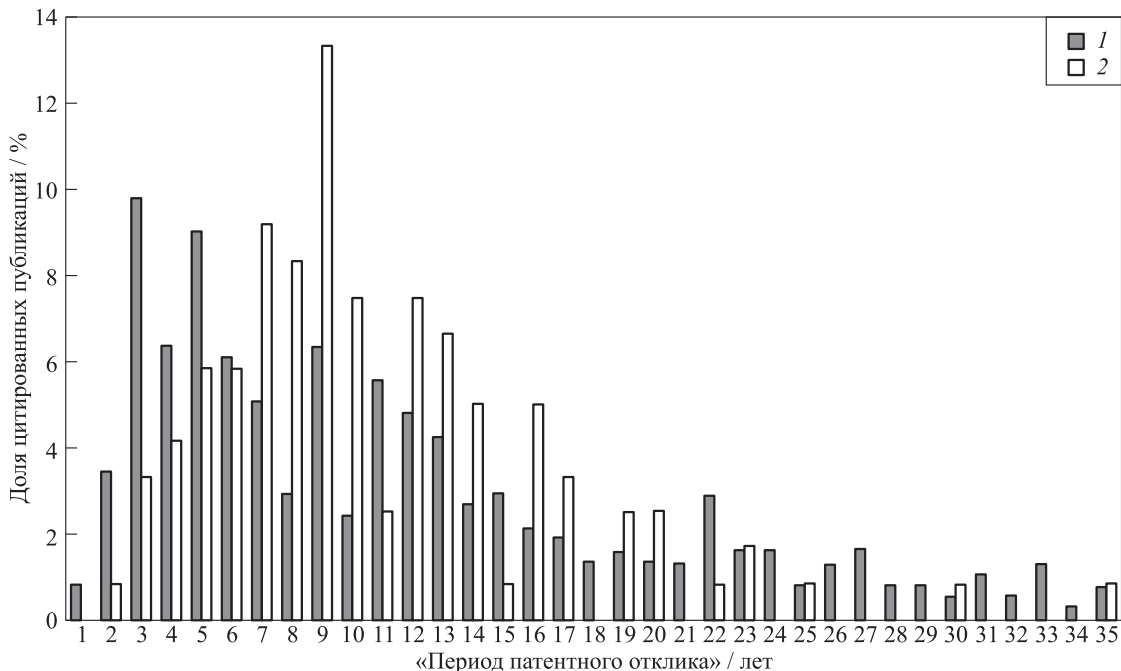


Рис. 7 Распределение «периодов патентного отклика» для класса G06 в качестве основного индекса: 1 — РФ; 2 — другие страны

G06. Расчеты, проведенные по полному массиву патентов класса G06 (в качестве основного индекса МПК), показывают, что максимум распределения значений равен 3 годам для патентообладателей из России и 9 годам — для зарубежных патентообладателей.

Рисунок 7 демонстрирует распределение «периодов патентного отклика» для класса G06 для массивов как российских, так и зарубежных патентообладателей (по оси абсцисс указана продолжительность «периодов патентного отклика» в годах; по оси ординат для каждого массива — доля от общего числа цитированных научных публикаций в данном массиве).

Были также получены распределения «периодов патентного отклика» для отдельных наиболее многочисленных по числу патентов подклассов G06, а именно: G06F, G06K, G06T (рис. 8 и 9).

Для подкласса G06F «Обработка цифровых данных с помощью электрических устройств» (см. рис. 8, а), распределения имеют максимумы в 6 и 9 лет соответственно.

Для подкласса G06K «Распознавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации» (см. рис. 8, б) распределения имеют по два максимума в 3 и 22 года и в 5 и 9 лет.

Для подкласса G06T «Обработка или генерация данных изображения. . .» (см. рис. 9), распределе-

ния имеют по два максимума в 5 и 9 лет и в 6 и 7 лет.

Отметим существенно разный характер полученных распределений для различных подклассов. Для выяснения причин различий в распределениях значений данного показателя требуется содержательный анализ изобретений и цитируемых в них научных публикаций, что выходит за рамки настоящей статьи.

5 Заключение

В ходе работ по оцениванию процессов трансфера знаний разработаны методы и технологии анализа активности патентования в РФ с использованием индикаторов, которые учитывают страну патентообладателя. Вычислены значения индикаторов ИЦ, которые демонстрируют значительные различия в интенсивности цитирования научных публикаций российскими и зарубежными авторами. Экспериментально показано, что интенсивность цитирования зависит не только от вида технологий (см. табл. 1), но и от страны патентообладателя (см. табл. 5 и 6).

Например, самая высокая цитируемость научных работ отмечена в описаниях изобретений российских патентообладателей для технологий подкласса G06F (Обработка цифровых данных с помощью электрических устройств) и G06K (Рас-

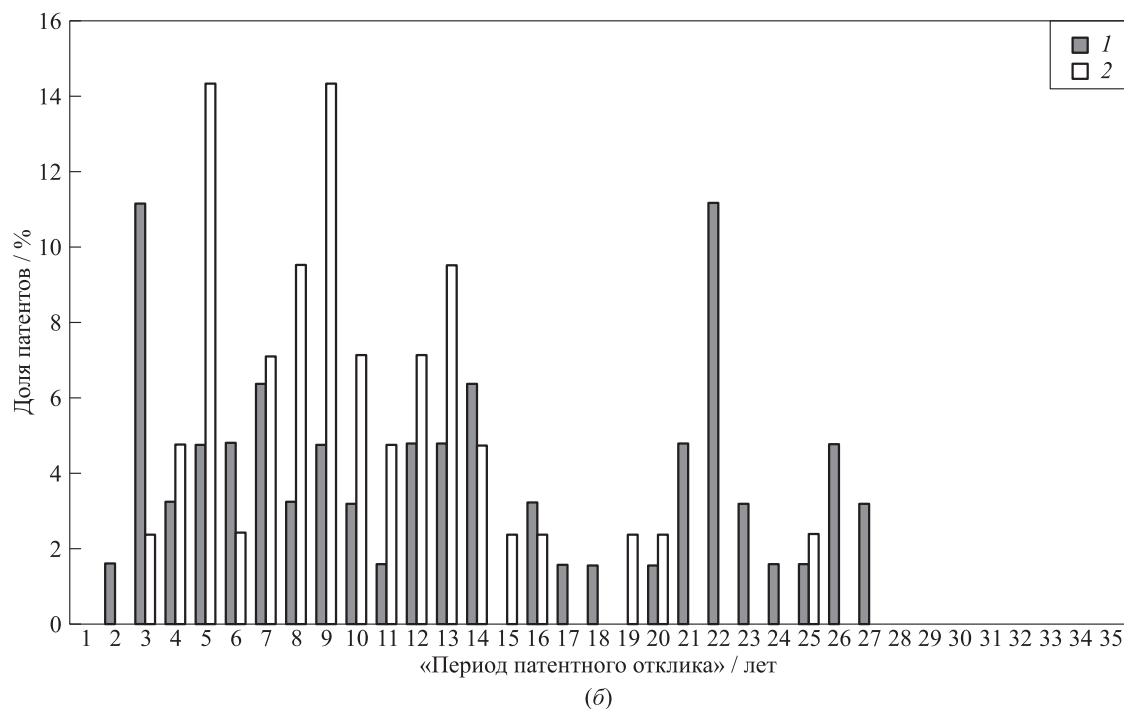
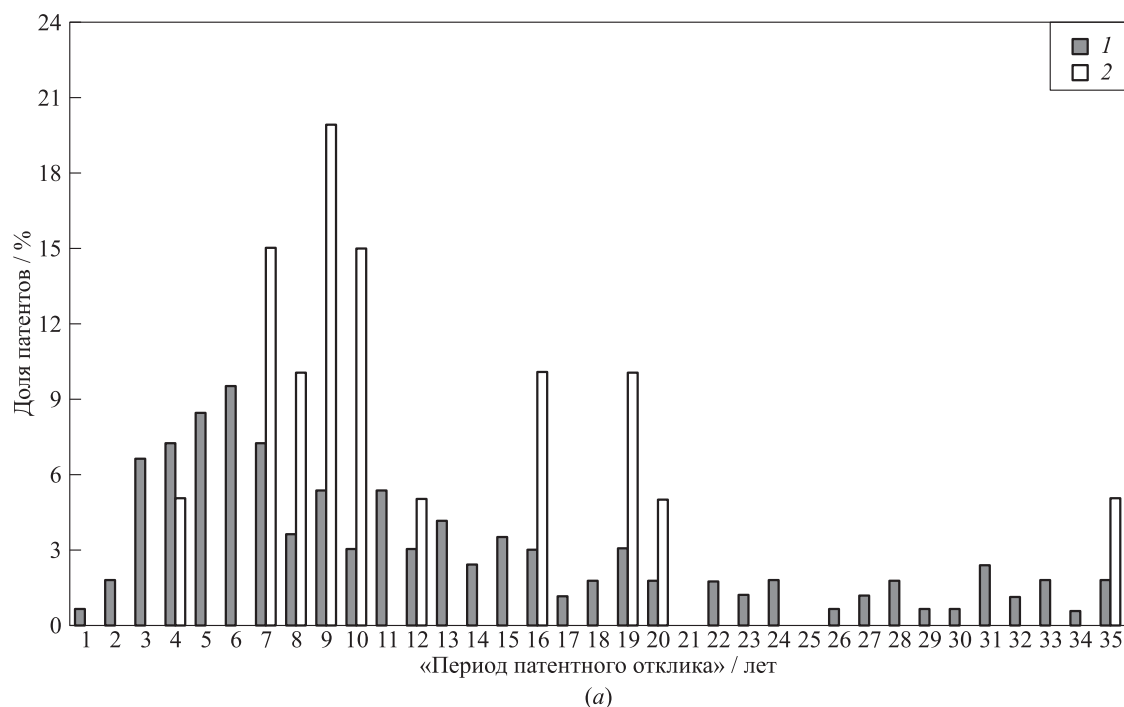


Рис. 8 Распределение «периодов патентного отклика» для подклассов G06F (а) и G06K (б) (основной индекс): 1 — РФ; 2 — другие страны

познавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации).

В то же время в описаниях изобретений зарубежных патентообладателей по интенсивности цитирования научных работ выделяется технология

G06T (Обработка или генерация данных изображения). При этом число патентов в подклассе G06T существенно меньше, чем в G06F и G06K.

Впервые были получены распределения значений «периода патентного отклика», которые дают возможность оценить динамику применения ре-

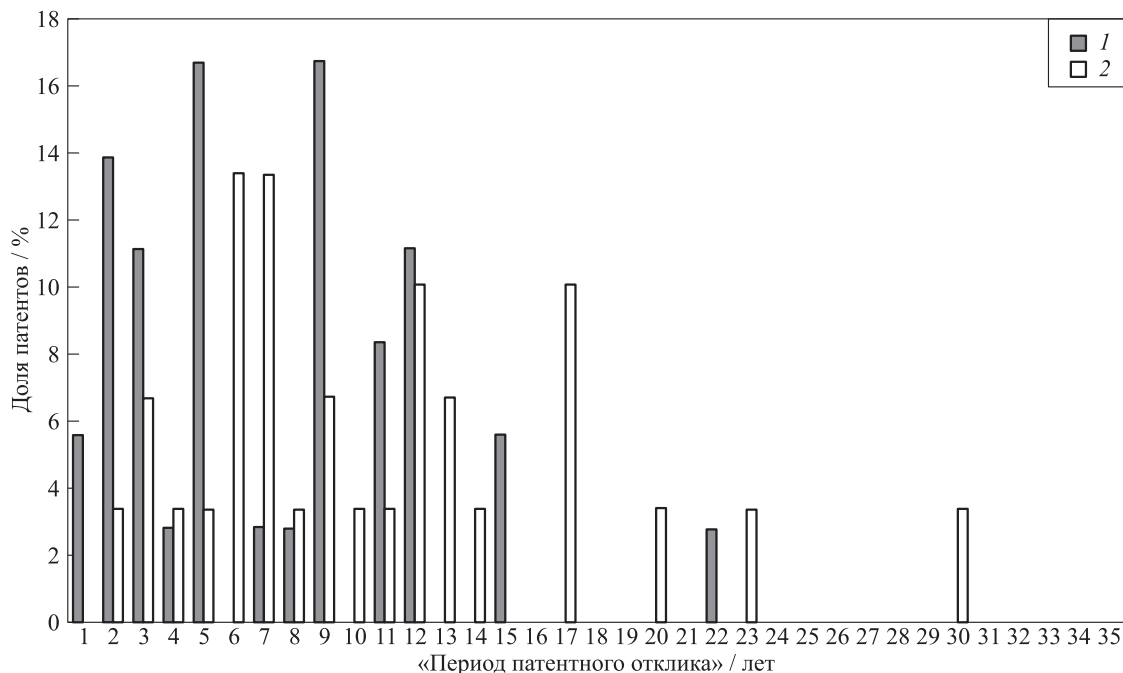


Рис. 9 Распределение времени между публикацией статьи и патента подкласса G06T (основной индекс) для патентообладателей из разных стран (в процентах к общему числу цитируемых изобретателями публикаций в каждой группе); по горизонтальной оси отложена разница в годах: 1 — РФ; 2 — другие страны

зультатов научных исследований в сфере разработки технологий. Расчеты, проведенные в данной работе, показывают, что для изобретений по ИКТ максимум распределения значений этого индикатора составляет 3 года для патентообладателей из России и 9 лет для зарубежных патентообладателей. Однако необходимо принимать во внимание и период времени от момента подачи заявки до публикации запатентованного изобретения. Если построить аналогичные распределения с учетом момента подачи заявки, то максимумы этих распределений сместятся влево.

Литература

- Giuri P., Mariani M., Brusoni S., Crespi G., Francoz D., Gambardella A., Garcia-Fontes W., Geuna A., Gonzales R., Harhoff D., Hoisl K., Le Bas C., Luzzi A., Magazzini L., Nesta L., Nomaler Ö., Palomares N., P. Patel, Romanelli M., Verspagen B. Inventors and invention processes in Europe: Results from the PatVal-EU survey // *Res. Policy*, 2007. Vol. 36. No. 8. P. 1107–1127.
- 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.
- Narin F., Noma E. Is technology becoming science? // *Scientometrics*, 1985. Vol. 7. No. 3–6. P. 369–381.
- Mansfield E. Academic research and innovation // *Res. Policy*, 1991. Vol. 20. No. 1. P. 1–12.
- Schmoch U. Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. No. 1. P. 193–211.
- Mansfield E. Academic research underlying industrial innovations: Sources, characteristics and financing // *Rev. Econ. Stat.*, 1995. Vol. 77. No. 1. P. 55–62.
- Narin F., Olivastro D. Linkage between patents and papers: An interim EPO/US comparison // *Scientometrics*, 1998. Vol. 41. No. 1–2. P. 51–59.
- Mansfield E. Academic research and industrial innovation: An update of empirical findings // *Res. Policy*, 1998. Vol. 26. No. 7–8. P. 773–776.
- 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.
- 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.
- European Commission. Third European Report on Science & Technology Indicators. — Luxembourg: Office for Official Publications of the European Communities, 2003. 451 p.
- 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 Proceedings: Book of abstracts. — Vienna: ARC GmbH, 2008. P. 196–200.
13. *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: Тр. IX Всеросс. науч. конф. RCDL'2007.* — Переславль: Университет города Переславля, 2007. С. 35–44.
 14. *Минин В. А., Зацман И. М., Кружков М. Г., Норекян 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.
 15. *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.
 16. *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.
 17. *Хавансков В. А., Шубников С. К.* Поиск и рубрицирование ссылок на цитируемые публикации в электронных библиотеках полнотекстовых описаний изобретений // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. XVI Всеросс. науч. конф. RCDL-2014.* — Дубна: ОИЯИ, 2014. С. 165–173.
 18. *Минин В. А., Зацман И. М., Хавансков В. А., Шубников С. К.* Индикаторы тематических взаимосвязей науки и технологий: от текста к числам // *Информатика и её применения*, 2014. Т. 8. Вып. 3. С. 114–125.
 19. *Минин В. А., Зацман И. М., Хавансков В. А., Шубников С. К.* Индикаторы тематических взаимосвязей науки и информационно-компьютерных технологий в начале XXI века // *Информатика и её применения*, 2015. Т. 9. Вып. 2. С. 111–120.
 20. *Зацман И. М., Вережкин Г. Ф.* Информационный мониторинг сферы науки в задачах программно-целевого управления // *Системы и средства информатики*, 2006. Вып. 16. С. 164–189.
 21. *Зацман И. М., Кожунова О. С.* Семантический словарь системы информационного мониторинга в сфере науки: задачи и функции // *Системы и средства информатики*, 2007. Вып. 17. С. 124–141.
 22. *Zatsman I., Kozhunova O.* Evaluating for institutional academic activities: Classification scheme for R&D indicators // 10th Conference (International) on Science and Technology Indicators: Book of abstracts. — Vienna: ARC GmbH, 2008. P. 428–431.
 23. *Zatsman I., Kozhunova O.* Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators // *Atlanta Conference on Science and Innovation Policy Proceedings / Eds. S. E. Cozzens*, P. Catalan. 2009. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.
 24. *Архипова М. Ю., Зацман И. М., Шульга С. Ю.* Индикаторы патентной активности в сфере информационно-коммуникационных технологий и методика их вычисления // *Экономика, статистика и информатика. Вестник УМО*, 2010. № 4. С. 93–104.
 25. *Zatsman I., Durnovo A.* Incompleteness problem of indicators system of research programme // 11th Conference (International) on Science and Technology Indicators: Book of abstracts. — Leiden: CWTS, 2010. P. 309–311.
 26. *Зацман И. М., Дурново А. А.* Моделирование процессов формирования экспертных знаний для мониторинга программно-целевой деятельности // *Информатика и её применения*, 2011. Т. 5. Вып. 4. С. 84–98.
 27. Российское агентство по патентам и товарным знакам (Роспатент): Годовой отчет 2000. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/otchet/otchet_2000_r6.
 28. Федеральная служба по интеллектуальной собственности (Роспатент): Годовой отчет 2014. http://www.rupto.ru/about/reports/2014_1#1.2.

Поступила в редакцию 19.04.16

INTENSITY OF CITATION OF SCIENTIFIC PUBLICATIONS IN INVENTIONS ON INFORMATION AND COMPUTER TECHNOLOGIES PATENTED IN RUSSIA BY DOMESTIC AND FOREIGN APPLICANTS

V. A. Minin, I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov

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

Abstract: The paper discusses the information relationship between science and technology and the methods for indicator assessment of transfer processes (transfer) of knowledge from different fields of research in the area of technological development. The proposed methods are designed to determine the values of the indicator of intensity of citation of scientific papers in the descriptions of the inventions patented in Russia by domestic and foreign applicants. A similar approach can be used to obtain indirect estimates of innovation potential of scientific research. The indicator values of intensity were calculated both in general and with the distribution by country of applicants. The paper presents the results of determining the values of the indicator. Full-text descriptions of inventions on class G06 of the International Patent Classification (Data Processing; Computing; Score) published by Rospatent in 2000–2012 were used as the source of information. The use of information resources of Rospatent was due to the fact that they are in the electronic form, i. e., available for automated processing. The result is the values of the indicator of intensity of citation of scientific works patented in the Russian Federation, divided into groups of domestic, foreign, and joint inventions. This specification allowed to estimate the activity of international technological cooperation and joint patenting in information and computer technologies (ICT) in Russia, as well as to determine the themes of cooperation in this area.

Keywords: citation of scientific papers; intensity of citation linkages between science and technologies; information technology; international patent classification; calculation of values of the indicator of intensity of citation

DOI: 10.14357/19922264160213

Acknowledgments

The work was financially supported by the Russian Foundation for Basic Research (project 16-07-00075).

References

1. Giuri, P., M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes, A. Geuna, R. Gonzales, D. Harhoff, K. Hoisl, C. Le Bas, A. Luzzi, L. Magazzini, L. Nesta, Ö. Nomaler, N. Palomares, P. Patel, M. Romanelli, and B. Verspagen. 2007. Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Res. Policy* 36(8):1107–1127.
2. 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.
3. Narin, F., and E. Noma. 1985. Is technology becoming science? *Scientometrics* 7(3–6):369–381.
4. Mansfield, E. 1991. Academic research and innovation. *Res. Policy* 20(1):1–12.
5. Schmoch, U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* 26(1):193–211.
6. Mansfield, E. 1995. Academic research underlying industrial innovations: Sources, characteristics and financing. *Rev. Econ. Stat.* 77(1):55–62.
7. Narin, F., and D. Olivastro. 1998. Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics* 41(1–2):51–59.
8. Mansfield, E. 1998. Academic research and industrial innovation: An update of empirical findings. *Res. Policy* 26(7–8):773–776.
9. 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.
10. Verbeek, A., K. Debackere, M. Luwel, P. Andries, E. Zimmermann, and D. Deleus. 2002. Linking science to tech-

- nology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(3):399–420.
11. European Commission. 2003. Third European Report on Science & Technology Indicators. Luxembourg: Office for Official Publications of the European Communities. 451 p.
 12. Van Looy, B., W. Hansen, H. Hollanders, and R. Tijssen. 2008. 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 Proceedings: Book of abstracts*. Vienna: ARC GmbH. 196–200.
 13. Zatsman, I. M., and S. K. Shubnikov. 2007. Printsipy obrabotki informatsionnykh resursov dlya otsenki innovatsionnogo potentsiala napravleniy nauchnykh issledovaniy [Principles of processing of information resources for assessment of innovative potential of fields of scientific research]. *Elektronnye Biblioteki: Perspektivnye Metody i Tekhnologii, Elektronnye Kollektzii: Tr. 9-y Vseross. nauch. konf. RCDL'2007* [Digital Libraries: Perspective Methods and Technologies, Electronic collections: 9th All-Russia Scientific Conference RCDL'2007 Proceedings]. Pereslavl'. 35–44.
 14. Minin, V. A., I. M. Zatsman, M. G. Kruzhkov, and T. P. Norekyan. 2013. Metodologicheskie osnovy sozdaniya informatsionnykh sistem dlya vychisleniya indikatorov tematiceskikh vzaimosvyazey nauki i tekhnologii [Methodological basis for the creation of information systems for the calculation of indicators of thematic science-technology linkages]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):70–81.
 15. Minin, V. A., I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov. 2013. Arkhitekturnye resheniya dlya sistem vychisleniya indikatorov tematiceskikh vzaimosvyazey nauki i tekhnologii [Architectural decisions for systems of calculation of indicators of thematic science-technology linkages]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):260–283.
 16. Zatsman, I. M., V. A. Havanskov, and S. K. Shubnikov. 2013. Metod izvlecheniya bibliograficheskoy informatsii iz polnotekstovykh opisaniy izobreteniy [Method of extraction of bibliographic information from full-text descriptions of inventions]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):52–65.
 17. Havanskov, V. A., and S. K. Shubnikov. 2014. Poisk i rubritirovanie ssylok na tsitiruemye publikatsii v elektronnykh bibliotekakh polnotekstovykh opisaniy izobreteniy [Search and classifying of cited publications in digital libraries of full-text descriptions of inventions]. *Elektronnye Biblioteki: Perspektivnye Metody i Tekhnologii, Elektronnye Kollektzii: Tr. 16-y Vseross. nauch. konf. RCDL'2014* [Digital Libraries: Perspective Methods and Technologies, Electronic Collections: 16th All-Russia Scientific Conference RCDL'2014 Proceedings]. Dubna. 165–173.
 18. Minin, V. A., I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov. 2014. Indikatory tematiceskikh vzaimosvyazey nauki i tekhnologii: Ot teksta k chislam [Indicators of thematic science-technology linkages: From text to numbers]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):114–125.
 19. Minin, V. A., I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov. 2015. Indikatory tematiceskikh vzaimosvyazey nauki i informatsionno-komp'yuternykh tekhnologii v nachale XXI veka [Indicators for thematic linkages between science and information and computer technologies at the beginning of the XXI century]. *Informatika i ee Primeneniya — Inform. Appl.* 9(2):111–120.
 20. Zatsman, I. M., and G. F. Verevkin. 2006. Informatsionny monitoring sfery nauki v zadachakh programmno-tselevogo upravleniya [Information monitoring in the field of science and problems of program-oriented management]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:164–189.
 21. Zatsman, I. M., and O. S. Kozhunova. 2007. Semanticheskii slovar' sistemy informatsionnogo monitoringa v sfere nauki: Zadachi i funktsii [The semantic dictionary of system for information monitoring in science: Tasks and functions]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 17:124–141.
 22. Zatsman, I., and O. Kozhunova. 2008. Evaluating for institutional academic activities: Classification scheme for R&D indicators. *10th Conference (International) on Science and Technology Indicators: Book of abstracts*. Vienna: ARC GmbH. 428–431.
 23. Zatsman, I., and O. Kozhunova. 2009. Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators. *2009 Atlanta Conference on Science and Innovation Policy Proceedings*. Eds. S. E. Cozzens and P. Catalan. Available at: <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf> (accessed January 24, 2016).
 24. Arkhipova, M. Yu., I. M. Zatsman, and S. Yu. Shul'ga. 2010. Indikatory patentnoy aktivnosti v sfere informatsionno-kommunikatsionnykh tekhnologii i metodika ikh vychisleniya [Indicators of patent activity in the sphere of information and communication technologies and technique of their calculation]. *Ekonomika, statistika i informatika. Vestnik UMO* [Economy, statistics and informatics. Herald of the UMO] 4:93–104.
 25. Zatsman, I., and A. Durnovo. 2010. Incompleteness problem of indicators system of research programme. *11th Conference (International) on Science and Technology Indicators: Book of abstracts*. Leiden: CWTS. 309–311.
 26. Zatsman, I. M., and A. A. Durnovo. 2011. Modelirovanie protsessov formirovaniya ekspertnykh znaniy dlya monitoringa programmno-tselevoy deyatel'nosti [Modeling of creation processes of expert knowledge for monitoring program-oriented activities]. *Informatika i ee Primeneniya — Inform. Appl.* 5(4):84–98.

27. Rossiyskoe agentstvo po patentam i tovarnym znakam (Rospatent) [The Russian Agency for Patents and Trademarks (Rospatent)]. 2000. Annual Report. Available at: http://www1.fips.ru/wps/wcm/connect/content_ru/ru/otchety/otchet_2000_r6 (accessed June 20, 2016).
28. Federal'naya sluzhba po intellektual'noy sobstvennosti (Rospatent) [The Federal service for intellectual property (Rospatent)]. 2014. Available at: http://www.rupto.ru/about/reports/2014_1#1 (accessed January 24, 2016).

Received April 19, 2016

Contributors

Minin Vladimir A. (b. 1941) — Doctor of Science in physics and mathematics, consultant, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; aleksiss@ya.ru

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; iz-ipi@i170.ipi.ac.ru

Havanskov Valerij A. (b. 1950) — scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44- 2 Vavilov Str., Moscow 119333, Russian Federation; havanskov@i170.ipi.ac.ru

Shubnikov Sergej K. (b. 1955) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44- 2 Vavilov Str., Moscow 119333, Russian Federation; sergeysh50@yandex.ru

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

Л. А. Мейханаджян¹

Аннотация: Рассматривается система $M/G/1/(r-1)$ с дисциплиной инверсионного порядка обслуживания и обобщенного вероятностного приоритета. Предполагается, что в момент поступления новой заявки в систему становится известной ее длина и, кроме того, в любой момент времени известна остаточная длина каждой заявки в системе. В момент поступления очередной заявки в непустую систему ее исходная длина сравнивается с остаточной длиной заявки на приборе, и в зависимости от результатов сравнения наступает одно из следующих событий: обе заявки покидают систему; только одна из заявок покидает систему (другая остается на приборе); обе заявки остаются в системе (одна попадает на прибор, другая — в очередь). Заявки, оставшиеся в системе, приобретают новую (случайную) длину в соответствии с заданным распределением, зависящим в общем случае от исходных длин заявок. Заявки, застающие систему полностью заполненной, теряются и не оказывают на нее никакого воздействия. В статье предложены математические соотношения для вычисления совместного стационарного распределения числа заявок в системе и остаточного времени обслуживания заявки на приборе, периода занятости системы, стационарного распределения времени ожидания и пребывания заявки длины x (в терминах преобразования Лапласа–Стилтьеса (ПЛС)).

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

DOI: 10.14357/19922264160214

1 Введение

В этой работе, являющейся продолжением работ [1, 2], будет рассматриваться та же однолинейная система массового обслуживания (СМО), что и в [1], но ограниченной емкости. Основным результатом работ [1, 2] состоит в нахождении совместного стационарного распределения вероятностей состояний системы $M/G/1$ с дисциплиной инверсионного порядка обслуживания и обобщенного вероятностного приоритета, а также основных стационарных вероятностных характеристик в терминах ПЛС. Сейчас же задача заключается в исследовании стационарных вероятностно-временных характеристик указанной системы в случае, когда присутствует ограничение на размер очереди.

2 Описание системы

Рассмотрим СМО с одним прибором, одной очередью для ожидающих заявок емкости $(r-1) < \infty$, $r \geq 2$, и входящим потоком заявок, который

для простоты будем называть здесь потоком пуассоновского типа. Отличие этого потока от пуассоновского заключается в следующем: интенсивность поступления заявок 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 функцией распределения (ФР) $B(x)$. Если же заявка поступает в систему в тот момент, когда система пуста, то исходное распределение времени обслуживания поступающей заявки является произвольным с ФР $\tilde{B}(x)$.

Далее для простоты изложения будем считать, что ФР $B(x)$ и $\tilde{B}(x)$ имеют непрерывные ограниченные плотности распределения $b(x) = B'(x)$ и $\tilde{b}(x) = \tilde{B}'(x)$, причем $\tilde{b} = \int_0^\infty x\tilde{b}(x) dx < \infty$ и $b = \int_0^\infty xb(x) dx < \infty$.

Обобщенный инверсионный порядок обслуживания с вероятностным приоритетом (LCFS BPP) заключается в следующем. Предполагается, что

* Работа выполнена при поддержке РФФИ (проект 15-07-03007).

¹ Российский университет дружбы народов, lameykanadzhyan@gmail.com

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:

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

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.

Для конечного накопителя необходимо также задать дисциплину принятия заявок в систему при отсутствии в нем свободных мест. Здесь для простоты изложения будет рассмотрен только тот случай, когда поступающая в заполненную систему заявка теряется. Заметим, что в этом случае принятая в систему заявка будет обязательно обслужена полностью. Для всех СМО с такой дисциплиной принятия заявок в систему при отсутствии в накопителе свободных мест стационарные вероятности $p_n(x_1, \dots, x_n)$ при $n < r$ совпадают с точностью до постоянной с аналогичными вероятностями для системы с бесконечным накопителем, различие заключается только в вероятностях $p_r(x_1, \dots, x_r)$. Однако несколько более сложно вычисляются стационарные распределения, связанные с временем пребывания заявки в системе, поскольку даже заявки, принятые в систему, могут покидать ее недообслуженными.

Далее будем предполагать, что система функционирует в стационарном режиме и $\bar{b} = \int_0^\infty x \bar{b}(x) dx < \infty$ и $b = \int_0^\infty x b(x) dx < \infty$. Отметим, что параметр $\rho = \lambda b$ для данной системы не является загрузкой в традиционном смысле и может существенно от нее отличаться.

3 Стационарные вероятностные характеристики

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 $\xi_1(t)$ которого 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

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

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

Обозначим через

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

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

стационарное распределение процесса $\eta(t)$. В силу сделанных в предыдущем пункте предположений

относительно параметров системы, можно показать (см., например, [3; 4, с. 273]), что существуют непрерывные и ограниченные плотности

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

$$1 \leq n \leq r.$$

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 [4] стационарные вероятности состояний в исходной и вспомогательной системах отличаются лишь на постоянный множитель (за исключением вероятности $p_r(x_1, \dots, x_r)$). Это дает возможность при составлении уравнений для $p_n(x_1, \dots, x_n)$, $n \geq 1$, воспользоваться вспомогательной системой и получить следующие соотношения:

$$-p'_1(x) = \tilde{\lambda} \tilde{b}(x) p_0 - \lambda p_1(x) +$$

$$+ \lambda \left(\int_0^\infty \int_0^\infty d(x|u, v) b(u) p_1(v) dudv + \right.$$

$$+ \int_0^\infty \int_0^\infty \int_0^\infty [d(x, y|u, v) b(u) p_1(v) +$$

$$\left. + d^*(y, x|u, v) b(u) p_1(v)] dydudv \right); \quad (1)$$

$$-p'_n(x_1, \dots, x_n) =$$

$$= \lambda \left(\int_0^\infty \int_0^\infty [d(x_2, x_1|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n) + \right.$$

$$+ d^*(x_1, x_2|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n)] dudv \left. - \right.$$

$$- \lambda p_n(x_1, \dots, x_n) +$$

$$+ \lambda \left(\int_0^\infty \int_0^\infty d(x_1|u, v) b(u) p_n(v, x_2, \dots, x_n) dudv + \right.$$

$$+ \int_0^\infty \int_0^\infty \int_0^\infty [d(x_1, y|u, v) b(u) p_n(v, x_2, \dots, x_n) +$$

$$+ d^*(y, x_1|u, v) b(u) p_n(v, x_2, \dots, x_n)] dydudv \left. \right),$$

$$1 \leq n \leq r-1; \quad (2)$$

$$-p'_r(x_1, \dots, x_n) =$$

$$= \lambda \left(\int_0^\infty \int_0^\infty [d(x_2, x_1|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n) + \right.$$

$$+ d^*(x_1, x_2|u, v) b(u) p_{n-1}(v, x_3, \dots, x_n)] dudv \left. \right). \quad (3)$$

Остановимся подробнее на выводе уравнения для плотности $p_r(x_1, \dots, x_r)$ (остальные уравнения получаются так же, как и в случае накопителя бесконечной емкости [1]). Рассмотрим моменты времени t и $(t + \Delta)$. Тогда для того, чтобы в момент времени $(t + \Delta)$ в системе находилось r заявок, причем на приборе заявка длины x_1 , а в очереди заявки длин x_2, \dots, x_r , нужно, чтобы произошло одно из следующих событий:

- в момент t в системе находилось $(r-1)$ 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 $p_{r-1}(t; v, x_3, \dots, x_r)$), 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 $(r-1)$ 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 $p_{r-1}(t; v, x_3, \dots, x_r)$), 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 r заявок, причем заявка на приборе имела длину $x_1 + \Delta$, первая заявка в очереди имела дли-

ну x_2, \dots , последняя заявка в очереди имела длину x_r (с плотностью вероятностей $p_r(t; x_1 + \Delta, x_2, \dots, x_r)$).

Вероятности других событий равны $o(\Delta)$. Применяя формулу полной вероятности, имеем:

$$p_r(t + \Delta; x_1, \dots, x_r) = \lambda \Delta \left(\int_0^\infty \int_0^\infty [d(x_2, x_1|u, v) b(u) p_{r-1}(t; v, x_3, \dots, x_r) + d^*(x_1, x_2|u, v) b(u) p_{r-1}(t; v, x_3, \dots, x_r)] dudv \right) + p_r(t; x_1 + \Delta, x_2, \dots, x_r),$$

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

К системе уравнений (1)–(3) нужно добавить начальные условия, которые удобно записать в виде:

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

$$p_n(\infty, x_2, \dots, x_r) = \lim_{X \rightarrow \infty} p_n(X, x_2, \dots, x_r) = 0, \quad 1 \leq n \leq r. \tag{5}$$

Как получаются соотношения (4) и (5), показано в [1]. Оставшаяся неизвестной стационарная вероятность p_0 отсутствия заявок в системе находится, как обычно, из условия нормировки:

$$\sum_{n=0}^r p_n = 1, \tag{6}$$

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

Как и в случае системы бесконечной емкости, полученные соотношения (1)–(6) позволяют теоретически последовательно по n находить стационарные плотности вероятностей $p_n(x_1, \dots, x_n)$. Однако на практике такие расчеты связаны с серьезными вычислительными сложностями.

Как показано, например, в [5], в практических случаях бывает достаточно знать только маргинальные плотности

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

Интегрируя (2) и (3) по x_2, \dots, x_r в пределах от нуля до бесконечности и вспоминая равенство (1),

получаем следующую систему интегродифференциальных уравнений для $p_n(x)$, $1 \leq n \leq r$:

$$-p'_n(x) = a_n(x) - \lambda p_n(x) + \int_0^\infty K_n(x, v) p_n(v) dv, \quad 1 \leq n \leq r - 1; \tag{7}$$

$$-p'_r(x) = a_r(x), \tag{8}$$

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

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

$$K_n(x, v) = \lambda \int_0^\infty b(u) du \left(d(x|u, v) + \int_0^\infty [d(x, y|u, v) + d^*(y, x|u, v)] dy \right), \quad 1 \leq n \leq r - 1.$$

Начальные условия для уравнений (7) и (8) по аналогии с (4) запишем в виде:

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

Решать систему (7) и (8) с начальными условиями (9) можно различными способами. Воспользуемся методом, предложенным в [1]. Прежде всего заметим, что из (8) немедленно следует, что

$$p_r(x) = \int_x^\infty a_r(u) du.$$

Решение уравнений (7) будем искать в виде:

$$p_n(x) = e^{\lambda x} q_n(x), \quad 1 \leq n \leq r - 1. \tag{10}$$

Подставляя в (7) вместо $p_n(x)$ ее выражение по формуле (10), получаем новое интегродифференциальное уравнение:

$$-q'_n(x) = e^{-\lambda x} a_n(x) + \int_0^\infty e^{\lambda v} e^{-\lambda x} K_n(x, v) q_n(v) dv, \quad 1 \leq n \leq r - 1.$$

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

$$q_n(y) = b_n(y) + \int_0^\infty G_n(y, v) q_n(v) dv, \quad 1 \leq n \leq r - 1, \tag{11}$$

где

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

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

Отметим, что свободный член $b_n(y)$ и ядро $G_n(y, v)$ интегрального уравнения являются неотрицательными функциями. Далее для расчета $q_n(y)$ можно применить подходящий метод решения интегральных уравнений Фредгольма 2-го рода (см., например, [6–8]).

В некоторых частных случаях решения уравнений (11) могут быть выписаны в явном виде. Например, это возможно в случае, когда известны сепарабельные аппроксимации для функций $d(x, y|u, v)$, $d^*(x, y|u, v)$, $d_0(x|u, v)$ и $d_0^*(x|u, v)$, т. е. разложения вида:

$$d(x, y|u, v) = \sum_{i=1}^{N_1} \alpha_{1i}(x) \beta_{1i}(y) \gamma_{1i}(u) \delta_{1i}(v);$$

$$d^*(x, y|u, v) = \sum_{i=1}^{N_2} \alpha_{2i}(x) \beta_{2i}(y) \gamma_{2i}(u) \delta_{2i}(v);$$

$$d_0(x|u, v) = \sum_{i=1}^{N_3} \alpha_{3i}(x) \gamma_{3i}(u) \delta_{3i}(v);$$

$$d_0^*(x|u, v) = \sum_{i=1}^{N_4} \alpha_{4i}(x) \gamma_{4i}(u) \delta_{4i}(v),$$

где N_1, N_2, N_3 и N_4 — некоторые натуральные числа, а $\alpha_{ij}(x), \beta_{ij}(x), \gamma_{ij}(x)$ и $\delta_{ij}(x)$ — некоторые известные функции. Тогда решение уравнения (11) при фиксированном n сводится к решению системы линейных уравнений относительно $(N_1 + N_2 + N_3 + N_4)$ неизвестных.

4 Стационарные временные характеристики

4.1 Стационарное распределение времени ожидания начала обслуживания

Для того чтобы найти показатели функционирования СМО, связанные с временем пребывания в системе, нужно прежде всего найти ПЛС периода занятости (ПЗ) системы.

Обозначим через $u_n(s; x)$, $1 \leq n \leq r$, ПЛС времени до того момента, когда в системе впервые останется $(n-1)$ заявок, при условии что на приборе

начала обслуживаться заявка (остаточной) длины x и в системе находится n заявок.

Учитывая, что по принятому соглашению поступающая в заполненную систему заявка сразу теряется, ПЛС $u_r(s; x)$ удовлетворяет уравнению:

$$u_r(s; x) = e^{-sx}. \quad (12)$$

Воспользовавшись свойствами ПЛС, получаем, что $u_n(s; x)$ равно:

- e^{-sx} , если до момента времени x окончания обслуживания заявки на приборе новая заявка не поступила (с вероятностью $e^{-\lambda x}$);
- e^{-st} , если в момент $0 < t < x$ поступила новая заявка и обе заявки покинули систему (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d_0(y, x-t)b(y) dy$);
- $e^{-st} u_n(s; v)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , одна из двух заявок (поступившая заявка или заявка на приборе) покинула систему, а оставшаяся приняла длину v и, значит, время до того момента, как в системе останется $(n-1)$ заявок, равно $u_n(s; v)$ (плотность вероятности данного события равна $\lambda e^{-\lambda t} \int_0^\infty d(v|y, x-t)b(y) dy$);
- $e^{-st} u_{n+1}(s; w) u_n(s; v)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , обе заявки остаются в системе (новая встает в очередь), причем длина новой заявки становится равной v , а на приборе — w (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d(v, w|y, x-t)b(y) dy$);
- $e^{-st} u_{n+1}(s; v) u_n(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , обе заявки остаются в системе (новая встает в очередь), причем длина новой заявки становится равной w , а на приборе — v (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d^*(v, w|y, x-t)b(y) dy$).

По формуле полной вероятности окончательно получаем:

$$u_n(s; x) = e^{-(s+\lambda)x} +$$

$$+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty d_0(y, x-t)b(y) dy +$$

$$+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty u_n(s; v) dv \int_0^\infty d(v|y, x-t)b(y) dy +$$

$$+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty u_{n+1}(s; w) dw \int_0^\infty u_n(s; v) dv \times$$

$$\begin{aligned} & \times \int_0^\infty d(v, w|y, x-t)b(y) dy + \int_0^x \lambda e^{-(\lambda+s)t} dt \times \\ & \times \int_0^\infty u_{n+1}(s; v) dv \int_0^\infty u_n(s; w) dw \times \\ & \times \int_0^\infty d^*(v, w|y, x-t)b(y) dy, \\ & 1 \leq n \leq r-1. \quad (13) \end{aligned}$$

Система уравнений (12)–(13) решается рекуррентно, начиная с $n = r - 1$.

Зная значения $u_n(s; x)$, можно найти основные стационарные временные характеристики заявок. Пусть в начальный момент в системе находится n заявок, $1 \leq n \leq r-1$, на приборе обслуживается заявка длины y и в этот момент в систему поступает заявка длины x . Обозначим через $w_n(s; x, y)$ ПЛС времени ожидания начала обслуживания этой заявки. В соответствии с дисциплиной обслуживания имеет место равенство:

$$\begin{aligned} w_n(s; x, y) = & \int_0^\infty \int_0^\infty d^*(v, w|x, y) dv dw + \\ & + \int_0^\infty d_0(v|x, y) dv + \int_0^\infty \int_0^\infty u_{n+1}(s; w) d(v, w|x, y) dv dw. \end{aligned}$$

Заметим, что вероятность того, что поступающая заявка длины x будет потеряна при поступлении в систему, равна:

$$\begin{aligned} \pi(x) = & \int_0^\infty \sum_{n=1}^{r-1} p_n(y) \left(d_0(x, y) + \int_0^\infty d_0^*(w|x, y) dw \right) dy + \\ & + \int_0^\infty p_r(y) dy. \end{aligned}$$

Тогда ПЛС $w(s)$ стационарного распределения времени ожидания начала обслуживания принятой в систему заявки определяется формулой:

$$\begin{aligned} w(s) = & \frac{1}{1-\pi} \left(p_0 + \right. \\ & \left. + \int_0^\infty \sum_{n=1}^{r-1} p_n(y) dy \int_0^\infty b(x)w_n(s; x, y) dx \right), \end{aligned}$$

где $\pi = \int_0^\infty \pi(x)b(x) dx$ — безусловная вероятность потери заявки.

4.2 Стационарное распределение времени пребывания заявки в системе

Распределение полного времени пребывания заявки в системе вычисляется несколько сложнее из-за того, что заявка, попавшая на прибор, может покинуть его и возвращаться на него обратно, менять свою длину, а также уйти из системы недообслуженной.

Остановимся на нахождении следующих характеристик, которые понадобятся в дальнейшем:

- стационарное распределение времени пребывания на приборе заявки, которая была обслужена до конца (с учетом возможных смен длин и прерываний), при условии что в момент поступления на прибор ее длина равнялась x , а в очереди было n , $0 \leq n \leq r - 1$, других заявок. Через $V_{1,n}(s; x)$ будем обозначать ПЛС этого распределения;
- стационарное распределение времени пребывания на приборе заявки, которая могла быть и не обслужена до конца (с учетом возможных смен длин и прерываний), при условии что в момент поступления на прибор ее длина равнялась x , а в очереди было n , $0 \leq n \leq r - 1$, других заявок. Через $V_{2,n}(s; x)$ будем обозначать ПЛС этого распределения.

Отметим, что здесь подразумевается, что время пребывания поступившей на прибор заявки включает все времена, на которые ее обслуживание было прервано.

Ввиду того что поступающая в заполненную систему заявка теряется, выпишем $V_{1,r-1}(s; x) = e^{-sx}$. Далее, воспользовавшись свойством ПЛС, найдем, что $V_{1,n}(s; x)$ равно:

- e^{-sx} , если до момента времени x окончания обслуживания заявки на приборе новая заявка не поступила (с вероятностью $e^{-\lambda x}$);
- $e^{-st}V_{1,n}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , изменила длину заявки на приборе на w , а сама покинула систему (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d_0^*(w|y, x-t)b(y) dy$);
- $e^{-st}V_{1,n+1}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , которая встала на первое место в очереди, причем новая заявка получила новую длину v , а заявка на приборе новую длину w (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d(v, w|y, x-t)b(y) dy$);

- $e^{-st}u_{n+2}(s; v)V_{1,n}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , которая встала на прибор, получив новую длину v , а заявка с прибора вытеснена на первое место в очереди и получила новую длину w (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d^*(v, w|y, x-t)b(y) dy$).

Воспользовавшись снова формулой полной вероятности, получаем, что уравнение для определения ПЛС $V_{1,n}(s; x)$ имеет следующий вид:

$$V_{1,n}(s; x) = e^{-(\lambda+s)x} + \int_0^\infty V_{1,n+1}(s; w)f(s; x, w) dw + \int_0^\infty V_{1,n}(s; w)g_{n+2}(s; x, w) dw, \quad 0 \leq n \leq r-2,$$

где

$$f(s; x, w) = \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty dv \int_0^\infty d(v, w|y, x-t)b(y) dy;$$

$$g_{n+2}(s; x, w) = \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty u_{n+2}(s; v) dv \times$$

$$\times \int_0^\infty d^*(v, w|y, x-t)b(y) dy +$$

$$+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty d_0^*(w|y, x-t)b(y) dy.$$

Уравнение для определения $V_{2,n}(s; x)$ получается аналогичным образом. Действительно, $V_{2,r-1}(s; x) = e^{-sx}$. Далее, ПЛС $V_{2,n}(s; x)$ равно:

- e^{-sx} , если до момента времени x окончания обслуживания заявки на приборе новая заявка не поступила (с вероятностью $e^{-\lambda x}$);
- e^{-st} , если в момент $0 < t < x$ поступила новая заявка и она вместе с заявкой на приборе покинула систему (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d_0(y, x-t)b(y) dy$);
- e^{-st} , если в момент времени $0 < t < x$ поступила новая заявка длины y , сама встала на прибор, а заявка с прибора покинула систему (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d_0(v|y, x-t)b(y) dy$);
- $e^{-st}V_{2,n}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , изменила длину заявки на приборе на w , а сама покинула систему (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d_0^*(w|y, x-t)b(y) dy$);

- $e^{-st}V_{2,n+1}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , которая встала на первое место в очереди, причем новая заявка получила новую длину v , а заявка на приборе новую длину w (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d(v, w|y, x-t)b(y) dy$);
- $e^{-st}u_{n+2}(s; v)V_{2,n}(s; w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , которая встала на прибор, получив новую длину v , а заявка с прибора вытеснена на первое место в очереди и получила новую длину w (с плотностью вероятностей $\lambda e^{-\lambda t} \int_0^\infty d^*(v, w|y, x-t)b(y) dy$).

Воспользовавшись снова формулой полной вероятности, получаем, что уравнение для определения ПЛС $V_{2,n}(s; x)$ имеет следующий вид:

$$V_{2,n}(s; x) = h(s, x) + \int_0^\infty V_{2,n+1}(s; w)f(s; x, w) dw + \int_0^\infty V_{2,n}(s; w)g_{n+2}(s; x, w) dw, \quad 0 \leq n \leq r-2,$$

где

$$h(s, x) = e^{-(\lambda+s)x} + \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty d_0(y, x-t)b(y) dy + \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty dv \int_0^\infty d_0(v|y, x-t)b(y) dy.$$

Решение полученных уравнений осуществляется рекуррентным образом, начиная с $n = r-1$. Естественно, ПЛС безусловных распределений получаются усреднением $V_{1,n}(s; x)$ и $V_{2,n}(s; x)$ по распределению длины заявки $B(x)$.

Наконец, перейдем к нахождению полного времени пребывания заявки в системе. Будем различать два случая: первый — когда заявка не может уходить из системы недообслуженной; второй — когда заявка на приборе может покинуть систему не обслуженной до конца. В обоих случаях, как обычно, полное время пребывания заявки в системе складывается из времени ожидания заявкой начала обслуживания и времени пребывания заявки на приборе (которое включает времена прерываний обслуживания).

В первом случае ПЛС стационарного распределения полного времени пребывания в системе поступающей заявки длины x обозначим через $V_1(s; x)$, во втором — через $V_2(s; x)$.

Рассмотрим первый случай.

Во-первых, заявка длины x может с вероятностью p_0 поступить в свободную систему, и тогда время ее пребывания в системе будет совпадать с временем ее пребывания на приборе (с учетом прерываний).

Во-вторых, с плотностью вероятностей $p_n(y)$ поступающая заявка длины x может застать в системе $1 \leq n \leq r - 1$ заявок, причем на приборе будет находиться заявка длины y . В этом случае возможны следующие варианты:

- либо с вероятностью $d_0(v|x, y) + d^*(v, w|x, y)$ поступающая заявка встанет на прибор, причем ее длина станет равной v и тогда полное время ее пребывания в системе будет совпадать с временем ее пребывания на приборе (с учетом прерываний);
- либо с вероятностью $d(v, w|x, y)$ поступающая заявка станет на первое место в очереди, получит новую длину v , а заявка на приборе — новую длину w ; при этом время пребывания в системе поступившей заявки будет равно сумме двух времен: времени до того момента, когда в системе снова станет n заявок, и времени пребывания на приборе (с учетом прерываний) заявки длины v .

Применяя формулу полной вероятности, приходим к следующему выражению для ПЛС $V_1(s; x)$ стационарного распределения полного времени пребывания принятой заявки в систему, в которой не допускается уход заявок недообслуженными:

$$\begin{aligned}
 V_1(s; x) = & \frac{1}{1 - \pi} \left(p_0 V_{1,0}(s; x) + \right. \\
 & + \int_0^{\infty} \sum_{n=1}^{r-1} p_n(y) \left[\int_0^{\infty} V_{1,n-1}(s; v) \left(d_0(v|x, y) + \right. \right. \\
 & \left. \left. + \int_0^{\infty} d^*(v, w|x, y) dw \right) dv \right] dy + \\
 & \left. \int_0^{\infty} \sum_{n=1}^{r-1} p_n(y) \left[\int_0^{\infty} \int_0^{\infty} u_{n+1}(s; w) \times \right. \right. \\
 & \left. \left. \times V_{1,n-1}(s; v) d(v, w|x, y) dv dw \right] dy \right). \quad (14)
 \end{aligned}$$

¹Например, если положить $d(x, y|u, v) = e^{-v} b(x) b(ye^{-v})$, $d^*(x, y|u, v) = 0$, $d(x|u, v) = 0$, $d_0(u, v) = 0$, $u, v > 0$, то среднее время до того момента, когда в системе останется $(r - 2)$ заявки, при условии что в начальный момент в системе было $(r - 1)$ заявок, без дополнительных ограничений на функцию $b(x)$ может быть равно бесконечности. При этом, учитывая пуассоновость входящего потока, с ненулевой вероятностью система переходит в состояние $(r - 2)$ и, вообще говоря, с ненулевой вероятностью может успеть выполнить до прихода очередной заявки любую находящуюся в ней работу (при условии ее конечности), т. е. полностью опустошиться.

Наконец, ПЛС $V_1(s)$ стационарного распределения полного времени пребывания в системе заявки произвольной длины получается усреднением $V_1(s; x)$ по распределению длины заявки $B(x)$ и равно $V_1(s) = \int_0^{\infty} V_1(s; x) b(x) dx$. Выражение для $V_2(s; x)$ получается путем замены в соответствующих местах формулы (14) $V_{1,n}(s; x)$ на $V_{2,n}(s; x)$.

5 Заключение

В заключение скажем несколько слов об условии существования стационарного режима. Для рассмотренной системы общего необходимого и достаточного условия его существования выписать не удастся. Оно зависит от конкретных параметров системы и в каждом отдельном случае нуждается в специальном исследовании. Конечность среднего времени обслуживания является только необходимым условием и, даже несмотря на присутствие ограничения на размер очереди, не является достаточным¹.

Литература

1. Мейханаджян Л. А., Милованова Т. А., Печинкин А. В., Разумчик Р. В. Стационарные вероятности состояний в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 16–26.
2. Мейханаджян Л. А., Милованова Т. А., Разумчик Р. В. Время ожидания в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом // Информатика и её применения, 2015. Т. 9. Вып. 2. С. 14–22.
3. Севастьянов Б. А. Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами // ТВП, 1957. Т. 2. Вып. 1. С. 106–116.
4. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
5. Meykhanadzhyan L., Razumchik R. New scheduling policy for estimation of stationary performance characteristics in single server queues with inaccurate job size information // 30th European Conference on Modelling and Simulation Proceedings. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2016. P. 710–716.
6. Jerri A. Introduction to integral equations with applications. — New York, NY, USA: John Wiley & Sons, 1999. 272 p.

7. Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. Numerical recipes: The art of scientific computing. — 3rd ed. — 2007. 1256 p.
8. Полянин А. Д., Манжиров А. В. Справочник по интегральным уравнениям. — Бока-Ратон — Лондон: Chapman & Hall, CRC Press, 2008. 1108 p.

Поступила в редакцию 19.04.16

STATIONARY CHARACTERISTICS OF THE FINITE CAPACITY QUEUEING SYSTEM WITH INVERSE SERVICE ORDER AND GENERALIZED PROBABILISTIC PRIORITY

L. A. Meykhanadzhyan

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

Abstract: Consideration is given to the $M/G/1/(r-1)$ queueing system with LIFO (last in, first out) preemptive generalized probabilistic priority policy. It is assumed that customer's service time becomes known upon its arrival at the system and at any time instant remaining service times of all customers present in the system are available. On arrival of a customer at a nonempty system, its service time is compared to the (remaining) service time of the customer in service and one of the following events occurs: both customers leave the system at once, one of the customers leaves the system (the other occupies the server), or both customers stay in the system (one occupies the server, the other — one place in the queue). Those customers which stay in the system acquire new service time according to a known distribution, which can depend on their initial service times. Arriving customers which find the queue full, leave the system and have no influence on it. Analytical expressions for the computation of the joint stationary distribution of the number of customers in the system and the remaining service time of the customer in the server, of the busy period and the stationary sojourn time (in terms of Laplace–Stieltjes transform) are proposed.

Keywords: queueing system; special discipline; LIFO; probabilistic priority

DOI: 10.14357/19922264160214

Acknowledgments

The work is supported by the Russian Foundation for Basic Research (project 15-07-03007).

References

- Meykhanadzhyan, L. A., T. A. Milovanova, A. V. Pechinkin, and R. V. Razumchik. 2014. Stationary veroyatnosti sostoyaniy v sisteme obsluzhivaniya s inversionnym poryadkom obsluzhivaniya i obobshchennym veroyatnostnym prioritetom [Stationary distribution in a queueing system with inverse service order and generalized probabilistic priority]. *Informatika i ee Primeneniya — Inform.Appl.* 8(3):16–26.
- Meykhanadzhyan, L. A., T. A. Milovanova, and R. V. Razumchik. 2015. Vremya ozhidaniya v sisteme obsluzhivaniya s inversionnym poryadkom obsluzhivaniya i obobshchennym veroyatnostnym prioritetom [Stationary waiting time in a queueing system with inverse service order and generalized probabilistic priority]. *Informatika i ee Primeneniya — Inform.Appl.* 9(2):14–22.
- Sevastianov, B. A. 1957. Ergodicheskaya teorema dlya markovskikh protsessov i ee prilozhenie k telefonnym sistemam s otkazami [An ergodic theorem for markov processes and its application to telephone systems with refusals]. *Teor. Veroyatnost. i Primenen.* [Probability Theory and Its Applications] 2(1):106–116.
- Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
- Meykhanadzhyan, L., and R. Razumchik. 2016. New scheduling policy for estimation of stationary performance characteristics in single server queues with inaccurate job size information. *30th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 710–716.
- Jerri, A. 1999. *Introduction to integral equations with applications*. New York, NY: John Wiley & Sons. 272 p.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. *Numerical recipes: The art of Scientific computing*. 3rd ed. 1256 p.
- Polyanin, A. D., and A. V. Manzhurov. 2008. *Handbook of integral equations*. Boca Raton — London: Chapman & Hall, CRC Press. 1108 p.

Received April 19, 2016

Contributor

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

Агаларов Мурад Яверович (р. 1987) — начальник отдела системного анализа ПАО Промсвязьбанк

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, заведующий сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Алексей Владимирович (р. 1995) — студент Московского физико-технического института

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Исаченко Роман Владимирович (р. 1994) — студент Московского физико-технического института

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, заведующий сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Марков Александр Сергеевич (р. 1993) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Мейханаджян Лусине Акобовна (р. 1990) — аспирант Российского университета дружбы народов

Миллер Григорий Борисович (р. 1980) — кандидат технических наук, научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Минин Владимир Александрович (р. 1941) — доктор физико-математических наук, консультант Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Монахов Михаил Михайлович (р. 1993) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Палионная София Игоревна (р. 1995) — студентка кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Серебрянский Сергей Михайлович (р. 1983) — старший преподаватель Троицкого филиала Челябинского государственного университета

Синицын Владимир Игоревич (р. 1968) — доктор физико-математических наук, доцент, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тырсин Александр Николаевич (р. 1961) — доктор технических наук, ведущий научный сотрудник Научно-инженерного центра «Надежность и ресурсы больших систем и машин» Уральского отделения Российской академии наук

Ульянов Владимир Васильевич (р. 1953) — доктор физико-математических наук, профессор факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Ушаков Владимир Георгиевич (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хавансков Валерий Александрович (р. 1950) — научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хохлов Юрий Степанович (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шубников Сергей Константинович (р. 1955) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация 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. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», формат которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном виде (в 2-х экз.) или в электронном виде (отсканированная копия подписанного документа).

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 ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 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;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://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 hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal”–kompozit [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
 - mathematical methods for studying complex systems and processes;
 - information systems and networks;
 - information technologies; and
 - architecture and software of computational complexes and networks.
1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 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;
 - Acknowledgments.
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 Publs. 352 p.
- Dissertation and Thesis:**
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
 - an article and additional materials have to be attached using the “attach” function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)
<http://www.ipiran.ru/english/journal.asp>