

Информатика и её применения

Том 9 Выпуск 2 Год 2015

СОДЕРЖАНИЕ

Коалиционно устойчивые эффективные равновесия в моделях коллективного поведения с обменом информацией	
Н. С. Васильев	2
Время ожидания в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом	
Л. А. Мейханаджян, Т. А. Милованова, Р. В. Разумчик	14
Моделирование нормальных процессов в стохастических системах со сложными трансцендентными нелинейностями	
И. Н. Сеницын, В. И. Сеницын, Э. Р. Корепанов	23
Нормальные условно-оптимальные фильтры Пугачёва для дифференциальных стохастических систем, линейных относительно состояния	
И. Н. Сеницын, Э. Р. Корепанов	30
Применение уравнения Пугачёва–Свешникова к решению задачи Бакстера о длительности выбросов	
С. В. Березин, О. И. Заяц	39
Предельные распределения для характеристик при усвоении данных наблюдений в стационарном режиме	
К. П. Беляев, Н. П. Тучкова	50
Сравнительный анализ применения эвристического и метаэвристического алгоритмов к задаче о школьном автобусе	
Е. М. Бронштейн, Д. М. Вагапова	56
Модели для представления и обработки референсных значений	
М. П. Кривенко	63
Согласование прогнозов при решении задач прогнозирования иерархических временных рядов	
М. М. Стенина, В. В. Стрижов	75
Непараметрическое оценивание многомерной плотности с помощью вейвлет-оценок одномерных проекций	
О. В. Шестаков	88
Ассоциативные портреты предметной области — инструмент автоматизированного построения систем big data для извлечения знаний: теория, методика, визуализация, возможное применение	
И. В. Галина, Е. Б. Козеренко, Ю. И. Морозова, Н. В. Сомин, М. М. Шарнин	93
Индикаторы тематических взаимосвязей отраслей науки и информационно-компьютерных технологий в начале XXI века	
В. А. Минин, И. М. Зацман, В. А. Хавансков, С. К. Шубников	111
Об авторах	124
Правила подготовки рукописей	126
Requirements for manuscripts	129

КОАЛИЦИОННО УСТОЙЧИВЫЕ ЭФФЕКТИВНЫЕ РАВНОВЕСИЯ В МОДЕЛЯХ КОЛЛЕКТИВНОГО ПОВЕДЕНИЯ С ОБМЕНОМ ИНФОРМАЦИЕЙ

Н. С. Васильев¹

Аннотация: Теория игр изучает процессы принятия решений в условиях неопределенности и конфликта интересов. В связи с развитием сетевых технологий практический интерес вызывают модели операций, в которых участники обмениваются информацией с целями принятия рациональных решений. В статье проведена аксиоматизация понятия стратегии. Обмен сведениями о стратегиях моделирует коллективные усилия игроков, направленные на уменьшение неопределенности, присущей этому процессу. В основе рационального выбора стратегий лежат два основных принципа — эффективности и равновесия (устойчивости), которые, как правило, противоречат один другому. Доказано, что надлежащий обмен информацией позволяет осуществить выбор устойчивой эффективной ситуации игры. Исследовано строение эффективных равновесных стратегий. Введено понятие коалиционно устойчивого решения игры, обобщающее понятие равновесия Нэша на случай, когда игроки могут вступать в коалиции. Доказана достижимость этого решения с помощью расширения игры введением дополнительного контролирующего игрока и надлежащего информационного обмена.

Ключевые слова: игра; стратегия; ситуация; исход игры; передача информации; динамика принятия решений; аксиоматизация; коалиция; кооперативная игра; характеристическая функция игры; наилучший гарантированный результат; стратегия наказания; равновесие Нэша; эффективность по Парето

DOI: 10.14357/19922264150201

1 Введение

Сложность формализации и исследования игровых задач, связанных с принятием рациональных решений, вынуждает ограничиваться изучением игр с малым числом игроков и с простейшими видами стратегий, например *стратегиями-константами*. Из-за этого классические модели не соответствуют приложениям, в которых наблюдаются весьма разнообразные способы коллективного поведения игроков. Нельзя заранее преуменьшать уровень рефлексии игроков. Требуется предусматривать самые общие классы стратегий. Эти классы должны включать функционалы, с помощью которых игроки осуществляют свой выбор в зависимости от имеющейся информации. Исследования в этом направлении ранее проводились школой Ю. Б. Гермейера. Речь идет о работах, посвященных изучению метаигр и игр с фиксированным порядком ходов [1–3]. Развитие понятия стратегии происходило в теории динамических игровых задач, в позиционных дифференциальных и многошаговых играх.

Библиографические ссылки на классические труды по теории игр и многочисленные примеры применяемых стратегий содержатся, например, в работах Дж. Неймана, Э. Цермело, Н. Н. Красов-

ского, А. И. Субботина, В. И. Горелика, А. Ф. Кононенко, Л. А. Петросяна, В. И. Жуковского [1–10]. К данному исследованию наиболее близки результаты Ю. Б. Гермейера и Н. С. Кукушкина, полученные для игр Γ^1 , Γ^2 и Γ^3 двух лиц с *фиксированным* порядком ходов и обменом информацией [3, 4].

Игры многих лиц N , $N > 2$, практически не изучены. Это связано с существенным усложнением задачи изучения рационального поведения игроков с ростом N . Развитие сетевых технологий и компьютерной техники стимулирует исследование моделей коллективного поведения [11]. Например, в игровой постановке задачи маршрутизации пакетной сети передачи данных число игроков измеряется десятками тысяч [12]. Все участники игры заинтересованы в рациональном разрешении конфликта интересов. Для этого они могут обмениваться информацией о применяемых стратегиях. Можно обеспечить достижимость ситуаций равновесия, если увеличить информированность игроков и тем самым расширить их классы стратегий. Так, с помощью использования стратегий-функций может проводиться численный поиск равновесий Нэша матричных игр в классе стратегий-констант [13].

В теории игр предложены два основных принципа рационального выбора решений, вообще говоря,

¹Московский государственный технический университет им. Н. Э. Баумана, nik8519@yandex.ru

противоречащих один другому. Это принципы равновесия Нэша и эффективности по Парето [14]. В работе поставлена и решена задача о возможности рационального выбора, удовлетворяющего одновременно обоим этим принципам. Для этого необходим обмен информацией между игроками.

Более сложное поведение игроков включает возможность кооперации. При «справедливом дележе» выигрыша, получаемого коалицией, ее участники могут рассчитывать на увеличение своих платежей [1, 6–8]. В статье введено понятие *коалиционной устойчивости* решения игры, «выдерживающего» возможность объединения игроков в коалиции. Доказано, что существуют эффективные коалиционно устойчивые равновесия игры.

2 Постановка игровой задачи

Пусть в игровой операции принимают участие игроки $i \in I = \{1, 2, \dots, n + 1\}$, интересы которых заключаются в максимизации своих критериев эффективности:

$$w_i = f_i(x_1, \dots, x_{n+1}), \quad x_i \in X_i, \quad i \in I. \quad (1)$$

Один из игроков, пусть это $n + 1$, отвечает природной неопределенности, «безразличие» которой моделируется с помощью функции $f_{n+1} = const$. Целью каждого i -го игрока является выбор контролируемого им фактора x_i из некоторого множества значений X_i , характеризующего возможности этого игрока в рассматриваемой операции. Не исключено, что некоторые *исходы* $x = (x_1, \dots, x_{n+1})$ игры запрещены. Тогда на выбор игроков дополнительно влияют ограничения вида $x \in Q_i$, в которых множества $Q_i \subset X \equiv X_1 \times X_2 \times \dots \times X_{n+1}$.

Согласно (1) результат (выигрыш) любого игрока зависит не только от его действий, но и от выбора параметров $x_j, j \neq i$, контролируемых остальными участниками игры. Поэтому ни один из игроков не имеет возможности точно прогнозировать величину получаемого им выигрыша и вынужден принимать решение о выборе значения своего параметра в условиях «хаоса» будущих результатов игры. С помощью фиксирования разнообразных *классов* стратегий моделируются возможные способы поведения игроков. Классы применяемых стратегий зависят от информированности игрока об обстановке игры и о поведении партнеров. В стратегию игроков входит возможность добровольной передачи информации о своем выборе стратегии, что влияет на классы стратегий, применяемых партнерами. Выбор стратегии называется ходом игрока. Ходы всегда делаются в некоторой очередности и должны приводить к некоторой *ситуации (исходу)* $x \in X$.

За счет введения дополнительного игрока, называемого «природой», обычно проводится объективизация обстановки проведения игровой операции, благодаря которой можно считать, что каждому игроку точно известны цели и возможности всех участников игры, заданной в нормальной форме (1). Руководствуясь принципом наилучшего гарантированного результата или располагая знанием вероятностного распределения, определенно-го на множестве параметров природы X_{n+1} , обычно устраняют природную неопределенность, должным образом модифицируя критерии остальных, «разумных» игроков [3]. Поэтому далее считаем, что $I = \{1, 2, \dots, n\}$ и функции выигрыша $f_i: X \rightarrow R, i \in I$, точно известны всем игрокам.

Каждый игрок может располагать некоторой информацией о факторах x_j , выбираемых партнерами, например возможен добровольный обмен этой информацией. Если игрок i знает x_j , то удобно считать, что значение фактора x_j ему сообщает сам выбирающий игрок j . Процесс (динамика) принятия решений с учетом информированности игроков формализуется в понятии стратегии.

Стратегией i -го игрока назовем функционал \tilde{x}_i вида:

$$\tilde{x}_i = x_i(x_{J_1(i)}, \tilde{x}_{J_2(i)}), \quad i \in I. \quad (2)$$

Для упрощения записи здесь введено обозначение $x_J = x_{J(i)}$ для вектора $(x_j, j \in J(i))$. Иными словами, стратегия (2) — это правило выбора значения контролируемого фактора $x_i \in X_i$ в зависимости от имеющейся информации о действиях других игроков — результатов их выборов. Индексные множества обладают свойствами

$$J_1(i) \cap J_2(i) = \emptyset, \\ i \notin J(i) = J_1(i) \cup J_2(i), \quad i = 1, 2, \dots, n,$$

означающими невозможность одновременной передачи информации о выбираемом значении $x_i \in X_i$ контролируемого фактора и о стратегии этого выбора \tilde{x}_i .

Выбор допустимых стратегий игроков $\tilde{x} \in \tilde{X}$ должен приводить к однозначно определенной ситуации игры x , определяющей их платежи (1). Составной частью применяемой стратегии может служить добровольная передача сведений о своем выборе [3]. Подобное поведение зачастую мотивировано близостью интересов взаимодействующих игроков и уменьшением неопределенности прогнозируемой ситуации игры с целями нахождения рациональных решений \tilde{x}^* . Классам стратегий \tilde{X}_i (2) отвечает *отношение предшествования* \tilde{D} , задаваемое на множестве ходов игроков. Каждому участнику конфликта приходится решать, когда сделать свой

ход — осуществить выбор стратегии и, возможно, сообщить его кому-либо из партнеров по игре, еще не определившихся со своим ходом. Передаваемые данные влияют на вид стратегий партнеров (см. (2)). Последние делают свой ход позже, учитывая получаемую информацию. В иерархических системах и в многошаговых играх классы \tilde{X}_i заданы правилами игры [15].

2.1 Динамика принятия решений

Очередность ходов в игре и классы применяемых стратегий формализуем с помощью некоторого бинарного антисимметричного отношения \tilde{D} , задание которого определяет классы применяемых стратегий (2). Если между игроками отсутствует обмен информацией, то берется тривиальное отношение $\tilde{D} = \emptyset$, означающее, что все игроки выполняют свои ходы одновременно и используют лишь стратегии-константы. В свою очередь, отношение \tilde{D} и соответствующие информационные обмены однозначно определены заданием классов стратегий (2). Наличие аргументов $\tilde{x}_{J_2(i)}$ функционала \tilde{x}_i (2) выделяет игроков $j, j \in J_2(i)$, которые делают свой ход *раньше* i -го игрока. Отношение предшествования наглядно представимо *диаграммой* \tilde{D} . В случае $(j, i) \in \tilde{D}$ будем говорить, что элемент j множества I «меньше» элемента $i \in I$, а элемент i «больше» j . Наличие дуги (j, i) равносильно тому, что игрок j ходит раньше игрока i . Транзитивное замыкание отношения \tilde{D} определяет частичный порядок на множестве игроков I [16]. Требование непротиворечивости процессов, связанных с обменом информацией и принятием решений, накладывает определенные ограничения на классы допустимых стратегий игроков (2), т. е. на вид диаграмм \tilde{D} .

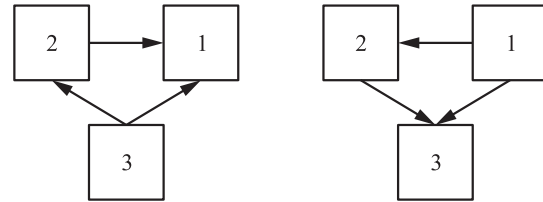
С классами функций (2) связано еще одно отношение D , *двойственное* к отношению \tilde{D} . С частичной упорядоченностью D происходит окончательный выбор игроками контролируемых факторов x_i как результат сделанных ходов. Выбор величин $x_j \in X_j, j \in J_1(i)$, должен произойти *раньше* x_i , ведь x_i — значение функции \tilde{x}_i (2). В паре $(j, i) \in D$, образующей дугу графа D , первый элемент, j , будем называть «меньшим», а второй, i , — «большим» по отношению D . Транзитивное замыкание отношения D является двойственным к \tilde{D} частичным порядком, в котором вычисляются значения параметров $x_j \in X_j$. Заметим, что пересечение отношений $\tilde{D} \cap D = \emptyset$ — тривиальный порядок на множестве I . Таким образом, вершины обеих диаграмм ходов отвечают игрокам. По стрелкам (j, i) передаются данные $\tilde{x}_j, j \in J_2(i)$, или $x_j, j \in J_1(i)$,

в зависимости от того, о какой из диаграмм \tilde{D} или D идет речь. Задание диаграмм \tilde{D} и D равносильно определению классов стратегий (2). Диаграмма \tilde{D} представляет стратегии игроков как функционалы. В двойственной к \tilde{D} диаграмме D стратегии рассматриваются как функции. Ситуация игры, отвечающая выбору стратегий из классов (2), определяется в результате вычислений (подстановок), проводимых в соответствии с этими диаграммами.

Пример 2.1. Построим диаграммы очередности ходов и выбора контролируемых параметров \tilde{D} и D для игры трех лиц, в которой игроки применяют следующие классы стратегий

$$\begin{aligned} \tilde{x}_1 &= x_1(\tilde{x}_2, \tilde{x}_3); \\ \tilde{x}_2 &= x_2(x_1, \tilde{x}_3); \\ \tilde{x}_3 &= x_3(x_1, x_2). \end{aligned}$$

Соответствующие диаграммы \tilde{D} и D отвечают полным взаимно обратным ($\tilde{D}^{-1} = D$) порядкам ходов, имеющим соответственно вид:



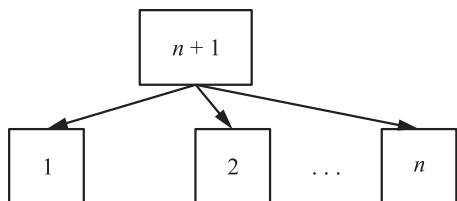
Замечание 2.1. Игроки могут сообщать не обязательно полную информацию о своих стратегиях. Для однозначности «восстановления» классов стратегий (2) по виду диаграмм \tilde{D} и D достаточно пометить их дуги теми частичными данными, которые по ним передаются. Если некоторые игроки блефуют, то пометка соответствующих дуг должна использовать другие обозначения, не совпадающие с переменными x и \tilde{x} .

Пример 2.2 (игра с «контролирующим» участником). Пусть контролирующий игрок $i = n + 1$ имеет целевую функцию $f_{n+1} = const$ и располагает знанием ходов \tilde{x}_j остальных игроков $i = 1, 2, \dots, n$. Он распорядится выбором параметра $x_{n+1} \in X_{n+1}$, где

$$\begin{aligned} X_{n+1} &= \bigcup_{j=1,2,\dots,n} \{x_{n+1} = (x_{n+1}^1, \dots, x_{n+1}^n) : \\ & x_{n+1}^i \in \{j\} \times X_j\} \cup 0, 0 \in R^n. \end{aligned}$$

При этом игрок $n + 1$ сообщает каждому участнику конфликта $i, i = 1, 2, \dots, n$, лишь i -ю координату выбираемого им вектора x_{n+1} . Выбор значения $x_{n+1} = 0$ используется в качестве признака «желательного» коллективного поведения игроков

$i = 1, 2, \dots, n$, например, приводящего к некоторому, вполне определенному исходу игры. Таким образом, по стрелкам $(n + 1, i)$ диаграммы D вида



передаются значения x_{n+1}^i . Подобные диаграммы D характерны для иерархических систем принятия решений [15]. Согласно сказанному выше, двойственное отношение $\tilde{D} = D^{-1}$. Итак, классы допустимых стратегий игроков содержат функции:

$$\begin{aligned} \tilde{x}_{n+1} &= x_{n+1}(\tilde{x}_1, \dots, \tilde{x}_n); \\ \tilde{x}_i &= x_i(x_{n+1}^i), \quad i = 1, 2, \dots, n. \end{aligned}$$

Теорема 2.1. *Критерием непротиворечивости процесса принятия решений для игры в нормальной форме (1) и классами стратегий $\tilde{X}_i = \{\tilde{x}_i\}$ (2) является ацикличность обеих диаграмм \tilde{D} и D . В этом случае выбор стратегий $\tilde{x} = \tilde{X} = \tilde{X}_1 \times \dots \times \tilde{X}_{n+1}$ однозначно определяет исход игры $x \in X$, так что определено отображение*

$$s: \tilde{X} \rightarrow X. \tag{3}$$

Доказательство. Пусть отображение вида (3) существует. Построим диаграммы \tilde{D} и D . Если хотя бы в одной из них найдем цикл C , то эта диаграмма не может определять отношение частичного порядка, в котором проводятся подстановки одних функций (2) в другие. Тогда вопреки допущению не определено отображение (3). Установим теперь достаточность условия ацикличности (отсутствие циклов) диаграмм \tilde{D} и D , воспользовавшись методом математической индукции. Для диаграмм, содержащих единственный узел, доказываемое утверждение очевидно. Предположим справедливость теоремы в случае всех ациклических диаграмм с не более чем $N - 1$ вершинами. Докажем, что тогда теорема верна для диаграмм с N , $N > 1$, вершинами. В ациклическом графе D существует вершина i_0 , отвечающая минимальному элементу отношения D [16]. Узел i_0 не имеет входных дуг. Если элемент i_0 является также минимальным по отношению к \tilde{D} , то игрок i_0 применяет некоторую стратегию-константу x_{i_0} (см. (2)). Подставив значение параметра x_{i_0} во все функции (2), $i \neq i_0$, исключим вершину i_0 из обеих диаграмм вместе с исходящими из нее дугами. Если теперь элемент i_0 не минимален по отношению к \tilde{D} , то по дугам $(j, i_0) \in \tilde{D}$ передаются стратегии \tilde{x}_j , $j \in J_2(i_0)$.

Для исключения узла i_0 из диаграммы D достаточно найти значение параметра x_{i_0} , вычислив значение функции $\tilde{x}_i = x_i(\tilde{x}_{J_2(i)})$, $i = i_0$. Аналогичные рассуждения применимы к начальному элементу \tilde{i}_0 диаграммы \tilde{D} , приводящие к исключению узла \tilde{i}_0 из диаграммы \tilde{D} . В полученных новых диаграммах \tilde{D}' и D' содержится не более $N - 1$ вершин. Осталось воспользоваться предположением индукции, что завершает доказательство.

Следствие 2.1. Транзитивное замыкание отношений \tilde{D} и D сохраняет непротиворечивость исходных классов стратегий.

Итак, в теореме 2.1 обоснована проведенная аксиоматизация понятия стратегии, использующая диаграммы отношений \tilde{D} и D , дуги которых помечены передаваемыми данными. Классы стратегий игроков назовем *допустимыми*, если диаграммы отношений \tilde{D} и D не содержат циклов. Стратегии из этих классов также называются *допустимыми*. Они вполне определены диаграммами \tilde{D} и D (см. пример 2.2).

В условиях теоремы 2.1 выбор всеми игроками своих стратегий (2), $\tilde{x}_i \in \tilde{X}_i$, $i \in I$, определяет исход игры $x = s(\tilde{x}) \in X$, $\tilde{x} \in \tilde{X}$, и, как следствие, выигрыши игроков, вычисляемые по формулам (3), (1), определяющие функции $\tilde{f}_i: \tilde{X} \rightarrow R$,

$$\tilde{f}_i(\tilde{x}) \triangleq f_i(s(\tilde{x})) = f_i(x), \quad i \in I. \tag{4}$$

Если имелись запрещенные исходы Q_i , то в «новой» игре (4), $\tilde{x} \in \tilde{X}$, они принимают вид $\tilde{x} \in s^{-1}Q_i$. Зависимость стратегий (2) от «параметров» \tilde{x}_j говорит о более высоком уровне рефлексии игроков по сравнению с игрой (1), $\tilde{x} \in X^0$, в которой X^0 — множество *стратегий-констант*. Дополнительная информированность $X_i^0 \subset \tilde{X}_i$ увеличивает возможности игроков по улучшению выигрышей.

В игре (4), $f' = \tilde{f}_i$, $X'_i = \tilde{X}_i$, в свою очередь, все или часть игроков могут использовать стратегии (2), $x = x'$, $\tilde{x} = \tilde{x}'$. Подобное усложнение классов применяемых стратегий приводит к играм, называемым *метаиграми* [1–3].

2.2 Рационализация выбора стратегий

В теории игр предпринята попытка разрешить конфликт между устойчивостью поведения игроков (*принципом равновесия*) и выгодностью для игроков принимаемых решений (*принципом эффективности*).

2.2.1 Принцип эффективности

Для операций (1) с векторным критерием $f: X \rightarrow R^{n+1}$ рациональный выбор игроками своих стратегий основывается на понятии *слабо*

эффективной (оптимальной по Слейтеру) ситуации x^* [8, 14, 17]. Это решение x^* таково, что не существует элемента $x \in X$, для которого $f(x) > f(x^*)$. Множество всех слабоэффективных решений игры S , как правило, не является одноэлементным. Поэтому возникает необходимость дальнейшего уточнения выбора игроков. Предпочтительней рассматривать подмножество $P \subset S$ эффективных (оптимальных по Парето) точек x^* , т. е. таких векторов, что не существует элемента $x \in X$, для которого выполнялось бы условие

$$f(x) \geq f(x^*), \quad f(x) \neq f(x^*).$$

При выполнении стандартных предположений непрерывности $f : X \rightarrow R^{n+1}$, компактности или конечности множества X доказываем, что множество P эффективных решений игры непустое. Подобные требования далее предполагаются выполненными. Из-за неоднозначности выбора $x^* \in P$ целесообразно еще более сузить множество P , ограничившись подмножеством вида $P_L = \{x \in P : f(x) \geq L\}$. Здесь вектор $L = L_I = (L_1, \dots, L_n)$ составлен из наилучших гарантированных результатов игроков

$$L_i(\tilde{X}_i) = \max_{\tilde{X}_i} \min_{\tilde{x}_j \in \tilde{X}_j, j \neq i} \tilde{f}_i(\tilde{x}), \quad i \in I,$$

в игре (1), (2). Дело в том, что никто из игроков «не согласится» получить выигрыш, меньший того, который он может сам себе гарантировать. Здесь и далее предполагаем, что если какой-либо экстремум не достигается, то берется ε -решение оптимизационной задачи с некоторой точностью $\varepsilon > 0$ по значению критерия.

Введем обозначение \tilde{P} для множества оптимальных по Парето стратегий $\tilde{x} \in \tilde{P}$, т. е. удовлетворяющих включению $s(\tilde{x}) \in P$. Аналогично понимаются множества $S_L, \tilde{S}_L, \tilde{P}_L$.

Лемма 2.1. Пусть все целевые функции $f_i > 0$. Тогда для любых классов допустимых стратегий множество $\tilde{P}_L \neq \emptyset$.

Доказательство. Рассмотрим какое-либо решение \tilde{x}^S оптимизационной задачи

$$\tilde{x}^S \in \text{Arg} \max_{\tilde{X}} \min_{i=1,2,\dots,n+1} \frac{\tilde{f}_i(\tilde{x})}{L_i}, \quad (5)$$

служащей критерием слабой эффективности [17]. Пусть вектор $\tilde{x}^\Gamma \in \tilde{X}$ составлен из гарантирующих стратегий игроков \tilde{x}_i^Γ , т. е. таких, что

$$\tilde{x}_i^\Gamma \in \text{Arg} \max_{\tilde{x}_i} \min_{\tilde{x}_j, j \neq i} \tilde{f}_i(\tilde{x}_1, \dots, \tilde{x}_{n+1}), \quad i \in I.$$

Тогда для всех допустимых значений индекса i выполнены неравенства $\tilde{f}_i(\tilde{x}^\Gamma) \geq L_i$. Ввиду (5) для набора стратегий игроков $\tilde{x}^S \in \tilde{S}$ выполнено

$$\min_i \frac{\tilde{f}_i(\tilde{x}^S)}{L_i} \geq \min_i \frac{\tilde{f}_i(\tilde{x}^\Gamma)}{L_i} \geq 1.$$

Следовательно, исход игры $x^S = s(\tilde{x}^S)$ содержится в множестве S_L . Если $x^S \notin P_L$, то результат x^S , по определению эффективности, может быть улучшен выбором вектора x^P , для которого $f(x^P) \geq f(x^S)$. По построению $x^P \in P_L$.

2.2.2 Принцип устойчивости

В качестве рационального выбора игроков понимается ситуация равновесия Нэша \tilde{x}^* , т. е. такой набор стратегий игроков, который определяется условием:

$$(\forall i)(\forall \tilde{x}_i) \tilde{f}_i(\tilde{x}_1^*, \dots, \tilde{x}_{i-1}^*, \tilde{x}_i, \tilde{x}_{i+1}^*, \dots, \tilde{x}_n^*) \leq \tilde{f}_i(\tilde{x}^*). \quad (6)$$

По смыслу соотношений (6) в ситуации \tilde{x}^* каждому игроку $i, i \in I$, не выгодно в одиночку отклоняться от своей стратегии \tilde{x}_i^* . Это обеспечивает равновесность ситуации \tilde{x}^* .

Теорема 2.2. Если классы допустимых стратегий таковы, что $\tilde{X}_i \subset \tilde{X}'_i, i = 1, 2, \dots, N$, то множества равновесий игр (1), (4), $\tilde{X}_i, u(1), (4), s = s', \tilde{X}'_i$, связаны включением $\tilde{X}^* \subset \tilde{X}'^*$.

Доказательство. Расширение классов стратегий за счет обмена информацией соответствует добавлению стрелок в диаграммы упорядоченности ходов \tilde{D} и D . При этом по теореме 2.1 нельзя нарушать свойство ацикличности получаемых графов. Поэтому достаточно провести доказательство методом математической индукции, проводимой по числу стрелок в диаграммах \tilde{D} или D . Пусть $\tilde{x}^* \in \tilde{X}^*$ и для определенности в диаграмму \tilde{D} включена новая дуга (j, i) . Тогда, по сравнению с исходной игрой, участник i получает дополнительную информацию о выборе стратегии \tilde{x}_j . Зафиксировав $\tilde{x}_j^*, j \neq i$, найдем стратегию \tilde{x}'_i , оптимальную по i -му критерию эффективности (1) при $\tilde{x}'_i \in \tilde{X}'_i$. Изучим ситуацию $\tilde{x}' = (\tilde{x}'_1, \dots, \tilde{x}'_{i-1}, \tilde{x}'_i, \tilde{x}'_{i+1}, \dots, \tilde{x}'_{n+1})$. Подстановка $\tilde{x}_j = \tilde{x}_j^*$ в функцию \tilde{f}'_i , исключая зависимость этой стратегии от переменной \tilde{x}_j , определяет некоторую стратегию i -го игрока \tilde{x}_i из класса \tilde{X}_i . Поэтому в ситуации равновесия Нэша $\tilde{x}^* \in \tilde{X}^*$ имеем

$$\begin{aligned} \tilde{f}_i(\tilde{x}^*) &= \max_{\tilde{X}_i} \tilde{f}_i(\tilde{x}_1^*, \dots, \tilde{x}_{i-1}^*, \tilde{x}_i, \tilde{x}_{i+1}^*, \dots, \tilde{x}_{n+1}^*) = \\ &= \max_{\tilde{X}'_i} \tilde{f}_i(\tilde{x}') \end{aligned}$$

и можно считать, что $\tilde{x}'_i = \tilde{x}_i^*$. Аналогичные равенства справедливы для всех $j \neq i$. Значит, $\tilde{x}^* \in \tilde{X}'^*$, т. е. $\tilde{X}^* \subset \tilde{X}'^*$, что и требовалось доказать.

Следующий результат служит обобщением теоремы Цермело из теории многошаговых антагонистических игр [17]. Игру назовем игрой с *полной информацией*, если в ней одно из отношений предшествования ходов \tilde{D} и D является полным.

Теорема 2.3. *В игре с полной информацией существует ситуация равновесия.*

Доказательство. Пусть для определенности \tilde{D} — полное отношение порядка. Воспользуемся методом динамического программирования. По условию теоремы у игроков существуют *абсолютно оптимальные* стратегии $\tilde{x}_i^a = x_i^a(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_{i-1})$, которые являются решениями следующих задач:

$$\begin{aligned} \tilde{x}_i^a = \operatorname{Arg} \max_{\tilde{x}_i} & \tilde{f}_i(\tilde{x}_1, \tilde{x}_2, \dots \\ & \dots, \tilde{x}_i, x_{i+1}(\tilde{x}_1, \dots, \tilde{x}_{i-1}, \tilde{x}_i), \dots \\ & \dots, x_{n+1}(\tilde{x}_1, \dots, \tilde{x}_{i-1}, \tilde{x}_i, \tilde{x}_{i+1}, \dots, \tilde{x}_n)). \end{aligned} \quad (7)$$

Тогда ситуация равновесия Нэша — это набор функций:

$$\begin{aligned} \tilde{x}^* = (\tilde{x}_1^a, \tilde{x}_2^a, \dots, \tilde{x}_{n+1}^a), \\ \tilde{x}_1^a \in X_1^0, \tilde{x}_1^a \equiv x_1^a, x_1^a \in X_1. \end{aligned}$$

Проверка формулы (6) не вызывает затруднений. Ситуацию равновесия $\tilde{x}^* \in \tilde{X}^*$ можно вычислить, последовательно решая задачи (7) при $i = n, n-1, \dots, 1$.

Не исключено, что в игре (1), (2) множество ситуаций равновесия пустое. Как показывают теоремы 2.2 и 2.3, за счет обмена информацией можно расширить классы допустимых стратегий так, что в новой игре появятся равновесия $\tilde{X}^* \neq \emptyset$. В статье [13] рассмотрен численный метод поиска ситуации равновесия бескоалиционной игры в классе смешанных стратегий-констант с помощью повторяющейся игры со смешанными стратегиями-функциями.

Применение принципа равновесия ограничивается возможной неэффективностью и неэквивалентностью для игроков имеющих ситуации равновесия. Эффективная ситуация равновесия вполне заслуживает названия *оптимального* решения игры.

3 Существование эффективных ситуаций равновесия

Поставим задачу выделения классов стратегий игроков (2), в которых существует это оптимальное решение.

3.1 Эффективные равновесия Нэша

Знание обстановки проведения операции позволяет игрокам самостоятельно, без обмена сведениями, находить свои гарантирующие стратегии. Если все игроки $i' \in I(i)$ знают стратегию \tilde{x}_i , которую выбирает игрок i , то они могут осуществить «наказание» партнера i , применяя стратегию *наказания* $\tilde{x}^{\text{Hi}} \equiv \tilde{x}_{I(i)}^{\text{Hi}}$ [3]:

$$\tilde{x}^{\text{Hi}} \in \operatorname{Arg} \min_{\tilde{x}_j, j \neq i} \tilde{f}_i(\tilde{x}_1, \dots, \tilde{x}_{n+1}), \quad \tilde{x}^{\text{Hi}} \in \prod_{j \neq i} \tilde{X}_j.$$

Пример 3.1. В примере 2.1 первого игрока наказывают выбором стратегий $\tilde{x}^{\text{H1}} = (\tilde{x}_2^{\text{H1}}, \tilde{x}_3^{\text{H1}})$, равных

$$\begin{aligned} \tilde{x}_2^{\text{H1}} = x_2^{\text{H1}}(x_1, \tilde{x}_3) & \in \operatorname{Arg} \min_{x_2} \tilde{f}_1(x_1, x_2, \tilde{x}_3); \\ \tilde{x}_3^{\text{H1}} = x_3^{\text{H1}}(x_1, x_2) & \in \operatorname{Arg} \min_{x_3} f_1(x_1, x_2, x_3). \end{aligned}$$

Теорема 3.1. *Пусть $f_i > 0$, $i \in I$, а диаграммы \tilde{D} и $D = \tilde{D}^{-1}$ отвечают полным порядкам на множестве I . Тогда существует эффективная ситуация равновесия Нэша \tilde{x}^* .*

Доказательство. В допустимом классе стратегий \tilde{X}_i (2) определим наилучший гарантированный результат $L_i = L_i[\tilde{X}_i]$ и гарантирующую стратегию \tilde{x}_i^Γ произвольного i -го игрока. По лемме 2.1 существует эффективная ситуация — элемент $x^\exists \in P_L$. Проведем необходимое обоснование для случая $n = 3$ игроков, как не ограничивающего общности. Тогда находимся в условиях и обозначениях из примеров 2.1 и 3.1. Дополнительно определим функцию $\bar{x}: X_1 \rightarrow X$ вида $\bar{x}(x_1) = s(x_1, \tilde{x}_2^{\text{H1}}, \tilde{x}_3^{\text{H1}})$ и результат игры $\bar{\bar{x}} = s(\tilde{x}_1^{\text{H2}}, \tilde{x}_2^{\text{H1}}, \tilde{x}_3^{\text{H2}})$ для указанной ситуации. Рассмотрим допустимые стратегии игроков:

$$\tilde{x}_1^* = \begin{cases} x_1^\exists, \tilde{x}_j = x_j^*, j = 2, 3; \\ \bar{\bar{x}}_1, \tilde{x}_2 = \tilde{x}_2^{\text{H1}}; \\ \tilde{x}_1^{\text{H2}}, \tilde{x}_2 \neq \tilde{x}_2^*; \tilde{x}_3 = \tilde{x}_3^*; \\ \tilde{x}_1^{\text{H3}}, \tilde{x}_2 = \tilde{x}_2^*, \tilde{x}_3 \neq \tilde{x}_3^*; \end{cases} \quad (8)$$

$$\tilde{x}_2^* = \begin{cases} x_2^\exists, x_1 = x_1^\exists, \tilde{x}_3 = \tilde{x}_3^*; \\ \bar{x}_2(x_1), x_1 \neq x_1^\exists, \tilde{x}_3 = \tilde{x}_3^*; \\ \tilde{x}_2^{\text{H3}}, \tilde{x}_3 \neq \tilde{x}_3^*; \end{cases} \quad (9)$$

$$\tilde{x}_3^* = \begin{cases} x_3^\exists, x_1 = x_1^\exists, x_2 = x_2^\exists; \\ \tilde{x}_3^{\text{H1}}, x_1 \neq x_1^\exists, x_1 \neq \bar{\bar{x}}_1, x_2 = \bar{x}_2(x_1); \\ \tilde{x}_3^{\text{H2}}, x_1 \neq x_1^\exists, x_1 \neq \bar{\bar{x}}_1, x_2 \neq \bar{x}_2(x_1); \\ \tilde{x}_3^{\text{H3}}, x_1 = \bar{\bar{x}}_1, x_2 = \bar{\bar{x}}_2. \end{cases} \quad (10)$$

Убедимся в том, что выбор стратегий (8)–(10) обладает всеми доказываемыми свойствами $\tilde{x}^* \in \tilde{X}^* \cap \tilde{P}_L$. Во-первых, согласно (8)–(10), справедливы неравенства:

$$\begin{aligned} \tilde{f}_i(\tilde{x}_1^*, \dots, \tilde{x}_{i-1}^*, \tilde{x}_i, \tilde{x}_{i+1}^*, \dots, \tilde{x}_{n+1}^*) &= \\ &= \tilde{f}_i(\tilde{x}_1^{ni}, \dots, \tilde{x}_{i-1}^{ni}, \tilde{x}_i, \tilde{x}_{i+1}^{ni}, \dots, \tilde{x}_{n+1}^{ni}) = \\ &= \min_{\tilde{x}_j, j \neq i} \tilde{f}_i(\tilde{x}_1, \dots, \tilde{x}_{n+1}) \leq \max_{\tilde{x}_i} \min_{\tilde{x}_j, j \neq i} f_i(\tilde{x}) = L_i \leq \\ &\leq \tilde{f}_i(\tilde{x}^*), \quad i = 1, 2, \dots, n. \end{aligned}$$

Поэтому \tilde{x}^* является равновесием Нэша. Во-вторых, $\tilde{f}(\tilde{x}^*) = f(x^*)$, $x^* \in P_L$. Значит, стратегия $\tilde{x}^* \in \tilde{P}_L$.

Теорема 3.2. В игре с контролирующим игроком существует эффективное равновесное решение.

Доказательство. По лемме 2.1 существует элемент $x^* \in P_L$. В условиях примера 2.2 контролирующий игрок знает, кто из игроков $j = 1, 2, \dots, n$ нарушает условие $\tilde{x}_j = \tilde{x}_j^*$, где $\tilde{x}_1^*, \dots, \tilde{x}_n^*$ — набор заранее фиксированных стратегий игроков $k = 1, 2, \dots, n$. Тогда он знает значение параметра $\bar{x}_j = \tilde{x}_j(j, x_j^*)$. Докажем, что искомым решением является следующий выбор допустимых стратегий игроков:

$$\begin{aligned} \tilde{x}_{n+1}^* &= (\tilde{x}_{n+1}^{1*}, \dots, \tilde{x}_{n+1}^{n*}); \\ \tilde{x}_{n+1}^{k*} &= x_{n+1}^{k*}(\tilde{x}_1, \dots, \tilde{x}_n) = \\ &= \begin{cases} (j, \bar{x}_j), (\exists! j) \tilde{x}_j \neq \tilde{x}_j^*, k \neq j; \\ (j, x_j^*), k = j; \\ 0, (\forall i) \tilde{x}_i = \tilde{x}_i^*, \end{cases} \quad k = 1, 2, \dots, n; \\ \tilde{x}_i^* &= x_i^*(x_{n+1}^i) = x_i^*(j, x_j) = \\ &= \begin{cases} x_i^{nj}(x_j), i \neq j; \\ x_i^*, x_{n+1}^i = 0 \vee i = j, \end{cases} \quad i = 1, 2, \dots, n. \end{aligned}$$

Игрок $n + 1$ сообщает игрокам $k \neq j$ величину $x_{n+1}^k = (j, \bar{x}_j)$. Поэтому становится возможным наказание «нарушителя» совместными действиями остальных игроков. В результате будет реализован такой исход игры x , для которого $f_j(x) \leq L_j \leq \tilde{f}_j(\tilde{x}^*) = f_j(x^*)$.

3.2 Коалиционно устойчивое эффективное решение

В ситуации равновесия Нэша никому из игроков в одиночку не выгодно отклоняться от своего решения. За счет коллективных действий игроки вполне могут рассчитывать на увеличение своих выигрышей. Изучен вопрос существования ситуаций равновесия, которые нецелесообразно нарушать с помощью формирования коалиций. Под

кооперативной игрой понимаем игру, в которой не исключено образование коалиций.

3.2.1 Характеристическая функция кооперативной игры

У всякой коалиции $K \subset I$ имеется критерий $f_K(x)$, с помощью которого оценивается качество коалиционного решения x_K , совместно принимаемого игроками $k \in K$. Рациональное поведение коалиции K заключается в стремлении по возможности максимизировать коалиционную функцию выигрыша $f_K(x)$. Обычно рассматривают функции f_K вида:

$$f_K(x) = \min_{k \in K} f_k(x); \quad (11)$$

$$f'_K(x) = \sum_{k \in K} f_k(x). \quad (12)$$

В отличие от (11), аддитивный критерий (12) отвечает кооперативной игре с неограниченными побочными платежами [3]. Игроки $k \in K$ делят между собой их суммарный выигрыш (12). Если решение о формировании коалиции принимается игроками в некоторой ситуации x^0 , для которой $f_k(x^0) > 0$, то вместо (11) целесообразно использовать платежные функции

$$f'_K(x^0, x) = \min_{k \in K} \frac{f_k(x)}{f_k(x^0)}. \quad (13)$$

Рациональное поведение коалиции K заключается в выборе эффективной по Парето стратегии $\tilde{x}_K^*(x_{I \setminus K})$ по векторному критерию \tilde{f} , если ей известен выбор параметра $x_{I \setminus K}$ дополняющей коалиции $I \setminus K$ [14, 17].

Введем характеристическую функцию игры $K \rightarrow L_K$, $K \in 2^I$, которая каждой коалиции ставит в соответствие ее наилучший гарантированный результат в классе стратегий-констант X^0 . Для коалиционных критериев (11) характеристическая функция игры равна

$$L_K = \max_{x_K} \min_{x_{I \setminus K}} \min_{k \in K} f_k(x). \quad (14)$$

При нормировке (13) вместо $f_k(x)$ в (14) следует взять величину $f_k(x)/f_k(x^0)$.

Кооперативная игра (12) имеет характеристическую функцию

$$L_K = \max_{x_K} \min_{x_{I \setminus K}} \sum_{k \in K} f_k(x). \quad (15)$$

Пусть коалиция K_1 , $K_1 \subset K$ представлена игроком K_1 , которому известен вектор x_{K_2} , выбираемый подкоалицией K_2 , $K_2 = K \setminus K_1$, коалиции K . Наилучшим условным гарантированным результатом коалиции K_1 назовем величину $L_{K_1}(x_{K_2})$ наилучшего гарантированного результата игрока K_1 .

Лемма 3.1. *Характеристическая функция (15) супераддитивна: $L_{K_1} + L_{K_2} \leq L_K$, а функция (14) удовлетворяет неравенству:*

$$\min \{L_{K_1}, L_{K_2}\} \leq L_K \leq \min \left\{ \max_{x_{K_2}} L_{K_1}(x_{K_2}), \max_{x_{K_1}} L_{K_2}(x_{K_1}) \right\}.$$

Найдем наилучший K -коалиционный результат $L_k(K)$ члена коалиции k , $k \in K$. Он равен $L_k(K) = \max_{x_K} \min_{x_{I \setminus K}} f_k(x)$. Аналогичная величина, вычисляемая для подкоалиции K' , $K' \subset K$, равна $L_{K'}(K) = \max_{x_K} \min_{x_{I \setminus K}} f_{K'}(x)$ и является наилучшим K -коалиционным результатом. Введенные величины связаны соотношениями

$$L_k \equiv L_k(\{k\}) \leq L_k(K); \\ L_{K'} \equiv L_{K'}(K') \leq L_{K'}(K).$$

Функцию $K' \rightarrow L_{K'}(K)$, $K' \subset K$, назовем K -характеристической. Для кооперативных игр (11), (12) I -характеристическая функция совпадает с обычной — (14), (15).

Следствие 3.1. *Характеристическая функция (14) удовлетворяет неравенству:*

$$\min_{k \in K} L_k \leq L_K \leq \min_{k \in K} L_k(K) \triangleq \bar{L}_K.$$

3.2.2 Коалиционная устойчивость результата игры

Большая выгодность коллективных действий по сравнению с индивидуальным поведением является причиной формирования коалиции. Распад коалиции происходит тогда, когда она становится невыгодной хотя бы одному из ее членов. Рассмотрим исходы игры $x^0 = s(\tilde{x})$, приводящие к выходу некоторых участников из коалиции K .

Определение 3.1. K -устойчивым назовем такой результат игры x^0 , что $\exists k_0 \in K f_{k_0}(x^0) \geq L_{k_0}(K)$.

Другими словами, исход игры x^0 устойчив относительно попытки формирования коалиции K . Результат игры I -устойчив, если x^0 является максимумом одной из платежных функций (1) $f_k : X \rightarrow R$. Выйти из полной коалиции I невозможно. Поэтому далее считаем $K \neq I$. В случае нормированного коалиционного критерия (13) в определении 3.1 нужно потребовать выполнения формулы $\exists k_0 \in K L_{k_0}(K) \leq 1$.

Если взять стратегии из теорем 3.1 и 3.2, в которые вместо вектора x^ϑ подставлен вектор x^0 , то они равновесны по Нэшу и вдобавок обеспечивают достижимость исхода x^0 . В любом равновесии \tilde{x}^* все игроки получают платежи, не меньшие их наилучших гарантированных результатов

$\tilde{f}_i(\tilde{x}^*) \geq L_i[\tilde{X}_i] \geq L_i[X_i^0] \equiv L_i$. Благодаря определению 3.1, $K = \{k_0\}$, имеем $f_{k_0}(x^0) \geq L_{k_0}(\{k_0\}) = L_{k_0}$, так что указанное неравенство выполнено для исхода игры x^0 .

Пусть дополняющая коалиция $I \setminus K$ имеет точную информацию о векторе x_K , выбираемом коалицией K . Тогда при наличии K -устойчивого исхода игры x^0 выбором стратегии наказания $\tilde{x}_{I \setminus K}^{\text{нк}} = X_{I \setminus K}^{\text{нк}}(x_K)$, $k \in K$, она может противодействовать попытке формирования коалиции K . В этом случае для некоторого игрока k_0 более привлекательным становится результат игры x^0 по сравнению с исходом $x = s(\tilde{x}_K, \tilde{x}_{I \setminus K}^{\text{нк}})$:

$$\exists k \in K \forall x_K \exists x_{I \setminus K} f_k(x_K, x_{I \setminus K}) \leq f_k(x^0).$$

Игрок k_0 не станет участвовать в коалиции K , поэтому она не сформируется.

Определение 3.2. Исход игры x^0 назовем коалиционно устойчивым, если он K -устойчив для любой коалиции $K \subset I$, $K \neq I$.

Коалиционно устойчивый результат игры x^0 обладает свойством

$$(\forall k) f_k(x^0) \geq L_k.$$

Он приводит к изоляционизму поведения игроков, препятствуя созданию коалиций.

Следствие 3.2.

1. В игре (1), (11) необходимым условием коалиционной устойчивости результата x^0 является $(\forall K \in 2^I) \wedge (K \neq I) (\exists k \in K) L_K \leq f_k(x^0)$, а достаточным условием служит выполнение $\forall K (K \in 2^I) \wedge (K \neq I) \Rightarrow L_K \leq f_K(x^0)$.
2. Пусть все K -характеристические функции в кооперативной игре (1), (12), $K \subset I$, супераддитивны. Тогда условие $\forall K (K \in 2^I) \wedge (K \neq I) \Rightarrow L_K \leq f_K(x^0)$ достаточно для коалиционной устойчивости результата игры.

Доказательство.

1. Если $(\forall K) \bar{L}_K \leq f_K(x^0)$, то по следствию 3.1 для любой коалиции K найдется такой игрок $k_0 \in K$, что справедливо неравенство

$$L_{k_0}(K) = \min_{k \in K} L_k(K) \triangleq \bar{L}_K \leq f_K(x^0) \triangleq \\ \triangleq \min_{k \in K} f_k(x^0) \leq f_{k_0}(x^0).$$

Достаточность доказана. По определению 3.2 исход игры x^0 удовлетворяет условию $(\forall K) \exists k_0 \in K f_{k_0}(x^0) \geq L_{k_0}(K)$. Снова воспользуемся следствием 3.1:

$$f_{k_0}(x^0) \geq L_{k_0}(K) \geq \min_{k \in K} L_k(K) \geq L_K.$$

Необходимое условие доказано.

2. Будем рассуждать от противного. Следствие неверно, $\exists K \forall k \in K f_k(x^0) < L_k(K)$. Просуммируем эти неравенства и воспользуемся супер-аддитивностью функции $K' \rightarrow L_{K'}(K)$:

$$f_K(x^0) = \sum_{k \in K} f_k(x^0) < \sum_{k \in K} L_k(K) \leq \leq L_K(K) = L_K.$$

Полученное неравенство $f_K(x^0) < L_K$ противоречит условию следствия.

Рассмотрим кооперативную игру из примера 2.2, в которой контролирующий игрок выбирает параметр из множества

$$X_{n+1} = \bigcup_{K: K \subset I, K \neq \emptyset} \{x_{n+1} = (x_{n+1}^1, \dots, x_{n+1}^n) : (\forall i) x_{n+1}^i \in K \times \{K\} \times X_K\} \cup \{0\},$$

где $0 \in R^n$, $X_K = \prod_{k \in K} X_k$.

Теорема 3.3. Пусть в игре (1) с участниками $k = 1, 2, \dots, n$ имеется коалиционно устойчивый исход x^0 . Тогда с помощью контролирующего игрока $n+1$ достигается коалиционно устойчивое эффективное решение.

Доказательство. Согласно определению 3.2 можно считать, что $x^0 \in P_L$. Пусть выбор ситуации \tilde{x}^* приводит к исходу игры x^0 . Контролирующему игроку известно множество «нарушителей» $K = \{k \in I : \tilde{x}_k \neq \tilde{x}_k^*\}$ результата игры x^0 . Согласно определению 3.2 он может найти того из участников $k_0 \in K_0(K)$ коалиции K , выигрыш которого не меньше K -коалиционного результата: $L_{k_0}(K) \leq \leq f_{k_0}(x^0)$. Об этом игроке k_0 он может проинформировать всех участников конфликта. Кроме того, игроку $n+1$ известен вектор \bar{x}_K с координатами $\bar{x}_k = \tilde{x}_k(k, K, x_K^0)$, $k \in K$, который он может также сообщить всем участникам дополняющей коалиции $k \in I \setminus K$. Проверим, что искомым решением игры является

$$\begin{aligned} \tilde{x}_{n+1}^* &= (\tilde{x}_{n+1}^{1*}, \dots, \tilde{x}_{n+1}^{n*}); \\ \tilde{x}_{n+1}^{k*} &= x_{n+1}^{k*}(\tilde{x}_1, \dots, \tilde{x}_n) = \\ &= \begin{cases} (k_0, K, \bar{x}_K), & k \in I \setminus K; \\ (k, K, x_K^0), & k \in K; \\ 0, & K = \emptyset; \end{cases} \end{aligned}$$

$$\begin{aligned} \tilde{x}_i^* &= x_i^*(x_{n+1}^i) = \\ &= \begin{cases} \begin{cases} x_i^{nk_0}(x_K), & i \in I \setminus K; \\ x_i^0, & i \in K, \end{cases} & x_{n+1}^i = (k, K, x_K); \\ x_i^0, & x_{n+1}^i = 0, \end{cases} \\ & \quad i \in I. \end{aligned}$$

Если $K \neq \emptyset$, то дополняющая коалиция, выбирая стратегию наказания участника коалиции k_0 , ограничивает его выигрыш величиной $L_{k_0}(K)$, меньшей $f_{k_0}(x^0)$:

$$\begin{aligned} \tilde{f}_{k_0}(\tilde{x}_K, \tilde{x}_{I \setminus K}^{nk_0}) &= f_{k_0}(s(\tilde{x}_K, \tilde{x}_{I \setminus K}^{nk_0})) \leq \\ &\leq \max_{x_K} \min_{x_{I \setminus K}} f_{k_0}(x) \triangleq L_{k_0}(K) \leq f_{k_0}(x^0). \end{aligned} \quad (16)$$

Выбор стратегии \tilde{x}^* реализует выполнение неравенства (16), препятствуя образованию коалиции. При этом $s(\tilde{x}^*) = x^0$. Теорема доказана.

Пример 3.2. Применим теоремы 3.2 и 3.3 к игре, в которой функции выигрыша игроков имеют вид

$$\begin{aligned} f_1(x) &= x_1 + x_2 - x_3; \\ f_2(x) &= x_1 - x_2 + x_3; \\ f_3(x) &= -x_1 + x_2 + x_3; \\ X_k &= [0, 1], \quad k = 1, 2, 3. \end{aligned}$$

С помощью определения 3.2 убедимся в коалиционной устойчивости эффективного по Парето вектора $x^0 = (1, 1, 1)$. Это так, потому что характеристическая функция (14) равна $L_{\{i\}} = 0$, $i = 1, 2, 3$, $L_{\{1,2\}} = L_{\{2,3\}} = 1/2$, $L_{\{1,3\}} = 0$, а выигрыши игроков $f_k(x^0) = 1$, $k = 1, 2, 3$. Другие эффективные ситуации равновесия Нэша, построенные по теореме 3.2, не являются коалиционно устойчивыми, например решение, которое строится по вектору $x^0 = (0, 1, 1)$.

Рассмотрим характеристическую функцию (15) $L'_{\{i\}} = 0$, $i = 1, 2, 3$, $L'_{\{i,j\}} = 2$, $i \neq j$. Исход $x^0 = (1, 1, 1)$ по-прежнему коалиционно устойчив. С помощью теоремы 3.3 можно построить коалиционно устойчивое эффективное решение этой игры.

Пример 3.3. Функции выигрыша игроков имеют вид:

$$\begin{aligned} f_1 &= \sum_{j=1}^n x_j, \quad f_k = (-1)x_{k-1} + \sum_{j \neq k-1} x_j, \\ & \quad k \in I, \quad X_k = [0, 1], \quad k = 2, \dots, n. \end{aligned}$$

Рассмотрим оптимальный по Парето вектор $x^0 = (1, \dots, 1)$ как исход кооперативной игры с характеристической функцией (14). Вычисления показывают, что

$$\begin{aligned} f_K(x^0) &= \begin{cases} n-1, & K \neq \{1\}; \\ n, & K = \{1\} \end{cases} \\ \bar{L}_K &= \begin{cases} |K| - 1, & (\exists k \in K) k-1 \in K; \\ |K| - 2, & (\forall k \in K) k-1 \in I \setminus K, \end{cases} \end{aligned}$$

где $|K|$ — число членов коалиции K и $L_K \leq \bar{L}_K$. Исход игры x^0 удовлетворяет достаточному условию коалиционной устойчивости из следствия 3.2. Тогда по теореме 3.3 в игре с контролирующим игроком существует коалиционно устойчивое эффективное решение.

4 Условия формирования коалиций

Вступление в некоторую коалицию является важной составляющей стратегии поведения игрока. Всякую коалицию представляет единственный игрок K , имеющий критерий эффективности f_K и гарантирующую стратегию $x_K^\Gamma \in X_K^0$. За аксиому примем следующее положение, являющееся необходимым условием формирования коалиции.

Аксиома 4.1. Если формируется коалиция K , то она удовлетворяет свойству:

$$(\forall K') K' \subset K \Rightarrow L_{K'} \leq \min_{x_{I \setminus K}} f_{K'}(x_K^\Gamma, x_{I \setminus K}). \quad (17)$$

Формула (17) запрещает распад создаваемой коалиции K на части $K' \neq K$. Формирование одноэлементной коалиции $K = \{k\}$ равносильно изоляционизму в поведении игрока k и применению им своей гарантирующей стратегии x_k^Γ . В этом случае игрока k можно исключить из игры, зная выбор его параметра $x_k = x_k^\Gamma$. Поэтому далее будем считать, что одноэлементные коалиции не создаются.

Как показывает неравенство (17), выбором гарантирующей стратегии x_K^Γ любая коалиция должна обеспечивать всем своим участникам выигрыши, не меньшие их собственных наилучших гарантированных результатов L_k .

Пусть игрок K' представляет коалицию K' и имеет критерий эффективности вида (11) или (12). Сформулируем аксиому поглощения.

Аксиома 4.2. Если $K' \in K$, $K \neq I$, то формируется коалиция $\bar{K} = K \setminus \{K'\} \cup K'$.

После поглощения коалиционные платежи и наилучшие гарантированные выигрыши сохраняются: $f_{\bar{K}} = f_K$, $L_{\bar{K}} = L_K$.

Замечание 4.1. В игре с побочными платежами (см. [3]) всегда можно считать выполненными обе аксиомы 4.1 и 4.2. Это обеспечивается за счет надлежащего дележа суммарного выигрыша (12).

Лемма 4.1. Если создается некоторая коалиция K , то существует K -коалиционно устойчивый исход игры. Доказательство. По аксиоме 4.1 для произвольной подкоалиции K' , $K' \subset K$, имеем:

$$L_{K'} \leq \min_{x_{I \setminus K}} f_{K'}(x_K^\Gamma, x_{I \setminus K}) \leq f_{K'}(x_K^\Gamma, x_{I \setminus K}^\Gamma).$$

По следствию 3.3 исход игры $x^0 = (x_K^\Gamma, x_{I \setminus K}^\Gamma)$ является K -коалиционно устойчивым.

С игрой (1), $\bar{X} = X^0$, естественно связать итерационный процесс поиска коалиционно устойчивого исхода игры (1), (2). На каждом шаге процесса при формировании коалиции K происходит замена K единственным игроком K . Игрок K становится представителем коалиции \bar{K} , если $K' \in K$ и игрок K' представляет коалицию K' (см. аксиому 4.2).

Теорема 4.1. Игровой операции (1) с множеством игроков I можно сопоставить игру $\bar{\Gamma}$, в которой часть игроков I объединены в попарно непересекающиеся коалиции K_r , $r = 1, 2, \dots, R$. В игре $\bar{\Gamma}$ существует эффективный коалиционно устойчивый исход.

Доказательство. Рассмотрим указанный итерационный процесс. Возьмем $x^0 = (x_1^\Gamma, \dots, x_n^\Gamma) \in X^0$, где множество

$$X^0 = \{x : \forall k \in I^0 f_k(x) \geq L_k\}.$$

Если исход x_0 не является коалиционно устойчивым, то по следствию 3.2 в игре $\Gamma^0 = \Gamma$, $I^0 = I$, существует максимальная по включению коалиция K_1 , для которой $L_{K_1} > f_{K_1}(x^0)$. Построим игру Γ^1 с множеством участников $I^1 = I \setminus K_1 \cup \{K_1\}$, где игрок K_1 представляет коалицию K_1 . В игре Γ^1 , как и выше, рассмотрим множество исходов

$$X^1 = \{x : \forall k \in I^1 f_k(x) \geq L_k\}.$$

Согласно (17) вектор $x^1 = (x_{K_1}^\Gamma, x_{I \setminus K_1}^0) \in X^1$. Результат игры x^1 либо коалиционно устойчив, либо это не так. Тогда по следствию 3.2 найдется максимальная по включению коалиция K_2 , для которой $L_{K_2} > f_{K_2}(x^0)$. Действуя аналогично, построим игру Γ^r , $r = 1, 2, \dots$. Ввиду конечности множества игроков этот процесс завершится на каком-то шаге R . По построению вектор x^R является коалиционно устойчивым исходом игры $\bar{\Gamma} = \Gamma^R$, причем $\bar{\Gamma} = \Gamma$, если $R = 0$. Если выбор x^R не эффективен, то найдется исход игры \bar{x} , для которого вполне не равенство $f(\bar{x}) \geq f(x^R)$. Отсюда следует, что обе ситуации \bar{x} , x^R удовлетворяют определению 3.2. Поэтому \bar{x} является искомым исходом игры.

5 Заключение

Проведена аксиоматизация понятия стратегии. Игровая задача может решаться в разных классах допустимых стратегий игроков, определяемых их информированностью. В теореме 2.1 обоснована непротиворечивость процесса принятия решений.

Под рациональным поведением игроков понимается их стремление к достижению либо эффективного по Парето решения, либо равновесия по Нэшу. В классе стратегий-констант ситуация равновесия, вообще говоря, не достижима, а эффективное решение определено неоднозначно и, вообще говоря, не устойчиво. Даже в играх с полной информацией эти принципы не всегда совместимы. В теореме 2.4 доказано, что при надлежащей информированности игроков в играх существуют эффективные равновесия. По теореме 3.1 введение контролирующего игрока упрощает информационные обмены, необходимые для достижимости решения игры. Несмотря на выявленную структуру рационального решения задачи, поиск решения сводится к минимаксным задачам, требующим привлечения серьезных вычислительных ресурсов [18].

В стратегию игроков входит возможность коалиционных действий. Следствие 3.2, теоремы 3.3 и 4.1 посвящены изучению свойства коалиционной устойчивости исхода игры. Эффективная коалиционно устойчивая ситуация игры сохраняется при коллективных действиях игроков.

Литература

- Howard N. Theory of meta-games // General Systems, 1966. Vol. XI. P. 187–200.
- Кукушкин Н. С. Точки равновесия в метаиграх // Ж. вычисл. матем. и матем. физ., 1974. Т. 14. № 2. С. 312–320.
- Гермейер Ю. Б. Игры с противоположными интересами. — М.: Наука, 1976. 326 с.
- Кукушкин Н. С. Роль взаимной информированности сторон в играх двух лиц с противоположными интересами // Ж. вычисл. матем. и матем. физ., 1972. Т. 8. № 4. С. 1029–1034.
- Красовский Н. Н., Субботин А. И. Позиционные дифференциальные игры. — М.: Наука, 1974. 458 с.
- Горелик В. А., Горелов М. А., Кононенко А. Ф. Анализ конфликтных ситуаций в системах управления. — М.: Радио и связь, 1991. 286 с.
- Петросян Л. А., Зенкевич Н. А., Семин Е. А. Теория игр. — М.: Высшая школа, 1998. 304 с.
- Жуковский В. И. Кооперативные игры при неопределенности и их приложения. — М.: Эдиториал УРСС, 1999. 336 с.
- Васин А. А., Морозов В. В. Теория игр и модели математической экономики. — М.: МАКС Пресс, 2005. 272 с.
- Колесник Г. В., Леонова Н. А. Теория игр в примерах и задачах. — Тверь: ТвГУ, 2012. 132 с.
- Опойцев В. И. Равновесие и устойчивость в моделях коллективного поведения. — М.: Наука, 1977. 248 с.
- Васильев Н. С. Использование принципа равновесия для управления маршрутизацией в транспортных сетях // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 29–36.
- Васильев Н. С. Численное решение бескоалиционных матричных игр // Наука и образование: Электронное научно-техническое издание, 2013. № 8. doi: 10.7463/0813.058774.
- Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982. 256 с.
- Моисеев Н. Н. Элементы теории оптимальных систем. — М.: Наука, 1975. 527 с.
- Скорняков Л. А. Элементы общей алгебры. — М.: Наука, 1983. 272 с.
- Карманов В. Г., Федоров В. В. Моделирование в исследовании операций. — М.: Твема, 1996. 102 с.
- Федоров В. В. Численные методы максимина. — М.: Наука, 1979. 280 с.

Поступила в редакцию 10.12.14

ON AVAILABILITY OF PARETO EFFECTIVE EQUILIBRIUM SITUATIONS IN COLLECTIVE BEHAVIOR MODELS WITH DATA EXCHANGE

N. S. Vasilyev

N. E. Bauman Moscow State Technical University, 5 Baumanskaya 2nd Str., Moscow 105005, Russian Federation

Abstract: Use of network technologies impels investigations of collective behavior models. Processes of decision making based on data exchange are of utmost interest. For this purpose, strategy axiomatization is proposed. Information exchange diminishes uncertainty in the processes and models collective efforts to achieve rational decisions. Rational behavior uses the principles of effectiveness and stability usually contradicting one another. Rational game solutions' structure is studied. It is discovered that data exchange allows achieving Pareto effective situation which is also the equilibrium one. A notion of coalitional stable game issue is introduced.

The situation prevents from forming coalitions and can simultaneously satisfy the property of Pareto effectiveness. It can also give Nash equilibrium if adequate players' strategies are used. An expansion of initial game by means of additional controlling player shows how the effective coalitional stable issue can be achieved.

Keywords: game; strategy; situation; game issue; information exchange; dynamics of decision making; axiomatization; coalition; cooperative game; characteristic function; the best guaranteed result; strategy of punishment; Pareto effectiveness; Nash equilibrium

DOI: 10.14357/19922264150201

References

- Howard, N. 1966. Theory of meta-games. *General Systems* 11:187–200.
- Kukushkin, N. S. 1974. Tochki ravnovesiya v metaigrakh [Equilibria points in meta-games]. *Zh. Vychisl. Mat. Mat. Fiz.* [Computational Mathematics and Mathematical Physics] 14(2):312–320.
- Germeyer, Yu. B. 1976. *Igry s neprotivopolozhnymi interesami* [Games with nonantagonistic interests]. Moscow: Nauka. 326 p.
- Kukushkin, N. S. 1972. Rol' vzaimnoy informirovannosti storon v igrah dvuh lits s neprotivopolozhnymi interesami [Role of partners mutual information in two person games with nonantagonistic interests]. *Zh. Vychisl. Mat. Mat. Fiz.* [Computational Mathematics and Mathematical Physics] 8(4):1029–1034.
- Krasovskiy, N. N., and A. I. Subbotin. 1974. *Pozitsionnye differentsial'nye igry* [Positional differential games]. Moscow: Nauka. 458 p.
- Gorelik, V. A., M. A. Gorelov, and A. F. Kononenko. 1991. *Analiz konfliktnykh situatsiy v sistemakh upravleniya* [Analysis of conflict situations in control systems]. Moscow: Radio i svyaz'. 286 p.
- Petrosyan, L. A., N. A. Zenkevich, and E. A. Semina. 1998. *Teoriya igr* [Theory of games]. Moscow: Vysshaya Shkola. 304 p.
- Zhukovskiy, V. I. 1999. *Kooperativnye igry pri neopredelennosti i ikh prilozheniya* [Cooperative games under uncertainty and its applications]. Moscow: Editorial URSS. 336 p.
- Vasin, A. A., and V. V. Morozov. 2005. *Teoriya igr i modeli matematicheskoy ekonomiki* [Theory of games and mathematical economics models]. Moscow: MAKS Press. 272 p.
- Kolesnik, G. V., and N. A. Leonova. 2012. *Teoriya igr v primerakh i zadachakh* [Theory of games in examples and tasks]. Tver': Tver' Gos. Univ. 132 p.
- Opoycev, V. I. 1977. *Ravnovesie i ustoychivost' v modelyakh kollektivnogo povedeniya* [Equilibrium and stability in collective behavior models]. Moscow: Nauka. 248 p.
- Vasil'ev, N. S. 2014. Ispol'zovanie printsipa ravnovesiya dlya upravleniya marshrutizatsiy v transportnykh setyakh [Equilibrium principle application to routing control in packet data-transmission networks]. *Informatika i ee Primeneniya — Inform. Appl* 8(1):29–36.
- Vasil'ev, N. S. 2013. Chislennoe reshenie beskoalitsionnykh matrichnykh igr [Numerical solution of matrix games without coalitions]. *Nauka i obrazovanie: Elektronnoe nauchno-tehnicheskoe izdanie* 8. doi: 10.7463/0813.058774.
- Podinovskiy, V. V., and V. D. Nogin. 1982. *Pareto optimal'nye resheniya mnogokriterial'nykh zadach* [Pareto optimal solutions in multicriteria problems]. Moscow: Nauka. 256 p.
- Moiseev, N. N. 1975. *Elementy teorii optimal'nykh sistem* [Elements of optimal systems theory]. Moscow: Nauka. 527 p.
- Skorniyakov, L. A. 1983. *Elementy obshchey algebry* [Elements of general algebra]. Moscow: Nauka. 272 p.
- Karmanov, V. G., and V. V. Fedorov. 1996. *Modelirovanie v issledovanii operatsiy* [Modelling in operations research]. Moscow: Tvema. 102 p.
- Fedorov, V. V. 1979. *Chislennyye metody maksimina* [Numerical methods of maximin]. Moscow: Nauka. 280 p.

Received December 10, 2014

Contributor

Vasilyev Nikolai S. (b. 1952) — Doctor of Science in physics and mathematics, professor, N. E. Bauman Moscow State Technical University, 5 Baumanskaya 2nd Str., Moscow 105005, Russian Federation; nik8519@yandex.ru

ВРЕМЯ ОЖИДАНИЯ В СИСТЕМЕ ОБСЛУЖИВАНИЯ С ИНВЕРСИОННЫМ ПОРЯДКОМ ОБСЛУЖИВАНИЯ И ОБОБЩЕННЫМ ВЕРОЯТНОСТНЫМ ПРИОРИТЕТОМ*

Л. А. Мейханаджян¹, Т. А. Милованова², Р. В. Разумчик³

Аннотация: Рассматривается система массового обслуживания (СМО) с одним прибором и бесконечным числом мест для ожидания, в которую поступает пуассоновский поток заявок. В системе реализован инверсионный порядок обслуживания с обобщенным вероятностным приоритетом, заключающийся в следующем. Предполагается, что в любой момент времени известна остаточная длина каждой заявки в системе. В момент поступления в систему новой заявки ее исходная длина сравнивается с остаточной длиной заявки на приборе, и в зависимости от результатов сравнения либо обе заявки покидают систему, либо только одна из них (вторая остается на приборе), либо обе остаются в системе (причем одна из них помещается на прибор). Если заявка остается в системе, она приобретает новую (случайную) длину. Предложены математические соотношения для вычисления распределения времени пребывания заявки в системе и периода занятости (ПЗ) (в терминах преобразования Лапласа–Стилтьеса (ПЛС)), а также некоторых временных характеристик.

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

DOI: 10.14357/19922264150202

1 Введение

Данная работа, являющаяся продолжением [1], посвящена нахождению математических соотношений для стационарных временных характеристик заявок, поступающих в одноканальную СМО неограниченной емкости с той же, что и в [1], дисциплиной — инверсионным порядком обслуживания с обобщенным вероятностным приоритетом (Last Come First Served with Generalized Probabilistic Priority, LCFS GPP). Исследования по системам со сложными, специальными дисциплинами обслуживания начались в середине XX в., и с тех пор накопился значительный объем научного материала. Некоторые из результатов, близких к теме настоящего исследования, можно найти в [2–8]. Среди недавних работ по специальным дисциплинам обслуживания стоит отметить [9, 10].

Поводом для написания настоящей статьи стала работа [1], в которой исследовались стационарные характеристики в одноканальной системе обслуживания с дисциплиной LCFS GPP, связанные с числом заявок в системе. Однако, как известно, если с точки зрения вероятностных характеристик (например, средней длины очереди) выбор дисциплины обслуживания может и не играть роли,

то с точки зрения временных характеристик специфика дисциплины обслуживания очень важна. Действительно, в случае дисциплины LCFS GPP поступающая заявка может как мгновенно покинуть систему (выбивая при этом заявку, обслуживаемую на приборе), так и вытеснить заявку с прибора или занять первое место в очереди. При этом каждый раз заявка, занимая прибор, может покинуть его недообслуженной и/или возвращаться на него конечное (но сколь угодно большое) число раз. Необходимо также добавить, что при наступлении каждого события, связанного с приходом очередной заявки, длины обслуживаемой заявки (если таковая имеется) и поступающей заявки могут меняться в соответствии с заданными законами распределения.

Данное описание не исчерпывает всех возможных событий, которые могут произойти с заявкой как только она поступила в систему, однако показывает, что важно уметь рассчитывать и временные характеристики системы.

Основной результат работы состоит в нахождении стационарного распределения времени пребывания заявки в системе из [1] в терминах ПЛС, а также нахождения ПЛС ПЗ. Нахождение моментов сводится к задаче нахождения приближенно-

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

¹ Российский университет дружбы народов, lameykanadzhyan@gmail.com

² Российский университет дружбы народов, tmilovanova77@mail.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, rrazumchik@gmail.com

го или точного решения неоднородных интегральных уравнений (в некоторых случаях — уравнений Фредгольма 2-го рода).

Статья организована следующим образом. Во втором разделе дается описание рассматриваемой системы. В разд. 3 приводятся формулы для ПЗ (в терминах ПЛС). В разд. 4 и 5 внимание уделяется нахождению стационарного распределения времени пребывания заявки на приборе и в системе соответственно. Выражения для некоторых временных характеристик приводятся в разд. 6. В заключении работы кратко резюмируются основные результаты и даются направления дальнейших исследований.

2 Описание системы

Рассмотрим СМО с входящим потоком заявок, который для простоты будем называть здесь потоком пуассоновского типа. Отличие этого потока от пуассоновского заключается в следующем: интенсивность поступления заявки равна λ , если на приборе имеется заявка, и $\tilde{\lambda}$, если система пуста.

Если в момент поступления заявки в систему на приборе имеется заявка, то исходное распределение времени обслуживания поступающей заявки является произвольным с 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 $\tilde{b}(x) = \tilde{B}'(x)$.

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

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

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

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

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

Будем считать для удобства изложения, что все ФР $D(x, y|u, v)$, $D^*(x, y|u, v)$, $D_0(x|u, v)$, $D_0^*(y|u, v)$, $D(y|u, v)$ и $D_0(u, v)$ имеют непрерывные ограниченные плотности:

$$d(x, y|u, v) = \frac{\partial^2 D(x, y|u, v)}{\partial x \partial y};$$

$$d^*(x, y|u, v) = \frac{\partial^2 D^*(x, y|u, v)}{\partial x \partial y};$$

$$d_0(x|u, v) = \frac{\partial D_0(x|u, v)}{\partial x};$$

$$d_0^*(y|u, v) = \frac{\partial D_0^*(y|u, v)}{\partial y}; \quad d(x|u, v) = \frac{\partial D(x|u, v)}{\partial x}.$$

Естественно, для любых u и v выполнено условие:

$$\begin{aligned} & \int_0^\infty \int_0^\infty [d(x, y|u, v) + d^*(x, y|u, v)] dx dy + \\ & + \int_0^\infty d(x|u, v) dx + d_0(u, v) = \\ & = D(\infty, \infty|u, v) + D^*(\infty, \infty|u, v) + \\ & + D(\infty|u, v) + d_0(u, v) = 1. \end{aligned}$$

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

Далее будем предполагать, что система функционирует в стационарном режиме. Достаточное условие его существования приведено в [1], где также были найдены аналитические выражения для $p_n(x)$ — стационарных плотностей вероятностей того, что в рассматриваемой системе находится n заявок и остаточная длина заявки, находящейся на

приборе, меньше x . Далее будем считать, что значения $p_n(x)$ известны.

3 Период занятости

Обозначим через $u(s, x)$ ПЛС длительности ПЗ, открываемого заявкой длины x . Для нахождения $u(s, x)$ воспользуемся свойствами ПЛС и рассмотрим все возможные события, которые могут произойти после поступления заявки, открывающей ПЗ. Так, ПЛС длительности ПЗ $u(s, x)$ 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 $\lambda e^{-\lambda t} \int_0^\infty d_0(y, x - t) b(y) dy$);
- $e^{-st} u(s, v)$, 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 открыла новый ПЗ системы, ПЛС длительности которого равно $u(s, v)$ (плотность вероятности данного события равна $\lambda e^{-\lambda t} \int_0^\infty d_0(v|y, x - t) b(y) dy$);
- $e^{-st} u(s, v) u(s, w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , обе заявки остаются в системе, причем длина одной из двух заявок (поступившей заявки или заявки на приборе) становится равной v , а второй — w (с плотностью вероятностей $\lambda e^{-\lambda t} \times \int_0^\infty [d(v, w|y, x - t) + d^*(v, w|y, x - t)] b(y) dy$).

Теперь по формуле полной вероятности получаем, что уравнение для нахождения ПЛС $u(s, x)$ имеет вид:

$$u(s, x) = a_1(s, x) + \int_0^\infty u(s, v) a_2(s, x, v) dv + \int_0^\infty u(s, v) dv \int_0^\infty u(s, w) a_3(s, x, v, w) dw, \quad (1)$$

где

$$a_1(s, x) = e^{-(s+\lambda)x} + \int_0^x \lambda e^{-(s+\lambda)t} dt \int_0^\infty d_0(y, x - t) b(y) dy;$$

$$a_2(s, x, v) = \int_0^x \lambda e^{-(s+\lambda)t} dt \int_0^\infty d_0(v|y, x - t) b(y) dy;$$

$$a_3(s, x, v, w) = \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty [d(v, w|y, x - t) + d^*(v, w|y, x - t)] b(y) dy.$$

Уравнение (1) представляет собой интегральное уравнение, решение которого, за исключением простейших частных случаев, не удастся получить в явном виде. Для его численного решения, вообще говоря, можно пользоваться известными численными методами (см., например, [11–13]).

4 Стационарное распределение времени пребывания заявки на приборе

Обратимся к задаче нахождения времени пребывания заявки на приборе. Поскольку вновь поступающие заявки могут вытеснять заявку с прибора (на первое место в очереди и вообще из системы), то время пребывания заявки на приборе можно считать по-разному, а именно: с учетом прерываний обслуживания и без.

4.1 Время пребывания заявки на приборе без учета времени прерываний обслуживания

Обозначим через $\varphi(s, x)$ ПЛС времени пребывания заявки длины x , поступившей на прибор без учета времени прерываний ее обслуживания. Для нахождения $\varphi(s, x)$, так же как и в случае ПЛС ПЗ, воспользуемся свойствами ПЛС и рассмотрим все возможные события, которые могут произойти с поступившей на прибор заявкой. Заявка будет пребывать на приборе время (в терминах ПЛС), равное:

- 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 (новая заявка и заявка на приборе) покинули 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 $\lambda e^{-\lambda t} \int_0^{\infty} d_0(v|y, x - t) b(y) dy$);
- $e^{-st} \varphi(s, w)$, 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 $\lambda e^{-\lambda t} \int_0^{\infty} d_0^*(w|y, x - t) b(y) dy$);
- $e^{-st} \varphi(s, w)$, если в момент времени $0 < t < x$ поступила новая заявка длины y , обе заявки (новая заявка и заявка на приборе) остались в системе, новая заявка получила длину v , заявка, бывшая на приборе, — длину w (с плотностью вероятностей $\lambda e^{-\lambda t} \times \int_0^{\infty} [d(v, w|y, x - t) + d^*(v, w|y, x - t)] b(y) dy$).

Воспользовавшись формулой полной вероятности, получаем, что ПЛС $\varphi(s, x)$ есть решение следующего интегрального уравнения:

$$\varphi(s, x) = b_1(s, x) + \int_0^{\infty} \varphi(s, w) b_2(s, x, w) dw, \quad (2)$$

где

$$\begin{aligned} b_1(s, x) &= e^{-(\lambda+s)x} + \\ &+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^{\infty} d_0(y, x - t) b(y) dy + \\ &+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^{\infty} dv \int_0^{\infty} d_0(v|y, x - t) b(y) dy; \\ b_2(s, x, w) &= \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^{\infty} d_0^*(w|y, x - t) b(y) dy + \\ &+ \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^{\infty} dv \int_0^{\infty} [d(v, w|y, x - t) + \\ &+ d^*(v, w|y, x - t)] b(y) dy. \end{aligned}$$

Уравнение (2) есть уравнение Фредгольма 2-го рода. Свободный член $b_1(s, x)$ и ядро $b_2(s, x, w)$ интегрального уравнения являются неотрицательными функциями. Численные методы решения интегральных уравнений Фредгольма 2-го рода хорошо известны (см., например, [11–13]). В данном случае, в принципе, можно воспользоваться итерационным методом, взяв в качестве начальной итерации нулевое приближение. Тогда итерации будут возрастающими. Однако практическая реализация метода является затруднительной ввиду того, что выражение для ядра уравнения является весьма сложным.

4.2 Время пребывания заявки на приборе с учетом времени прерываний обслуживания

Обозначим через $\psi(s, x)$ ПЛС времени пребывания заявки длины x , поступившей на прибор, с учетом времени прерываний ее обслуживания. Действуя, как и ранее, приходим к тому, что время пребывания (в терминах ПЛС) 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 $\lambda e^{-\lambda t} \int_0^{\infty} d_0(v|y, x - t) b(y) dy$);
- $e^{-st} \psi(s, w)$, 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 $\lambda e^{-\lambda t} \int_0^{\infty} d_0^*(w|y, x - t) b(y) dy$);
- $e^{-st} \psi(s, w)$, 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 $\lambda e^{-\lambda t} \int_0^{\infty} d(v, w|y, x - t) b(y) dy$);
- $e^{-st} u(s, v) \psi(s, w)$, 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 $\psi(s, x)$ имеет следующий вид:

$$\psi(s, x) = b_1(s, x) + \int_0^{\infty} \psi(s, w) b_3(s, x, w) dw, \quad (3)$$

где

$$\begin{aligned}
 b_3(s, x, w) = & \\
 = & \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty d_0^*(w|y, x-t) b(y) dy + \\
 + & \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty dv \int_0^\infty d(v, w|y, x-t) b(y) dy + \\
 + & \int_0^x \lambda e^{-(\lambda+s)t} dt \int_0^\infty u(s, v) dv \int_0^\infty d^*(v, w|y, x-t) b(y) dy.
 \end{aligned}$$

Как видно, вывод уравнения (3) для $\psi(s, x)$ практически полностью повторяет вывод уравнения (2) для $\varphi(s, x)$. Единственное отличие состоит в том, что при расчете $\psi(s, x)$ необходимо учитывать то время, которое проходит с момента, когда заявка вытесняется с прибора на первое место в очереди, и до момента, когда заявка снова занимает прибор. Это время в точности совпадает с длиной ПЗ, открываемого вновь поступающей заявкой, которая прерывает обслуживание заявки на приборе. Уравнение (3) также является уравнением Фредгольма 2-го рода, и к нему применимы те же замечания, что и к уравнению (2).

5 Стационарное распределение времени пребывания заявки в системе

Найдем в терминах ПЛС $\chi(s, x)$ стационарное распределение времени пребывания в системе поступающей заявки длины x .

Во-первых, заявка длины x может с вероятностью p_0 поступить в свободную систему и тогда время ее пребывания будет совпадать с полным временем ее обслуживания.

Во-вторых, с плотностью вероятностей $\sum_{n=1}^\infty p_n(y)$ она может застать систему занятой, причем на приборе будет находиться заявка длины y . В этом случае возможны следующие варианты:

- либо с вероятностью $d_0(x, y) + \int_0^\infty d_0^*(w|x, y) dw$ поступившая заявка покинет систему и при этом ее время пребывания в системе будет равно нулю;
- либо с вероятностью $d_0(v|x, y) + d^*(v, w|x, y)$ она сама встанет на прибор, причем ее длина станет равной v и тогда время ее пребывания в системе будет совпадать с полным временем ее обслуживания с учетом прерываний;

– либо с вероятностью $d(v, w|x, y)$ она станет на первое место в очередь, получит новую длину v , а заявка на приборе — новую длину w ; при этом время пребывания в системе поступившей заявки будет равно сумме двух времен: длины ПЗ, открываемого заявкой длины w , и полного времени обслуживания заявки длины v с учетом прерываний.

Применяя формулу полной вероятности, приходим к следующему выражению для $\chi(s, x)$:

$$\begin{aligned}
 \chi(s, x) = & p_0 \psi(s, x) + \\
 + & \int_0^\infty \sum_{n=1}^\infty p_n(y) \left[d_0(x, y) + \int_0^\infty d_0^*(w|x, y) dw + \right. \\
 + & \int_0^\infty \psi(s, v) \left(d_0(v|x, y) + \int_0^\infty d^*(v, w|x, y) dw \right) dv + \\
 + & \left. \int_0^\infty \int_0^\infty \psi(s, v) u(s, w) d(v, w|x, y) dv dw \right] dy.
 \end{aligned}$$

Наконец, ПЛС $\chi(s)$ стационарного распределения времени пребывания в системе заявки произвольной длины получается усреднением по распределению длины заявки $B(x)$ и равно

$$\chi(s) = \int_0^\infty \chi(s, x) b(x) dx.$$

Для определения моментов времени пребывания заявки в системе достаточно воспользоваться свойством ПЛС. Расчетные формулы ввиду громоздкости здесь не приводятся.

6 Некоторые временные характеристики

Покажем, как, используя результаты, полученные в предыдущих разделах, можно найти выражения для некоторых вероятностных характеристик поступающих в систему заявок. Ограничимся следующими характеристиками:

- вероятность $\pi(x)$ того, что заявка будет потеряна при поступлении при условии, что ее исходная длина равнялась x ;
- вероятность $\hat{q}_i(x)$, $i \geq 0$, того, что заявка не будет потеряна при поступлении, не будет обслужена до конца и за время пребывания в системе сменит длину i раз при условии, что ее исходная длина равнялась x ;

- вероятность $q_i^*(x)$, $i \geq 0$, того, что заявка не будет потеряна при поступлении, будет обслужена до конца и за время пребывания в системе сменит длину i раз при условии, что ее исходная длина равнялась x ;
- среднее время пребывания в системе заявки в расчете на единицу времени обслуживания.

Очевидно, что заявка, поступившая в систему в момент, когда на приборе оканчивается обслуживание или нет заявок, не будет потеряна. Тогда из самого определения введенной дисциплины обслуживания следует, что

$$\pi(x) = \int_0^\infty \sum_{n=1}^\infty p_n(y) \left(d_0(x, y) + \int_0^\infty d_0^*(w|x, y) dw \right) dy.$$

Для нахождения вероятностей $\hat{q}_i(x)$, $i \geq 0$, заметим, что заявка, заставшая в момент поступления прибор свободным, сразу начинает обслуживаться и может покинуть систему не обслужившись, ни разу не сменив свою исходную длину. Обозначим через $q_i(x)$ вероятность того, что заявка не будет обслужена до конца, сменит длину i , $i \geq 0$, раз при условии, что ее исходная длина равнялась x , а прибор в момент ее поступления был свободен. Можно показать, что для $q_i(x)$ имеют место следующие рекуррентные соотношения:

$$q_0(x) = \int_0^x \lambda e^{-\lambda t} dt \int_0^\infty d_0(y, x-t) b(y) dy + \int_0^x \lambda e^{-\lambda t} dt \int_0^\infty dv \int_0^\infty d_0(v|y, x-t) b(y) dy;$$

$$q_i(x) = \int_0^x \lambda e^{-\lambda t} dt \int_0^\infty q_{i-1}(w) dw \times \int_0^\infty d_0^*(w|y, x-t) b(y) dy + \int_0^x \lambda e^{-\lambda t} dt \int_0^\infty dv \times \int_0^\infty q_{i-1}(w) dw \int_0^\infty [d(v, w|y, x-t) + d^*(v, w|y, x-t)] b(y) dy.$$

Напротив, если поступающая заявка застанет прибор занятым и останется в системе, то она обязательно хотя бы один раз сменит свою длину (ввиду предположения о непрерывности функций $D(x, y|u, v)$, $D^*(x, y|u, v)$, $D_0(x|u, v)$, $D_0^*(y|u, v)$ и $D(y|u, v)$). Воспользовавшись теперь формулой

полной вероятности, получим следующую формулу для расчета $\hat{q}_i(x)$:

$$\begin{aligned} \hat{q}_0(x) &= q_0(x)p_0; \\ \hat{q}_i(x) &= q_i(x)p_0 + \\ &+ \int_0^\infty \sum_{n=1}^\infty p_n(y) \left[\int_0^\infty q_{i-1}(v) d_0(v|x, y) dv + \right. \\ &+ \int_0^\infty q_{i-1}(v) dv \int_0^\infty [d(v, w|x, y) + \\ &\left. + d^*(v, w|x, y)] dw \right] dy, \quad i \geq 1. \end{aligned}$$

Аналогичные рассуждения используются для нахождения вероятностей $q_i^*(x)$. Обозначим через $\tilde{q}_i(x)$ вероятность того, что заявка будет обслужена до конца, сменит длину i , $i \geq 0$, раз при условии, что ее исходная длина равнялась x , а в момент ее поступления прибор был свободен. Рекуррентные соотношения, которым удовлетворяют $\tilde{q}_i(x)$, имеют вид:

$$\begin{aligned} \tilde{q}_0(x) &= e^{-\lambda x}, \\ \tilde{q}_i(x) &= \int_0^x \lambda e^{-\lambda t} dt \int_0^\infty \tilde{q}_{i-1}(w) dw \times \\ &\times \int_0^\infty d_0^*(w|y, x-t) b(y) dy \int_0^\infty \lambda e^{-\lambda t} dt \int_0^\infty dv \times \\ &\times \int_0^\infty \tilde{q}_{i-1}(w) dw \int_0^\infty [d(v, w|y, x-t) + \\ &+ d^*(v, w|y, x-t)] b(y) dy, \quad i \geq 1. \end{aligned}$$

Учитывая, что заявка, заставшая в момент поступления прибор занятым (и оставшаяся в системе), обязательно хотя бы один раз сменит свою длину до того момента, когда будет обслужена до конца, по формуле полной вероятности получаем следующую рекуррентную формулу для расчета $q_i^*(x)$, $i \geq 0$:

$$\begin{aligned} q_0^*(x) &= \tilde{q}_0(x)p_0; \\ q_i^*(x) &= \tilde{q}_i(x)p_0 + \\ &+ \int_0^\infty \sum_{n=1}^\infty p_n(y) \left[\int_0^\infty \tilde{q}_{i-1}(v) d_0(v|x, y) dv + \right. \\ &+ \int_0^\infty \tilde{q}_{i-1}(v) dv \int_0^\infty [d(v, w|x, y) + d^*(v, w|x, y)] dw \left. \right] dy. \end{aligned}$$

Если число раз, которое заявка сменила свою длину прежде, чем покинуть систему обслуженной (недообслуженной), не имеет значения, то для

нахождения соответствующей вероятности достаточно просуммировать $q_i^*(x)$ ($\hat{q}_i(x)$) по всем значениям i . Заметим, что данные вероятности можно найти и другим способом, выписав для них по отдельному уравнению и решив их.

Обозначим через W_x время пребывания в системе заявки, которая при поступлении в систему имеет длину x . Тогда средний штраф за предоставление поступающей заявке x единиц времени обслуживания определяется как $E(W_x)/x^1$. Используя результаты разд. 5, нетрудно выписать соответствующую расчетную формулу

$$\frac{E(W_x)}{x} = -\frac{1}{x} \chi'(0, x),$$

откуда безусловное значения штрафа получается усреднением по распределению длины заявки $B(x)$.

7 Заключение

В настоящей работе предложен метод нахождения (в терминах ПЛС) ПЗ и стационарного распределения времени пребывания заявки в системе $M/G/1$ с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом. Укажем на проблему расчета временных характеристик по полученным соотношениям, что является предметом дальнейших исследований. Во всех случаях неизвестные величины (например, моменты) являются решениями неоднородных интегральных уравнений. Как показывают численные эксперименты, качество получаемых решений сильно зависит от вида функций $d(u, v)$, $d^*(x|u, v)$, $d(x|u, v)$, $d(x, y|u, v)$ и $d^*(x, y|u, v)$ и применение традиционных методов решения (метода конечных сумм, метода конечных приближений и др.) зачастую не дает удовлетворительных результатов. Численные методы Монте Карло позволяют находить приближенные решения (как показывают сравнения с результатами имитационного моделирования) за конечное время, однако ограничения, присущие этим методам, не позволяют использовать их для произвольных функций $d(u, v)$, $d^*(x|u, v)$, $d(x|u, v)$, $d(x, y|u, v)$ и $d^*(x, y|u, v)$.

Литература

1. Мейханаджян Л. А., Милованова Т. А., Печинкин А. В., Разумчик Р. В. Стационарные вероятности состояний

2. Schrage L. A proof of the optimality of the shortest remaining processing time discipline // *Oper. Res.*, 1968. Vol. 16. P. 687–690.
3. Нагоненко В. А. О характеристиках одной нестандартной системы массового обслуживания. I; II // *Изв. АН СССР. Технич. кибернет.*, 1981. № 1. С. 187–195; № 3. С. 91–99.
4. Нагоненко В. А., Печинкин А. В. О большой загрузке в системе с инверсионным обслуживанием и вероятностным приоритетом // *Изв. АН СССР. Технич. кибернет.*, 1982. № 1. С. 86–94.
5. Печинкин А. В. Об одной инвариантной системе массового обслуживания // *Math. Operationsforsch. und Statist. Ser. Optimization*, 1983. Vol. 14. № 3. P. 433–444.
6. Нагоненко В. А., Печинкин А. В. О малой загрузке в системе с инверсионным порядком обслуживания и вероятностным приоритетом // *Изв. АН СССР. Технич. кибернет.*, 1984. № 6. С. 82–89.
7. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
8. Печинкин А. В., Стальченко И. В. Система $MAP/G/1/\infty$ с инверсионным порядком обслуживания и вероятностным приоритетом, функционирующая в дискретном времени // *Вестник Российского ун-та дружбы народов. Сер. Математика. Информатика. Физика*, 2010. № 2. С. 26–36.
9. Nair J., Wierman A., Zwart B. Tail-robust scheduling via limited processor sharing // *Perform. Evaluation*, 2010. Vol. 67. No. 11. P. 978–995.
10. Wierman A., Zwart B. Is tail-optimal scheduling possible? // *Oper. Res.*, 2012. Vol. 60. No. 5. P. 1249–1257.
11. Jerri A. Introduction to integral equations with applications. — New York, NY, USA: John Wiley & Sons, 1999. 272 p.
12. Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. Numerical recipes. — 3rd ed. — The Art of Scientific Computing, 2007. 1235 p.
13. Polyanin, A. D., Manzhirov A. V. Handbook of integral equations. Boca Raton — London: Chapman & Hall / CRC Press, 2008. 1108 p.
14. Kleinrock L. Queueing systems: Vol. II — Computer applications. — New York, NY, USA: Wiley Interscience, 1976. 576 p.
15. Avi-Itzhak B., Brosh E., Levy H. SQF: A slowdown queueing fairness measure // *Perform. Evaluation*, 2007. Vol. 64. No. 9. P. 1121–1136.

Поступила в редакцию 28.04.15

¹Отношение W_x/x показывает, во сколько раз время пребывания заявки в системе отличается от ее исходной длины. В предположении, что более длинные заявки должны находиться в системе дольше, чем короткие, моменты случайной величины W_x/x могут использоваться для оценки справедливости дисциплины обслуживания [14, 15].

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

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

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

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

Abstract: The paper considers a single-server queueing system with a buffer of infinite capacity. Customers arrive according to a Poisson process. Service discipline is LIFO (Last In, First Out) with generalized probabilistic priority. It is assumed that at any instant, the remaining service time of each customer present in the system is known. Upon arrival of a new customer, its service time is compared with the remaining service time of the customer in service. As a result of the comparison, one of the following occurs: both customers leave the system; one customer leaves the system and the other occupies the server; and both customers stay in the system (one of the two occupies the server). These actions are governed by probabilistic functions. Whenever a customer remains in the system, it acquires a new (random) service time. The paper proposes the methods for calculating customer's sojourn time distribution and busy period (in terms of Laplace–Stieltjes transform) and several performance characteristics.

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

DOI: 10.14357/19922264150202

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 13-07-00223).

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.
- Schrage, L. 1968. A proof of the optimality of the shortest remaining processing time discipline. *Oper. Res.* 16:687–690.
- Nagonenko, V. A. 1981. O kharakteristikakh odnoy nestandartnoy sistemy massovogo obsluzhivaniya [On the characteristics of one nonstandard queueing system]. I; II. *Izv. AN SSSR. Tekhnich. Kibernet.* [Proceedings of the Academy of Sciences of the USSR. Technical Cybernetics] (1):187–195; (3):91–99.
- Nagonenko, V. A., and A. V. Pechinkin. 1982. O bol'shoy zagruzke v sisteme s inversionnym obsluzhivaniem i veroyatnostnym prioritetom [On high load in the system with an inversion procedure service and probabilistic priority]. *Izv. AN SSSR. Tekhnich. Kibernet.* [Proceedings of the Academy of Sciences of the USSR. Technical Cybernetics] (1):86–94.
- Pechinkin, A. V. 1983. Ob odnoy invariantnoy sisteme massovogo obsluzhivaniya [On an invariant queueing system]. *Math. Operationsforsch. und Statist. Ser. Optimization* 14(3):433–444.
- Nagonenko, V. A., and A. V. Pechinkin. 1984. O maloy zagruzke v sisteme s inversionnym poryadkom obsluzhivaniya i veroyatnostnym prioritetom [On low load in the system with an inversion procedure service and probabilistic priority]. *Izv. AN SSSR. Tekhnich. Kibernet.* [Proceedings of the Academy of Sciences of the USSR. Technical Cybernetics] (6):82–89.
- Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
- Pechinkin, A. V., and I. V. Stalchenko. 2010. Sistema MAP/G/1/∞ s inversionnym poryadkom obsluzhivaniya i veroyatnostnym prioritetom, funktsioniruyushchaya v diskretnom vremeni [The MAP/G/1/∞ discrete-time queueing system with inverse service order and probabilistic priority]. *Vestnik Rossiyskogo Universiteta Druzhby Narodov. Ser. Matematika. Informatika. Fizika* [Bulletin of Peoples' Friendship University of Russia. Ser. Mathematics. Information Sciences. Physics] 2:26–36.
- Nair, J., A. Wierman, and B. Zwart. 2010. Tail-robust scheduling via limited processor sharing. *Perform. Evaluation* 67(11):978–995.
- Wierman, A., and B. Zwart. 2012. Is tail-optimal scheduling possible? *Oper. Res.* 60(5):1249–1257.
- Jerri, A. 1999. *Introduction to integral equations with applications*. New York, NY: John Wiley & Sons. 272 p.

12. Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B. P. Flannery. 2007. *Numerical recipes*. 3rd ed. The Art of Scientific Computing. 1235 p.
13. Polyanin, A. D., and A.V. Manzhirov. 2008. *Handbook of integral equations*. Boca Raton – London: Chapman & Hall / CRC Press. 1108 p.
14. Kleinrock, L. 1976. *Queueing systems: Vol. II — Computer applications*. New York, NY: Wiley Interscience. 576 p.
15. Avi-Itzhak, B., E. Brosh, and H. Levy. 2007. SQF: A slowdown queueing fairness measure. *Perform. Evaluation* 64(9):1121–1136.

Received April 28, 2015

Contributors

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

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

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, 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; associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ieee.org

МОДЕЛИРОВАНИЕ НОРМАЛЬНЫХ ПРОЦЕССОВ В СТОХАСТИЧЕСКИХ СИСТЕМАХ СО СЛОЖНЫМИ ТРАНСЦЕНДЕНТНЫМИ НЕЛИНЕЙНОСТЯМИ*

И. Н. Сеницын¹, В. И. Сеницын², Э. Р. Корепанов³

Аннотация: Рассматривается развитие методов аналитического и статистического моделирования нормальных стохастических процессов (СтП) на случай непрерывных и дискретных стохастических систем (СтС) (в том числе на многообразиях) с винеровскими и пуассоновскими шумами и со сложными трансцендентными нелинейностями (СТН). Даны типовые представления скалярных и векторных СТН. Получены уравнения методов нормальной аппроксимации (МНА) и статистической линеаризацией (МСЛ). Представлено алгоритмическое обеспечение МНА (МСЛ) для СтС с СТН. Приведены тестовые примеры. Рассмотрены возможные обобщения полученных результатов.

Ключевые слова: аналитическое и статистическое моделирование; метод нормальной аппроксимации (МНА); метод статистической линеаризации (МСЛ); многочлены Эрмита; нормальный стохастический процесс; сложные трансцендентные нелинейности (СТН); стохастические системы (СтС)

DOI: 10.14357/19922264150203

1 Введение

Рассмотрим развитие методов аналитического и статистического моделирования нормальных СтП, приведенных в [1–4], на случай непрерывных и дискретных СтС, в том числе и на многообразиях с СТН. В разд. 2 даны типовые представления трансцендентных функций и соответствующих СТН. Раздел 3 посвящен уравнениям МНА и МСЛ. Алгоритмическое обеспечение МНА (МСЛ) для СтС с СТН представлено в разд. 4. Тестовые примеры приведены в разд. 5. В заключении даны выводы и рассмотрены некоторые обобщения полученных результатов.

2 Трансцендентные функции и нелинейности

Как известно [5, 6], трансцендентной аналитической функцией (ТАФ) в узком смысле слова называется мероморфная функция в плоскости комплексного переменного y , отличная от рациональной. В частности, сюда относятся целые ТАФ, целые функции, отличные от многочленов, например показательная функция e^y , тригонометрические функции $\sin y$ и $\cos y$, гиперболические

функции $\operatorname{sh} y$ и $\operatorname{ch} y$. Примерами собственно мероморфных ТАФ могут служить функции $\operatorname{tg} y$, $\operatorname{ctg} y$, $\operatorname{th} y$ и $\operatorname{cth} y$.

В рамках теории элементарных функций к трансцендентным вычислительным операциям относятся операции взятия тригонометрических (гиперболических) или обратных функций, логарифмирования и потенцирования [5].

В широком смысле под ТАФ понимается всякая аналитическая функция, отличная от алгебраической, для вычисления значений которой помимо алгебраических операций над аргументом необходимо применить предельный переход в той или иной форме. Примерами предельных переходов могут быть различные интегральные, интегродифференциальные и другие операторные преобразования.

Примерами СТН, получаемых посредством отрезков сумм элементарных ТАФ, могут служить следующие:

$$\varphi^{\text{CTH}}(Y, t) = \sum_{r=1}^n l_{rt} \varphi_r^{\text{TH}}(Y); \quad (1)$$

$$\varphi^{\text{CTH}}(Y, t) = \sum_{r=1}^n l_{rt} \varphi_r^{\text{TH}}(Y) \varphi_r^{\text{AH}}(Y), \quad (2)$$

* 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, ekorepanov@ipiran.ru

а также дробно-рациональные представления:

$$\varphi^{\text{CTH}}(Y, t) = \frac{\sum_{r=1}^{n'} l'_{rt} \varphi_r^{\text{TH}}(Y)}{\sum_{r=1}^{n''} l''_{rt} \varphi_r^{\text{TH}}(Y)}; \quad (3)$$

$$\varphi^{\text{CTH}}(Y, t) = \frac{\sum_{r=1}^{n'} l'_{rt} \varphi_r^{\text{TH}}(Y) \varphi_r^{\text{AH}}(Y)}{\sum_{r=1}^{n''} l''_{rt} \varphi_r^{\text{TH}}(Y) \varphi_r^{\text{AH}}(Y)}, \quad (4)$$

где $\varphi_r^{\text{TH}}(Y)$, $\varphi_r^{\text{TH}}(Y)$ и $\varphi_r^{\text{TH}}(Y)$ — элементарные ТАФ; l_{rt} , l'_{rt} и l''_{rt} — коэффициенты, зависящие от времени t ; $\varphi_r^{\text{AH}}(Y)$, $\varphi_r^{\text{AH}}(Y)$ и $\varphi_r^{\text{AH}}(Y)$ — алгебраические нелинейности (многочлены, степенные, иррациональные, дробно-рациональные и другие функции).

Другими примерами СТН являются нелинейности, получаемые путем соответствующего преобразования аргумента:

$$\varphi^{\text{CTH}}(Y, t) = \varphi^{\text{AH}}(\psi^{\text{TH}}(Y, t), t); \quad (5)$$

$$\varphi^{\text{CTH}}(Y, t) = \varphi^{\text{TH}}(\psi^{\text{AH}}(Y, t), t), \quad (6)$$

где φ^{AH} , ψ^{AH} , $\varphi^{\text{TH}}(Y, t)$ и $\psi^{\text{TH}}(Y, t)$ — элементарные алгебраические и трансцендентные нелинейности.

Операция интегрирования ТАФ может вывести из класса элементарных ТАФ, т.е. интеграл от элементарной ТАФ не всегда может быть выражен через элементарные функции (алгебраические или трансцендентные). В результате приходится обращаться к специальным функциям [5–8]. Простейшими примерами специальных функций могут служить интегралы от элементарных ТАФ, а также функций, получающихся из них с помощью конечного числа вычислительных операций и операций дифференцирования [5]. Сюда относятся и функции, обратные указанным.

В качестве примеров скалярных СТН векторного аргумента $Y = [Y_1 \dots Y_p]^T$ рассмотрим следующие:

$$\varphi^{\text{CTH}}(Y, t) = \sum_{r=1}^n \prod_{h=1}^H l_{rh,t} \varphi_{rh}^{\text{TH}}(Y_h); \quad (7)$$

$$\varphi^{\text{CTH}}(Y, t) = \frac{\sum_{r=1}^{n'} \prod_{h=1}^{H'} l'_{rh,t} \varphi_{rh}^{\text{TH}}(Y_h)}{\sum_{r=1}^{n''} \prod_{h=1}^{H''} l''_{rh,t} \varphi_{rh}^{\text{TH}}(Y_h)}. \quad (8)$$

В случае векторных и матричных СТН формулы (1)–(8) имеют место для соответствующих компонент.

В настоящей статье ограничимся рассмотрением трансцендентных нелинейностей (ТН), получающихся согласно (1)–(6) только из элементарных

трансцендентных функций. Случаи специальных аналитических и разрывных функций будут предметом дальнейшего рассмотрения.

3 Уравнения методов нормальной аппроксимации и статистической линеаризации со сложными трансцендентными нелинейностями

Как известно [9–11], уравнения конечномерных непрерывных нелинейных систем со стохастическими возмущениями путем расширения вектора состояния СтС могут быть записаны в виде следующего векторного стохастического дифференциального уравнения Ито:

$$dY_t = a(Y_t, t) dt + b(Y_t, t) dW_0 + \int_{R_0} c(Y_t, t, v) P^0(dt, dv), \quad Y(t_0) = Y_0. \quad (9)$$

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

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

В (9) принято: \int_{Δ} — число скачков пуассоновского СтП в интервале времени $\Delta = (t_1, t_2]$; $\nu_P(t, A)$ — интенсивность пуассоновского СтП $P(t, A)$; A — некоторое борелевское множество пространства R_0^q с выколотым началом. Начальное значение Y_0 представляет собой случайную величину (с.в.), не зависящую от приращений $W_0(t)$ и $P(t, A)$ на интервалах времени, следующих за t_0 , $t_0 \leq t_1 \leq t_2$, для любого множества A .

Для аддитивных нормальных (гауссовских) и обобщенных пуассоновских возмущений уравнение (9) имеет вид:

$$\dot{Y} = a(Y_t, t) + b_0(t)V, \quad V = \dot{W}, \quad Y(t_0) = Y_0. \quad (10)$$

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

Для компонент $\varphi(Y_t, t) = \{a_h, b_{kj}, c_h\}$ функций a , b и c , являющихся СТН, примем представления (1)–(6). Если предположить существование конечных вероятностных моментов второго порядка для моментов времени t_1 и t_2 , то уравнения МНА примут следующий вид [9–11]:

– для характеристических функций:

$$\left. \begin{aligned} g_1^N(\lambda; t) &= \exp \left[i\lambda^T m_t - \frac{1}{2} \lambda^T K_t \lambda \right]; \\ g_{t_1, t_2}^N(\lambda_1, \lambda_2; t_1, t_2) &= \\ &= \exp \left[i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \lambda \right], \end{aligned} \right\} \quad (11)$$

где

$$\bar{\lambda} = [\lambda_1^T \lambda_2^T]^T; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T;$$

$$\bar{K}_2 = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix};$$

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

$$\dot{m}_t = a_1(m_t, K_t, t), \quad m_0 = m(t_0); \quad (12)$$

$$\dot{K}_t = a_2(m_t, K_t, t), \quad K_0 = K(t_0); \quad (13)$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = K(t_1, t_2) a_{21}(m_{t_2}, K_{t_2}, t_2)^T, \\ K(t_1, t_1) = K_{t_1}. \quad (14)$$

Здесь приняты следующие обозначения:

$$m_t = M_{\Delta_y}^N Y_t; \quad Y_t^0 = Y_t - m_t; \quad K_t = M_{\Delta_y}^N Y_t^0 Y_t^{0T};$$

$$K(t_1, t_2) = M_{\Delta_y}^N Y_{t_1}^0 Y_{t_2}^{0T};$$

$$a_1 = a_1(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t);$$

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

$$a_{21} = a_{21}(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t) Y_t^{0T};$$

$$a_{22} = a_{22}(m_t, K_t, t) = M_{\Delta_y}^N \sigma(Y_t, t);$$

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

где $M_{\Delta_y}^N$ — символ вычисления математического ожидания для нормальных распределений (11).

Для стационарных СтС нормальные стационарные СтП — если они существуют, то $m_t = \bar{m}$, $K_t =$

\bar{K} , $K(t_1, t_2) = k(\tau)$ ($\tau = t_1 - t_2$), — определяются уравнениями [9–11]:

$$a_1(\bar{m}, \bar{K}) = 0; \quad a_2(\bar{m}, \bar{K}) = 0; \quad (15)$$

$$\dot{k}_\tau(\tau) = a_{21}(\bar{m}, \bar{K}) \bar{K}^{-1} k(\tau), \quad k(0) = \bar{K} \\ (\forall \tau > 0), \quad k(\tau) = k(-\tau)^T \quad (\forall \tau < 0). \quad (16)$$

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

Уравнения МНА в случае СтС (10) переходят в уравнения МСЛ [9–11], если принять

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

$$b(Y_t, t) = b_0(t), \quad \sigma(Y_t, t) = b_0(t) \nu(t) b_0(t)^T = \sigma_0(t);$$

$$k_1^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) a_0(m_t, K_t, t)^T \right]^T; \quad (17)$$

$$\dot{m}_t = a_1(m_t, K_t, t), \quad m_0 = m(t_0); \quad (18)$$

$$\dot{K}_t = k_1^a(m_t, K_t, t) K_t + K_t k_1^a(m_t, K_t, t)^T + \sigma_0(t); \\ K_0 = K(t_0); \quad (19)$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = K(t_1, t_2) K_{t_2} k_1^a(m_{t_2}, K_{t_2}, t_2)^T, \\ K(t_1, t_2) = K_{t_1}. \quad (20)$$

Для стационарных СтС (10) при условии асимптотической устойчивости матрицы $k_1^a(\bar{m}, \bar{K})$ в основе МСЛ лежат уравнения (15) и (16), записанные в виде:

$$a_1(\bar{m}, \bar{K}) = 0; \quad (21)$$

$$k_1^a(\bar{m}, \bar{K}) \bar{K} + \bar{K} k_1^a(\bar{m}, \bar{K})^T + \bar{\sigma}_0 = 0; \quad (22)$$

$$\dot{k}_\tau(\tau) = k_1^a(\bar{m}, \bar{K}) k(\tau), \quad k(0) = \bar{K} \\ (\forall \tau > 0), \quad k(\tau) = k(-\tau)^T \quad (\forall \tau < 0). \quad (23)$$

Уравнения (11)–(14) лежат в основе МНА для СтС (9), а уравнения (17)–(20) — в основе МСЛ для СтС (10). Для определения стационарных СтП согласно МНА служат соотношения (15) и (16), а МСЛ — (21)–(23).

Теперь рассмотрим дискретную СтС с СТН, описываемую уравнениями вида [4, 10]:

$$Y_{k+1} = a_k(Y_k) + b_k(Y_k) V_k^d \quad (k = 1, 2, \dots). \quad (24)$$

Здесь Y_k — $(p \times 1)$ -мерный вектор состояния, $Y_k \in \Delta_y$ (Δ_y — многообразие состояний); функции $a_k(Y_k)$ и $b_k(Y_k)$ имеют размерности $(p \times 1)$ и $(p \times m)$ соответственно; через V_k^d обозначен векторный дискретный шум, обладающий интенсивностью ν_k^d . В случае аддитивного шума, когда $b_k(Y_k) = b_{0k}$, уравнение (24) примет вид:

$$Y_{k+1} = a_k Y_k + b_{0k} V_k^d. \quad (25)$$

В основе МНА лежат следующие соотношения и уравнения [4, 10]:

$$g_{1k}^N(\lambda) = \exp\left(i\lambda m_k - \frac{1}{2} \lambda^T K_k \lambda\right);$$

$$g_{k_1 k_2}^N = \exp\left(i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \bar{\lambda}\right);$$

$$m_{k+1} = a_{1k} = M_{\Delta_y}^N a_k, \quad m_1 = M_{\Delta_y}^N Y_1;$$

$$K_{k+1} = a_{2k} = M_{\Delta_y}^N [a_k a_k^T] - [M_{\Delta_y}^N a_k] [M_{\Delta_y}^N a_k^T] + M_{\Delta_y}^N [b_k \nu_k^d b_k^T], \quad K_1 = M_{\Delta_y}^N Y_1^0 Y_1^{0T};$$

$$K(l, h) = a_{3k} = M_{\Delta_y}^N Y_l^0 a_h (Y_h)^T,$$

$$K(l, l) = K_l \quad (l < h), \quad K(l, h) = K(h, l)^T \quad (l > h).$$

В основе МСЛ для (25) после статистической линеаризации функции $a_k(Y_k)$ согласно

$$a_k(Y_k) = a_{0k}(m_k, K_k) + k_{1k}^a(m_k, K_k) Y_k^0$$

будут лежать уравнения [4, 10]:

$$m_{k+1} = a_{0k}, \quad m(1) = m_1;$$

$$K_{k+1} = k_{1k}^a K_k (k_{1k}^a)^T + b_{0k} \nu_k^d b_{0k}^T, \quad K(1) = K_1;$$

$$K(l, h+1) = K(l, h) (k_{1h}^a)^T,$$

$$K(l, l) = K_l \quad (l < h), \quad K(l, h) = K(h, l)^T \quad (l > h).$$

Для определения стационарных СтП согласно МНА и МСЛ с характеристиками

$$m_k = \bar{m}; \quad K_k = \bar{K}; \quad K(l, h) = \bar{k}(r) \quad (r = h - l)$$

используются уравнения:

$$\bar{m} = a_{1k}(\bar{m}, \bar{K});$$

$$\bar{K} = a_{2k}(\bar{m}, \bar{K});$$

$$\bar{K} = k_{1k}^a \bar{K} (k_{1k}^a)^T + b_{0k} \nu_k^d b_{0k}^T;$$

$$\bar{k}(r+1) = k(r) (k_{1k}^a)^T, \quad \bar{k}(0) = \bar{K}.$$

Как следует из уравнений МНА, необходимо уметь вычислять следующие интегралы:

$$I_{0k}^a = I_{0k}^a(m_k, K_k) = M_{\Delta_y}^N a_k(Y_k); \quad (26)$$

$$I_{1k}^a = I_{1k}^a(m_k, K_k) = M_{\Delta_y}^N a_k(Y_k) Y_k^{0T}; \quad (27)$$

$$I_{0k}^\sigma = I_{0k}^\sigma(m_k, K_k) = M_{\Delta_y}^N \sigma(Y_k)$$

$$(\sigma(Y_k) = b_k \nu_k^d b_k^T).$$

Для МСЛ достаточно вычислить интеграл (26), причем интеграл (27) вычисляется по формуле:

$$k_{1k}^a = k_{1k}^a(m_k, K_k) = \left[\frac{\partial}{\partial m_k} I_{0k}^a(m_k, K_k)^T \right]^T.$$

4 Алгоритмическое обеспечение аналитического и статистического моделирования

Как следует из (21), для МНА необходимо уметь вычислять следующие интегралы:

$$I_0^a = I_0^a(m_t, K_t, t) = a_1(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t); \quad (28)$$

$$I_1^a = I_1^a(m_t, K_t, t) = a_{21}(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t) Y_t^{0T};$$

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

Для МСЛ достаточно вычислить интеграл (28), причем интеграл I_1^a вычисляется по формуле [9–11]:

$$k_{1k}^a = k_{1k}^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) I_0^a(m_t, K_t, t)^T \right]^T.$$

При аналитическом моделировании для элементарных одномерных ТН скалярного аргумента, а также их суперпозиции составлены таблицы формул (см., например, [9–11]). В [11] приведены также таблицы формул для двух-, трех-, четырех- и n -мерных аргументов. Соответствующие формулы для круговых ТН даны в [12, 13].

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

В [14] изложены алгоритмы аналитического и статистического моделирования распределений (в том числе нормальных) в нелинейных СтС на многообразиях. Алгоритмы аналитического, статистического моделирования для СтС с СТН, а также смешанные алгоритмы различной степени точности относительно шага интегрирования также представлены в [14].

5 Примеры

1. Для СТН вида

$$\varphi_1^{\text{СТН}}(Y) = \sin(pY^2 + 2qY + r),$$

$$\varphi_2^{\text{СТН}}(Y) = \cos(pY^2 + 2qY + r)$$

интегралы (28) с учетом [7, 8] имеют следующий вид:

$$I_{01}(m, D) = \frac{1}{\sqrt{2\pi D}} A_1, \quad I_{02}(m, D) = \frac{1}{\sqrt{2\pi D}} A_2,$$

где

$$\begin{aligned} A_1 &= \\ &= \frac{\sqrt{\pi}}{\sqrt[4]{a^2 + p^2}} \exp \left[\frac{a(b^2 - ac) - (aq^2 - 2bpq + cp^2)}{a^2 + p^2} \right] \times \\ &\times \sin \left[\frac{1}{2} \operatorname{actg} \frac{p}{a} - \frac{p(q^2 - pr) - (b^2p - 2abq + a^2r)}{a^2 + p^2} \right]; \\ A_2 &= \\ &= \frac{\sqrt{\pi}}{\sqrt[4]{a^2 + p^2}} \exp \left[\frac{a(b^2 - ac) - (aq^2 - 2bpq + cq^2)}{a^2 + p^2} \right] \times \\ &\times \cos \left[\frac{1}{2} \operatorname{actg} \frac{p}{a} - \frac{p(q^2 - pr) - (b^2p - 2abq + a^2r)}{a^2 + p^2} \right] \\ &\left(a = \frac{1}{2D}, \quad b = -\frac{m}{2D}, \quad c = \frac{m^2}{2D} \right). \end{aligned}$$

2. Для некоторых СТН при $m = 0$, $D \neq 0$ на основе [7, 8] получены следующие выражения для интегралов (28):

(а) $\varphi^{\text{СТН}}(Y) = Y^{-1} \sin aY \quad (a > 0)$;

$$I_0^\varphi(0, D) = \frac{a}{\sqrt{2\beta}} \sum_{j=0}^{\infty} \frac{(-1)^j}{j(2j+1)} \left(\frac{a}{2\beta} \right)^{2j} \left(\beta^2 = \frac{1}{2D} \right);$$

(б) $\varphi^{\text{СТН}}(Y) = Y^2 \exp(-aY^{-2})$; $\mu = \frac{1}{2D}$;

$$I_0^\varphi(0, D) = \frac{1}{2\sqrt{2D}} \mu^{-3/2} (1 + 2\sqrt{a\mu}) \exp(-2\sqrt{a\mu});$$

(в) $\varphi^{\text{СТН}}(Y) = Y^{-2} \exp(-aY^{-2})$; $\mu = \frac{1}{2D}$;

$$I_0^\varphi(0, D) = \frac{1}{\sqrt{2aD}} \exp(-2\sqrt{a\mu});$$

(г) $\varphi^{\text{СТН}}(Y) = \operatorname{sh}^2 aY \quad (a > 0)$;

$$I_0^\varphi(0, D) = \frac{1}{2\sqrt{2\mu D}} \left(\exp \frac{a^2}{\mu} - 1 \right); \quad \mu = \frac{1}{2D};$$

(д) $\varphi^{\text{СТН}}(Y) = \operatorname{ch}^2 aY \quad (a > 0)$;

$$I_0^\varphi(0, D) = \frac{1}{2\sqrt{2\mu D}} \left(\exp \frac{a^2}{\mu} + 1 \right); \quad \mu = \frac{1}{2D}.$$

3. Рассмотрим скалярную дифференциальную СтС (10) при $a(Y_t) = \varphi^{\text{СТН}}(Y_t)$. Уравнения МСЛ (18), (19) имеют следующий вид:

$$\dot{m} = a_1(m, D), \quad m(t_0) = m_0;$$

$$\dot{D} = a_2(m, D), \quad D(t_0) = D_0.$$

Воспользуемся известным результатом из теории вычисления интегралов путем разложения по многочленам Эрмита [7, 8]:

$$\int_{-\infty}^{\infty} e^{-\xi^2} \varphi(\xi) d\xi = \sum_{j=1}^n w_j \varphi(\xi_j) + R_n.$$

Здесь $H_n = H_n(\xi)$ — многочлен Эрмита; ξ_j — j -й нуль H_n ; w_j и R_n — весовые коэффициенты и остаточный член, вычисляемые по формулам:

$$w_j = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(\xi_j)]^2};$$

$$R_n = \frac{n! \sqrt{\pi}}{2^n (2n)!} \varphi^{2n}(\xi) \quad (-\infty < \xi < \infty).$$

Опуская промежуточные вычисления, приведем окончательные формулы для вычисления правых частей уравнений МСЛ:

$$a_1(m, D) = \sum_{j=1}^n w_j \bar{a}(m, D; \xi_j);$$

$$\bar{a}(m, D; \xi_j) = \frac{a(m + \xi_j \sqrt{2D})}{\sqrt{2\pi D}};$$

$$a_2(m, D) = \sum_{j=1}^n w_j \bar{\sigma}(m, D; \xi_j);$$

$$\bar{\sigma}(m, D; \xi_j) =$$

$$= \frac{2a(m + \xi_j \sqrt{2D})(m + \xi_j \sqrt{2D}) + \sigma(m + \xi_j \sqrt{2D})}{\sqrt{2\pi D}}.$$

Ограничиваясь небольшим числом многочленов Эрмита H_n , получаем алгоритмы МСЛ для различных СТН, приведенных в примерах 1 и 2.

6 Заключение

Рассмотрены дифференциальные и разностные СтС (в том числе и на многообразиях) с винеровскими и пуассоновскими шумами и с СТН. Такие модели описывают поведение многих современных нано- и квантовооптических технических средств информатики. Приводятся уравнения МНА и МСЛ для моделирования нестационарных и стационарных нормальных процессов. Рассматриваются методы вычисления типовых интегралов для одно- и многомерных СТН скалярного и векторного аргумента, получающихся из суперпозиции элементарных ТН. Обсуждается алгоритмическое обеспечение аналитического и статистического моделирования. Особый интерес представляет раз-

вите алгоритмов для ТН на основе специальных функций. Приводятся тестовые примеры.

Результаты допускают обобщение на случай интегродифференциальных и операторных СтС с СТН, в том числе с автокоррелированными шумами.

Литература

1. Сеницын И. Н., Сеницын В. И. Аналитическое моделирование нормальных процессов в стохастических системах со сложными нелинейностями // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 2–4.
2. Сеницын И. Н., Сеницын В. И., Сергеев И. В., Белоусов В. В., Шоргин В. С. Математическое обеспечение аналитического моделирования стохастических систем со сложными нелинейностями // Системы и средства информатики, 2014. Т. 24. № 3. С. 4–29.
3. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р. Моделирование нормальных процессов в стохастических системах со сложными иррациональными нелинейностями // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 2–8.
4. Сеницын И. Н., Сеницын В. И., Сергеев И. В., Корепанов Э. Р., Белоусов В. В., Шоргин В. С. Математическое обеспечение моделирования нормальных процессов в стохастических системах со сложными иррациональными нелинейностями // Системы и средства информатики, 2015. Т. 25. № 2. С. 3–19.
5. Попов Б. А., Теслер Г. С. Вычисление функций на ЭВМ: Справочник. — Киев: Наукова Думка, 1984. 599 с.
6. Кудрявцев Л. Д., Соломенцев Е. Д. Трансцендентная функция // Математическая энциклопедия / Гл. ред. И. М. Виноградов. — М.: Советская энциклопедия, 1984. С. 425.
7. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. — М.: ГИФМЛ, 1963. 1100 с.
8. Справочник по специальным функциям / Под ред. М. Абрамовича, И. Стигана. — М.: Наука, 1979. 832 с.
9. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. [Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering. — Chichester, New York, NY, USA: John Wiley, 1987. 549 p.]
10. Пугачев В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
11. Сеницын И. Н., Сеницын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
12. Сеницын И. Н. Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 12–18.
13. Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Конашенкова Т. Д. Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем // Системы и средства информатики, 2012. Вып. 22. № 1. С. 29–40.
14. 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.

Поступила в редакцию 25.02.15

MODELING OF NORMAL PROCESSES IN STOCHASTIC SYSTEMS WITH COMPLEX TRANSCENDENTAL NONLINEARITIES

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: Development of methods for analytical and statistical modeling for discrete and continuous stochastic systems (StS) with Wiener and Poisson noises and with complex transcendental nonlinearities (CTN) is given. Typical representation of scalar and vector CTN is considered. Equations for the normal approximation method (NAM) and the method of statistical linearization (MSL) are deduced. Also, NAM and MSL for StS with CTN algorithms are given. Test examples are presented. Some generalizations are given.

Keywords: analytical and statistical modeling; complex transcendental nonlinearities (CTN); Hermite polynomials; method of statistical linearization (MSL); normal approximation method (NAM); stochastic systems (StS)

DOI: 10.14357/19922264150203

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 15-07-02244).

References

1. Sinitsyn, I. N., and V. I. Sinitsyn. 2014. Analiticheskoe modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi nelineynostyami [Analytical modeling of normal processes in stochastic systems with complex nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):2–4.
2. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, V. V. Belousov, and V. S. Shorgin. 2014. Matematicheskoe obespechenie analiticheskogo modelirovaniya stokhasticheskikh sistem so slozhnymi nelineynostyami [Mathematical software for analytical modeling of stochastic systems with complex nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(3):4–29.
3. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2015. Modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi irratsional'nymi nelineynostyami [Modeling of normal processes in stochastic systems with complex irrational nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):2–8.
4. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, E. R. Korepanov, V. V. Belousov, and V. S. Shorgin. 2015. Matematicheskoe obespechenie modelirovaniya normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi irratsional'nymi nelineynostyami [Mathematical software for modeling of normal processes in stochastic systems with complex irrational nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):3–19.
5. Popov, B. A., and G. S. Tesler. 1984. *Vychislenie funktsiy na EVM: Spravochnik* [Computing of functions: Handbook]. Kiev: Naukova Dumka. 599 p.
6. Kudryavtsev, L. D., and E. D. Solomentsev. 1984. Transepdentnaya funktsiya [Transcendental function]. *Matematicheskaya entsiklopediya* [Mathematical encyclopedia]. Ed. I. M. Vinogradov. Moscow: Sovetskaya Entsiklopediya. 425.
7. Gradshteyn, I. S., and I. M. Ryzhik. 1963. *Tablitsy integralov, summ, ryadov i proizvedeniy*. [Tables of integrals, sums, series, and products]. Moscow: GIFML. 1100 p.
8. Abramovich, M., and I. Stigan, eds. 1979. *Spravochnik po spetsial'nym funktsiyam* [Handbook on special functions]. Moscow: Nauka. 832 p.
9. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York, NY: John Wiley. 549 p.
10. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
11. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. Lektsii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 488 p.
12. Sinitsyn, I. N. 2012. Matematicheskoe obespechenie dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem, osnovannoe na parametrizatsii raspredeleniy [Mathematical software for analysis of nonlinear multichannel circular stochastic systems based on parametrization of distributions]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):12–18.
13. Sinitsyn, I. N., E. R. Korepanov, V. V. Belousov, and T. D. Konashenkova. 2012. Razvitie matematicheskogo obespecheniya dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem [Development of mathematical software for analysis of nonlinear multichannel circular stochastic systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):29–40.
14. 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.

Received February 25, 2015

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; VSinitsyn@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

НОРМАЛЬНЫЕ УСЛОВНО-ОПТИМАЛЬНЫЕ ФИЛЬТРЫ ПУГАЧЁВА ДЛЯ ДИФФЕРЕНЦИАЛЬНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ, ЛИНЕЙНЫХ ОТНОСИТЕЛЬНО СОСТОЯНИЯ*

И. Н. Сеницын¹, Э. Р. Корепанов²

Аннотация: Рассматриваются вопросы аналитического синтеза нормальных условно-оптимальных фильтров Пугачёва (НФП) для обработки информации в дифференциальных негауссовских стохастических системах (СтС), линейных относительно состояния (условия Липцера–Ширяева). Особое внимание уделено синтезу НФП для СтС при условиях Липцера–Ширяева на основе аппроксимации апостериорного распределения нормальным и квазилинейным НФП, основанным на статистической линейаризации нелинейных функций, зависящих от наблюдений. Для СтС высокой размерности путем выбора структурных функций, отражающих аналитическую природу наблюдаемой системы, можно синтезировать НФП, простыми в компьютерной реализации и для работы в режиме реального времени. Изложенные алгоритмы положены в основу модуля инструментального программного обеспечения «StS-Filter». Даны тестовые примеры. Приводятся некоторые обобщения.

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

DOI: 10.14357/19922264150204

1 Введение

Многие практические задачи обработки информации в статистических научных исследованиях основаны на использовании теории фильтрации процессов в СтС, линейных относительно состояния [1–6]. Первые работы в этом направлении для гауссовских систем выполнены Липцером и Ширяевым [7], а для негауссовских на основе субоптимальной фильтрации — Пугачёвым и Сеницыным (см. обзор [1]). В [5, 6] изучены вопросы синтеза и устойчивости фильтров для линейных СтС с аддитивными и мультипликативными негауссовскими шумами. Статья посвящена вопросам аналитического синтеза и устойчивости НФП для нелинейных негауссовских СтС, линейных относительно состояния.

2 Дифференциальные стохастические системы, линейные относительно состояния

Рассмотрим нелинейную дифференциальную СтС [1]:

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

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

заданную на многообразиях $\Delta = \Delta^{x,y}$ и Δ^V . Здесь X_t и Y_t — векторы состояния и наблюдения размерности n_x и n_y ; $V = \dot{W}$, W — векторный процесс с независимыми приращениями, состоящий из винеровской $W_0(t)$ и пуассоновской частей:

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

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 центрированной 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 $(n_2 \times 1)$, $(n_x \times n_v)$, $(n_y \times 1)$

* 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, ekoeranov@ipiran.ru

и $(n_y \times n_v)$, удовлетворяющие следующим условиям Липцера–Ширяева [7]:

- функции φ и φ_1 линейны относительно состояния X_t :

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

- функции ψ и ψ_1 не зависят от состояния X_t :

$$\psi(X_t, Y_t, t) = \psi(Y_t, t); \quad \psi_1(X_t, Y_t, t) = \psi_1(Y_t, t).$$

Предполагается, что уравнения СтС (1) и (2) понимаются в смысле Ито и имеют решение в среднем квадратическом (с.к.) [1]. Систему (1)–(2) будем называть гауссовской, если $V = \dot{W}_0$, а X_0 и Y_0 — гауссовские. Важный частный случай (1) и (2) составляют уравнения с аддитивными шумами, когда

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

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

$$\left. \begin{aligned} V &= [V_1^T V_2^T]^T, \quad \nu = \begin{bmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{bmatrix}; \\ \psi V &= \psi' V_1; \quad \psi_1 V = \psi_1' V_2. \end{aligned} \right\} \quad (4)$$

Замечание 2. Для случая, когда уравнения СтС линейны и содержат гауссовские и негауссовские аддитивные и мультипликативные шумы в уравнениях состояния и наблюдения, соответствующие уравнения приведены в [5, 6].

3 Дифференциальные фильтры Липцера–Ширяева

Для гауссовской СтС (1)–(2) при условиях Липцера–Ширяева известные следующие точные уравнения нелинейной фильтрации по критерию минимума с.к. ошибки [1, 7]:

$$\begin{aligned}\dot{\hat{X}}_t &= [a_1(Y_t, t)\hat{X}_t + a_0(Y_t, t)] + [R_t b_1(Y_t, t)]^T + \\ &+ (\psi \nu_0 \psi_1^T)(Y_t, t) (\psi_1 \nu_0 \psi_1^T)^{-1}(Y_t, t) \left\{ \dot{Y}_t - \right. \\ &\left. - [b_1(Y_t, t)\hat{X}_t + b_0(Y_t, t)] \right\}, \quad \hat{X}_{t_0} = \hat{X}_0; \quad (5)\end{aligned}$$

$$\begin{aligned}\dot{R}_t &= a_1(Y_t, t)R_t + R_t a_1(Y_t, t)^T + (\psi \nu_0 \psi^T)(Y_t, t) - \\ &- [R_t b_1(Y_t, t)]^T + (\psi \nu_0 \psi_1^T)(Y_t, t) (\psi_1 \nu_0 \psi_1^T)^{-1} \times \\ &\times (Y_t, t) [b_1(Y_t, t)R_t + (\psi_1 \nu_0 \psi^T)(Y_t, t)], \\ R_{t_0} &= R_0, \quad (6)\end{aligned}$$

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

Как и в случае линейной фильтрации при аддитивных шумах [1], уравнения дифференциального фильтра Липцера–Ширяева (ФЛШ) [4, 7] 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 (5) и (6). Однако в противоположность линейной фильтрации для ФЛШ нельзя вычислить R_t заранее, до получения результатов наблюдений, так как от последних зависят коэффициенты уравнения (2). Поэтому ФЛШ в данном случае должен выполнять интегрирование обоих уравнений (5) и (6). 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

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

Таким образом, имеет место следующее утверждение.

Теорема 1. Пусть в гауссовской системе (1)–(2) при условиях Липцера–Ширяева диффузионная матрица $\sigma_1 = \sigma_1(Y_t, t) = \psi_1 \nu_0 \psi_1^T(Y_t, t)$ не вырождена. Тогда с.к. оптимальный фильтр определяется уравнением (5), причем его точность оценивается согласно (6).

Замечание 3. Очевидно, что ФЛШ будет совпадать с обобщенным фильтром Калмана–Бьюси, фильтрами второго порядка, гауссовыми фильтрами [1, 4, 8], если записать условия Липцера–Ширяева в виде:

$$\left. \begin{aligned} a_1(Y_t, t)X_t + a_0(Y_t, t) &= \\ = \bar{\varphi}(\hat{X}_t, Y_t, t) + \bar{\varphi}_x(\hat{X}_t, Y_t, t)^T (X_t - \hat{X}_t); \\ b_1(Y_t, t)X_t + b_0(Y_t, t) &= \\ = \bar{\varphi}_1(\hat{X}_t, Y_t, t) + \bar{\varphi}_{1x}(\hat{X}_t, Y_t, t)^T (X_t - \hat{X}_t). \end{aligned} \right\} \quad (7)$$

Здесь

$$\begin{aligned}\bar{\varphi}_x(\hat{X}_t, Y_t, t) &= a_1(Y_t, t); \\ \bar{\varphi}_{1x}(\hat{X}_t, Y_t, t) &= b_1(Y_t, t); \\ \bar{\varphi} &= a_1(Y_t, t)\hat{X}_t + a_0(Y_t, t); \\ \bar{\varphi}_1 &= b_1(Y_t, t)\hat{X}_t + b_0(Y_t, t)\end{aligned}$$

и необходимо учесть, что $\bar{\varphi}_x$ и $\bar{\varphi}_{1x}$ не зависят от X_t .

4 Нормальные фильтры Пугачёва для дифференциальных стохастических систем, линейных относительно состояния

Следуя [1], будем искать фильтр для оценки \hat{X}_t в виде следующего уравнения:

$$d\hat{X} = \alpha_t \xi(\hat{X}_t, Y_t, t) dt + \beta_t \eta(\hat{X}_t, Y_t, t) dY_t + \gamma_t dt, \quad (8)$$

где $\xi = \xi(\hat{X}_t, Y_t, t)$ и $\eta = \eta(\hat{X}_t, Y_t, t)$ — некоторые функции текущих значений наблюдаемого процесса Y_t , оценки \hat{X}_t и времени t ; α_t, β_t и γ_t — некоторые функции времени.

Если бы коэффициенты α_t, β_t и γ_t в (8) были известными функциями времени, то уравнение (8) 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 (8) как функции времени из условия минимума с.к. ошибки $M \left| \hat{X} - X_t \right|^2 = \min$ при всех $t > t_0$. Это приводит к теории условно-оптимального фильтра Пугачева (ФП), когда в уравнение фильтра задаются заранее и оптимизируются только коэффициенты этого уравнения. Итак, мы 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 Пугачёву [1], 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 [1]. Значит, надо определить такую оптимальность фильтра, при которой возможно решение задачи. Будем считать оптимальным такой допустимый фильтр, который на каждом бесконеч-

но 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 (7) в любой момент $t \geq t_0$, обеспечивающих минимум с.к. ошибки фильтрации $M \left| \hat{X}_s - X_s \right|^2$ 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 [1]. Общая теория ФП по с.к. критерию развита для уравнений (1), (2) и подробно изложена в [1]. 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 [1, 4, 8].

Применяя теорию ФП [1] к нормальным процессам в гауссовской СтС (1), (2) при условиях Липцера—Ширяева и $V = V_0$, приходим к НФП вида (8). Входящие в (8) коэффициенты определяются следующими уравнениями:

$$\alpha_t m_1 + \beta_t m_2 + \gamma_t = m_0; \quad (9)$$

$$m_0 = M_{\Delta}^N[\varphi]; \quad m_1 = M_{\Delta}^N[\xi]; \quad m_2 = M_{\Delta}^N[\eta]; \quad (10)$$

$$\beta_t = \kappa_{02} \kappa_{22}^{-1}; \quad (11)$$

$$\kappa_{02} = M_{\Delta}^N \left[(X_t - \hat{X}_t) \varphi_1^T \eta^T \right] + M_{\Delta}^N [\psi \nu_0 \psi_1^T \eta^T]; \quad (12)$$

$$\kappa_{22} = M_{\Delta}^N \eta \psi_1 \nu(t) \psi_1^T \eta^T. \quad (13)$$

$$\alpha_t \kappa_{11} + M_{\Delta}^N \left[(\hat{X}_t - X_t) (\xi^T \alpha_t^T + \gamma_t^T) \frac{\partial \xi^T}{\partial x} \right] = \kappa'_{01} - \beta_t \kappa'_{21}; \quad (14)$$

$$\kappa'_{21} = M_{\Delta}^N [\eta \varphi_1 - m_2] \xi; \quad (15)$$

$$\begin{aligned} \kappa'_{01} = & \kappa_{01} + M_{\Delta}^N \left[(X_t - \hat{X}_t) \frac{\partial \xi^T}{\partial t} \right] + \\ & + M_{\Delta}^N \left\{ (X_t - \hat{X}_t) \varphi_1^T + \right. \\ & \left. + \psi \nu_0 \psi_1^T - \beta_t \eta \psi_1 \nu_0 \psi_1^T \right\} \left(\frac{\partial}{\partial y} + \eta^T \beta_t^T \frac{\partial}{\partial x} \right) \xi^T + \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} M_{\Delta}^N (X_t - \hat{X}_t) \left\{ \text{tr} \left[\psi_1 \nu_0 \psi_1^T \left(\frac{\partial}{\partial y} + \right. \right. \right. \\
 & \quad \left. \left. \left. + 2\eta^T \beta_t^T \frac{\partial}{\partial x} \right) \frac{\partial^T}{\partial y} \right] + \right. \\
 & \quad \left. + \text{tr} \left[\beta_t \eta \psi_1 \nu_0 \psi_1^T \eta^T \beta_t^T \frac{\partial}{\partial x} \frac{\partial^T}{\partial x} \right] \right\} \xi^T; \quad (16)
 \end{aligned}$$

$$\begin{cases} \kappa_{11} = M_{\Delta}^N [\xi - m_1] \xi^T; \\ \kappa_{21} = M_{\Delta}^N [\varphi_1 - m_2] \xi^T; \\ \kappa_{01} = M_{\Delta}^N [\varphi - m_0] \xi^T. \end{cases} \quad (17)$$

Точность НФП определяется уравнением:

$$\begin{aligned}
 \dot{R}_t = & M_{\Delta}^N \left[(X_t - \hat{X}_t) \varphi (X_t, Y_t, t)^T + \right. \\
 & \left. + \varphi (X_t, Y_t, t) (X_t^T - \hat{X}_t^T) - \beta_t \eta (Y_t, \hat{X}_t, t) \times \right. \\
 & \left. \times \psi_1 (X_t, Y_t, t) \nu_0 \psi_1 (X_t, Y_t, t)^T \eta (Y_t, \hat{X}_t, t)^T \beta_t^T + \right. \\
 & \left. + \psi (X_t, Y_t, t) \nu_0 \psi (X_t, Y_t, t)^T \right]. \quad (18)
 \end{aligned}$$

Таким образом, справедливо следующее утверждение.

Теорема 2. Пусть для гауссовской системы (1)–(2) при условиях Липцера–Ширяева выполнены условия невырожденности матрицы κ_{22} (13) и конечности величин κ_{ij} ($i, j = 0, 1, 2$). Тогда НФП определяется уравнением (8), а α_t, β_t и γ_t — уравнениями (9)–(18).

Для негауссовской СтС в уравнениях теоремы 2 следует заменить ν_0 на ν согласно (3), а в выражении (16) учесть два дополнительных интегральных члена:

$$\begin{aligned}
 \kappa'_{01} = & \kappa_{01} + M_{\Delta}^N \left[(X_t - \hat{X}_t) \frac{\partial \xi^T}{\partial t} \right] + \\
 & + M_{\Delta}^N \left\{ (X_t - \hat{X}_t) \left[\varphi_1^T - \int_{R_0^g} c(\rho)^T \nu_P(t, \rho) d\rho \psi_1^T \right] + \right. \\
 & \left. + \psi \nu_0 \psi_1^T - \beta_t \eta \psi_1 \nu_0 \psi_1^T \right\} \left(\frac{\partial}{\partial y} + \eta^T \beta_t^T \frac{\partial}{\partial x} \right) \xi^T + \\
 & + \frac{1}{2} M_{\Delta}^N \left[(X_t - \hat{X}_t) \right] \times \\
 & \times \left\{ \text{tr} \left[\psi_1 \nu_0 \psi_1^T \left(\frac{\partial}{\partial y} + 2\eta^T \beta_t^T \frac{\partial}{\partial x} \right) \frac{\partial^T}{\partial y} \right] + \right. \\
 & \left. + \text{tr} \left[\beta_t \eta \psi_1 \nu_0 \psi_1^T \eta^T \beta_t^T \frac{\partial}{\partial x} \frac{\partial^T}{\partial x} \right] \right\} \xi^T + \\
 & + M_{\Delta}^N \int_{R_0^g} \left[X_t - \hat{X}_t + (\psi - \beta_t \eta \psi_1) c(\rho) \times \right.
 \end{aligned}$$

$$\begin{aligned}
 & \left. \times \xi (Y_t + \psi c(\rho), \hat{X}_t + \beta_t \eta \psi_1 c(\rho), t) - \right. \\
 & \left. - \xi^T \right]^T \nu_P(t, d\rho) d\rho, \quad (19)
 \end{aligned}$$

где функции φ, φ_1, ψ и ψ_1 удовлетворяют условиям Липцера–Ширяева. В результате имеем следующее утверждение.

Теорема 3. Пусть для негауссовской СтС (1)–(2) в условиях Липцера–Ширяева матрица κ_{22} (13) невырождена, а интегралы (9), (13), (15), (17) и (19) конечны. Тогда НФП определяется уравнением (8), а коэффициенты α_t, β_t и γ_t — утверждениями (9), (11) и (14).

Замечание 4. Теория НФП (оценивания состояния и параметров систем) не позволяет получить нормальные с.к. оптимальные фильтры. Можно получить только ФП, которые в общем случае хуже оптимальных, но зато легко реализуемы. Однако, если с.к. оптимальная оценка \hat{X}_t вектора X_t удовлетворяет уравнению допустимого фильтра (8) при каких-либо коэффициентах времени α_t, β_t и γ_t , то уравнения теорем 2 и 3, 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).

Замечание 5. 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 (8), зависящими лишь от соответствующих компонент вектора \hat{X}_t . Так, например, взяв ξ и η в (8), 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.

5 Дифференциальный нормальный фильтр Пугачёва на основе нормальной аппроксимации апостериорного распределения

Рассмотрим сначала гауссовскую СтС (1)–(2). 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 субоптимальной фильтрации

при аппроксимации апостериорного распределения вектора X_t нормальным распределением будем иметь следующие стохастические дифференциальные уравнения, определяющие \hat{X}_t и R_t [1]:

$$\dot{\hat{X}}_t = f(\hat{X}_t, Y_t, R_t, t) + h(\hat{X}_t, Y_t, R_t, t) \left[\dot{Y}_t - f^{(1)}(\hat{X}_t, Y_t, R_t, t) \right]; \quad (20)$$

$$\begin{aligned} \dot{R}_t = & \left\{ f^{(2)}(\hat{X}_t, Y_t, R_t, t) - \right. \\ & \left. - h(\hat{X}_t, Y_t, R_t, t) \psi_1 \nu_0 \psi_1^T(Y_t, t) h(\hat{X}_t, Y_t, R_t, t)^T \right\} + \\ & + \sum_{r=1}^{n_y} \rho_r(\hat{X}_t, Y_t, R_t, t) \left[\dot{Y}_r - f_r^{(1)}(\hat{X}_t, Y_t, R_t, t) \right], \quad (21) \end{aligned}$$

где

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

$$\begin{aligned} f^{(1)}(\hat{X}_t, Y_t, R_t, t) &= \left\{ f_r^{(1)}(\hat{X}_t, Y_t, R_t, t) \right\} = \\ &= M_{\Delta}^N \varphi(Y_t, \nu, t) = b_1(Y_t, t) \hat{X}_t + b_0(Y_t, t); \end{aligned}$$

$$\begin{aligned} h(\hat{X}_t, Y_t, R_t, t) &= M_{\Delta}^N [x \varphi_1(Y_t, x, t)^T + \\ &+ \psi \nu_0 \psi_1^T(Y_t, x, t)] - \hat{X}_t f^{(1)}(\hat{X}_t, Y_t, R_t, t)^T \times \\ &\times (\psi_1 \nu_0 \psi_1^T)^{-1}(Y_t, t) = \left[R_t b_1(Y_t, t)^T + \right. \\ &\left. + (\psi \nu_0 \psi_1^T)(Y_t, t) \right] (\psi \nu_0 \psi_1^T)^{-1}(Y_t, t); \end{aligned}$$

$$\begin{aligned} f^{(2)}(\hat{X}_t, Y_t, R_t, t) &= M_{\Delta}^N \left\{ (x - \hat{X}_t) \varphi(Y_t, x, t)^T + \right. \\ &+ \varphi(Y_t, x, t) (x^T - \hat{X}_t^T) + \psi \nu_0 \psi^T(Y_t, x, t) \left. \right\} = \\ &= \left[R_t b_1(Y_t, t)^T + (\psi \nu \psi_1^T)(Y_t, t) \right] (\psi_1 \nu_0 \psi_1^T)^{-1} \times \\ &\times (Y_t, t) [b_1(Y_t, t) R_t + (\psi_1 \nu_0 \psi^T)(Y_t, t)]; \end{aligned}$$

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

Здесь функции a_r — r -й элемент матрицы-строки

$$A_{\varphi_1} = (\varphi_1^T - \hat{\varphi}_n^T) (\psi_1 \nu_0 \psi_1^T)^{-1};$$

$B_{kr}^{\psi \psi_1}$ — элемент k -й строки и r -го столбца матрицы

$$B^{\psi \psi_1} = (\psi \nu_0 \psi_1^T) (\psi_1 \nu_0 \psi_1^T)^{-1};$$

b_r — r -й столбец матрицы $B^{\psi \psi_1}$:

$$b_r^{\psi \psi_1} = \left[b_{1r}^{\psi \psi_1} \dots b_{n_x r}^{\psi \psi_1} \right]^T.$$

Количество уравнений метода нормальной аппроксимации (МНА) одномерного апостериорного распределения определяется по формуле:

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

За начальные значения \hat{X}_t и R_t при интегрировании уравнений (20) и (21), естественно, следует принять условные математическое ожидание и ковариационную матрицу величины X_0 относительно Y_0 :

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

Замечание 6. Если нет информации об условном распределении X_0 относительно Y_0 , то начальные условия можно взять в виде:

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

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

Сравнивая уравнения (20) и (21) с уравнениями ФЛШ, имеем следующий результат.

Теорема 4. Для гауссовской СтС (1)–(2) при условиях Липцера–Ширяева НФП (на основе нормальной аппроксимации апостериорной плотности) и ФЛШ совпадают.

Для негауссовский СтС (1)–(2) при условиях Липцера–Ширяева, учитывая два дополнительных интегральных члена в (19), приходим к уравнениям (20), (21). При этом коэффициенты α_t и γ_t определяются (9), (14), (19) после нахождения β_t по формуле (11). Таким образом, имеем следующий результат.

Теорема 5. Пусть для негауссовской системы (1)–(2) при условиях Липцера–Ширяева выполнены условия невырожденности матрицы κ_{22} и конечности величин (9)–(19) ($i, j = 0, 1, 2$). Тогда НФП определяется уравнением (8), а коэффициенты α_t , β_t и γ_t — уравнениями (9)–(19).

6 Квазилинейный нормальный фильтр Пугачёва

Особое практическое значение имеет случай (1)–(2) при условиях Липцера–Ширяева с аддитивными (в общем случае негауссовскими) шумами. Следуя [1], 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 уравнения (1) и (2) приводятся к эквивалентной гауссовской системе, линейной относительно X_t^0 и Y_t^0 и нелинейной относительно m_t^x и m_t^y :

$$\dot{X}_t = \bar{a}Y_t + \bar{a}_1X_t + \bar{a}_0 + \bar{\psi}V; \quad (22)$$

$$\dot{Y}_t = \bar{b}Y_t + \bar{b}_1X_t + \bar{b}_0 + \bar{\psi}_1V. \quad (23)$$

Здесь

$$\left. \begin{aligned} \bar{a} &= k_y^{a_1x} + k_y^{a_0}; & \bar{b} &= k_y^{b_1x}; \\ \bar{a}_1 &= k_x^{a_1x}; & \bar{b}_1 &= k_x^{b_1x}; \\ a_0 &= (k_{0y}^{a_0} - k_{1y}^{a_0})m_t^y; & b_0 &= (k_{0y}^{b_0} - k_{1y}^{b_0})m_t^y. \end{aligned} \right\} \quad (24)$$

Правые части (24) зависят от вероятностных моментов первого и второго порядка и определяются из следующей линейной дифференциальной системы для вектора $Z_t = [X_t^T Y_t^T]^T$:

$$\begin{aligned} \dot{\bar{m}}_t^z &= c\bar{m}_t^z + c_0; \\ K_t^z &= cK_t^z + K_t^z c^T + l\nu l^T, \end{aligned}$$

где

$$c = \begin{bmatrix} \bar{a}_1 & \bar{a} \\ \bar{b}_1 & \bar{b} \end{bmatrix}; \quad c_0 = \begin{bmatrix} \bar{a}_0 \\ \bar{b}_0 \end{bmatrix}; \quad l = \begin{bmatrix} \bar{\psi}_t \\ \bar{\psi}_{1t} \end{bmatrix}.$$

Применяя теорию квазилинейной фильтрации [1] к уравнениям (22) и (23), получим следующие уравнения квазилинейного фильтра:

$$\begin{aligned} \dot{\hat{X}}_t &= \bar{a}Y_t + \bar{a}_1\hat{X}_t + \bar{a}_0 + \\ &+ \beta_t \left[\dot{Y}_t - (\bar{b}Y_t + \bar{b}_1\hat{X}_t + \bar{b}_0) \right]; \quad (25) \end{aligned}$$

$$\beta_t = R_t \bar{b}_1^T + (\bar{\psi}\nu\bar{\psi}_1^T) (\bar{\psi}_1\nu\bar{\psi}_1^T)^{-1}; \quad (26)$$

$$\begin{aligned} \dot{R}_t &= \bar{a}_1 R_t + R_t \bar{a}_1^T + \bar{\psi}\nu\bar{\psi}^T - \\ &- (R_t \bar{b}_1^T + \bar{\psi}\nu\bar{\psi}_1) (\bar{\psi}_1\nu\bar{\psi}_1^T)^{-1} (\bar{b}R_t + \bar{\psi}_1\nu\bar{\psi}_1^T). \quad (27) \end{aligned}$$

Таким образом, имеем следующий результат.

Теорема 6. Пусть уравнения негауссовской СтС (1)–(2) при условиях Липцера–Ширяева и с аддитивными шумами допускают применение метода статистической линеаризации (МСЛ). Тогда уравнения квазилинейного НФП имеют вид (25)–(27).

Замечание 7. Согласно [5, 6], из уравнения (27) получаются приближенные условия асимптотической устойчивости в терминах равномерной наблюдаемости и управляемости в следующем виде:

$$\begin{aligned} 0 \leq R_t(R_0, t_0) &\leq u_A(t, t_0)R_0u_A(t, t_0) + \\ &+ \int_{t_0}^t u_A(t, \tau)\sigma_{11}u_A(t, \tau)^T d\tau, \end{aligned}$$

где $u_A(t, \tau)$ — фундаментальная матрица однородного уравнения Риккати, причем

$$A = \bar{a}_1 - \beta_t \bar{b}_1 - R_t \bar{b}_1^T \sigma_{22}^{-1} b_1 - \sigma_{12} \sigma_{22}^{-1} \bar{b}_1;$$

$$\sigma_{11} = \bar{\psi}_{10t} \nu \bar{\psi}_{10}^T;$$

$$\sigma_{12} = \bar{\psi}_{10t} \nu \bar{\psi}_{20t}^T; \quad \sigma_{22} = \bar{\psi}_{20t} \nu \bar{\psi}_{20t}^T.$$

7 Заключение

1. Рассмотрены вопросы аналитического синтеза условно-оптимальных фильтров Пугачёва для обработки информации в дифференциальных негауссовских СтС, линейных относительно состояния (условия Липцера–Ширяева). Особое внимание уделено синтезу фильтров для СтС при условиях Липцера–Ширяева путем аппроксимации апостериорного распределения нормальным и квазилинейным фильтрами, основанным на статистической линеаризации нелинейных функций, зависящих от наблюдений.
2. Для СтС высокой размерности путем выбора структурных функций, отражающих аналитическую природу наблюдаемой системы, можно синтезировать простые в компьютерной реализации фильтры для работы в режиме реального времени.
3. Изложенные алгоритмы положены в основу модуля математического обеспечения инструментального программного обеспечения «StS-Filter». Тестирование проведено на основе примеров [1, 4–6, 8].
3. Полученные результаты позволяют синтезировать дифференциальным ФП, во-первых, для случая широкополосных шумов в уравне-

ниях (1) и (2), если аналогично [6] заменить широкополосный белый шум эквивалентным нормальным (гауссовским) белым шумом и, вторых, для случая автокоррелированных шумов. Результаты также обобщаются на случай дискретных СтС, линейных относительно состояния, если воспользоваться [3].

Приложение

Тестовые примеры

1. Найти приближенно оптимальный алгоритм оценивания состояния X_t системы, описываемой скалярным уравнением

$$\dot{X}_t = -\theta X_t + V_1, \quad (\text{П.1})$$

и неизвестного параметра θ по результатам наблюдения процесса

$$\dot{Y}_t = X_t + V_2, \quad (\text{П.2})$$

где V_2 — белый шум, независимый от V_1 .

Заменим параметр θ стохастическим процессом Θ_t , определяемым уравнением $\dot{\Theta}_t = 0$, и примем за расширенный вектор состояния пару $[X_t \Theta_t]^T$. Тогда уравнения НФП будут иметь вид:

$$\begin{aligned} \dot{\hat{X}}_t &= -\hat{X}_t \hat{\Theta}_t - R_{12} + \nu_2^{-1} R_{11} (\dot{Y}_t - \hat{X}_t); \\ \dot{\hat{\Theta}}_t &= \nu_2^{-1} R_{12} (\dot{Y}_t - \hat{X}_t); \\ \dot{R}_{11} &= \nu_1 - 2 (\hat{\Theta}_t R_{11} + \hat{X}_t R_{12}) - \nu_2^{-1} R_{11}^2; \\ \dot{R}_{12} &= -\hat{\Theta}_t R_{12} - \hat{X}_t R_{22} - \nu_1^{-1} R_{11} R_{12}; \\ \dot{R}_{22} &= -\nu_2^{-1} R_{12}^2, \end{aligned}$$

где R_{11} , R_{12} и R_{22} — апостериорные дисперсии и ковариация ошибок оценок \hat{X}_t и $\hat{\Theta}_t$ соответственно. За начальные значения \hat{X}_t , $\hat{\Theta}_t$, R_{11} , R_{12} и R_{22} следует принять соответствующие априорные величины, причем $\hat{\Theta}_0$, R_{220} и R_{120} всегда приходится брать произвольно, так как априорной информации о параметре θ обычно нет, за исключением, может быть, информации о возможном диапазоне его значений. Этот фильтр совпадает с обобщенным фильтром Калмана [1, 4, 8].

2. Найдем НФП для системы (П.1), (П.2). За основу класса допустимых фильтров примем первые два уравнения МНА и соответственно положим, что

$$\xi(\hat{X}_t, \hat{\Theta}_t, t) = [\hat{X}_t \hat{\Theta}_t \hat{X}_t]^T \eta(\hat{X}_t, \hat{\Theta}_t, t) = 1.$$

Тогда класс допустимых фильтров будет представлять собой систему двух уравнений:

$$\begin{aligned} \dot{\hat{X}}_t &= \alpha_{11} \hat{X}_t \hat{\Theta}_t + \alpha_{12} \hat{X}_t + \beta_1 \dot{Y}_t + \gamma_1; \\ \dot{\hat{\Theta}}_t &= \alpha_{21} \hat{X}_t \hat{\Theta}_t + \alpha_{22} \hat{X}_t + \beta_2 \dot{Y}_t + \gamma_2. \end{aligned}$$

Здесь

$$\begin{aligned} \kappa_{22} &= \nu_2; \\ \kappa_{02} &= M_N \left[(X_t - \hat{X}_t) X_t (\Theta_t - \hat{\Theta}_t) X_t \right]^T = \\ &= [m_{2000} - m_{1010} m_{1100} - m_{1001}]^T; \end{aligned}$$

$\beta_1 = \nu_2^{-1} (m_{2000} - m_{1010})$; $\beta_2 = \nu_2^{-1} (m_{1100} - m_{1001})$, где $m_{pqrs} = M_N X_t^p \Theta_t^q \hat{X}_t^r \hat{\Theta}_t^s$. Уравнения, определяющие оптимальные коэффициенты α_{11} , α_{12} , α_{21} , α_{22} , γ_1 и γ_2 , имеют следующий вид:

$$\begin{aligned} m_{0011} \alpha_{11} + m_{0010} \alpha_{12} + \gamma_1 &= -m_{1100} - \beta_1 m_{1000}; \\ m_{0011} \alpha_{21} + m_{0010} \alpha_{22} + \gamma_2 &= -\beta_1 m_{1000}; \\ (2m_{0022} - m_{1012} - m_{0011}^2) \alpha_{11} + \\ &+ (2m_{0021} - m_{1012} - m_{0010} m_{0011}) \alpha_{12} + \\ &+ (m_{0031} - m_{1021}) \alpha_{21} + (m_{0030} - m_{0020}) \alpha_{22} + \\ &+ (m_{0011} - m_{1001}) \gamma_1 + (m_{0020} - m_{1010}) \gamma_2 = \\ &= -m_{1111} - 2m_{1100} m_{0011} + \beta_1 (m_{2001} - 2m_{1011} + \\ &+ m_{1000} m_{0011}) + \beta_2 (m_{2010} - m_{1020}) - \nu_2 \beta_1^2 m_{0001} + \\ &+ \nu_2 \beta_1 \beta_2 (m_{1000} - 2m_{0010}); \\ (2m_{0021} - m_{1011} - m_{0010} m_{0011}) \alpha_{11} + \\ &+ (2m_{0020} - m_{1010} - m_{0010}^2) \alpha_{12} + \\ &+ (m_{0010} - m_{1000}) \gamma_1 = -m_{1110} + m_{1100} m_{1010} + \\ &+ \beta_1 (m_{2000} - 2m_{1010} + m_{1000} m_{0010}) - \nu_2 \beta_1^2; \\ (m_{0013} - m_{0112}) \alpha_{11} + (m_{0012} - m_{0111}) \alpha_{12} + \\ &+ (2m_{0022} - m_{0121} - m_{0011}^2) \alpha_{21} + \\ &+ (2m_{0021} - m_{0120} - m_{0010} m_{0001}) \alpha_{22} + \\ &+ (m_{0002} - m_{0101}) \gamma_1 + (m_{0011} - m_{0110}) \gamma_2 = \\ &= \beta_1 (m_{1101} - m_{1002}) + \beta_2 (m_{1110} - 2m_{1011} + \\ &+ m_{1000} m_{0011}) + \nu_2 \beta_1 \beta_2 (m_{0100} - 2m_{0001}) - \nu_2 \beta_2^2 m_{0010}; \\ (m_{0012} - m_{0111}) \alpha_{11} + (m_{0011} - m_{0110}) \alpha_{12} + \\ &+ (m_{0021} - m_{0011} m_{0010}) \alpha_{21} + \\ &+ (m_{0020} - m_{0010}^2) \alpha_{22} + (m_{0001} - m_{0100}) \gamma_1 = \\ &= \beta_1 (m_{1100} - m_{1001}) - \beta_2 (m_{1010} - m_{1000} m_{0010}) - \nu_2 \beta_1 \beta_2. \end{aligned}$$

Числовые значения получаются путем решения МНА нормальной апостериорной плотности с помощью инструментального программного обеспечения «StS-Filter».

3. Построим теперь НФП для оценивания одного только неизвестного параметра θ . Для определения класса допустимых фильтров включим неизвестную функцию времени \hat{X}_t в уравнение для $\dot{\hat{\Theta}}_t$ предыдущего примера в оптимизируемые коэффициенты. Тогда получим следующее уравнение класса допустимых фильтров:

$$\begin{aligned} \dot{\hat{\Theta}}_t &= \alpha_t \hat{\Theta}_t + \beta_t \dot{Y}_t + \gamma_t; \\ \beta_t &= \nu_2^{-1} (m_{110} - m_{101}); \\ \alpha_t &= \frac{-m_{111} + m_{110} m_{001} + \beta_t (m_{101} - m_{100} m_{001})}{m_{002}^2 - m_{001}^2}; \\ \gamma_t &= -\alpha_t m_{001} - \beta_t m_{100}. \end{aligned}$$

Литература

1. Синицын И. Н. Фильтры Калмана и Пугачева. — 2-е изд. — М.: Логос, 2007. 776 с.
2. Корепанов Э. Р. Стохастические информационные технологии на основе фильтров Пугачева // Информатика и её применения, 2011. Т. 5. Вып. 2. С. 36–57.
3. Синицын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
4. Синицын И. Н., Синицын В. И. Лекции по нормальной и эллипсоидальной аппроксимации в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 476 с.
5. Синицын И. Н., Корепанов Э. Р. Устойчивые линейные условно оптимальные фильтры и экстраполяторы для стохастических систем с мультипликативными шумами // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 70–75.
6. Синицын И. Н., Корепанов Э. Р. Синтез устойчивых линейных фильтров и экстраполяторов Пугачева для стохастических систем с мультипликативными широкополосными шумами // Системы и средства информатики, 2015. Вып. 25. № 1. С. 108–126.
7. Липцер Р. Ш., Шираев А. Н. Статистика случайных процессов. — М.: Наука, 1974. 476 с.
8. Ройтенберг Я. Н. Автоматическое управление. — 3-е изд., перераб. и доп. — М.: Наука, 1992. 576 с.

Поступила в редакцию 31.03.15

NORMAL PUGACHEV FILTERS 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 applied theory of analytical synthesis of normal conditionally optimal (Pugachev) filters (NPF) in state linear non-Gaussian stochastic systems (StS) is presented. Special attention is paid to NPF for differential StS satisfying Liptzer–Shirayev conditions based on the normal approximation of *a posteriori* density and quasi-linear NPF based on statistical linearization of nonlinear functions depending on observations. For StS of high dimension and real-time problems, NPF are more effective than the suboptimal filters. The NPF algorithms are the basis of the “StS-Filters” software tool. Test examples are given.

Keywords: Liptzer–Shirayev filter (LSF); Liptzer–Shirayev 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: 10.14357/19922264150204

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 15-07-02244).

References

1. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd ed. Moscow: Logos. 776 p.
2. 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.
3. 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.
4. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 488 p.
5. Sinitsyn, I. N., and E. R. Korepanov. 2015. Ustoychivye lineynye uslovno optimal'nye fil'try i ekstrapolyatory dlya stokhasticheskikh sistem s mul'tiplikativnymi shumami [Stable linear conditionally optimal filters and extrapolators for stochastic systems with multiplicative noises]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):70–75.
6. Sinitsyn, I. N., and E. R. Korepanov. 2015. Sintez ustoychivyykh fil'trov i ekstrapolyatorov Pugacheva dlya stokha-

- sticheskikh sistem s mul'tiplikativnymi shumami [Synthesis of stable conditionally optimal filters and extrapolators for stochastic systems with multiplicative wide band noises]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(1):108–126.
7. Liptser, R. Sh., and A. N. Shiryaev. 1974. *Statistika sluchaynykh protsessov* [Statistics of stochastic processes]. Moscow: Nauka, 476 p.
8. Roytenberg, Ya. N. 1992. *Avtomaticheskoe upravlenie* [Automatic control]. 3rd ed. Moscow: Nauka. 576 p.

Received March 31, 2015

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

ПРИМЕНЕНИЕ УРАВНЕНИЯ ПУГАЧЁВА–СВЕШНИКОВА К РЕШЕНИЮ ЗАДАЧИ БАКСТЕРА О ДЛИТЕЛЬНОСТИ ВЫБРОСОВ

С. В. Березин¹, О. И. Заяц²

Аннотация: Вычисляется закон распределения длительности выбросов случайного процесса за подвижную границу, которая перемещается равномерно с заданной скоростью. На выбросы исследуется процесс скошенного броуновского движения (СБД), являющийся одним из важнейших типовых процессов современной стохастической динамики. Этот процесс моделирует динамику броуновской частицы, которая встречает на своем пути полупрозрачный, частично отражающий экран. Задача решается методом уравнения Пугачёва–Свешникова, которое определенным образом модифицируется с учетом специфических свойств СБД. Указанное уравнение решается аналитически путем сведения к соответствующей краевой задаче Римана. Также описанный метод позволяет получить ряд дополнительных практически интересных характеристик СБД, таких как вероятность невыхода за экран, закон распределения времени первого достижения экрана, моменты времени пребывания за экраном и ряд других.

Ключевые слова: марковский процесс; уравнение Пугачёва; уравнение Пугачёва–Свешникова; краевая задача Римана; стохастическая механика; скошенное броуновское движение; время пребывания

DOI: 10.14357/19922264150205

1 Введение

В приложениях теории случайных процессов часто приходится иметь дело с различного рода функционалами от траекторий этих процессов. Важнейшей разновидностью таких функционалов является время пребывания процесса в заданной области. Оно определяется следующим образом. Пусть задан векторный случайный процесс $\mathbf{U}(\tau)$, принимающий значения в n -мерном евклидовом пространстве \mathbb{R}^n . Выберем область пребывания $\Omega \subset \mathbb{R}^n$ и зафиксируем интервал наблюдения $[0, t]$. Тогда временем пребывания процесса \mathbf{U} в Ω будет называться лебегова мера множества тех моментов наблюдения, для которых \mathbf{U} не покидал Ω :

$$T(t) = \text{mes} \{ \tau \in [0, t] \mid \mathbf{U}(\tau) \in \Omega \}. \quad (1)$$

Время пребывания, рассматриваемое как функция от t , представляет собой некоторый случайный процесс. Задача получения закона его распределения (или хотя бы моментов этого распределения) возникает во многих приложениях. Так, в теории надежности область Ω отвечает нормативному, штатному режиму работы системы, а выход из нее — переходу к аварийному, форс-мажорному режиму. При «жесткой» трактовке условий надежности первый же выброс за пределы Ω влечет за собой отказ системы, прекращение ее работы. Однако условия

надежности можно ставить и в более «мягкой» форме, когда отдельные выбросы из Ω допускаются, но их суммарная длительность должна быть ограничена.

Функционалы вида (1) играют важную роль и в стохастической финансовой математике. Платежные функции ряда нестандартных (экзотических) опционов выражаются через время пребывания курсовой стоимости выше или ниже определенного уровня [1]. Такие же функционалы возникают при решении задач теории массового обслуживания методом диффузионной аппроксимации [2]. В этом случае приходится вычислять время пребывания винеровского процесса на отрезках единичной длины, ограниченных натуральными числами.

Отметим, что в перечисленных, а также во многих других, подобных им приложениях обычно требуется найти закон распределения T аналитически, так как этот закон часто служит лишь промежуточным результатом для последующих вычислений. Для процессов \mathbf{U} общего вида решить такую задачу крайне затруднительно, и приходится ограничивать класс рассматриваемых процессов. В частности, для гауссовских процессов разработана асимптотическая теория, основанная на предположении, что выбросы за пределы Ω являются редкими событиями [3].

Между тем задачу можно, в принципе, решить точно, если считать \mathbf{U} непрерывным марковским

¹ Санкт-Петербургский политехнический университет Петра Великого, servberезin@yandex.ru

² Санкт-Петербургский политехнический университет Петра Великого, zay.oleg@gmail.com

процессом [4]. Время пребывания T тогда можно представить в виде решения дифференциального уравнения

$$\dot{T} = I_{\Omega}(\mathbf{U}), \quad T(0) = 0, \quad (2)$$

где

$$I_{\Omega}(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega; \\ 0, & \mathbf{x} \notin \Omega \end{cases}$$

обозначает индикаторную функцию множества Ω . Расширенный процесс (\mathbf{U}, T) также будет являться марковским, а распределение одной из его компонент T можно найти, решая соответствующее уравнение Колмогорова.

Последнее обычно оказывается достаточно сложным по своему виду, поэтому применяют и альтернативный подход, основанный на дискретизации времени, переходе к схеме случайных блужданий и предельном переходе, когда шаг блужданий стремится к нулю [5].

В настоящее время аналитическое выражение плотности распределения $f(y; t)$ времени пребывания T получено лишь для небольшого числа одномерных процессов. При переходе к векторным процессам приходится довольствоваться асимптотическими и приближенными формулами [6].

Наиболее детально изучена задача о времени пребывания одномерного марковского процесса U на положительной полуоси, когда (1) имеет вид:

$$T(t) = \text{mes} \{ \tau \in [0, t] \mid U(\tau) > 0 \}. \quad (3)$$

Впервые такую задачу для винеровского процесса решил в 1939 г. П. Леви [7], который открыл знаменитый закон арксинуса

$$f(y; t) = \frac{1}{\pi \sqrt{y(t-y)}} \quad (0 < y < t). \quad (4)$$

На этот результат самое пристальное внимание обратил В. Феллер. Он усовершенствовал вывод (4), рассмотрел ряд смежных задач, подробно проинтерпретировал этот закон, проверил его экспериментально и изложил весь круг связанных с ним вопросов в известной монографии [5]. С учетом всего сказанного задачу о времени пребывания (3) на положительной полуоси иногда называют задачей Феллера для процесса U .

Помимо случая стандартного винеровского процесса решение этой задачи в настоящее время получено также для винеровского процесса с постоянным сносом. На эту тему имеется целый цикл публикаций, включающий работы Дж. Акахори [8], А. Дассиоса [9], М. Йора [10], Л. Такача [11], А. Пехтля [12], ориентированные, в основном, на приложения к стохастической финансовой математике и опубликованные в конце 1990-х гг.

В настоящей статье будет рассмотрена иная, более сложная задача, касающаяся времени пребывания

$$T(t) = \text{mes} \{ \tau \in [0, t] \mid U(\tau) > b\tau \} \quad (b > 0)$$

заданного одномерного процесса U на полуоси $(b\tau, +\infty)$, граница которой движется равномерно с постоянной скоростью $b \geq 0$. Идея постановки подобной задачи принадлежит Г. Бакстеру [13], который впервые решил ее для винеровского процесса еще в 1956 г.

Отметим, что задача Феллера для винеровского процесса с постоянным сносом [8–12] фактически эквивалентна этой задаче Бакстера для стандартного винеровского процесса [13]. Винеровский процесс с постоянным сносом отличается от последнего как раз на линейную функцию времени. Пересечение винеровским процессом с таким сносом нулевого уровня означает выброс самого винеровского процесса за подвижный линейно растущий барьер. В обоих случаях речь идет об одних и тех же выбросах.

Как было отмечено ранее, обычно применяемые методы решения задач на время пребывания основаны на использовании либо уравнения Колмогорова, либо теории случайных блужданий. Насколько известно авторам, получить с их помощью решение задачи Феллера для каких-либо других процессов U , помимо разобранных в работах [7–13], до последнего времени не удавалось, между тем задачи эти интересны и значимы для многих важных приложений. Поэтому актуальной является разработка альтернативных методов решения.

2 Уравнение Пугачёва–Свешникова

Одним из авторов настоящей статьи ранее был предложен метод решения задачи Феллера, основанный на использовании уравнения Пугачёва–Свешникова [14]. Последнее представляет собой специальный частный случай хорошо известного уравнения Пугачёва для характеристической функции непрерывного марковского процесса [15]. При преобразовании уравнения Пугачёва к форме, предложенной Свешниковым, оно приобретает вид сингулярного интегродифференциального уравнения типа свертки. Вполне естественно, что такое специальное преобразование требует некоторого сужения класса рассматриваемых нелинейных систем.

Сам Свешников первоначально ограничился лишь системами, включающими нелинейности релейного типа [16, 17]. Свое уравнение Свешников

предложил решать приближенно в качестве альтернативы метода стохастической линеаризации. Впоследствии было установлено, что уравнение Пугачёва–Свешникова сохраняет силу не только в классе систем релейного типа, но также и для некоторых кусочно-линейных систем, в частности для систем, линейных в полупространствах [18] и четвертях пространства [19]. Кроме того, выяснилось, что это уравнение допускает не только приближенное, но в целом ряде случаев также и точное аналитическое решение путем сведения к решению соответствующей краевой задачи типа Римана.

Если исследуемый случайный процесс U является компонентой векторного марковского процесса, описываемого системой кусочно-линейных стохастических дифференциальных уравнений, линейной в полупространствах или четвертьпространствах, то после добавления к этой системе уравнения типа (2), задающего время пребывания процесса U на положительной полуоси, приходим к кусочно-линейной системе именно того вида, который был подробно разобран в [18, 19]. Это, в свою очередь, позволяет решать и соответствующую задачу Бакстера. Действительно, эта задача для процесса U сводится к решению задачи Феллера для вспомогательного процесса V , задаваемого условием $V(t) = U(t) - bt$. Если в уравнениях движения исходной задачи выразить U через V , то вновь полученная система по-прежнему будет принадлежать к классу кусочно-линейных систем вида [18, 19], а значит, допускает применение метода уравнения Пугачёва–Свешникова.

Указанным методом в работе [14] была решена задача Феллера для винеровского процесса, процесса Г. Уленбека – Л. Орнштейна, а также процесса Т. Кохи – Дж. Динза [20], играющего важную роль в стохастической механике. Впоследствии независимо от работ [8–12] методом [14] была решена также и рассмотренная в указанных работах задача Феллера для винеровского процесса с постоянным сносом [21]. В последнее время также получено решение задачи Феллера для некоторого обобщения процесса Кохи–Динза [19].

Далее вначале рассмотрим применение описанного метода к решению классической задачи Бакстера [13], касающейся самого винеровского процесса, а затем обратимся к более сложному примеру.

3 Решение классической задачи Бакстера

Пусть $W(t)$ обозначает стандартный винеровский процесс. Получим закон распределения сум-

марного времени, которое для $0 \leq \tau \leq t$ проведет процесс $U(\tau) = aW(\tau)$ выше возрастающего по линейному закону переменного уровня $b\tau$, где a и b – заданные положительные постоянные.

Положим

$$V(t) = \frac{\sqrt{2}}{a} (U(t) - bt), \quad \mu = \frac{b}{\sqrt{2}a}.$$

Тогда задача сведется к исследованию двумерного марковского процесса

$$\left. \begin{aligned} dV &= -2\mu dt + \sqrt{2} dW; \\ dT &= \frac{1}{2} (1 + \text{sign } V) dt \end{aligned} \right\} \quad (5)$$

при нулевых начальных условиях

$$V(0) = 0; \quad T(0) = 0.$$

Вторая компонента T процесса (5) дает длительность выбросов его первой компоненты V за нулевой уровень, т. е. совпадает с длительностью выбросов исходного процесса U за линейную границу $b\tau$.

Тем самым задача Бакстера для процесса U сведена к задаче Феллера для винеровского процесса V с постоянным сносом -2μ . Последняя задача подробно разобрана в [21], что позволяет получить следующее обобщение закона арксинуса.

Утверждение 1. *Плотность распределения времени пребывания $T(t)$ процесса $U(\tau)$ выше подвижной границы, растущей по линейному закону $b\tau$, дается формулой*

$$f(y; t) = \left(\frac{e^{-\mu^2 y}}{\sqrt{\pi y}} - \mu \text{Erfc}(\mu\sqrt{y}) \right) \times \left(\frac{e^{-\mu^2(t-y)}}{\sqrt{\pi(t-y)}} + \mu \text{Erfc}(-\mu\sqrt{t-y}) \right), \quad (6)$$

причем здесь

$$\text{Erfc } x = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-s^2} ds$$

обозначает дополнительную функцию ошибок.

Вполне естественно, что найденное распределение совпадает с решением Бакстера [13], а при $\mu = 0$ оно переходит в закон арксинуса (4).

Метод получения плотности распределения (6), основанный на решении уравнения Пугачёва–Свешникова, изложен в [21] и допускает обобщение на ряд процессов более общего вида. Один из примеров такого решения приводится в следующих разделах статьи и основывается на понятии СБД, которое по этой причине необходимо разобрать более подробно.

4 Скошенное броуновское движение

Скошенное броуновское движение впервые возникло в статье [22] как обобщение процесса броуновского движения с отражением. Позже было доказано, что СБД описывает движение броуновской частицы при наличии полупроницаемого частично отражающего экрана. Этот процесс не только интересен с математической точки зрения, но и весьма полезен для экономических, биологических, астрономических и разнообразных физических приложений [23].

Существует несколько подходов к определению СБД [23], из них для целей настоящей статьи наиболее удобен тот, который использует стохастическое дифференциальное уравнение, включающее локальное время. Рассмотрим одномерное прямолинейное движение частицы. В точке с координатой x расположим полупроницаемый экран. До момента достижения экрана движение частицы будем считать броуновским. Если, достигнув экрана, частица имела положительную скорость, то она свободно пересечет его с вероятностью β , а с вероятностью $1 - \beta$ упруго отразится от экрана. При противоположном направлении выхода частицы на экран вероятности пересечения и отражения соответственно равны $1 - \beta$ и β , т. е. являются дополнительными для соответствующих вероятностей в прямом направлении. В такой ситуации будем называть *скошенным броуновским движением* сильное решение стохастического дифференциального уравнения (СДУ) [23]

$$dU = (2\beta - 1)h^2 dL_U^x + h dW, \quad (7)$$

где W — стандартный винеровский процесс, а L_U^x — процесс симметричного локального времени [22] для U на уровне x , задаваемый формулой:

$$L_U^x(t) = \lim_{\varepsilon \rightarrow +0} \frac{1}{2\varepsilon} \text{mes} \{ \tau \in [0, t] \mid U(\tau) \in [x - \varepsilon, x + \varepsilon] \}.$$

Отметим, что данное выше определение СБД корректно в силу доказанной в [24] теоремы существования и единственности решения уравнения (7).

Уравнение (7) отличается от соответствующего уравнения для винеровского процесса лишь слагаемым, включающим локальное время. Добавляя подобные же слагаемые и в другие, более сложные уравнения, получаем возможность определить различные обобщения СБД. Одним из простейших таких обобщений является *СБД с постоянным сносом* [25], получаемое добавлением слагаемого с локальным временем в уравнение броуновского движения с постоянным сносом:

$$dU = c dt + (2\beta - 1)h^2 dL_U^x + h dW. \quad (8)$$

Это слагаемое, как и ранее, моделирует полупроницаемый частично отражающий экран, помещенный в точку x .

При формулировке задачи Бакстера применительно к процессу (8) будем теперь считать, что положение частично проницаемого экрана меняется во времени t по линейному закону $x(t) = a + bt$, где $a > 0$ и $b \geq 0$. Тогда сама задача Бакстера (задача определения длительности выбросов процесса U за такой подвижный экран) сводится к анализу процесса

$$\left. \begin{aligned} dU &= c dt + (2\beta - 1)h^2 dL_U^{a+bt} + h dW; \\ dT &= \frac{1}{2}(1 + \text{sign}(U - a - bt)) dt \end{aligned} \right\} \quad (9)$$

при начальных условиях $U(0) = 0$ и $T(0) = 0$. Необходимо изучить характеристики компоненты T расширенного процесса (U, T) .

Интересными частными случаями являются $\beta = 0$ и 1 , которые соответствуют упруго отражающему экрану и экрану-ловушке, пропускающему только частицы, движущиеся с положительной скоростью. В случае $\beta = 1/2$ экран вообще не влияет на частицу и U является винеровским процессом (броуновским движением) с постоянным сносом [21].

Вводя обозначения

$$\begin{aligned} V &= \frac{\sqrt{2}}{h} (U - a - bt); & \mu &= \frac{b - c}{\sqrt{2}h}; \\ \eta &= 2\beta - 1; & v_0 &= -\frac{a\sqrt{2}}{h} < 0, \end{aligned}$$

получим

$$\left. \begin{aligned} dV &= -2\mu dt + 2\eta dL_V^0 + \sqrt{2} dW; \\ dT &= \frac{1}{2}(1 + \text{sign} V) dt \end{aligned} \right\} \quad (10)$$

при начальных условиях

$$V(0) = v_0; \quad T(0) = 0.$$

Система уравнений (10) относится к классу кусочно-линейных стохастических систем, линейных в полупространствах, но отличается от стандартных уравнений этого класса [14] наличием локального времени пребывания на границе областей линейности. Несмотря на указанное усложнение постановки задачи, ее решение по-прежнему может быть получено методом, описанным в [14].

5 Решение задачи в изображениях

Повторяя вывод уравнения Пугачёва–Свешникова [18] с учетом дополнительного слагаемого, со-

держашего локальное время, получим следующую модификацию указанного уравнения.

Утверждение 2. *Характеристическая функция $E(z_1, z_2; t)$ системы ординат $(V(t), T(t))$ процесса (10) подчиняется уравнению*

$$\frac{\partial E(z_1, z_2; t)}{\partial t} = - \left(z_1^2 + 2i\mu z_1 - \frac{iz_2}{2} \right) E(z_1, z_2; t) + \frac{z_2}{2\pi} \text{v.p.} \int_{-\infty}^{+\infty} \frac{E(s, z_2; t)}{s - z_1} ds + 2i\eta z_1 \Phi(z_2; t) \quad (11)$$

при начальном условии $E(z_1, z_2; 0) = e^{iv_0 z_1}$, где функция $\Phi(z_2; t)$ дается интегралом

$$\Phi(z_2; t) = \frac{1}{2\pi} \text{v.p.} \int_{-\infty}^{+\infty} E(s, z_2; t) ds. \quad (12)$$

Последнее слагаемое в правой части (11), отсутствовавшее в статьях [14, 18, 19, 21], появилось из-за наличия в уравнениях движения (10) локального времени. Отметим, что интерпретация интеграла (12) в смысле главного значения по Коши принципиально важна. Как будет показано позже, в классическом смысле этот интеграл расходится, так как имеет неинтегрируемую особенность на бесконечности.

Метод решения уравнения (11) аналогичен [14] и основывается на переходе с помощью формул Ю. В. Сохоцкого к краевой задаче Римана теории функций комплексного переменного. Формулы Сохоцкого имеют вид:

$$\left. \begin{aligned} F^+(z_1, z_2; t) - F^-(z_1, z_2; t) &= E(z_1, z_2; t); \\ F^+(z_1, z_2; t) + F^-(z_1, z_2; t) &= \\ &= \frac{1}{\pi i} \text{v.p.} \int_{-\infty}^{\infty} \frac{E(s, z_2; t)}{s - z_1} ds. \end{aligned} \right\} \quad (13)$$

Здесь $(z_1, z_2) \in \mathbb{R}^2$, а функции F^+ и F^- являются аналитическими по аргументу z_1 соответственно в верхней и нижней полуплоскостях расширенной комплексной плоскости.

Применяя преобразование Лапласа по t , обозначая аргумент этого преобразования через p , а изображения — той же буквой, что и оригиналы, но с волной сверху, с помощью (13) перейдем от уравнения (11) относительно E к краевой задаче Римана относительно изображений \tilde{F}^+ и \tilde{F}^- по Лапласу.

Утверждение 3. *Задача Коши для уравнения (11) эквивалентна краевой задаче Римана, заключающейся в нахождении пары функций \tilde{F}^\pm , аналитических*

по z_1 соответственно в верхней и нижней полуплоскостях и удовлетворяющих при $\text{Im } z_1 = 0$ краевому условию

$$\begin{aligned} (z_1^2 + 2i\mu z_1 + p - iz_2) \tilde{F}^+ &= \\ = (z_1^2 + 2i\mu z_1 + p) \tilde{F}^- + 2i\eta z_1 \tilde{\Phi}(z_2; p) + e^{iv_0 z_1}, \end{aligned} \quad (14)$$

где $\tilde{F}^\pm(z_1, z_2; p)$ и $\tilde{\Phi}(z_2; p)$ обозначают изображения функций $F^\pm(z_1, z_2; t)$ и $\Phi(z_2; t)$ соответственно, причем все изображения существуют при $\text{Re } p > 0$.

Обратим внимание, что задача (14) отличается от аналогичных краевых задач [14, 18, 19, 21] появлением в краевом условии слагаемого, содержащего $\tilde{\Phi}$, связь которого с искомыми краевыми значениями \tilde{F}^\pm задается формулами (12) и (13).

Повторение рассуждений статьи [14] позволяет фактически решить задачу (14).

Утверждение 4. *Краевая задача Римана (14) эквивалентна обратной задаче подбора функций \tilde{G}_0 и \tilde{G}_1 , исходя из условий аналитичности по аргументу z_1 правых частей каждого из выражений*

$$\left. \begin{aligned} \tilde{F}^+(z_1, z_2; p) &= \\ = \frac{\tilde{G}_0(z_2; p) + (\tilde{G}_1(z_2; p) + i\eta \tilde{\Phi}(z_2; p)) z_1}{z_1^2 + 2i\mu z_1 + p - iz_2}; \\ \tilde{F}^-(z_1, z_2; p) &= \\ = \frac{\tilde{G}_0(z_2; p) + (\tilde{G}_1(z_2; p) - i\eta \tilde{\Phi}(z_2; p)) z_1 - e^{iv_0 z_1}}{z_1^2 + 2i\mu z_1 + p} \end{aligned} \right\} \quad (15)$$

соответственно в верхней и нижней полуплоскостях.

Обозначая корни знаменателей (15) через $\kappa^\pm = i(-\mu \pm \sqrt{\mu^2 + p - iz_2})$ и $\nu^\pm = i(-\mu \pm \sqrt{\mu^2 + p})$, находим условия аналитичности \tilde{F}^\pm в виде:

$$\left. \begin{aligned} \tilde{G}_0(z_2; p) + (\tilde{G}_1(z_2; p) + i\eta \tilde{\Phi}(z_2; p)) \kappa^+ &= 0; \\ \tilde{G}_0(z_2; p) + (\tilde{G}_1(z_2; p) - i\eta \tilde{\Phi}(z_2; p)) \nu^- &= \\ &= e^{iv_0 \nu^-}. \end{aligned} \right\} \quad (16)$$

Исключение \tilde{G}_0 из (15) и (16) дает

$$\left. \begin{aligned} \tilde{F}^+(z_1, z_2; p) &= \frac{\tilde{G}_1(z_2; p) + i\eta \tilde{\Phi}(z_2; p)}{z_1 - \kappa^-}; \\ \tilde{F}^-(z_1, z_2; p) &= \frac{\tilde{G}_1(z_2; p) - i\eta \tilde{\Phi}(z_2; p)}{z_1 - \nu^+} + \\ &+ \frac{e^{iv_0 \nu^-} - e^{iv_0 z_1}}{z_1^2 + 2i\mu z_1 + p}; \\ (\nu^- - \kappa^+) \tilde{G}_1(z_2; p) - \\ - i\eta (\nu^- + \kappa^+) \tilde{\Phi}(z_2; p) &= e^{iv_0 \nu^-}. \end{aligned} \right\} \quad (17)$$

Далее с помощью формул (13) и (17) находим выражение для \tilde{E} и подставляем его в (12). Разрешая полученные уравнения относительно $\tilde{\Phi}$, получаем

$$\tilde{\Phi}(z_2; p) = -i\tilde{G}_1(z_2; p).$$

Отметим, что интегралы от \tilde{F}^\pm по z_1 , вообще говоря, расходятся в обычном смысле, но тем не менее существуют в смысле главного значения, а именно это и предполагалось в формуле (12). В итоге приходим к окончательным выражениям для введенных изображений:

$$\left. \begin{aligned} \tilde{G}_1(z_2; p) &= \frac{e^{iv_0\nu^-}}{(1-\eta)\nu^- - (1+\eta)\varkappa^+}; \\ \tilde{F}^+(z_1, z_2; p) &= \frac{(1+\eta)\tilde{G}_1(z_2; p)}{z_1 - \varkappa^-}; \\ \tilde{F}^-(z_1, z_2; p) &= \frac{(1-\eta)\tilde{G}_1(z_2; p)}{z_1 - \nu^+} + \frac{e^{iv_0\nu^-} - e^{iv_0z_1}}{z_1^2 + 2i\mu z_1 + p}. \end{aligned} \right\} \quad (18)$$

Далее изображение искомой характеристической функции очевидным образом находится по первой формуле Сохоцкого (13).

Утверждение 5. Изображение \tilde{E} характеристической функции системы компонент $(V(t), T(t))$ процесса (10) выражается в виде:

$$\tilde{E}(z_1, z_2; p) = \left[\frac{1+\eta}{z_1 - \varkappa^-} - \frac{1-\eta}{z_1 - \nu^+} \right] \tilde{G}_1(z_2; p) - \frac{e^{iv_0\nu^-} - e^{iv_0z_1}}{z_1^2 + 2i\mu z_1 + p}, \quad (19)$$

где $\tilde{G}_1(z_2; p)$ определяется согласно (18).

Полагая здесь $z_1 = 0$, получаем изображение маргинальной характеристической функции T .

Следствие 1. Преобразование Лапласа для характеристической функции искомого времени пребывания T имеет вид:

$$\begin{aligned} \tilde{E}_T(z_2; p) &= \\ &= \frac{(1-\eta)\varkappa^- - (1+\eta)\nu^+}{(1-\eta)\nu^- - (1+\eta)\varkappa^+} \frac{e^{iv_0\nu^-}}{\varkappa^- \nu^+} - \frac{e^{iv_0\nu^-} - 1}{p}. \end{aligned} \quad (20)$$

Функция (20) представляет собой двукратное преобразование плотности вероятности $f(y; t)$: по Фурье (по аргументу y) и по Лапласу (по аргументу t). Получению соответствующих оригиналов посвящен следующий раздел.

6 Решение задачи в оригиналах

Полученное в предыдущем разделе изображение (20) характеристической функции \tilde{E}_T за исключением табличного последнего слагаемого, пропорционального $1/p$, представляет собой произведение экспоненты от аргумента $\sqrt{\mu^2 + p}$ и рациональной функции от аргументов $\sqrt{\mu^2 + p}$ и $\sqrt{\mu^2 + p - iz_2}$. Такая структура двукратного преобразования Фурье–Лапласа гарантирует получение явного выражения оригинала $f(y; t)$ в квад-

ратурах. Действительно, пользуясь методом двукратного преобразования Лапласа, подробно описанным в [21], общими свойствами преобразования Лапласа и обобщенной теоремой умножения Эфроса [26], можно свести исходную задачу обращения к задаче обращения двукратного преобразования Лапласа, представляющего собой рациональную дробь от аргументов $q = -iz_2$ и p , которое уже является табличным. Опуская промежуточные выкладки, приходим к следующему конечному результату.

Утверждение 6. Плотность вероятности времени пребывания $f(y; t)$ второй компоненты процесса (9) при $0 < y < t$ дается выражением:

$$\begin{aligned} f(y; t) &= \left[1 - \frac{1}{2} \left(e^{-2v_0\mu^-} \operatorname{Erfc} \left(-\frac{v_0 - 2|\mu|t}{2\sqrt{t}} \right) + \right. \right. \\ &\quad \left. \left. + e^{2v_0\mu^+} \operatorname{Erfc} \left(-\frac{v_0 + 2|\mu|t}{2\sqrt{t}} \right) \right) \right] \delta(y) + \\ &\quad + \frac{e^{-\mu^2 t + v_0\mu}}{4\pi(y(t-y))^{3/2}} \left[\frac{1-\eta}{1+\eta} \int_0^{+\infty} \int_{-v_0}^{+\infty} \chi^+(s_1, s_2) \times \right. \\ &\quad \times e^{\mu(s_1+v_0) + ((\mu(3\eta-1))/(1+\eta))s_2} ds_1 ds_2 + \\ &\quad \left. + \frac{1+\eta}{1-\eta} \int_0^{+\infty} \int_{-v_0}^{+\infty} \chi^-(s_1, s_2) \times \right. \\ &\quad \left. \times e^{((\mu(3\eta+1))/(1-\eta))(s_1+v_0) - \mu s_2} ds_1 ds_2 \right]; \end{aligned}$$

$$\begin{aligned} \chi^\pm(s_1, s_2) &= \\ &= \frac{1 \pm \operatorname{sign} [(1+\eta)(s_1+v_0) - (1-\eta)s_2]}{2} \times \\ &\quad \times s_1 s_2 e^{-s_1^2/(4(t-y)) - s_2^2/(4y)}, \end{aligned} \quad (21)$$

где $\delta(y)$ обозначает дельта-функцию Дирака, а $\mu^\pm = (|\mu| \pm \mu)/2$ — положительную и отрицательную части числа μ .

В частном случае, когда $\eta = 0$ и $v_0 = 0$, дельтообразное слагаемое в (21) исчезает, а двукратный интеграл преобразуется к виду обычного закона арксинуса (6).

Другой важный частный случай, отвечающий $\mu = 0$ и нулевому начальному условию $v_0 = 0$, описывает ситуацию, при которой постоянный снос совпадает со скоростью движения экрана. При этом интеграл (21) выражается в конечном виде через элементарные функции

$$f(y; t) = \frac{(1-\eta^2)t}{\pi\sqrt{y(t-y)}((1+\eta)^2 t - 4\eta y)} \quad (0 < y < t).$$

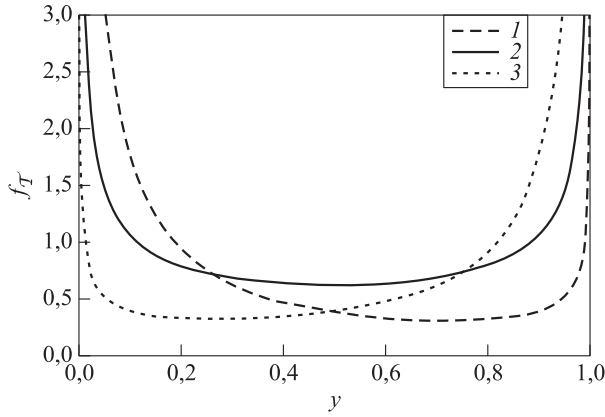


Рис. 1 График $f_T(y)$ при различных значениях η ($\mu = 0$): 1 – $\eta = -0,5$; 2 – 0; 3 – $\eta = 0,5$

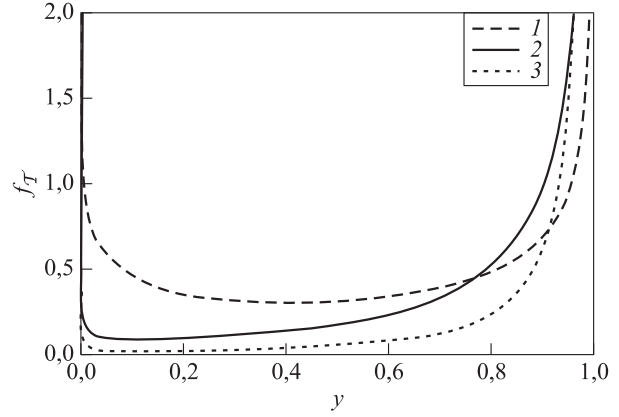


Рис. 2 График $f_T(y; t)$ при различных значениях η ($\mu = -1, t = 1$): 1 – $\eta = -0,5$; 2 – 0; 3 – $\eta = 0,5$

Последняя формула, разумеется, при $\eta = 0$ переходит в известный закон арксинуса (4).

Перейдем к безразмерному времени $T = T/t$, выраженному в долях интервала наблюдения t , его плотность вероятности $f_T(y) = tf(ty; t)$ не зависит от t :

$$f_T(y) = \frac{1 - \eta^2}{\pi \sqrt{y(1-y)}((1+\eta)^2 - 4\eta y)} \quad (0 < y < 1). \quad (22)$$

График f_T представлен на рис. 1.

Нетрудно убедиться, что доля среднего время, проведенного частицей за движущимся полупрозрачным частично отражающим экраном, дается равенством

$$M[T] = \frac{\eta + 1}{2} = \beta.$$

Этот факт в некотором смысле отражает эргодическую природу рассматриваемого процесса, коль скоро параметр β был определен ранее как вероятность проникновения частицы через экран.

Рассмотрев более общий случай $v_0 = 0, \mu \neq 0$. Получим:

$$f(y; t) = \frac{e^{-\mu^2 t}}{4\pi(y(t-y))^{3/2}} \left[\frac{1-\eta}{1+\eta} \int_0^{+\infty} \int_0^{+\infty} \chi^+(s_1, s_2) \times \right. \\ \times e^{\mu s_1 + ((\mu(3\eta-1))/(1+\eta))s_2} ds_1 ds_2 + \\ \left. + \frac{1+\eta}{1-\eta} \int_0^{+\infty} \int_0^{+\infty} \chi^-(s_1, s_2) \times \right. \\ \left. \times e^{((\mu(3\eta+1))/(1-\eta))s_1 - \mu s_2} ds_1 ds_2 \right].$$

Соответствующий график для плотности вероятности безразмерного времени $f_T(y; t) = tf(ty; t)$ представлен на рис. 2.

Обратим внимание, что в отличие от выражения (22), плотность $f_T(y; t)$ теперь уже зависит от t , что очень хорошо иллюстрируется формулой (24), которая будет получена в следующем разделе настоящей статьи.

7 Дальнейшие результаты

В этом разделе будет приведен ряд полезных и качественно интересных результатов, которые удастся получить без трудоемкого обращения приведенных выше изображений Лапласа.

Рассмотрим независимый от процесса (10) случайный экспоненциальный момент времени τ_λ , где $\lambda > 0$. Нетрудно показать (см., например, [27]), что характеристическая функция $E_*(z_1, z_2; \lambda)$ случайного вектора $(V(\tau_\lambda), T(\tau_\lambda))$ легко выражается через $\tilde{E}(z_1, z_2; p)$, а именно:

$$E_*(z_1, z_2; \lambda) = \lambda \tilde{E}(z_1, z_2; \lambda). \quad (23)$$

Указанное взаимно однозначное соответствие позволяет, по существу, не делать различия между изображением по Лапласу $\tilde{E}(z_1, z_2; p)$ характеристической функции системы величин $(V(t), T(t))$ и характеристической функцией $E_*(z_1, z_2; \lambda)$, вычисленной для величин $(V(\tau_\lambda), T(\tau_\lambda))$.

Полагая $z_2 = 0$ в (19), находим:

$$E_{*V}(z_1; \lambda) = E_*(z_1, 0; \lambda) = \\ = \frac{\lambda}{z_1^2 + 2i\mu z_1 + \lambda} \left(\frac{\eta z_1 e^{v_0(\mu + \sqrt{\mu^2 + \lambda})}}{i(\eta\mu - \sqrt{\mu^2 + \lambda})} + e^{iv_0 z_1} \right).$$

Если теперь положить в (19) $z_1 = z_2 = 0$, то, воспользовавшись интерпретацией \tilde{F}^\pm как односторонних преобразований Фурье [18], приходим к паре формул

$$P\{V(\tau_\lambda) \geq 0\} = \pm \lambda \tilde{F}^\pm(0, 0; \lambda),$$

что позволяет сформулировать следующее

Утверждение 7. Вероятность $P\{U(\tau_\lambda) > a + b\tau_\lambda\} = P\{V(\tau_\lambda) > 0\}$ того, что ордината $U(\tau_\lambda)$ окажется выше растущей по линейному закону границы, дается формулой:

$$P\{U(\tau_\lambda) > a + b\tau_\lambda\} = \frac{(1 + \eta)\lambda e^{v_0(\mu + \sqrt{\mu^2 + \lambda})}}{2(\sqrt{\mu^2 + \lambda} + \mu)(\sqrt{\mu^2 + \lambda} - \eta\mu)}.$$

Отсюда, устремляя параметр λ к нулю, легко находим асимптотику вероятности $P\{U(t) > a + bt\}$ при $t \rightarrow \infty$. Полученный результат полностью согласуется со здравым смыслом. Действительно, при $\eta = -1$ имеем отражающий экран, и поэтому $\lim_{t \rightarrow \infty} P\{U(t) > a + bt\} = 0$, т. е. частица останется ниже экрана при любом μ . Пусть теперь $\eta \neq -1$, тогда будем иметь:

$$\lim_{t \rightarrow \infty} P\{U(t) > a + bt\} = \begin{cases} \beta, & \mu = 0; \\ 1, & \mu < 0; \\ 0, & \mu > 0, \end{cases} \quad (24)$$

что также физически вполне объяснимо: при $\mu < 0$ частица рано или поздно обгонит экран, при $\mu > 0$ частица никогда его не догонит, а при $\mu = 0$ частица окажется выше экрана с вероятностью β (именно так и было введено β выше). Отметим, что последний факт может быть использован при решении задачи идентификации параметра β по опытным данным.

Формула (20) позволяет также найти важный в практическом отношении закон распределения времени первого достижения экрана θ . Действительно, пусть $\eta \neq -1$. Тогда, анализируя выражение (20) для \tilde{E}_T , легко заметить, что единственным слагаемым, не исчезающим при $z_2 \rightarrow \infty$, является последнее, которое вообще не зависит от z_2 . Этому слагаемому отвечает дельтаобразная плотность, сосредоточенная в нуле, которая, в свою очередь, дает вероятность $P\{T(\tau_\lambda) = 0\}$. Отсюда получаем:

$$P\left\{\sup_{\tau \in [0, \tau_\lambda]} (U(\tau) - a - b\tau) < 0\right\} = P\{\theta > \tau_\lambda\} = P\{T(\tau_\lambda) = 0\} = 1 - e^{v_0(\mu + \sqrt{\mu^2 + \lambda})}, \quad (25)$$

и это выражение, естественно, совпадает со случаем броуновского движения с постоянным сносом [26]. Из физических соображений легко понять, что до момента первого достижения процессом СБД со сносом экрана этот процесс ведет себя в точности,

как аналогичный процесс без экрана, а значит, время достижения границы вообще не должно зависеть от η . Поэтому выражение (25) сохраняет силу также и при $\eta = -1$. Эти рассуждения приводят к важному выводу.

Утверждение 8. Вероятность $P\left\{\sup_{\tau \in [0, \tau_\lambda]} (U(\tau) - a - b\tau) < 0\right\}$ невыхода процесса $U(t)$ в течение времени τ_λ за растущую по линейному закону границу дается формулой (25).

Учитывая связь (23) между изображением характеристической функции \bar{E} и характеристической функцией E_* , после обратного преобразования Лапласа получаем аналогичное утверждение для произвольного фиксированного момента времени t .

Утверждение 9. Вероятность $P\left\{\sup_{\tau \in [0, t]} (U(\tau) - a + b\tau) < 0\right\}$ невыхода процесса $U(t)$ в течение времени t за растущую по линейному закону границу дается формулой:

$$P\left\{\sup_{\tau \in [0, t]} (U(\tau) - a - b\tau) < 0\right\} = P\{\theta > t\} = 1 - \frac{1}{2} \left(e^{-2v_0\mu^-} \operatorname{Erfc}\left(-\frac{v_0 - 2|\mu|t}{2\sqrt{t}}\right) + e^{2v_0\mu^+} \operatorname{Erfc}\left(-\frac{v_0 + 2|\mu|t}{2\sqrt{t}}\right) \right).$$

Полученное выражение, как легко заметить, фактически уже содержалось в представлении (21).

Найдем также выражение для математического ожидания времени $T(\tau_\lambda)$. Дифференцируя $E_*(z_2; \lambda)$ по z_2 и полагая $z_2 = 0$, имеем:

$$\begin{aligned} \tilde{m}(\lambda) &= M[T(\tau_\lambda)] = \\ &= \frac{(1 + \eta)e^{v_0(\mu + \sqrt{\mu^2 + \lambda})}}{2(\sqrt{\mu^2 + \lambda} + \mu)(\sqrt{\mu^2 + \lambda} - \eta\mu)} = \\ &= \frac{P\{U(\tau_\lambda) > a + b\tau_\lambda\}}{\lambda}. \end{aligned}$$

Отсюда с учетом $M[\tau_\lambda] = 1/\lambda$ сразу же получаем равенство

$$\frac{M[T(\tau_\lambda)]}{M[\tau_\lambda]} = P\{U(\tau_\lambda) > a + b\tau_\lambda\}, \quad (26)$$

которое можно интерпретировать как некоторую обобщенную эргодичность.

Как легко проверить, равенство (26) можно обобщить на случай векторного процесса $U(t)$ и области Ω произвольного вида следующим образом.

Утверждение 10. Для произвольного непрерывного процесса $U(t)$ и произвольной области Ω справедливо равенство:

$$\frac{M[\text{mes}\{\tau \in [0, \tau_\lambda] \mid U(\tau) \in \Omega\}]}{M[\tau_\lambda]} = P\{U(\tau_\lambda) \in \Omega\}.$$

Следует отметить, что все полученные в данном разделе характеристики, относящиеся к показательному моменту времени τ_λ , в принципе, могут быть найдены также и для произвольного фиксированного момента времени t . Соответствующее преобразование Лапласа удастся обратить методом, описанным при выводе представления (21). Оставляя в стороне более детальный математический и физический анализ построенного выше решения, отметим, что изложенный метод позволяет решить и целый ряд других, более сложных, интересных как с прикладной, так и с вероятностной точки зрения задач.

8 Заключение

Метод уравнения Пугачёва–Свешникова позволяет успешно решать целый ряд задач, касающихся времени пребывания одномерного процесса на полуоси. Для реализации метода процесс, исследуемый на выбросы, должен описываться системой стохастических дифференциальных уравнений, которая линейна либо во всем пространстве, либо в полупространствах. Допускается также линейность и в четвертьпространствах, но такая задача более сложна математически. В предыдущих работах авторов полуось, представляющая область пребывания процесса, имела фиксированную постоянную границу. В настоящей статье эта граница равномерно движется с заданной скоростью. Кроме того, в уравнения движения добавлены дополнительные члены, соответствующие локальному времени процесса.

В статье детально изучено распределение длительности выбросов типового процесса СБД с постоянным сносом за равномерно движущуюся границу. Найденное распределение представляет собой некоторое обобщение классического закона арксинуса. Попутно получен ряд дополнительных характеристик движения: вероятность невыхода процесса за подвижную границу, распределение времени ее первого достижения и т. п.

Построенное аналитическое решение представляет самостоятельный интерес как эталонное точное аналитическое решение типовой задачи статистической динамики. Кроме того, оно может быть использовано для оценки погрешности и тес-

тирования существующих приближенных методов решения уравнения Пугачёва [15].

Литература

1. Люу Ю. Д. Методы и алгоритмы финансовой математики / Пер. с англ. — М.: Бином, 2007. 752 с. (Lyu Y. D. Financial engineering and computation. — 1st ed. — Cambridge: Cambridge University Press, 2001. 627 p.)
2. Cohen J. W., Hooghiemstra G. Brownian excursion, the $M/M/1$ queue and their occupation times // Math. Oper. Res., 1981. Vol. 6. No. 4. P. 608–629.
3. Berman S. M. Sojourn and extremes of stochastic processes. — Belmont: CRC Press, 1992. 320 p.
4. Бородин А. Н. Случайные процессы. — СПб.: Лань, 2013. 640 с.
5. Феллер В. Введение в теорию вероятностей и ее приложения / Пер. с англ. — М.: Мир, 1967. Т. 2. 765 с. (Feller W. An introduction to probability theory and its applications. — New York, NY, USA: Wiley, 1966. Vol. II. 626 p.)
6. Korpas A. K. Occupation times of continuous Markov processes. Ph.D. Thesis. — Bowling Green: Bowling Green State University, 2006. 92 p.
7. Lévy P. Sur une probléme de Marcinkiewicz // Comptes rendus Academie sciences Paris, 1939. T. 208. P. 319–321, errata p. 776.
8. Akahori J. Some formulae for a new type of path-dependent option // Ann. Appl. Probab., 1995. Vol. 5. No. 2. P. 383–388.
9. Dassios A. The distribution of the quantile of a Brownian motion with drift and the pricing of related path-dependent options // Ann. Appl. Probab., 1995. Vol. 5. No. 2. P. 389–398.
10. Yor M. The distribution of Brownian quantiles // J. Appl. Probab., 1995. Vol. 32. P. 405–416.
11. Takács L. On a generalization of the arc-sine law // Ann. Appl. Probab., 1996. Vol. 6. No. 3. P. 1035–1040.
12. Pechtl A. Distribution of occupation times of Brownian motion with drift // J. Appl. Math. Decision Sci., 1999. Vol. 3. P. 41–62.
13. Baxter G. Wiener process distributions of the arcsine law type // Proc. Am. Math. Soc., 1956. Vol. 7. P. 738–741.
14. Заяц О. И. Об аналитическом решении задачи Феллера о длительности выбросов // Труды СПбГТУ. Прикладная математика, 1996. № 461. С. 92–100.
15. Пугачёв В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1985 (1-е изд.), 1990 (2-е изд.).
16. Свешников А. А. Применение теории непрерывных марковских процессов к решению нелинейных задач прикладной гироскопии // Тр. V Междунар. конф. по нелинейным колебаниям. — Киев: ИМ АН УССР, 1970. Т. 3. С. 659–665.

17. Свешников А. А., Ривкин С. С. Вероятностные методы в прикладной теории гироскопов. — М.: Наука, 1974. 536 с.
18. Заяц О. И. Применение уравнения Пугачёва—Свешникова к исследованию кусочно-линейных стохастических систем, линейных в полупространствах // Научно-технические ведомости СПбГПУ. Физико-математические науки, 2013. № 4-1. С. 128–142.
19. Заяц О. И., Бerezин С. В. Применение уравнения Пугачёва—Свешникова к исследованию кусочно-линейных стохастических систем, линейных в четвертях пространства // Научно-технические СПбГПУ. Информатика. Телекоммуникации. Управление, 2013. № 6. С. 87–101.
20. Caughey T. K., Dienes J. K. Analysis of non-linear first order system with a white noise input // J. Appl. Phys., 1961. Vol. 32. No. 11. P. 2476–2479.
21. Заяц О. И. Решение задачи Феллера для винеровского процесса с постоянным сносом // Труды СПбГТУ. Прикладная математика, 1999. № 477. С. 67–72.
22. Ito K., McKean H. P. Brownian motion on a half-line // Illinois J. Math., 1963. Vol. 7. P. 181–231.
23. Lejay A. On the constructions of the skew Brownian motion // Probab. Survveys, 2006. Vol. 3. P. 413–466.
24. Le Gall J.-F. One-dimensional stochastic differential equations involving the local times of the unknown process // Stochastic analysis and application / Eds. A. Truman, D.W. Williams. — Lecture notes in mathematics ser. — Berlin–Heidelberg: Springer, 1984. Vol. 1095. P. 51–82.
25. Appuhamillage T., Bokil V., Thomann E., Waymire E., Wood B. Occupation and local times for skew Brownian motion with applications to dispersion across an interface // Ann. Appl. Probab., 2011. Vol. 21. No. 1. P. 183–214.
26. Лаврентьев М. А., Шабат Б. В. Методы теории функции комплексного переменного. — М.: Наука, 1965. 716 с.
27. Бородин А. Н., Салминен П. Справочник по броуновскому движению: факты и формулы / Пер. с англ. — СПб.: Лань, 2000. 640 с. (Borodin A. N., Salminen P. Handbook of Brownian motion. Facts and formulae. Probability and its applications. Basel: Birkhäuser, 1996. 462 p.)

Поступила в редакцию 02.02.15

APPLICATION OF THE PUGACHEV–SVESHNIKOV EQUATION TO THE BAXTER OCCUPATION TIME PROBLEM

S. V. Berezin and O. I. Zayats

Institute of Applied Mathematics and Mechanics, Peter the Great St. Petersburg State Polytechnic University, 29 Politekhnicheskaya Str., St. Petersburg 195251, Russian Federation

Abstract: The Baxter problem, that is, an occupation (sojourn) time above a moving barrier, for a skew Brownian motion is considered. The latter is known as a model of a semipermeable barrier which permits either movement through it or reflection to the opposite direction with a specified probability. The Pugachev–Sveshnikov equation for a continuous Markov process is used to obtain an analytic solution of the problem. The generic method to solve the Pugachev–Sveshnikov equation for occupation-time type problems involves its reduction to a certain Riemann boundary value problem. This problem is solved, and the occupation time probability density function is derived. Along the way, some additional characteristics of the skew Brownian motion are obtained such as the first passage time, nonexceedance probability, occupation time moments, and some other characteristics.

Keywords: Markov process; Pugachev equation; Pugachev–Sveshnikov equation; Riemann boundary value problem; stochastic mechanics; skew Brownian motion; occupation time; sojourn time

DOI: 10.14357/19922264150205

References

1. Lyuu, Y. D. 2001. *Financial engineering and computation*. 1st ed. Cambridge: Cambridge University Press. 627 p.
2. Cohen, J. W., and G. Hooghiemstra. 1981. Brownian excursion, the $M/M/1$ queue and their occupation times. *Math. Oper. Res.* 6(4):608–629.
3. Berman, S. M. 1992. *Sojourn and extremes of stochastic processes*. Belmont: CRC Press. 320 p.
4. Borodin, A. N. 2013. *Sluchaynye protsessy* [Stochastic processes]. St. Petersburg: Lan'. 640 p.
5. Feller, W. 1966. *An introduction to probability theory and its applications*. New York, NY: Wiley. Vol. II. 626 p.
6. Korpas, A. K. 2006. Occupation times of continuous Markov processes. Ph.D. Thesis. Bowling Green: Bowling Green State University. 92 p.
7. Lévy, P. 1939. Sur une problème de Marcinkiewicz. *Comptes rendus Academie sciences Paris* 208:319–321, errata p. 776.

8. Akahori, J. 1995. Some formulae for a new type of path-dependent option. *Ann. Appl. Probab.* 5(2):383–388.
9. Dassios, A. 1995. The distribution of the quantile of a Brownian motion with drift and the pricing of related path-dependent options. *Ann. Appl. Probab.* 5(2):389–398.
10. Yor, M. 1995. The distribution of Brownian quantiles. *J. Appl. Probab.* 32:405–416.
11. Takács, L. 1996. On a generalization of the arc-sine law. *Ann. Appl. Probab.* 6(3):1035–1040.
12. Pechtl, A. 1999. Distribution of occupation times of Brownian motion with drift. *J. Appl. Math. Decision Sci.* 3:41–62.
13. Baxter, G. 1956. Wiener process distributions of the arcsine law type. *Proc. Am. Math. Soc.* 7:738–741.
14. Zayats, O.I. 1996. Ob analiticheskom reshenii zadachi Feller'a o dilitel'nosti vybrosov [On analytic solution of the Feller problem of upward excursions]. *Trudy SPbGTU. Prikladnaya Matematika* [SPb Polytechnic Univeristy "Applied Mathematics" Proceedings] 461:92–100.
15. Pugachev, V. S., and I. N. Sinityn. 1987. *Stochastic differential systems. Analysis and filtering*. Chechester. 549 p.
16. Sveshnikov, A. A. 1970. Primenenie teorii nepreryvnykh markovskikh protsessov k resheniyu nelineynykh zadach prikladnoy giroskopii [Application of the theory of continuous Markov processes to solution of nonlinear problems of applied gyroscopy]. *Tr. V Mezhdunar. konf. po nelineynym kolebaniyam* [5th Conference (International) on Nonlinear Vibrations Proceedings]. Kiev. 3:659–665.
17. Sveshnikov, A. A., and S. S. Rivkin. 1974. *Veroyatnostnye metody v prikladnoy teorii giroskopov* [Probabilistic methods in the theory of gyroscopy]. Moscow: Nauka. 536 p.
18. Zayats, O. I. 2013. Primenenie uravneniya Pugacheva–Sveshnikova k issledovaniyu kusochno-lineynykh stokhasticheskikh sistem, lineynykh v poluprostranstvakh [Analysis of piecewise linear stochastic systems in half-spaces by means of the Pugachev–Sveshnikov equation]. *Nauchno-Tekhnicheskije Vedomosti SPbGPU, Fiziko-Matematicheskie Nauki* [St. Petersburg State Polytechnical University J. Physics and Mathematics] 4(1):128–142.
19. Zayats, O. I., and S. V. Berezin. 2013. Primenenie uravneniya Pugacheva–Sveshnikova k issledovaniyu kusochno-lineynykh stokhasticheskikh sistem, lineynykh v chetvertyakh prostranstva [Analysis of piecewise linear stochastic systems in quarter-spaces by means of the Pugachev–Sveshnikov equation]. *Nauchno-tekhnicheskije Vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University J. Computer Science. Telecommunication and Control Systems] 6:87–101.
20. Caughey, T. K., and J. K. Dienes. 1961. Analysis of nonlinear first order system with a white noise input. *J. Appl. Phys.* 32(11):2476–2479.
21. Zayats, O. I. 1999. Reshenie zadachi Feller'a dlya vinerovskogo protsessa s postoyannym snosom [Solution of the Feller problem for a Wiener process with a constant drift]. *Trudy SPbGTU. Prikladnaya Matematika* [SPb Polytechnic Univeristy "Applied Mathematics" Proceedings]. 477:67–72.
22. Ito, K., and H. P. McKean. 1963. Brownian motion on a half-line. *Illinois J. Math.* 7:181–231.
23. Lejay, A. 2006. On the constructions of the skew Brownian motion. *Probab. Surveys* 3:413–466.
24. Le Gall, J.-F. 1984. One-dimensional stochastic differential equations involving the local times of the unknown process. *Stochastic analysis and application*. Eds. A. Truman and D. W. Williams. Lecture notes in mathematics ser. Berlin–Heidelberg: Springer. 1095:51–82.
25. Appuhamillage, T., V. Bokil, E. Thomann, E. Waymire, and B. Wood. 2011. Occupation and local times for skew Brownian motion with applications to dispersion across an interface. *Annals Appl. Probab.* 21(1):183–214.
26. Lavrent'ev, M. A., and B. V. Shabat. 1965. *Metody teorii funktsii kompleksnogo peremennogo* [Methods of the theory of functions of a complex variable]. Moscow: Nauka. 716 p.
27. Borodin, A. N., and P. Salminen. 1996. *Handbook of Brownian motion. Facts and formulae. Probability and its applications*. Basel: Birkhäuser. 462 p.

Received February 2, 2015

Contributors

Berezin Sergey V. (b. 1986) — senior scientist, Institute of Applied Mathematics and Mechanics, Peter the Great St. Petersburg State Polytechnic University, 29 Politekhnikeskaya Str., St. Petersburg 195251, Russian Federation; servberezin@yandex.ru

Zayats Oleg I. (b. 1952) — Candidate of Science (PhD) in physics and mathematics, associate professor, Institute of Applied Mathematics and Mechanics, Peter the Great St. Petersburg State Polytechnic University, 29 Politekhnikeskaya Str., St. Petersburg 195251, Russian Federation; zay.oleg@gmail.com

ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ДЛЯ ХАРАКТЕРИСТИК ПРИ УСВОЕНИИ ДАННЫХ НАБЛЮДЕНИЙ В СТАЦИОНАРНОМ РЕЖИМЕ*

К. П. Беляев¹, Н. П. Тучкова²

Аннотация: Рассматривается линейная задача усвоения данных наблюдений в гидродинамическую модель. Для такой задачи, сформулированной в терминах функционирования марковской цепи, исследуется сходимость переходных распределений цепи к стационарному распределению, находятся достаточные условия этой сходимости и строится функциональное уравнение для характеристической функции этого распределения. Далее рассматривается схема серий, зависящих от параметра. Исследуется сходимость стационарных распределений цепи, когда параметр стремится к нулю. Показывается, что в такой схеме в стационарном режиме марковской цепи существует предельное распределение параметров, и доказывается, что оно будет гауссовым, находятся его среднее и дисперсия. Обсуждается, как данная схема может быть применена для оценки характеристик при оперативном усвоении данных и прогнозировании состояния среды.

Ключевые слова: методы усвоения данных наблюдений; стационарные распределения цепей Маркова; асимптотическое распределение цепей при малом значении параметра

DOI: 10.14357/19922264150206

1 Введение

Современная оперативная океанография, оперативный прогноз погоды и ряд других областей основаны на сочетании численного моделирования процесса и усвоения данных наблюдений в модель. Задача усвоения состоит в том, чтобы оптимальным образом сочетать модельный расчет и наблюдения, сохранить баланс энергии, массы, тепла и других входящих в модель параметров и при этом сделать модельный расчет как можно более близким к наблюдениям в смысле заданной метрики. Далее модель, стартуя с оптимального в этом смысле усвоенного или скорректированного поля, должна дать лучший прогноз, чем в случае без усвоения данных. При этом качество как самой модели, так и метода усвоения может быть количественно оценено по степени близости расчета и/или прогноза к наблюдаемым данным. Важна также техническая сторона дела, а именно: срок обработки данных, используемые ресурсы компьютера, память, время работы процессора, скорость передачи информации от получения данных до доставки потребителю, доступ для визуализации результатов и многое другое. Все эти проблемы прямо или косвенно решаются в задачах оперативного усвоения данных наблюдений.

Схема линейного усвоения данных достаточно проста, она реализуется в системе уравнений следующего вида:

$$\left. \begin{aligned} X_n &= F(X_{n-1}); \\ X_{n+1} &= X_n + K(Y_n - HX_n). \end{aligned} \right\} \quad (1)$$

Здесь и далее используются следующие обозначения:

- $X_n, n = 1, 3, 5, \dots$, — вектор состояния модели, $X_n \in R^r$, т. е. $X_n = (x_n^1, \dots, x_n^r)$. Например, в моделях динамики океана X_n может состоять из компонент температуры, солёности, скорости течений и т. п. Для определенности этот вектор в статье рассматривается в каждой точке пространства области, состоящей из N точек сетки, где задана модель;
- F — оператор модели, т. е. векторная функция из $R^r \mapsto R^r$.
- вектор $Y \in R^{N_{\text{obs}}}$ — вектор наблюдаемых величин, его длина на шаге n обозначается как N_{obs} . Обычно $N_{\text{obs}} < r$, так как не все расчетные величины в модели могут быть наблюдаемы. Например, в моделях циркуляции океана

* Работа выполнена при поддержке РФФИ (проекты 14-05-00363 и 14-07-00037).

¹Институт океанологии им. П. П. Ширшова Российской академии наук; Федеральный университет штата Баия, Сальвадор, Бразилия, kb@sail.msk.ru

²Вычислительный центр им. А. А. Дородницына Российской академии наук, tuchkova@ccas.ru

обычно наблюдаемыми могут быть температура, соленость, уровень океана и очень редко компоненты скорости течений;

- матрица K в (1) — это $(r \times N_{\text{obs}})$ -матрица, она имеет в литературе название весовой матрицы (Kalman gain matrix). Ее физический смысл состоит в том, чтобы передать вес наблюдения в точку расчетной сетки, т. е. она передает значения из пространства наблюдений в пространство модели;
- матрица H в (1) — это матрица проекционного оператора, имеющая размерность $N_{\text{obs}} \times r$ и осуществляющая обратную операцию, т. е. передающая сигнал от пространства модели в пространство наблюдений.

Таким образом, матрица KH будет квадратной $(r \times r)$ -матрицей. Расчет начинается с заданного поля X_0 .

Такая ассимиляционная схема описана в многочисленных книгах и статьях, например в работах [1–3]. Схемы усвоения различаются конкретным выбором модели F и/или матрицы K . В частности, хорошо известный фильтр Калмана [4] относится к данной схеме. Есть и другие методы определения матрицы K : например, в [5] матрица K строится на основе решений параболических дифференциальных уравнений типа Фоккера–Планка.

В схеме (1) последовательность X_n образует цепь Маркова. Вектор наблюдений Y_n предполагается случайным на каждом шаге с заданным распределением. Возникает задача построения распределения случайного вектора X_n и изучение его предельного поведения при $n \rightarrow \infty$. Это достаточно сложная и до конца не решенная проблема. В [5] приведены некоторые решения для линейной модели $F(X) = \Lambda X$, где Λ — это $(r \times r)$ -матрица.

Значительно более сложной задачей будет определение совместного распределения векторов $X_{n_1}, X_{n_2}, \dots, X_{n_k}$ для заданного набора индексов (n_1, n_2, \dots, n_k) . В работе [6] показано, что при физически обоснованных и проверяемых условиях совместное распределение векторов $X_{n_1}, X_{n_2}, \dots, X_{n_k}$ стремится к совместному распределению векторов, определяющих диффузионный процесс, для случая $X_{n_k} = X_{0+n_k dt}$ при $n_k \rightarrow \infty dt \rightarrow 0$, когда временной интервал $[0, T]$ разбивается точками $[0, n_1, n_2, \dots, n_k, \dots, T]$.

Важный частный случай имеет место, когда матрица K мала. Так, например, происходит в случае ансамблевого фильтра Калмана (EnKF), в схеме объективного анализа [1, 4], когда матрица K строится на основе статистики аномалий, т. е. разницы между расчетным и средним многолетним значением какого-нибудь модельного поля, например

поля температуры или уровня океана. Если реальная изменчивость не очень велика, то модельная разница между конкретным расчетом и средним многолетним для температуры в точке составляет около $2\text{--}3^\circ\text{C}$ и в этом случае матрица K по порядку величины будет около 10^{-4} в соответствующих единицах. Поэтому имеет смысл изучить предельное поведение распределений X_n при малых K , т. е. когда K можно представить как $K_l = \rho_l K$, $l = 1, 2, 3, \dots$, и устремить $\rho_l \rightarrow 0$. Такого рода задачи активно рассматриваются в теории массового обслуживания при изучении пределов характеристик при малой и большой нагрузке.

Цели настоящей работы заключаются в том, чтобы получить условия стационарного режима цепи Маркова (1) и найти распределение этого режима; получить предельное распределение характеристик в стационарном режиме при $K_l = \rho_l K$, $\rho_l \rightarrow 0$. Метод доказательства теоремы является оригинальным и близок к опубликованному ранее в работе [7].

2 Основные формулировки

Пусть X_n и Y_n , $n = 1, 2, 3, \dots$, — случайные величины, определенные на заданном вероятностном пространстве. Для определенности представим Y_n как $Y_n = HZ_n$, где случайная величина $Z_n \in R^r$. Как следует из формул (1), матрица KH имеет размерность $r \times r$. Пусть $G_n(x) = \text{Pr}(X_n < x)$ и $\Gamma_n(x) = \partial/\partial x[\text{Pr}(Z_n < x)]$ — их соответствующие функции распределения и плотности.

Лемма 1. Пусть $n \rightarrow \infty$ и $\Gamma_n(x) > 0$. Для того чтобы существовало стационарное распределение X_n , $\lim P(X_n < x) = G(x)$, достаточно выполнения следующего условия: для любых x и y существует решение

$$y = F(x), \quad KHx = y. \quad (2)$$

Доказательство. Покажем, что для любой пары состояний цепи x, y существует такой момент времени k , что переходные вероятности

$$P(X_{n+k} = x | X_n = y) > 0. \quad (3)$$

Действительно, определим $P(X_{n+k} = x)$ в состоянии x для некоторых n и k . Если $n+k$ нечетно, то из условия (2) следует, что есть состояние $\bar{x} = F^{-1}(x)$ такое, что $P(X_{n+k-1} = \bar{x}) > 0$. Тогда если $X_{n+k-2} = y$, то для того чтобы цепь в момент $n+k$ оказалась в состоянии x , требуется, чтобы $KHZ_{n-1} = \bar{x} - y - KH y$, вероятность чего больше нуля по условию (2). Аналогично доказывается и для четных $n+k$ шагов. Следовательно, любые два состояния цепи являются смежными в том смысле, что за конечное число шагов можно попасть из одного состояния в другое с ненулевой вероятностью.

Отсюда следует существование стационарного распределения цепи Маркова (см. [8]).

Замечание 1. Доказательство проведено для дискретных цепей, но легко обобщается на случай непрерывных цепей Маркова. Для определенности в дальнейшем все распределения будем считать непрерывными.

Замечание 2. Условия (2) и условие $\Gamma_n(x) > 0$ означают, что данные и модель не противоречат друг другу. Понятно, например, что если модель динамики океана дает температуру воды $y = 20^\circ\text{C}$ и при этом наблюдения показывают $x = 2^\circ\text{C}$, то вероятности перехода (3) будут практически нулевыми, а решение уравнения $y = F(x)$, хотя теоретически и возможно, физически не реализуемо. Поэтому при практических расчетах важно, чтобы модель адекватно описывала физику, а данные измерялись в соответствии с этой моделью.

Рассмотрим теперь вложенную цепь Маркова только для нечетных номеров шага по времени $X_{2k} = X_{2k-1} + K(Y_{2k} - HX_{2k-1})$.

Лемма 2. Пусть выполнены условия леммы 1 и пусть $\pi(\omega)$ обозначает характеристическую функцию случайной величины X_{2k} в стационарном режиме, т. е.

$$\pi(\omega) = \lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} e^{i\omega x} dP(X_{2k} = x).$$

Как обычно, $i = \sqrt{-1}$.

Тогда $\pi(\omega)$ удовлетворяет функциональному уравнению:

$$\pi(\omega) = \pi[(I + KH)\omega]\psi((KH)^{-1}\omega), \quad (4)$$

где $\psi(\omega)$ — характеристическая функция случайной величины $Z = \lim_{k \rightarrow \infty} Z_{2k}$ в стационарном режиме и I — единичная матрица размерности $r \times r$.

Обозначим $\pi^{(s)}(0) = \pi_s$, $\psi^{(s)}(0) = \psi_s$.

Доказательство. Из уравнения Колмогорова–Чепмена следует соотношение для переходных вероятностей:

$$dP(X_{2k} = x) = \int_{-\infty}^{\infty} dP(X_{2k-1} = y) dP(KHZ_{2k-1} = x - y - KH y).$$

Умножая обе части этого равенства на $e^{i\omega x}$ и интегрируя по x от $-\infty$ до ∞ , после простых преобразований приходим к соотношению:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{i\omega x} dP(X_{2k} = x) &= \\ &= \int_{-\infty}^{\infty} e^{i(I+KH)y\omega} dP(X_{2k-1} = y) \times \\ &\quad \times \int_{-\infty}^{\infty} e^{i\omega x} dP(X_{2k-1} = (KH)^{-1}x). \end{aligned}$$

Из этого равенства переходом к пределу при $k \rightarrow \infty$ доказывается формула (4). В силу единственности стационарного распределения цепи Маркова уравнение (4) справедливо для всей цепи.

В дальнейшем X, Z будут обозначать соответствующие предельные случайные величины: $X = \lim_{k \rightarrow \infty} X_n$, $Z = \lim_{k \rightarrow \infty} Z_n$.

Сходимость рассматривается по вероятности: $P = (|X - \lim_{k \rightarrow \infty} X_n| > \epsilon) \rightarrow 0 \forall \epsilon > 0$, где $|X|$ обозначает любую из векторных норм в R^r .

Замечание 3. Равенство (4) может рассматриваться как скалярное или как векторное, где

$$\pi(\omega) = (\pi_1(\omega_1), \pi_2(\omega_2), \dots, \pi_r(\omega_r));$$

$$\pi_i(\omega_i) = \lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} e^{i\omega y_i} dP(X_n = y), \quad i = 1, 2, \dots, r,$$

причем все координаты кроме y_i свертываются.

Замечание 4. Как видно из (4), стационарное распределение зависит от распределения случайной величины Z и от передачи сигнала от Z к $X = \lim_{k \rightarrow \infty} X_n$ через матрицу KH и не зависит напрямую от модели F . Это несколько удивительно, но физически оправдано. Действительно, неприводимая цепь Маркова аналогична динамической системе с трением, в которой предельный режим если существует, то не зависит от начального состояния, а определяется внешним воздействием. В данном случае со временем состояния цепи «перемешиваются», а следовательно, в пределе стремятся к равномерному распределению, независимо от уравнений модели. Поэтому стационарный режим будет определяться внешним воздействием и передачей этого воздействия к характеристикам цепи Маркова.

Следствие. Моменты стационарного состояния цепи можно найти по формулам:

$$\begin{aligned} \pi_s &= \pi_s(I+KH)^s + \pi_{s-1}(I+KH)^{s-1}(KH)^{-1}\psi_1 + \dots \\ &\quad \dots + (KH)^{-s}\psi_s. \end{aligned}$$

Здесь и далее обозначено для вектора $a \in R^r$ и матрицы $B: R^r \mapsto R^r$

$$aB_s = a_i \left(\sum_{j=1}^r B_{ij} \right), \quad i = 1, 2, \dots, r.$$

В частности,

$$\begin{aligned} \pi_1 &= (I + KH)^{-2} \psi_1; \\ \pi_2 &= (I + KH)^{-2} [2\pi_1 \psi_1 + (I + KH)^{-1} \psi_2]. \end{aligned}$$

В дальнейшем будем рассматривать распределение исходной цепи Маркова в стационарном режиме. Без ограничения общности будем предполагать, что случайная величина Z имеет нулевое среднее $EZ = 0$ или $\psi_1 = 0$. Понятно, что из-за линейности второго из соотношений (1) X можно представить как $EX + X'$, $EX' = 0$ и EX определяется из (1) для ненулевого EZ .

Далее рассматривается схема серий для величин X , Z и K , зависящих от индекса l : X_l , Z_l и K_l , $l = 1, 2, \dots$. Вводится также параметр ρ_l так, что $\rho_l \rightarrow 0$. Очевидно, что характеристическая функция $\pi(\omega)$ также зависит от l , что для краткости не показывается.

Лемма 3. Пусть $K_l = \rho_l K$, $Z_l \rightarrow 0$ так, что $\rho_l^{-1} EZ_l^2 < \infty$, $\rho_l \rightarrow 0$.

Тогда для любого $s > 0$ верно утверждение:

$$\rho_l^{s+1} \pi_s \rightarrow 0. \quad (5)$$

Доказательство. Докажем методом математической индукции. Для $s = 0$ утверждение (5) очевидно. Пусть оно справедливо для некоторого $s > 0$. Дифференцируя обе части (4) $s + 2$ раза, получаем:

$$\begin{aligned} \pi^{(s+2)}(0) &= \pi^{(s+2)}(0)(I + KH)^{s+2} + \\ &+ \frac{(s+1)(s+2)}{2} \pi^{(s)}(0)(I + KH)^s \psi^{(2)}(0)(KH)^{-2} + \\ &+ A_s, \quad (6) \end{aligned}$$

где A_s содержит слагаемые с производными $\pi(\omega)$ порядка меньше s . Для получения формулы (6) использовано условие $\psi' = 0$.

Уравнение (6) можно преобразовать и получить следующее равенство:

$$\begin{aligned} \pi_{s+2}(KH)(I^{s+1} + (I + KH)I^s + \dots + (I + KH)^{s+1}) = \\ = \frac{(s+1)(s+2)}{2} \pi_s(I + KH)^s \psi_2(KH)^{-2} + A_s. \end{aligned}$$

В схеме серий это равенство переписывается как

$$\begin{aligned} \pi_{s+2}(K_l H)(I^{s+1} + (I + K_l H)I^s + \dots \\ \dots + (I + K_l H)^{s+1}) = \\ = \frac{(s+1)(s+2)}{2} \pi_s(I + K_l H)^s \psi_2(K_l H)^{-2} + A_s. \end{aligned}$$

Из условий леммы 3 следует:

$$\begin{aligned} \pi_{s+2} \rho_l^2 (KH)(I^{s+1} + (I + \rho_l KH)I^s + \dots \\ \dots + (I + \rho_l KH)^{s+1}) = \\ = \frac{(s+1)(s+2)}{2} \pi_s \rho_l (I + \rho_l KH)^s \psi_2(KH)^{-2} + \\ + A_s. \quad (7) \end{aligned}$$

Умножая обе части (7) на ρ_l^{s+1} , приходим к следующему равенству:

$$\begin{aligned} \pi_{s+2} \rho_l^{s+3} (KH)(I^{s+1} + (I + \rho_l KH)I^s + \dots \\ \dots + (I + \rho_l KH)^{s+1}) = \\ = \frac{(s+1)(s+2)}{2} \pi_s \rho_l^{s+1} (I + \rho_l KH)^s (\rho_l \psi_2)(KH)^{-2} + \\ + A_s \rho_l^{s+1}. \end{aligned}$$

Если $\rho_l \rightarrow 0$, то правая часть этого уравнения из предположения индукции и условий леммы 3 стремится к 0. Следовательно, и левая часть этого уравнения стремится к нулю, что и требовалось доказать.

Введем обозначение $\lambda_s = \lim_{\rho_l \rightarrow 0} \pi_s \rho_l^s$.

Лемма 4. Если выполнены условия леммы 3, то для любого четного s

$$\lambda_s = (s-1)!! \psi_0^s (KH)^{-2},$$

где $\psi_0 = \lim_{\rho_l \rightarrow 0} \rho_l^{-1} \psi_2 (2KH)^{-1}$,

и для нечетных s

$$\lambda_s = 0.$$

Доказательство. Из равенства (7) можно получить

$$\begin{aligned} \pi_s \rho_l^2 (KH)(I^{s-1} + (I + (\rho_l KH)I^{s-2} + \dots \\ \dots + (I + \rho_l KH)^{s-1}) = \\ = \frac{s(s-1)}{2} \pi_{s-2} \rho_l^{-1} (I + \rho_l KH)^{s-2} \psi_2(KH)^{-2} + \\ + A_{s-2}. \quad (8) \end{aligned}$$

Умножая обе части (8) на ρ_l^{s-2} и принимая во внимание утверждение леммы 3, приходим к следующему равенству:

$$\lambda_s s = \frac{s(s-1)}{2} \lambda_{s-2} [\rho_l^{-1} \psi_2(KH)] (KH)^{-2}. \quad (9)$$

Так как $\lambda_0 = 1, \lambda_1 = 0$, утверждение леммы 4 непосредственно следует из формулы (9).

Теперь сформулируем основное утверждение данной работы.

Теорема. Предельное распределение случайной величины X при условиях лемм 2 и 3 будет задаваться формулой:

$$\lim_{\rho_l \rightarrow 0} P(\rho_l^{-1}(X - EX) < x) = \Phi(x; (2KH)^{-2}\psi_0).$$

Как обычно, $\Phi(x; \sigma^2)$ обозначает гауссово распределение с нулевым средним и дисперсией σ^2 .

Доказательство следует из разложения Тейлора для характеристической функции случайной величины X :

$$\pi((\rho_l \omega)) = E e^{i\omega \rho_l X} = \sum_{j=0}^{\infty} \frac{(i\omega)^j \rho_l^j \pi_j}{j!}.$$

Поскольку

$$\sum_{j=0}^{\infty} \frac{(i\omega)^j \rho_l^j \pi_j}{j!} \rightarrow \sum_{j=0}^{\infty} \frac{(i\omega)^j}{j!} \lambda_j \text{ при } \rho_l \rightarrow 0,$$

то

$$\begin{aligned} \sum_{j=0}^{\infty} \frac{(i\omega)^j}{j!} \lambda_j &= \\ &= \sum_{j=2k, k=0}^{\infty} \frac{(i\omega)^{2k}}{(2k)!} (2k-1)!! \psi_0^{2k} (KH)^{-2k} = \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!!} (\omega \psi_0)^{2k} (KH)^{-2k} = \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(k)!} \left(\frac{\omega \psi_0}{2}\right)^{2k} (KH)^{-2k}. \end{aligned}$$

Последнее соотношение определяет характеристическую функцию гауссовой случайной величины с дисперсией $(2KH)^2$, что и требовалось показать.

Замечание 5. При доказательстве использовалось существование высших моментов распределений случайных величин X и Z в каждой серии. Это ограничение несущественно, так как в окончательном ответе требуется только существование их вторых моментов.

Замечание 6. Условие $\rho_l^{-1} E Z_l^2 < \infty, l \rightarrow 0$, является стандартным (см., например, [8]) для существования предельного распределения Гаусса.

Литература

1. *Jazwinski A. H.* Stochastic processes and filtering theory. — New York, NY, USA: Academic Press, 1970. 376 p.
2. *Ghil M., Malnotte-Rizzoli P.* Data assimilation in meteorology and oceanography // *Adv. Geophys.*, 1991. Vol. 33. P. 141–266.
3. *Evensen G.* Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics // *J. Geophys. Res.*, 1994. Vol. 6. P. 1014–1062.
4. *Evensen G.* The ensemble Kalman filter: Theoretical formulation and practical implementation // *Ocean Dyn.*, 2003. Vol. 53. P. 343–367.
5. *Belyaev K., Tanajura C. A. S., O'Brien J. J.* A data assimilation technique with an ocean circulation model and its application to the tropical Atlantic // *Appl. Math. Model.*, 2001. Vol. 25. P. 655–670.
6. *Tanajura C. A. S., Belyaev K.* A sequential data assimilation method based on the properties of diffusion-type process // *Appl. Math. Model.*, 2009. Vol. 33. P. 2165–2174.
7. *Belyaev K., Nazarov L.* Limit theorems for characteristics of a queuing system with batch processing // *Theory Prob. Appl.*, 1995. Vol. 40. No. 4. P. 73–78.
8. *Gikhman I., Skorokhod A.* An introduction to the theory of random processes. — New York, NY, USA: Dover Publ. Inc., 1996. 519 p.

Поступила в редакцию 19.02.15

ON A LIMIT DISTRIBUTION OF CHARACTERISTICS IN STATIONARY REGIME FOR THE LINEAR ASSIMILATION PROBLEM

K. P. Belyaev¹ and N. P. Tuchkova²

¹Shirshov Institute of Oceanology, Russian Academy of Sciences, 36 Nakhimovsky Pr., Moscow 119299, Russian Federation

²Dorodnicyn Computing Centre, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: A commonly used linear assimilation problem when the model state vector is corrected by observed data through the system of linear equations is considered. This problem is formulated as a Markov chain problem. For this problem, convergence of transitional probability of the corresponding Markov chain is investigated and the sufficient conditions of this convergence are found out. A special case of series depending on a parameter when this parameter goes to zero is discussed and the limit theorem about convergence to Gaussian distribution for analysis of state vector characteristics for this case is proved. The mean value and variance of this distribution are determined. The paper discusses how these results can be applied to practical operational assimilation and forecasting.

Keywords: data assimilation methods; Markov chains stationary distributions; asymptotic distribution of chains at a small value of a parameter

DOI: 10.14357/19922264150206

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (projects 14-05-00363 and 14-07-00037).

References

1. Jazwinski, A. H. 1970. *Stochastic processes and filtering theory*. New York, NY: Academic Press. 376 p.
2. Ghil, M., and P. Malnotte-Rizzoli. 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.* 33:141–266.
3. Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.* 6:1014–1062.
4. Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* 53:343–367.
5. Belyaev, K., C. A. S. Tanajura, and J. J. O'Brien. 2001. A data assimilation technique with an ocean circulation model and its application to the tropical Atlantic. *Appl. Math. Model.* 25:655–670.
6. Tanajura, C. A. S., and K. Belyaev. 2009. A sequential data assimilation method based on the properties of diffusion-type process. *Appl. Math. Model.* 33:2165–2174.
7. Belyaev, K., and L. Nazarov. 1995. Limit theorems for characteristics of a queuing system with batch processing. *Theory Prob. Appl.* 40(4):73–78.
8. Gikhman, I., and A. Skorokhod. 1996. *An introduction to the theory of random processes*. New York, NY: Dover Publ. Inc. 519 p.

Received February 19, 2015

Contributors

Belyaev Konstantin P. (b. 1955) — Doctor of Science in physics and mathematics; leading scientist, Shirshov Institute of Oceanology, Russian Academy of Sciences, 36 Nakhimovsky Pr., Moscow 119299, Russian Federation; kosbel55@gmail.com

Tuchkova Natalia P. (b. 1955) — Candidate of Science (PhD) in physics and mathematics; senior scientist, Dorodnicyn Computing Centre, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; tuchkova@ccas.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРИМЕНЕНИЯ ЭВРИСТИЧЕСКОГО И МЕТАЭВРИСТИЧЕСКОГО АЛГОРИТМОВ К ЗАДАЧЕ О ШКОЛЬНОМ АВТОБУСЕ*

Е. М. Бронштейн¹, Д. М. Вагапова²

Аннотация: Рассматривается задача о школьном автобусе, которая заключается в обеспечении доставки школьников по окончании занятий из школы по их остановкам. Целью является минимизация длины максимального из маршрутов. Представлен краткий обзор работ по данной тематике. Приведена постановка задачи и формализация. Описан эвристический алгоритм, предложенный авторами ранее. Также описан двухэтапный алгоритм на основе метаэвристики муравьиной колонии: после первоначальной кластеризации остановок, на которых высаживаются школьники, к каждому кластеру применяется алгоритм муравьиной колонии с различными значениями параметров. Представлены результаты сравнения эффективности предложенных алгоритмов, а также результаты работы программ для двух алгоритмов.

Ключевые слова: маршрутизация; задача о школьном автобусе; алгоритм муравьиной колонии; кластеризация

DOI: 10.14357/19922264150207

1 Введение

Задача о школьном автобусе (School Bus Routing Problem, SBPR) заключается в доставке школьников после окончания занятий до их остановок за минимальное время. Впервые задача была сформулирована в [1]. Задача о школьном автобусе входит в общее семейство задач маршрутизации (Vehicle Routing Problem, VPR). Обзор точных и эвристических методов решения задач маршрутизации представлен в [2]. Задача маршрутизации с одним транспортным средством без дополнительных ограничений является задачей коммивояжера (Traveling Salesman Problem, TSP).

Рассматриваемые задачи относятся к классу NP-трудных, поэтому для их решения целесообразно использовать эвристические и метаэвристические алгоритмы. Простой эвристический алгоритм решения задачи о школьном автобусе предложен в работе [3]. В статье разработано решение данной задачи с помощью двухэтапного подхода на основе метаэвристического алгоритма муравьиной колонии (Ant Colony Optimization, ACO). На первом этапе производится кластеризация остановок, на которых высаживаются школьники, а затем к каждому кластеру применяется алгоритм муравьиной колонии. Для расширения области поиска решений итерационно производится смещение границ кластеров. Число таких итераций равно числу оста-

новок. Впервые двухэтапный подход такого типа для решения задачи маршрутизации был предложен в [4].

В [5] предложен новый подход к решению задачи о школьном автобусе в городе. Рассматривается многомерная целевая функция. Задача разбивается на две подзадачи. Первая — это распределение студентов по автобусам, и вторая — построение автобусных маршрутов. Для первой подзадачи используется алгоритм районирования.

Решение задачи о школьном автобусе с помощью модифицированного алгоритма муравьиной колонии предложено в [6]. Задача решается на реальных данных г. Богота (Колумбия). Используется двухэтапный подход: сначала производится кластеризация остановок согласно их географическому расположению (с севера на юг, затем с запада на восток), а потом применяется модифицированный алгоритм муравьиной колонии для задачи коммивояжера.

Алгоритм муравьиной колонии является достаточно эффективной метаэвристикой для нахождения приближенных решений различных задач дискретной оптимизации. Результаты, полученные с помощью разработанного варианта двухэтапного подхода, сравниваются с результатами, полученными с помощью эвристического алгоритма из работы [3].

* Работа выполнена при поддержке РФФИ (проект 13-01-00005).

¹Уфимский государственный авиационный технический университет, bro-efim@yandex.ru

²Уфимский государственный авиационный технический университет, vagapova-dm@mail.ru

2 Постановка задачи

Более подробное описание задачи приведено в [3]. По окончании занятий требуется развезти школьников по их остановкам на нескольких автобусах. При этом школьников следует доставить за минимально возможное время. Это условие будет выполнено, если маршрут с максимальной длиной будет минимальным среди всех способов доставки. Предполагается, что автобусы движутся с постоянной скоростью.

Для построения математической модели введем соответствующие величины.

1. Для каждой из N остановок задана величина D_n ($n = 1, \dots, N$) — число школьников, которых надо доставить на эту остановку. Для единообразия школу считаем 0-й остановкой.
2. Каждый из K автобусов характеризуется числом T^k ($k = 1, \dots, K$) посадочных мест.
3. Матрица $C = \|c(i, j)\|$ содержит минимальные расстояния (от i -й остановки до j -й). Элементы матрицы C удовлетворяют неравенству треугольника.
4. Необходимо найти величины x_n^k ($n = 1, \dots, N$; $k = 1, \dots, K$) — число школьников, которых необходимо доставить k -м автобусом на n -ю остановку.

Опишем ограничения задачи.

На каждую остановку, кроме школы, должно быть доставлено неотрицательное число школьников:

$$x_n^k \geq 0 \quad (n = 1, \dots, N; k = 1, \dots, K), \text{ целые.} \quad (1)$$

На каждую остановку необходимо доставить всех школьников, для которых это необходимо:

$$\sum_{k=1}^K x_n^k = D_n \quad (n = 1, \dots, N). \quad (2)$$

Ограничения на вместимость автобусов:

$$\sum_{n=1}^N x_n^k \leq T^k \quad (k = 1, \dots, K). \quad (3)$$

Добавлением виртуальных школьников в количестве $\sum_{k=1}^K T^k - \sum_{n=1}^N D_n$ (в случае положительности этой величины), для которых пунктом доставки является школа, можно добиться того, что ограничения (3) выполняются в форме равенств. Таким образом, полагаем, что общее число школьников равно числу мест в автобусах: $\sum_{n=1}^N D_n = \sum_{k=1}^K T^k$.

Пусть $R^k = \{i: x_i^k \geq 1\}$ — множество остановок, на которых должен высадить пассажиров k -й автобус. Следует отметить, что величинами x_n^k маршруты автобусов однозначно не определяются. Пусть $P(R^k)$ — минимальная длина маршрута, проходящего через все остановки из R^k , у которого начальный пункт — школа.

Целевая функция:

$$\max_k \{P(R^k)\} \rightarrow \min. \quad (4)$$

Результатом решения задачи (1)–(4) является график доставки, при котором школьник, покидающий автобус последним, проедет минимально возможное расстояние. Очевидно, что задача (1)–(4) имеет решение (возможно, не единственное).

В работе [3] путем добавления новых переменных данная задача сведена к линейной частично целочисленной форме, для задач малой размерности приведено точное решение, а также предложен эвристический алгоритм, описанный далее.

3 Эвристический алгоритм

Для простоты полагаем, что все автобусы имеют одну и ту же вместимость T .

Найдем максимальное из расстояний в графе от нулевой вершины (школы) до всех вершин. Пусть соответствующий путь имеет вид $0-i_1-\dots-i_k$.

Заполняем автобус по следующим правилам.

Если $\sum_{s=1}^k D_{i_s} \geq T$, то автобус доставит школьников на остановки, расположенные на пути $0-i_1-\dots-i_k$, начиная от последней в пределах вместимости (может оказаться, что на одну из остановок будут доставлены этим автобусом не все школьники, которым это необходимо), иначе найдем остановки i_{s^*}, j^* , на которых достигается $\min_{s,j} \{c(i_k, j): j \notin \{i_0, \dots, i_k\}; c(i_s, j) + c(j, i_{s+1}) - c(i_s, i_{s+1}): s = 0, \dots, k-1, j \notin \{i_0, \dots, i_k\}\}$, где $c(i, j)$ — минимальное расстояние между остановками, $i_0 = 0$.

Если число школьников, которых необходимо доставить на остановки i_1, \dots, i_k, j^* , не меньше T , то назначаем автобус, который доставит школьников в пределах вместимости автобусов на остановки пути $0-i_1-\dots-i_k-j^*$, начиная с последней, если минимум достигается на величинах первого вида, и пути $0-i_1-\dots-i_{s^*}-j^*-i_{s^*+1}-\dots-i_k$ если минимум достигается на величинах второго типа.

В противном случае процесс добавления вершин к пути продолжается аналогично до заполнения автобуса. В результате как число свободных автобусов, так и число остановок, на которые еще

не доставлены все школьники, уменьшаются. Далее автобусы заполняются последовательно по тому же алгоритму.

4 Двухэтапный алгоритм на основе метаэвристики муравьиной колонии

Алгоритм муравьиной колонии был предложен М. Дориго [7]. Алгоритм имитирует поведение муравьев в колонии. Отдельный муравей не способен выжить, однако вся колония достигает высокого уровня самоорганизации благодаря «социальности» муравьев. Цель муравьев — найти кратчайший путь от жилища до источника пищи. В ряде работ (см., например, [8–14]) были предложены различные модификации алгоритма. В данной работе воспользуемся классическим вариантом данного алгоритма.

Введем следующие обозначения:

$A_{ij}(t)$ — количество феромона на дуге (i, j) перед t -й итерацией (начальные значения $A_{ij}(1)$ принимаются одинаковыми положительными величинами для всех дуг);

$\tau_{ij}^k(t)$ — количество феромона, которое k -й муравей оставляет на дуге (i, j) на t -й итерации;

$B_{ij}(t) = \sum_{k=1}^m \tau_{ij}^k(t)$ — количество феромона, которое все муравьи оставляют на дуге (i, j) на t -й итерации (m — число муравьев).

Для достижения цели муравьи пользуются четырьмя принципами:

1. Положительная обратная связь.

$$\tau_{ij}^k(t) = \begin{cases} \frac{Q}{L^k(t)}, & \text{если } (i, j) \in T^k(t); \\ 0 & \text{иначе,} \end{cases} \quad (5)$$

где $T^k(t)$ — маршрут, пройденный k -м муравьем на t -й итерации; $L^k(t)$ — длина маршрута $T^k(t)$; Q — масштабирующий параметр, одного порядка с длиной оптимального маршрута (в частности, за Q можно принять длину самого длинного пути от школы до всех остановок).

2. Отрицательная обратная связь. Феромон испаряется со временем. После каждой итерации количество феромона определяется по следующей формуле:

$$A_{ij}(t+1) = (1 - \rho)A_{ij}(t) + B_{ij}(t), \quad (6)$$

где ρ — коэффициент испарения, $\rho \in [0, 1]$.

3. Случайность. Каждый муравей выбирает дугу случайным образом. Вероятность перехода k -го муравья от i -й остановки к j -й на итерации t вычисляется по следующей формуле:

$$\rho_{ij}^k(t) = \begin{cases} \frac{[A_{ij}(t)]^\alpha [1/D_{ij}]^\beta}{\sum_{l \in J_i^k} [A_{il}(t)]^\alpha [1/D_{il}]^\beta}, & \text{если } j \in J_i^k; \\ 0 & \text{иначе,} \end{cases} \quad (7)$$

где D_{ij} — расстояние от i -й до j -й остановки; α — параметр, регулирующий влияние уровня феромона на вероятность $p_{ij}^k(t)$, при $\alpha = 0$ муравей будет выбирать следующую остановку, только исходя из расстояния до нее (жадный алгоритм); β — параметр, регулирующий влияние расстояния между остановками на вероятность $p_{ij}^k(t)$, при $\beta = 0$ муравей будет выбирать следующую остановку, только исходя из уровня феромона, что приведет к быстрой сходимости к некоторому субоптимальному решению; J_i^k — список остановок, еще не посещенных k -м муравьем.

По формуле (7) вычисляются вероятности перехода муравья на каждую из непосещенных остановок, выбор следующей остановки производится случайно в соответствии с вычисленными вероятностями.

4. Многократность (множество муравьев). Длинные дуги будут содержать меньшее количество феромона, чем короткие, в соответствии с (5). Муравьи будут стремиться к дугам с большим количеством феромонов и тем самым будут предпочитать короткие дуги длинным.

Двухэтапный алгоритм решения задачи о школьном автобусе имеет следующую структуру. Первый этап заключается в кластеризации пунктов (остановок). На втором этапе для построения маршрутов будет использоваться алгоритм муравьиной колонии с различными значениями параметров α , β , ρ . Заметим, что для формирования маршрута одного автобуса запускается колония муравьев.

Первый этап — кластеризация.

Целью данного этапа является распределение школьников по маршрутам (автобусам). Для этого разобьем всю плоскость, на которой лежат остановки (рис. 1), на кластеры (штриховые линии — границы кластеров), каждый кластер соответствует одному автобусу. Выбираем начальный луч, проходящий через некоторую остановку. Двигаемся по часовой стрелке (на рис. 1 обозначена пунктиром). Кластером является минимальный сектор, для ко-

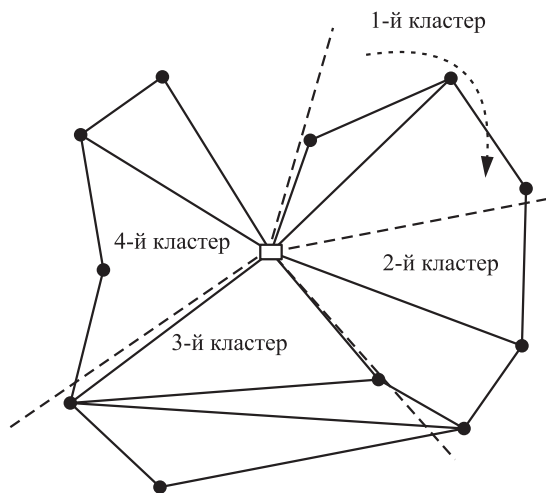


Рис. 1 Схема разбиения множества остановок на кластеры

того число школьников, которых надо доставить на остановки, принадлежащие этому кластеру, не меньше вместимости транспортного средства. Затем формируются следующие кластеры. Если упомянутое число школьников в кластере больше вместимости автобуса, то граничная остановка кластера войдет в следующий кластер с соответственно уменьшенным числом школьников. Для построения начальных лучей при разбиении на кластеры последовательно перебираются все остановки.

Приведем пример, который показывает целесообразность изменения начального луча. Пусть три остановки имеют координаты: $A(1, 0)$, $B(1, 1)$, $C(0, 1)$; школа расположена в точке $O(0, 0)$; вместимость автобуса — 20 школьников; автобусов — два; на остановки A и C надо доставить по 10 школьников; на остановку B — 20; прямые дороги с двусторонним движением соединяют пары остановок (OA) , (AB) , (BC) , (CO) . Если в качестве начального принять луч OB , то в один кластер попадает остановка B , а в другой — остановки A и C . Расстояние, необходимое для доставки одного из школьников во втором кластере, равно 3. Если же в качестве начального луча взять OA , то в один кластер попадут остановки A и B (с частичной доставкой), а во второй — B (с частичной доставкой) и C . Максимальное расстояние доставки равно 2.

Второй этап — маршрутизация.

Целью данного этапа является построение маршрутов движения автобусов, т. е. установление порядка посещения остановок автобусами в каждом кластере (минимальных гамильтоновых путей). Следует отметить, что маршруты в общем случае могут выходить за границы кластеров. Для построения маршрутов используется алгоритм муравьи-

ной колонии. Автобус доставляет всех школьников на каждую из остановок кластера, кроме последней, а оставшихся школьников на эту остановку доставляет автобус, обслуживающий следующий кластер (возможно, не один). Ниже приводится описание алгоритма.

1. Цикл по числу кластеров $u = (1, \text{число автобусов})$.
2. Инициализация параметров алгоритма — $\alpha, \beta, \rho, Q, A_{ij}(1)$.
3. Размещение муравьев в начальной вершине (школе). Число муравьев равно числу остановок на данном маршруте.
4. Выбор начального кратчайшего маршрута и определение длины этого пути L^{u*} .
5. Цикл по времени жизни колонии $t = (1, \text{время жизни колонии})$.
6. Цикл по числу муравьев $k = (1, \text{число остановок на маршруте } (m))$.
7. Построение маршрута $T^{ku}(t)$ по формуле (7) и расчет длины $L^{ku}(t)$.
8. Конец цикла по муравьям.
9. Проверка всех $L^{ku}(t)$ на лучшее решение по сравнению с L^{u*} .
10. В случае если решение $L^{ku}(t)$ лучше, обновление L^{u*} и T^{u*} .
11. Цикл по всем дугам графа.
12. Обновление следов феромона на дуге по правилам (5) и (6).
13. Конец цикла по дугам.
14. Конец цикла по времени жизни колонии.
15. Вывод кратчайшего маршрута T^{u*} и его длины L^{u*} .
16. Конец цикла по кластерам.

5 Вычислительный эксперимент

Двухэтапный алгоритм, описанный в разд. 4, и эвристический алгоритм, описанный в разд. 3, были реализованы в среде Borland DELPHI 6. Вычисления проводились на персональном компьютере (процессор: Intel(R) Core(TM) 2 Duo CPU T5750 2,00 GHz, ОЗУ (RAM): 3 ГБ, ОС: Windows Vista Home Premium 32). Число автобусов принималось равным 10 или 20. Вместимость автобусов принималась равной 20, 30 или 60. Число остановок — 50, 60, 70 или 80. Параметр α принимался равным 0, 1, 2 или 6; β — 0, 1, 2, 6; ρ — 0,2, 0,5, 0,7; $Q = 10\,000$; начальный уровень феромона $A_{ij}(1) = 100$. Время жизни колонии — 200. С помощью датчика псевдослучайных чисел генерировались значения координат каждой остановки (расстояние бралось евклидово), а также число школьников на каждой

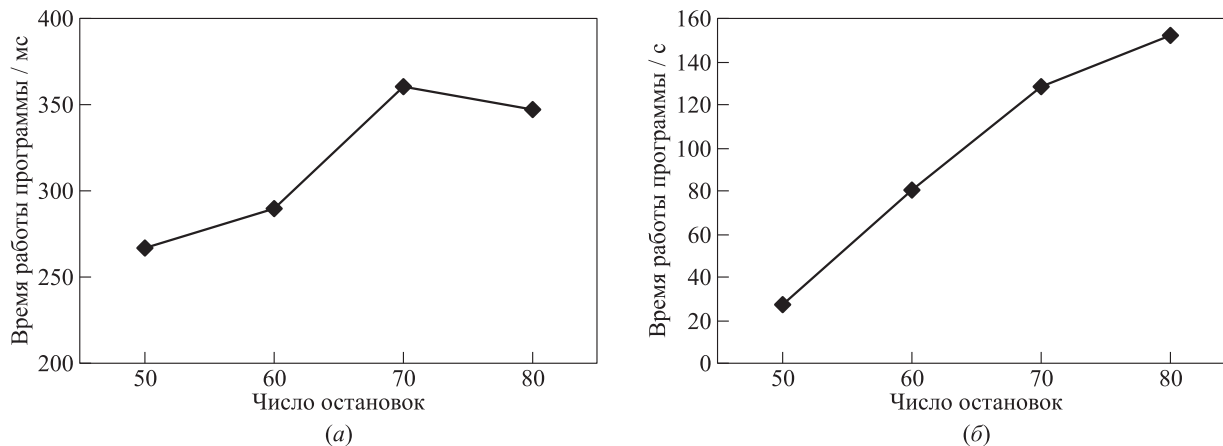


Рис. 2 Время работы эвристического (а) и двухэтапного (б) алгоритмов

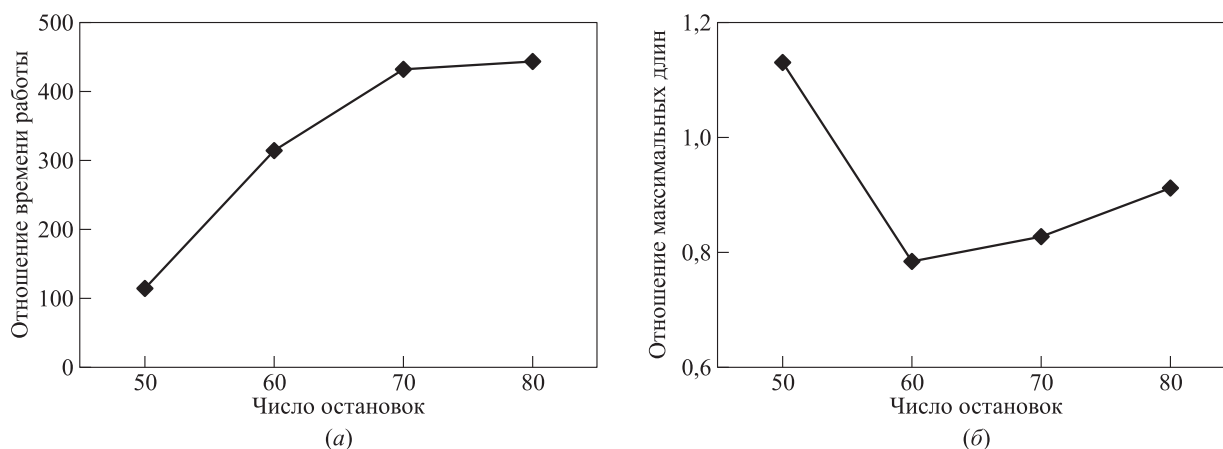


Рис. 3 Отношение времени работы двухэтапного алгоритма к времени работы эвристического алгоритма (а) и отношение максимальных длин маршрутов, полученных двухэтапным и эвристическим алгоритмами (б)

остановке с учетом того, что общее число школьников равно общему числу посадочных мест. Для каждого числа остановок генерировалось 10 примеров.

Рисунки 2 и 3 иллюстрируют зависимость средних значений величин, полученных в ходе вычислительного эксперимента, от числа остановок.

Отметим, что приводится время расчета (до 4 мин) для одного набора параметров алгоритма муравьиной колонии и для одного расположения начального луча на первом этапе, а именно для тех, при которых получен лучший результат. При переборе всех наборов величин время вычислений доходит до 10 ч.

6 Заключение

В работе рассматривается задача о школьном автобусе, разработан двухэтапный алгоритм решения, основанный на алгоритме муравьиной колонии. Производится сравнение его эффективности с простым эвристическим алгоритмом.

Вычислительный эксперимент показал, что

- (1) время работы эвристического алгоритма слабо изменяется с ростом числа остановок;
- (2) время работы двухэтапного алгоритма и его отношение к времени работы эвристического алгоритма резко возрастают с ростом числа остановок;
- (3) отношение максимальных длин маршрутов находится в промежутке 0,75–1,15 и слабо изменяется с ростом числа остановок.

Таким образом, простой эвристический алгоритм сопоставим по эффективности полученного результата с двухэтапным, но гораздо менее трудоемок.

Литература

1. Newton R. M., Thomas W. H. Design of school Bus Routes by computer // Socio-Economic Planning Science, 1969. Vol. 3. P. 75–85.

2. Archetti C. Matheuristics for routing problems. http://www.sintef.no/contentassets/cfb19ab9b7c74d03904c7746ee1d8e77/matheuristics_routing_verolog2014_new.pdf.
3. Бронштейн Е. М., Ваганова Д. М., Назмутдинова А. В. О построении семейства маршрутов доставки школьников за минимальное время // Автоматика и телемеханика, 2014. № 7. С. 43–51.
4. Fisher M. L., Jaikumar R. A generalized assignment heuristic for vehicle routing // Networks, 1981. Vol. 11. No. 2. P. 109–124.
5. Bowerman R., Hall B., Calamai P. A multi-objective optimization approach to urban school Bus Routing: Formulation and solution method // Transportation Research Part A: Policy and Practice, 1995. Vol. 29. No. 2. P. 107–123.
6. Arias-Rojas J. S., Jimenez J. F., Montoya-Torres J. R. Solving of school bus routing problem by ant colony // Revista EIA, 2012. Vol. 9. No. 17. P. 193–208.
7. Dorigo M. Optimization, learning and natural algorithms. PhD Thesis. — Milano, Italy: Politecnico di Milano, 1992.
8. Gambardella L. M., Dorigo M. Ant-Q: A reinforcement learning approach to the traveling salesman problem // 12th Conference (International) on Machine Learning. — Tahoe City: Morgan Kaufmann, 1995. P. 252–260.
9. Dorigo M., Maniezzo V., Colomi A. The Ant System: Optimization by a colony of cooperating agents // IEEE Trans. Systems Man Cybernetics. Part B, 1996. Vol. 26. No. 1. P. 29–41.
10. Dorigo M., Gambardella L. M. Ant Colony System: A cooperative learning approach to the traveling salesman problem // IEEE Trans. Evol. Comput., 1997. Vol. 1. No. 1. P. 53–66.
11. Stützle T., Hoos H. MAX-MIN Ant System and local search for the traveling salesman problem // IEEE International Conference on Evolutionary Computation. — Indianapolis: IEEE, 1997. P. 309–314.
12. Штовба С. Д. Муравьиные алгоритмы // Экспонента Про. Математика в приложениях, 2003. № 4. С. 70–75.
13. Курейчик В. М., Кажаров А. А. О некоторых модификациях муравьиного алгоритма // Известия Южного федерального университета. Технические науки, 2008. Т. 81. № 4. С. 7–12.
14. Слестников С. А. Применение метаэвристических алгоритмов для задачи маршрутизации транспорта // Экономика и математические методы, 2014. Т. 50. № 1. С. 117–126.

Поступила в редакцию 19.02.15

COMPARATIVE ANALYSIS OF APPLICATION OF HEURISTIC AND METAHEURISTIC ALGORITHMS TO THE SCHOOL BUS ROUTING PROBLEM

E. M. Bronshtein and D. M. Vaganova

Ufa State Aviation Technical University, 12 K. Marx Str., Ufa 450000, Russian Federation

Abstract: This paper considers the school bus routing problem, which is to ensure delivery of students after lessons from school to their stops. The objective function is to minimize the maximum length of the routes. A short review of the literature on this theme is provided. The problem definition and formalization is given. The heuristic algorithm proposed by the authors earlier is described. A two-step algorithm based on ant colony metaheuristics is described. The algorithm consists of initial clustering of stops at which students drop off, and subsequent ant colony optimization with different parameters, which is applied to each cluster. The results of comparing the efficiency of the proposed algorithms and the performance of the program for two algorithms are presented.

Keywords: vehicle routing problem; school bus routing problem; ant colony optimization; clustering

DOI: 10.14357/19922264150207

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 13-01-00005).

References

1. Newton, R. M., and W. H. Thomas. 1969. Design of school Bus Routes by computer. *Socio-Economic Planning Science* 3:75–85.
2. Archetti, C. Matheuristics for routing problems. Available at: http://www.sintef.no/contentassets/cfb19ab9b7c74d03904c7746ee1d8e77/matheuristics_routing_verolog2014_new.pdf (accessed March 10, 2015).
3. Bronshtein, E. M., D. M. Vaganova, and A. V. Nazmutdinova. 2014. O postroenii semeystva marshrutov dostavki shkol'nikov za minimal'noe vremya [About creating a set of delivery routes of students in a minimal time]. *Avtomati-*

- ka i Telemekhanika* [Automation and Remote Control] 7:43–51.
4. Fisher, M. L., and R. A. Jaikumar. 1981. Generalized assignment heuristic for vehicle routing. *Networks* 11(2):109–124.
 5. Bowerman, R., B. Hall, and P. A. Calamai. 1995. Multi-objective optimization approach to urban school Bus Routing: Formulation and solution method. *Transportation Research Part A: Policy and Practice* 29(2):107–123.
 6. Arias-Rojas, J. S., J. F. Jimenez, and J. R. Montoya-Torres. 2012. Solving of school bus routing problem by ant colony. *Revista EIA* 9(17):193–208.
 7. Dorigo, M. 1992. *Optimization, learning and natural algorithms*. PhD Thesis. Milan, Italy: Politecnico di Milano.
 8. Gambardella, L. M., and M. Dorigo. 1995. Ant-Q: A reinforcement learning approach to the traveling salesman problem. *12th Conference (International) on Machine Learning*. Tahoe City. 252–260.
 9. Dorigo, M., V. Maniezzo, and A. Coloni. 1996. The Ant System: Optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybernetics B* 26(1):29–41.
 10. Dorigo, M., and L. M. Gambardella. 1997. Ant Colony System: A cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evol. Comput.* 1(1):53–66.
 11. Stützle, T., and H. Hoos. 1997. MAX-MIN Ant System and local search for the traveling salesman problem. *IEEE Conference (International) on Evolutionary Computation*. Indianapolis. 309–314.
 12. Shtovba, S. D. 2003. Murav'inye algoritmy [Ant colony algorithms]. *Exponenta Pro. Matematika v prilozheniyakh* [Exponenta-Pro. Mathematics in Applications] (4):70–75.
 13. Kureychik, V. M., and A. A. Kazharov. 2008. O nekotorykh modifikatsiyakh murav'inogo algoritma [About some modifications of the ant colony algorithm]. *Izvestiya Yuzhnogo Federal'nogo Universiteta. Tekhnicheskie Nauki* [News of South Federal University. Technical Sciences] 81(4):7–12.
 14. Slastnikov, S. A. 2014. Primenenie metaevristicheskikh algoritmov dlya zadachi marshrutizatsii transporta [Application of metaheuristic algorithms for vehicle routing problem]. *Ekonomika i Matematicheskie Metody* [Economics and Mathematical Methods] 50(1):117–126.

Received February 19, 2015

Contributors

Bronshtein Efim M. (b. 1946) — Doctor of Science in physics and mathematics; professor, Ufa State Aviation Technical University, 12 K. Marx Str., Ufa 450000, Russian Federation; bro-efim@yandex.ru

Vagapova Diana M. (b. 1987) — applicant, Ufa State Aviation Technical University, 12 K. Marx Street, Ufa 450000, Russian Federation; vagapova-dm@mail.ru

МОДЕЛИ ДЛЯ ПРЕДСТАВЛЕНИЯ И ОБРАБОТКИ РЕФЕРЕНСНЫХ ЗНАЧЕНИЙ

М. П. Кривенко¹

Аннотация: Рассмотрена задача моделирования референсных значений — результатов наблюдения определенного типа величин, полученных от одного лица или группы лиц, в соответствии с заявленным описанием. Для этих целей предлагается использовать смесь нормальных распределений, которая может эффективно служить как средство аппроксимации реальных данных и при этом быть доступной с точки зрения теоретического анализа. При оценивании параметров компонентов смеси распределений главную роль играет метод максимального правдоподобия и его воплощение в виде EM (expectation-maximization) алгоритма. Для подбора числа компонентов смеси предлагается использовать критерий отношения правдоподобия и метод на основе расстояния между распределениями типа хи-квадрат. Их свойства исследуются с помощью бутстреп-метода. В качестве экспериментов рассматривается описание эмпирического распределения данных о пациентах, включающих возраст и измерения PSA (Prostate-Specific Antigen). Предложенные решения имеют явные преимущества: высокую детализацию по возрастам, сглаживание результатов наблюдений для различных по объему возрастным групп, возможность формировать предположения о характере зависимости между возрастом и уровнем PSA.

Ключевые слова: смесь нормальных распределений; оценивание числа компонентов смеси; референсные значения

DOI: 10.14357/19922264150208

1 Введение

Практика диагностической медицины основывается, в частности, на данных, собранных во время обследования пациентов и при проведении клинических и дополнительных исследований. Они интерпретируются путем сравнения с ранее установленными и рекомендованными (принятыми для имеющихся условий диагностирования) данными. Если состояние пациента похоже на то, которое типично для конкретного заболевания, медицинский работник может сформировать диагноз на основе только этих данных (диагностика путем подтверждения). Этот диагноз становится более правдоподобным, если наблюдаемые симптомы и признаки не соответствуют образцам, характеризующим набор альтернативных заболеваний (диагностика путем исключения).

Данные, соответствующие определенному состоянию пациентов, называются референсными значениями. Сопутствующие понятия: референсная популяция, референсные индивидуумы, референсный интервал.

На языке задач классификации референсные значения есть обучающая выборка, относящаяся к определенному классу. Диагностирование пациента соответствует задаче классификации, понима-

емой как сопоставление данных предопределенным классам. Если деление на классы отсутствует или является нечетким, актуальной становится задача классификации как группирования. В литературе по распознаванию образов эти два значения термина *классификация* определяются как распознавание образов без учителя (классификация как группирование) и с учителем (классификация как сопоставление).

Базовой книгой, в которой классификация рассматривается как группирование, является [1]. В ней описывается методология классификации, охватывающая сбор данных, выбор и кодирование специфических свойств, нахождение сходства, конструирование иерархических групп, их формирование с помощью статистических методов. В [1, п. 7.3.2] также вводится понятие кластерного анализа как статистической процедуры, распределяющей объекты по группам (кластерам).

Постановка задачи кластерного анализа является крайне общей, и, как следствие, ее решение сопровождается рядом проблем:

- формальные статистические оценки, такие как критерии значимости или точные распределения, редки в кластеризации; чаще приходится полагаться на интуицию и здравый смысл;

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

- при решении прикладных задач трудно формально определить понятие кластера, но при этом у людей присутствует достаточно твердое представление о том, каков кластер. Соответственно, должны быть доступны эффективные разведочные методы, которые представляют данные различными способами так, чтобы пользователь мог сформировать собственные суждения о присутствии и конструкции кластеров;
- существует много методов кластеризации, но явно недостаточно способов выбора среди них. В рамках статистического подхода далеки от решений проблемы принятия вероятностных моделей, оценивания числа кластеров, надежности принадлежности объектов кластерам;
- существующие теории и методы не всегда соответствуют масштабам многих проблем кластеризации. Обеспечивающие группирование данных структуры данных и алгоритмы работают на сотнях объектов, но часто возникает необходимость в системах классификации для тысяч изображений, миллионов пациентов и миллиардов текстов;
- задача выбора системы признаков часто решается эвристическим образом, поэтому при наличии избыточного числа признаков надо быть готовым к тому, что необходимая классификация может быть «потеряна», а при неправильном выборе признаков может получаться классификация, которая не интересна с практической точки зрения.

Кластерный анализ референсных значений оказывается эффективным подходом не только при выявлении скрытых закономерностей в реальных данных, но и при аппроксимации их распределения, служащего для обеспечения классификации в смысле сопоставления.

2 Задачи и типовые методы обработки референсных значений

Согласно [2, гл. 5] и исследованиям автора данной работы можно выделить следующие группы задач статистической обработки референсных значений:

- разбиение (стратификация) референсных значений на подходящие группы;
- описание распределений каждой группы;
- выделение выбросов (аномальных данных);

- определение референсных пределов;
- обновление (актуализация) информации о референсных значениях;
- согласование референсных значений [3, 4];
- представление данных и диагностирование некоторого пациента с помощью имеющихся наборов референсных значений.

Обычно за основу при обработке референсных значений берутся следующие модели:

- параметрическая, нормальное распределение;
- непараметрическая.

Параметрический подход при формировании модели данных означает, что априорные предположения об объекте анализа формулируются в виде параметрического семейства распределений. При описании референсных значений вариант нормального распределения является наиболее распространенным и в то же время чаще всего критикуемым (см., например, [5, п. 2.5.1]).

Отказ от априорных предположений в виде нормального распределения приводит к непараметрической модели, в рамках которой никакой редукции данных при обработке референсных значений не осуществляется, они сами используются для построения статистического вывода.

В данной работе для описания референсных значений предлагается использовать модель смеси нормальных распределений. Она за счет усложнения позволяет лучше описывать реальные распределения данных, но, оставаясь параметрической моделью, обеспечивать значительное сжатие информации об объекте анализа. При этом остается в силе преимущество относительно простых аналитических решений типовых вероятностных задач в силу того, что основой по-прежнему остается нормальное распределение.

3 Модель смеси нормальных распределений

Модель конечной смеси нормальных распределений (далее для краткости — просто смеси распределений) означает, что плотность распределения $f(u)$ представима в виде:

$$f(u) = \sum_{j=1}^k p_j \varphi(u, \vartheta_j).$$

В этой модели неизвестными являются все или часть следующих характеристик (параметров смеси): число k элементов смеси; вероятности p_j появления элементов смеси (веса элементов смеси); параметры ϑ_j нормального распределения.

Оценивание параметров смеси осуществляется с помощью метода максимального правдоподобия $L(k, p, \vartheta) \rightarrow \max_{k, p, \vartheta}$, где $L(k, p, \vartheta) \equiv L(x, k, p, \vartheta) =$

$$= \prod_{i=1}^n \sum_{j=1}^k p_j \varphi(x_i, \vartheta_j) - \text{функция правдоподобия.}$$

Одна из главных трудностей при решении данной задачи связана с оцениванием целочисленного параметра k — числа элементов анализируемой смеси. Дело в том, что принцип максимального правдоподобия сам по себе не дает решения задачи нахождения оценки для k , имеющей практический смысл. Поэтому обычно используется следующая схема оценивания параметров смеси. Сначала задается верхняя граница для возможного числа элементов смеси k_{\max} и находятся оценки параметров p и ϑ для последовательности отдельных значений $k = 1, \dots, k_{\max}$. Затем тем или иным способом, отличным от максимизации правдоподобия, подбирается «наилучшее» значение $\hat{k} \in \{1, \dots, k_{\max}\}$, которое и выступает в качестве оценки числа элементов смеси. Таким образом, далее при оценивании параметров элементов смеси будет предполагаться, что число элементов смеси k является фиксированным, а неизвестными параметрами остаются только p и ϑ . Тогда согласно принципу максимального правдоподобия центральным становится решение оптимизационной задачи вида:

$$\ln L(x, k, p, \vartheta) = \sum_{i=1}^n \ln \left(\sum_{j=1}^k p_j \varphi(x_i, \vartheta_j) \right) \rightarrow \max_{p, \vartheta}. \quad (1)$$

Наиболее работоспособную общую схему процедур, позволяющих находить решения задачи (1), обычно называют EM-алгоритмом. Указанный алгоритм имеет достаточно богатую историю, изложение которой дано в [6, разд. 1.13 и 1.18; 7, введение; 8, разд. 1.8].

Введем в рассмотрение так называемые апостериорные вероятности q_{ij} принадлежности наблюдения x_i к j -му элементу смеси. Если известны значения параметров p_j и ϑ_j , $j = 1, \dots, k$, то при наблюдаемом значении x_i , $i = 1, \dots, n$, апостериорная вероятность принадлежности этого значения к j -му элементу смеси принимает вид:

$$q_{ij}(x_i, p_j, \vartheta_j) = \frac{p_j \varphi(x_i, \vartheta_j)}{\sum_{l=1}^k p_l \varphi(x_i, \vartheta_l)}.$$

Из этого определения следует, что для всех допустимых i и j $q_{ij} \geq 0$ и $\sum_{j=1}^k q_{ij} = 1$.

Теперь для нахождения оценки $(\hat{p}, \hat{\vartheta})$ параметров смеси распределений появляется возможность дать краткую запись EM-алгоритма, а именно:

- (1) положить $t = 0$ и задаться начальными значениями $(p^{(0)}, \vartheta^{(0)})$ для параметров смеси;
- (2) вычислить значения $q_{ij}^{(t)}(x_i, p_j^{(t)}, \vartheta_j^{(t)})$ (E-шаг);
- (3) определить значения $(p^{(t+1)}, \vartheta^{(t+1)})$ из условия максимизации функции правдоподобия (M-шаг);
- (4) положить $(\hat{p}, \hat{\vartheta}) = (p^{(t+1)}, \vartheta^{(t+1)})$ и завершить процесс нахождения оценки параметров смеси, если построенные оценки $(p^{(t+1)}, \vartheta^{(t+1)})$ являются «подходящими», или положить $t = t + 1$ и перейти к шагу 2 в противном случае.

Применение EM-алгоритма в описанном виде встречает как общие для итерационных процедур трудности (не определены способ задания начального приближения на первом шаге и критерий завершения работы на последнем шаге алгоритма), так и порождает специфические проблемы (устремление функции правдоподобия в бесконечность при реализации M-шага алгоритма).

Далее кратко дадим возможные, апробированные на практике варианты решения возникающих в связи с этим задач; подробнее решение указанных вопросов рассмотрено, например, в [9, п. 2.2.2].

Задание начальных приближений. Для этого в настоящее время предложено большое число разнообразных алгоритмов, которые группируются в рамках следующих общих подходов:

- случайный выбор;
- применение отличных от EM-алгоритма, обычно более «быстрых» процедур кластерного анализа;
- построение разбиения объектов на кластеры в пространстве сниженной размерности.

В данной работе в одномерном случае применялось приближение на основе равновероятных элементов смеси, а в многомерном случае — сначала оценивание параметров смеси для одномерного случая (например, для одного из показателей или для первой главной компоненты), а потом с помощью матрицы апостериорных вероятностей переход к формированию параметров смеси в многомерном случае. С позиций качества получаемых итоговых оценок параметров элементов смеси эти подходы оказываются практически идентичными. Кроме того, применялся прием оценивания параметров смеси от большего числа элементов смеси к меньшему с исключением отдельных компонентов смеси при формировании начальных приближений для оценки с меньшим количеством.

Завершение работы EM-алгоритма. Обычным приемом при задании критерия окончания итерационных алгоритмов является сравнение значений переменных, получающихся на последующих шагах поиска решения задачи. Сравнение аргументов поиска экстремума функции правдоподобия в рассматриваемом случае — достаточно сложная задача хотя бы по той причине, что в нем должны участвовать разнотипные переменные (веса, средние и ковариационные матрицы). При этом решение необходимо принимать простое: продолжать или не продолжать итерационный процесс. Поэтому в качестве контролируемой величины обычно берется значение максимума функции правдоподобия. Несмотря на недостатки этого подхода, обычно используется следующее правило: как только относительное приращение максимального значения функции правдоподобия на последовательных шагах итерационного процесса становится меньше заданного порога, процесс оценивания прекращается.

Реализация M-шага при нарушении условия ограниченности функции правдоподобия. Существенным недостатком EM-алгоритма является возможность обработки только ограниченной по величине функции правдоподобия. Но для реальных данных отдельные кластеры могут оказаться вырожденными (либо из-за их дискретности, либо из-за того, что размерность параметрического пространства оказывается меньше принятой для совокупности кластеров), что противоречит указанному требованию для применимости EM-алгоритма. Это приводит к необходимости использования уточненной (невыврожденной) модели смеси нормальных многомерных распределений, в которой введены ограничения на множество возможных значений ковариационных матриц элементов смеси (в одномерном случае это ограничение снизу на значение дисперсии). Задание ограничения снизу на собственные значения ковариационных матриц элементов смеси необходимо для предотвращения устремления функции правдоподобия в бесконечность и потери (превышения) порядка при выполнении операций с числами с плавающей точкой, но оно требует задание этого ограничения. С одной стороны, значение должно быть достаточно большим, чтобы обеспечить выполнение указанных условий, с другой стороны, неразумное увеличение этого значения может привести к снижению качества кластерного анализа (например, слишком большие значения приведут к потере индивидуальности кластеров).

Охарактеризуем возможности модели смеси распределений в соответствии с приведенным ра-

нее списком задач статистической обработки референсных значений:

- смеси распределений являются одной из базовых моделей кластерного анализа данных, что в данном случае составляет содержание задачи разбиения (стратификации) референсных значений на подходящие группы. В этой ситуации смесь используется не только для аппроксимации распределения реальных данных, но и для представления кластерной структуры этих данных. Понятно, что интерпретация кластеров является самостоятельной задачей и решается в основном специалистами в предметной области;
- выделение выбросов, а также определение референсных пределов может осуществляться путем решения задачи байесовской классификации данных, причем не только в случае одномерных референсных значений;
- обновление (актуализация) информации о референсных значениях требует наличия всех ранее сформированных данных и решается путем переоценки параметров смеси, что при наличии современных средств вычислительной техники не должно приводить к каким-либо проблемам;
- согласование референсных значений на базе смеси распределений пока сопряжено с аналитическими трудностями, но вполне успешно может быть реализовано с помощью непараметрических методов анализа;
- диагностирование с помощью имеющихся наборов референсных значений решается с помощью байесовской классификации.

Новые возможности открываются в задаче разбиения референсных значений на группы при использовании многомерных (в частности, двухмерных) смесей распределения.

Далее наряду с общим случаем, когда в записи нормального распределения μ_j означает вектор средних, а Σ_j — ковариационную матрицу, интерес будут представлять частные случаи одномерных данных, когда $\vartheta_j = (\mu_j, \sigma_j^2)$, а также двухмерных, когда $\vartheta_j = (\mu_{j1}, \mu_{j2}, \sigma_{j1}^2, \sigma_{j2}^2, \rho_j)$.

В двухмерном случае плотность нормального распределения имеет вид:

$$\varphi(x_1, x_2, (\mu_{j1}, \mu_{j2}, \sigma_{j1}^2, \sigma_{j2}^2, \rho_j)) = \frac{1}{2\pi\sigma_{j1}\sigma_{j2}} \times \times \frac{1}{\sqrt{1-\rho_j^2}} \exp\left\{-\frac{u_{j1}^2 - 2\rho_j u_{j1}u_{j2} + u_{j2}^2}{2(1-\rho_j^2)}\right\}, \quad (2)$$

где $u_{j1} = (x_1 - \mu_{j1})/\sigma_{j1}$, $u_{j2} = (x_2 - \mu_{j2})/\sigma_{j2}$ и параметры в общей и частной записи нормального распределения связаны как

$$\Sigma_j = \begin{pmatrix} \sigma_{j1}^2 & \sigma_{j1}\sigma_{j2}\rho_j \\ \sigma_{j1}\sigma_{j2}\rho_j & \sigma_{j2}^2 \end{pmatrix}.$$

Совместное распределение можно разложить на произведение условного и безусловного распределений, каждое из которых является также нормальным, а именно:

$$\begin{aligned} \varphi(x_1, x_2, (\mu_{j1}, \mu_{j2}, \sigma_{j1}^2, \sigma_{j2}^2, \rho_j)) &= \\ &= \varphi(x_1|x_2, \tilde{\mu}_j, \tilde{\sigma}_j^2) \varphi(x_2, \mu_{j2}, \sigma_{j2}^2), \end{aligned} \quad (3)$$

где

$$\tilde{\mu}_j = \mu_{j1} + \rho_j \sigma_{j1} \frac{x_2 - \mu_{j2}}{\sigma_{j2}}; \quad (4)$$

$$\tilde{\sigma}_j^2 = \sigma_{j1}^2 (1 - \rho_j^2). \quad (5)$$

Вернемся к описанию смеси двумерных нормальных распределений

$$\begin{aligned} f(u_1, u_2) &= \\ &= \sum_{j=1}^k p_j \varphi(u_1, u_2, (\mu_{j1}, \mu_{j2}, \sigma_{j1}^2, \sigma_{j2}^2, \rho_j)). \end{aligned} \quad (6)$$

Оно может стать источником информации о распределении одного показателя, например описываемого с помощью переменной u_1 , при фиксированном значении другого (переменная u_2). Для этого наряду с (2) надо иметь частное распределение переменной u_2 . Непосредственное интегрирование (6) дает

$$f(u_2) = \sum_{j=1}^k p_j \varphi(u_2, (\mu_{j2}, \sigma_{j2}^2)).$$

Теперь из (2) с учетом (3) получаем:

$$\begin{aligned} f(u_1|u_2) &= \frac{f(u_1, u_2)}{f(u_2)} = \\ &= \frac{\sum_{j=1}^k p_j \varphi(u_1|u_2, \tilde{\mu}_j, \tilde{\sigma}_j^2) \varphi(u_2, \mu_{j2}, \sigma_{j2}^2)}{\sum_{l=1}^k p_l \varphi(u_2, (\mu_{l2}, \sigma_{l2}^2))} = \\ &= \sum_{j=1}^k \tilde{p}_j \varphi(u_1|u_2, \tilde{\mu}_j, \tilde{\sigma}_j^2), \end{aligned}$$

где

$$\tilde{p}_j = \frac{p_j \varphi(u_2, \mu_{j2}, \sigma_{j2}^2)}{\sum_{l=1}^k p_l \varphi(u_2, (\mu_{l2}, \sigma_{l2}^2))}. \quad (7)$$

Таким образом, доказано, что условное распределение для смеси (6) есть опять же смесь нормальных распределений с параметрами, определяемыми (7), (4) и (5).

4 Оценивание числа элементов смеси

Нахождение подходящего числа элементов смеси \hat{k} является важной, но крайне сложной задачей; ее обычно относят к группе задач выбора модели (определенная модель соответствует некоторому значению k).

В арсенале соответствующих инструментов анализа данных можно выделить следующие:

- графический анализ;
- информационные критерии;
- множественная проверка статистических гипотез.

Графический анализ. Несмотря на то что принцип максимального правдоподобия сам по себе не позволяет получить решение задачи оценивания числа элементов смеси, информацию о его реальном «подходящем» значении из поведения максимума функции правдоподобия получить все же можно. Обычно исследуемая функция «существенно изменяется» лишь до некоторого значения \hat{k} , которое и можно принять за оценку искомого числа элементов смеси. Попытка формализации понятия «существенно изменяется» приводит к необходимости указать некоторое конкретное значение порога, при этом качество принятого решения должно оцениваться с помощью некоторого внешнего показателя, например с помощью вероятности ошибки классификатора, для которого при аппроксимации распределений элементов классов принимается модель смеси распределений.

Информационные критерии. Дальнейшим развитием приема анализа степени изменения $\max_{p, \vartheta} L(k, p, \vartheta)$ являются идеи подбора модели с помощью некоторого показателя (меры) адекватности модели, являющегося компромиссом между приспособленностью модели и ее сложностью. Перебирая различные варианты модели (различные значения k), отдают предпочтение тому из них, которому соответствует экстремум выбранного показателя.

Сложность модели смеси нормальных d -мерных распределений характеризуется количеством параметров модели, а именно:

$$v(k) = (k - 1) + k \left[d + \frac{d(d + 1)}{2} \right].$$

Рассмотрим далее критерии, связанные с различными мерами адекватности модели.

Информационный критерий Акаике (Akaike) основывается на асимптотической теории распределения функции правдоподобия (см. [6, п. 6.8.2]) и определяется показателем следующего вида:

$$AIC(k) = -2 \ln L(k, \hat{p}, \hat{\vartheta}) + 2v(k).$$

Байесовский критерий основывается на аппроксимации интегрального правдоподобия (см. [6, п. 6.9.3]) и определяется показателем следующего вида:

$$BIC(k) = -2 \ln L(k, \hat{p}, \hat{\vartheta}) + v(k) \ln n.$$

Классификационные критерии основываются на предположении, что модель смеси описывает отделяемые друг от друга кластеры. Представим логарифм функции правдоподобия как

$$\ln L(k, p, \vartheta) = C(k, p, \vartheta) + E(k, p, \vartheta),$$

где

$$C(k, p, \vartheta) = \sum_{j=1}^k \sum_{i=1}^n q_{ij} \ln p_j \varphi(x_i, \vartheta_j);$$

$$E(k, p, \vartheta) = - \sum_{j=1}^k \sum_{i=1}^n q_{ij} \ln q_{ij}, \quad E(k, p, \vartheta) \geq 0.$$

Тогда получаем:

$$C(k, p, \vartheta) = \ln L(k, p, \vartheta) - E(k, p, \vartheta). \quad (8)$$

Если компоненты смеси хорошо отделимы, то матрица апостериорных вероятностей $Q = (q_{ij})$ определяет разбиение элементов x_1, \dots, x_n и $E(k, p, \vartheta) \approx 0$. В противном случае $E(k, p, \vartheta)$ принимает большие значения. Как следствие, энтропия классификационной матрицы Q может нести информацию о числе элементов смеси. Заметим, что уравнение (8) показывает, что классификационное правдоподобие $C(k, p, \vartheta)$ может рассматриваться как компромисс между степенью соответствия модели смеси данным, измеренной с помощью логарифма правдоподобия $\ln L(k, p, \vartheta)$, и способностью модели смеси порождать разбиение данных, измеренной с помощью энтропии классификации $E(k, p, \vartheta)$. Поэтому основой критерия может стать классификационное правдоподобие в виде $C(k, \hat{p}, \hat{\vartheta})$. Аппроксимация интегрального классификационного правдоподобия в духе байесовского критерия (см. [6, п. 6.11.4]) приводит к следующему показателю:

$$AWE(k) = -2C(k, \hat{p}, \hat{\vartheta}) + 2v(k)(1,5 + \ln n).$$

Из комментариев в перечисленных разделах [6] и опыта автора данной работы (см., в частности, [10,

разд. 6.3]) следует, что перечисленные критерии скорее иллюстрируют наличие кластерной структуры в данных, нежели дают оценку для числа кластеров — числа элементов смеси.

Множественная проверка статистических гипотез.

Придать строгость задаче выбора числа элементов смеси можно, если сформулировать ее на языке проверки гипотез. При этом возникает набор отдельных задач проверки частных гипотез, связанных тем или иным образом с конкретным значением числа элементов смеси, и задача объединения отдельных решений в одно общее. В совокупности это приводит к проблеме множественной проверки статистических гипотез.

Проблему обоснования множественного выбора подходящего целочисленного значения для оценки \hat{k} можно заменить выбором одного из решений более простых задач сравнения двух предположений: число элементов смеси равно k_1 или k_2 , где $k_1 < k_2$. Так как при анализе данных исследователь чаще всего следует принципу минимизации размерности параметрического пространства, последнюю задачу упрощают и рассматривают случай $k_2 = k_1 + 1$. При этом путем последовательного (начиная со значения $k_1 = 1$) перебора предположений ищется такое минимальное число элементов смеси, более которого не имеет смысл усложнять модель.

Если H_k — гипотеза о том, что число элементов смеси равно k , то задачу оценивания числа элементов смеси будем решать путем анализа последовательности альтернатив A_k , включающих нулевую гипотезу H_k и конкурирующую гипотезу H_{k+1} . В качестве оценки \hat{k} возьмем такое минимальное значение k , для которого при анализе альтернатив A_1, \dots, A_{k-1} предпочтение было отдано конкурирующей гипотезе, а при анализе A_k — нулевой гипотезе.

В рамках принципа максимального правдоподобия для обоснования выбора частного решения естественно обратиться к критерию отношения правдоподобия, статистика которого удобно представить в виде:

$$s_k = -2 \ln \frac{\max_{p, \vartheta} L(k, p, \vartheta)}{\max_{p, \vartheta} L(k+1, p, \vartheta)}.$$

Тогда нулевой гипотезе будет отдаваться предпочтение при малых значениях этой статистики, а конкурирующей гипотезе — при больших.

К сожалению, даже в случае нормального распределения элементов смеси не выполняются условия, при которых статистика критерия отношения правдоподобия в подходящем виде име-

ет привычное асимптотическое χ^2 -распределение с числом степеней свободы, равным разнице в числе параметров при двух конкурирующих гипотезах. Имеются многочисленные примеры (см., например, [11]), в которых демонстрируется существенная зависимость поведения критерия отношения правдоподобия от способов реализации EM-алгоритма (например, от различных стратегий начальных приближений и правил останова итерационного процесса нахождения оценок параметров смесей). Имеющиеся попытки аппроксимации распределения статистики критерия отношения правдоподобия обосновываются общими соображениями или методом моделирования в частных случаях (в аккумулярованном виде данный материал представлен в [6, разд. 6.5]).

В создавшейся ситуации приходится прибегать к процедурам, в основе применения которых лежит использование бутстреп-метода. Рассмотрим одну из них: параметрическую процедуру оценки качества принятого решения о числе элементов смеси.

Для реализации слабого критерия значимости необходимо знать распределение $F_{S_k|H_k}$ статистики S_k критерия при условии нулевой гипотезы H_k . Тогда в терминах критического уровня значимости решения будут приниматься по величине $\alpha_k^* = 1 - F_{S_k|H_k}(s_k^*)$, где s_k^* — наблюдаемое значение статистики критерия.

Построить теоретически распределение $F_{S_k|H_k}$ вряд ли возможно, поэтому необходимо прибегнуть к бутстреп-методу. Это означает, что вместо $\alpha_k^* = 1 - F_{S_k|H_k}(s_k^*)$ рассматривается бутстреп-оценка

$$\alpha_k^B = 1 - F_{S_k^B|H_k}^B(s_k^*),$$

где $F_{S_k^B|H_k}^B$ — бутстреп-распределение.

В рассматриваемом случае бутстреп-распределение $F_{S_k^B|H_k}^B$ может быть получено только методом моделирования, схема которого при фиксированном значении k такова:

- нахождение оценок $\hat{p}(k)$ и $\hat{\vartheta}(k)$ параметров p и ϑ по исходной выборке при условии, что число элементов смеси равно k ;
- генерация нужного количества наблюдаемых значений статистики критерия отношения правдоподобия, состоящая из шага генерации выборки из смеси распределений с параметрами $\hat{p}(k)$ и $\hat{\vartheta}(k)$ и шага подсчета значения статистики s_k (включающего опять же оценивание параметров p и ϑ , но уже по сгенерированной выборке и для двух различных значений числа элементов смеси, а именно: k и $k + 1$);
- нахождение бутстреп-оценки α_k^B .

Описанный критерий отношения правдоподобия воплощал идею графических методов последовательного анализа значений функции правдоподобия. В качестве дополнительного в данной работе предлагается подход, который прямо следует из цели привлечения модели смеси распределений для описания имеющихся референсных значений. Пусть референсные значения отвечают некоторому распределению, которое аппроксимируется смесью нулевых гипотез \tilde{H}_k о том, что распределение данных есть $f_k(u)$, при конкурирующей гипотезе, что оно иное. Для этого можно использовать, в принципе, любой критерий, основанный на расстоянии между распределениями. Далее будет рассматриваться статистика χ^2 , которая контролирует согласованность гипотетических вероятностей $p_l(\tilde{H}_k)$ и эмпирических частот v_l попадания в l -ю ячейку (разряд, класс), а именно:

$$\chi^2(\tilde{H}_k) = n \sum_l \frac{(v_l - p_l(\tilde{H}_k))^2}{p_l(\tilde{H}_k)}.$$

Гипотетические вероятности $p_l(\tilde{H}_k)$ зависят от ряда параметров, характеризующих смесь из k элементов, поэтому вместо них подставляются оценки, найденные с помощью EM-алгоритма. Таким образом может быть подсчитано значение $\hat{\chi}^2(\tilde{H}_k)$. Нулевая гипотеза будет приниматься при малых значениях статистики $\hat{\chi}^2(\tilde{H}_k)$, а отвергаться при больших.

Принятие решений основывается на величине

$$\alpha(\hat{\chi}^2(\tilde{H}_k)) = 1 - F_{\chi^2(\tilde{H}_k)|\tilde{H}_k}(\hat{\chi}^2(\tilde{H}_k)).$$

При использовании бутстреп-метода вместо неизвестного распределения статистики критерия $F_{\chi^2(\tilde{H}_k)|\tilde{H}_k}(\hat{\chi}^2(\tilde{H}_k))$ рассматривается его бутстреп-аналог, строящийся путем статистических испытаний. При этом схема моделирования при фиксированном значении k такова:

- нахождение оценок $\hat{p}(k)$ и $\hat{\vartheta}(k)$ параметров p и ϑ по исходной выборке;
- генерирование бутстреп-выборки для значений статистики $\chi^2(\tilde{H}_k)$, состоящее из повторений следующих шагов: генерация выборки из смеси распределений с параметрами $\hat{p}(k)$ и $\hat{\vartheta}(k)$, подсчет по ней набора значений v_l^B , нахождение оценок $\hat{p}^B(k)$ и $\hat{\vartheta}^B(k)$ параметров p и ϑ по сгенерированной выборке, вычисление с помощью $\hat{p}^B(k)$ и $\hat{\vartheta}^B(k)$ набора $p_l^B(\tilde{H}_k)$, получение с помощью наборов v_l^B и $p_l^B(\tilde{H}_k)$ очередного значения $(\chi^2)^B(\tilde{H}_k)$;

– нахождение бутстреп-оценки $\alpha^B(\chi^2(\tilde{H}_k))$ для $\alpha(\chi^2(\tilde{H}_k))$ в виде относительной частоты наступления в ходе генерирования бутстреп-выборки события $(\chi^2)^B(\tilde{H}_k) > \chi^2(\tilde{H}_k)$.

Рассмотренные два подхода (на основе отношения правдоподобия и с использованием расстояния) к подбору числа элементов смеси реализуют множественную проверку гипотез при необходимости принять единое решение: для отдельных гипотез имеются индивидуальные критерии и проблема состоит в их комбинировании. При этом возникает следующая проблема: если для каждого отдельного критерия задать уровень значимости α , то с ростом числа гипотез вероятность одного или более ошибочного отвержения гипотез при условии их справедливости может существенно вырасти (см. [12, разд. 9]). Таким образом, говорить о том, что игнорирование множественности приводит к процедуре, контролирующей принятие решений с уровнем α , ошибочно. К сожалению, решения, предложенные в [12], не подходят для рассматриваемой задачи. Основная причина кроется в наличии зависимости между решениями частных задач о конкретном числе элементов смеси, причем эта связь имеет весьма сложный характер. В данной работе для того, чтобы получить представление о том, на что можно рассчитывать при множественной проверке гипотез, рассматривается следующая процедура принятия решения об определенном значении \hat{k} числа элементов смеси: \hat{k} таково, что для $k = 1, 2, \dots, \hat{k} - 1$ гипотеза о том, что число компонент смеси равно k , отвергается, а для $k = \hat{k}$ — принимается. Для исследования статистических свойств этой процедуры предлагается опять же обращаться к бутстреп-методу.

5 Эксперименты

Объектом анализа были данные для 363 мужчин, включающие их возраст и измерения PSA, опубликованные в [5, приложение 5.1 гл. 5].

В соответствии с общепринятой градацией были выделены возрастные группы, для которых проводилось оценивание основных характеристик референсных значений. Проверка независимости проводилась с помощью χ^2 -критерия подтверждения независимости в таблице сопряженности значений показателей «возраст» и «PSA». Далее приведены основные результаты:

1. Диапазон возрастов — меньше 50, объем группы — 68; характеристики PSA: среднее = 1,0; стандартное отклонение (CO) = 1,0; минимальное зна-

чение = 0,2; максимальное значение = 6,3; количество различных = 23;

критический уровень значимости для проверки независимости «возраст» и «PSA» = 79%.

2. Диапазон возрастов — [50, 60), объем группы — 114;

характеристики PSA: среднее = 1,2; CO = 1,1; минимальное значение = 0,2; максимальное значение = 6,7; количество различных = 30;

критический уровень значимости для проверки независимости «возраст» и «PSA» = 49%.

3. Диапазон возрастов — [60, 70), объем группы — 126;

характеристики PSA: среднее = 2,0; CO = 1,8; минимальное значение = 0,2; максимальное значение = 11,1; количество различных = 45;

критический уровень значимости для проверки независимости «возраст» и «PSA» = 2%.

4. Диапазон возрастов — не меньше 70, объем группы — 55;

характеристики PSA: среднее = 2,2; CO = 1,9; минимальное значение = 0,2; максимальное значение = 7,9; количество различных = 33;

критический уровень значимости для проверки независимости «возраст» и «PSA» = 76%.

Для сравнения характеристики PSA по всей совокупности данных следующие: среднее = 1,6; CO = 1,6; минимальное значение = 0,2; максимальное значение = 11,1; количество различных = 56; критический уровень значимости для проверки независимости «возраст» и «PSA» = 0%.

В приведенных результатах обращают на себя внимание два момента: высокая дискретность исходных данных (среди 363 значений PSA различных всего 56), подтверждение общеизвестного факта зависимости значений PSA от возраста.

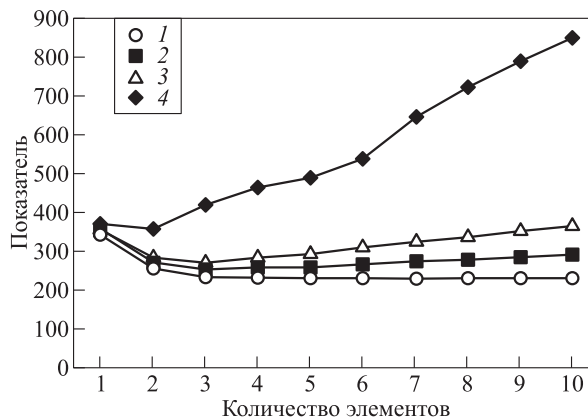


Рис. 1 Зависимость информационных показателей от числа элементов смеси: 1 — $-2 \ln L(k)$; 2 — AIK; 3 — BIC; 4 — AWE

Таблица 1 Результаты множественной проверки гипотез на основе критерия отношения правдоподобия

k	$\max_{p, \vartheta} \ln L(k, p, \vartheta)$	s_k	$\alpha_k^B, \%$	$\text{Ave}\{s_k\}$	$E\{S_k\}$ ожидаемое
1	-172,84	85,08	0	3,29	3,00
2	-130,30	23,98	0	2,95	3,00
3	-118,31	3,31	23	2,12	3,00
4	-116,65	1,16	45	1,56	3,00
5	-116,08	0,02	29	0,15	3,00
6	-116,07	-0,00	93	0,03	3,00
7	-116,07	-0,00	91	-0,00	3,00
8	-116,07	-0,00	90	0,00	3,00
9	-116,07	-0,00	85	-0,01	3,00
10	-116,07	-0,00	90	-0,00	3,00

Для экспериментов с моделированием распределения одномерных данных о PSA была выбрана 2-я возрастная группа, достаточно представительная по объему и без признаков зависимости показателей «возраст» и «PSA».

Возможности подбора числа элементов смеси с помощью графических методов и критериев информационного типа продемонстрированы на рис. 1:

- минимумы для AIC ($k = 3$), BIC ($k = 3$) и AWE ($k = 2$) выражены слабо;
- принимаемое решение о \hat{k} не подтверждается какими-либо количественными характеристиками его качества.

Множественная проверка гипотез на основе критерия отношения правдоподобия дала результаты, представленные в табл. 1. В ней для $k = 1, 2, \dots, 10$ приведены значения α_k^B и средние значения $\text{Ave}\{s_k\}$ статистики S_k при условии нулевой гипотезы H_k , которые оценивались с помощью 10^3 экспериментов. В последнем столбце приведены ожидаемые значения $E\{S_k\}$, полученные из предположения, что асимптотическое для статистики отношение правдоподобия χ^2 -распределение имеет число степеней свободы, равное разнице в числе параметров при двух конкурирующих гипотезах.

Из полученных результатов можно сделать следующие выводы:

- оценкой для числа элементов смеси может быть значение 3, причем даже для малых уровней значимости в 1% этой оценке соответствует отвержение нулевой гипотезы о том, что $k = 2$, в пользу конкурирующей, что $k = 3$, но принятие нулевой гипотезы о том, что $k = 3$, при конкурирующей, что $k = 4$;
- сравнение двух последних столбцов не оставляет надежды на использование асимптотических

результатов о распределении статистики критерия отношения правдоподобия.

Для того чтобы представить, на что можно рассчитывать при множественной проверке гипотез, задаваясь конкретным уровнем значимости α в частном случае, анализировалась следующая процедура. Пусть при последовательной проверке частных гипотез ($k = 1, 2, \dots$) число первых отвергнутых нулевых гипотез равно N_{H_0} , а гипотеза с индексом $N_{H_0} + 1$ принимается. В качестве оценки для числа элементов смеси примем получившееся N_{H_0} . Если исходная гипотеза заключается в том, что $\hat{k} = 3$, то интересно связать вероятность $\tilde{\alpha}$ отвергнуть эту гипотезу, если она верна, с конкретным значением α частного критерия. Для этого было проведено 10^3 бутстреп-экспериментов и оценены значения $\tilde{\alpha}$ (табл. 2).

Заметим, что умеренные значения α частного критерия порождают практически те же значения для множественной проверки гипотез, а уменьшение α ($\alpha = 2,5$ и $1,0$) приводит, вообще говоря, к росту $\tilde{\alpha}$.

Высокое качество принимаемых решений относительно числа элементов смеси подтверждается данными табл. 3, где приведены выборочные характеристики оценки $\hat{k} = N_{H_0}$.

Таблица 2 Результаты бутстреп-оценки связи уровня значимости частной и множественной проверки гипотез на основе критерия отношения правдоподобия

$\alpha, \%$	$\tilde{\alpha}, \%$
1,0	5,9
2,5	4,2
5,0	5,4
10,0	10,3
20,0	20,4

Таблица 3 Результаты бутстреп-оценки моментов оценки числа элементов смеси на основе критерия отношения правдоподобия

Моменты	$\alpha, \%$				
	1,0	2,5	5,0	10,0	20,0
Среднее	2,95	2,99	3,04	3,10	3,23
CO	0,06	0,04	0,06	0,11	0,24

Таблица 4 Результаты бутстреп-оценки моментов оценки числа элементов смеси на основе расстояния χ^2 между распределениями

Моменты	$\alpha, \%$				
	1,0	2,5	5,0	10,0	20,0
Среднее	2,20	2,35	2,49	2,67	2,92
CO	0,24	0,29	0,42	0,56	0,93

Описанные выше действия по исследованию выборочных свойств оценки числа элементов смеси были повторены для критерия, основанного на расстоянии χ^2 между распределениями. Эксперименты по множественной проверке гипотез на основе указанного критерия позволяют сделать следующие выводы:

- оценкой для числа элементов смеси опять же может быть значение 3, так как нулевая гипотеза о числе элементов смеси, равном 2, отвергается с двухпроцентным уровнем значимости, но уже для $k = 3$ нулевая гипотеза принимается при критическом уровне значимости в 62%;
- асимптотические результаты о распределении статистики критерия не применимы.

Выводы о свойствах описанной процедуры множественной проверки гипотез при заданном уровне значимости частного критерия несколько иные:

- значения α частного критерия порождают большие значения вероятности $\tilde{\alpha}$ (так $\tilde{\alpha} = 78\%$ для $\alpha = 1\%$; $\tilde{\alpha} = 66\%$ для $\alpha = 2,5\%$; $\tilde{\alpha} = 56\%$ для $\alpha = 5\%$; $\tilde{\alpha} = 43\%$ для $\alpha = 10\%$; $\tilde{\alpha} = 33\%$ для $\alpha = 20\%$);
- качество оценки $\hat{k} = N_{\overline{H_0}}$ снижается (см. табл. 4 для критерия, основанного на расстоянии между распределениями, и табл. 3).

Таким образом, данные о PSA для мужчин в возрасте от 50 до 60 лет могут быть описаны с помощью смеси трех нормальных распределений $\varphi(u, (\mu, \sigma^2))$, а именно:

$$f(u) = 0,459\varphi(u, (0,527, 0,050)) + 0,446\varphi(u, (1,307, 0,252)) + 0,095\varphi(u, (3,829, 2,211)) .$$

Проведенный анализ относится к определенному возрастному диапазону, выбранному достаточно произвольно. Последнее можно исключить, обобщая задачу анализа данных за счет перехода от одномерного случая (описание условного распределения показателя «PSA» при определенных значениях «возраст») к двумерному (описание совместного распределения показателей «PSA» и «возраст»). Теперь объектом анализа совершенно обоснованно может стать вся совокупность исходных данных.

Подбор числа элементов смеси с помощью графических методов и информационных критериев не дал четких результатов, в частности минимум АИК достигается при $k = 5$, минимум BIC — при $k = 3$, минимум AWE — при $k = 2$. Множественная проверка гипотез на основе критерия отношения правдоподобия привела к оценке $\hat{k} = 4$ (для уровня значимости в 2% этой оценке соответствует отвержение нулевой гипотезы о том, что $k = 3$, в пользу конкурирующей, что $k = 4$, но принятие нулевой гипотезы о том, что $k = 4$, при конкурирующей, что $k = 5$). Анализ выборочных распределений с помощью статистики χ^2 не проводился, так как соответствующий критерий требует адаптации к условию многомерных данных.

Таким образом, совместное распределение показателей «PSA» (соответствует u_1) и «возраст» (соответствует u_2) может быть описано с помощью смеси четырех нормальных распределений $\varphi(x_1, x_2, (\mu_{j1}, \mu_{j2}, \sigma_{j1}^2, \sigma_{j2}^2, \rho_j))$, а именно:

$$f(u) = 0,447\varphi(u_1, u_2, (0,591, 55,800, 0,068, 83,600, 0,147)) + 0,345\varphi(u_1, u_2, (1,410, 60,400, 0,286, 89,600, 0,126)) + 0,187\varphi(u_1, u_2, (3,750, 63,400, 2,370, 68,900, 0,211)) + 0,022\varphi(u_1, u_2, (5,520, 72,400, 12,100, 25,000, -0,989)).$$

Отсюда находятся условные распределения, с помощью которых можно, в частности, получить референсные интервалы для отдельных значений возраста.

На рис. 2 приведены графики оценок правых 95%-ных границ референсных интервалов как функции возраста: 1 — построенный с помощью условных распределений и 2 — на основе рекомендаций British Association of Urological Surgeons (BAUS). Заметим, что рекомендации в зависимости от источника могут отличаться (см. для сравнения упомянутый [13] и, например, [14]). Важно, чтостораживающих отличий от рекомендаций не замечено. При этом оценки на основе условных распределений имеют явные преимущества: высокую детализацию по возрастам, сглаживание результатов наблюдений для различных по объему возрастных групп, возможность формировать предпо-

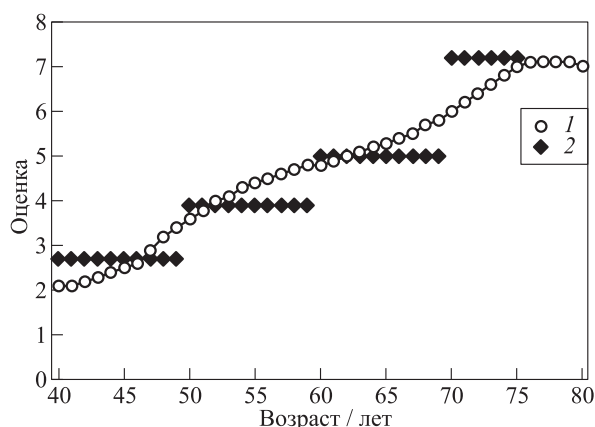


Рис. 2 Зависимость правых 95%-ных референсных границ от возраста, найденных с помощью оценок условных распределений (1) и рекомендованных BAUS (2)

жения о характере зависимости между возрастом и уровнем PSA.

6 Заключение

Модель смеси нормальных распределений может эффективно использоваться как средство аппроксимации реальных данных и при этом быть доступной с точки зрения теоретического анализа. Для оценивания параметров модели могут применяться различные методы: графические, метод моментов, максимального правдоподобия, методы функции ошибок (методы минимизации расстояния), байесовские. Среди них метод максимального правдоподобия и его воплощение в виде EM-алгоритма занимают главенствующее место, потому что позволяют решать следующие вопросы:

- не привлекать дополнительной априорной информации к уже не простой постановке задачи описания данных с помощью параметрической модели;
- автоматически и достаточно эффективно решать задачу оценивания параметров в широком диапазоне значений числа элементов смеси.

Очевидным кандидатом для проверки гипотез, связанных с моделью смеси, является критерий отношения правдоподобия. Но в реальной обстановке не очень больших объемов данных возникают проблемы в описании распределения статистики

критерия хотя бы в условиях одной из принимаемых гипотез. В связи с этим приходится рассчитывать только на бутстеп-метод, что предъявляет повышенные требования к эффективности реализации EM-алгоритма.

Литература

1. Sokal R. R., Sneath P. H. A. Principles of numerical taxonomy. — San Francisco — London: W. H. Freeman and Co., 1963. 359 p.
2. Burtis C. A., Bruns D. E. Tietz fundamentals of clinical chemistry and molecular diagnostics. — St. Louis, MO, USA: Elsevier Health Sciences, 2014. 1104 p.
3. Sivkov A., Keshishev N., Krivenko M., Kovchenko G., Nikonova L. Comparison of chromogranin-A levels determined by different test systems in patients with prostate diseases // Eur. Urol. Suppl., 2014. Vol. 13. No. 5. P. 143.
4. Кривенко М. П. Сравнительный анализ процедур регрессионного анализа // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 70–78.
5. Harris E. K., Boyd J. C. Statistical bases of reference values in laboratory medicine. — New York, NY, USA: Marcel Dekker, 1995. 361 p.
6. McLachlan G., Peel D. Finite mixture models. — New York, NY, USA: Wiley & Sons, 2000. 456 p.
7. Королев В. Ю. EM-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений: Теоретический обзор. — М.: ИПИ РАН, 2007. 94 с.
8. McLachlan G., Krishnan T. The EM algorithm and extensions. — Hoboken, NJ, USA: Wiley & Sons, 2008. 359 p.
9. Кривенко М. П. Прикладные методы оценивания распределения многомерных данных малой выборки. — М.: ИПИ РАН, 2011. 146 с.
10. Васильев В. Г., Кривенко М. П. Методы автоматизированной обработки текстов. — М.: ИПИ РАН, 2008. 305 с.
11. Seidel W., Mosler K., Alker M. A cautionary note on likelihood ratio tests in mixture models // Ann. Inst. Statist. Math., 2000. Vol. 52. No. 3. P. 481–487.
12. Lehmann E. L., Romano J. P. Testing statistical hypotheses. — New York, NY, USA: Springer, 2005. 792 p.
13. PSA Measurements. — British Association of Urological Surgeons (BAUS). March 2014. <http://www.baus.org.uk>.
14. Brosnan S. A. Prostate-specific antigen testing // Medscape, January 13, 2015. <http://emedicine.medscape.com/article/457394-overview>.

Поступила в редакцию 19.02.15

MODELS FOR REPRESENTATION AND TREATMENT OF REFERENCE VALUES

M. P. Krivenko

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

Abstract: The article considers the problem of modeling reference values — results of a certain type of quantities obtained from a single individual or a group of individuals corresponding to a stated description. For this purpose, the article proposes to use a mixture of normal distributions, which can effectively serve as a means of approximating the actual data and to be accessible from the standpoint of theoretical analysis. In estimating the parameters of mixture of distributions, the major role is played by the maximum likelihood method and its embodiment in the form of the expectation-maximization (EM) algorithm. For assessing the number of mixture components, the article suggests to use the likelihood ratio test and a method based on the chi-square distance between the distributions. Their properties are investigated using the bootstrap method. As an experiment, the article considers the description of the empirical distribution of patient data, including the age and measurements of PSA (Prostate-Specific Antigen). The proposed solutions have clear advantages: high detail by age, smoothing the results of observations for age groups which are different in size, and the opportunity to form assumptions about the nature of the relationship between age and PSA.

Keywords: mixture of normal distributions; assessing the number of components in mixture models; reference values

DOI: 10.14357/19922264150208

References

1. Sokal, R. R., and P. H. A. Sneath. 1963. *Principles of numerical taxonomy*. San Francisco — London: W. H. Freeman and Co. 359 p.
2. Burtis, C. A., and D. E. Bruns. 2014. *Tietz fundamentals of clinical chemistry and molecular diagnostics*. St. Louis, MO: Elsevier Health Sciences. 1104 p.
3. Sivkov, A., N. Keshishev, M. Krivenko, G. Kovchenko, and L. Nikonova. 2014. Comparison of chromogranin-A levels determined by different test systems in patients with prostate diseases. *Eur. Urol. Suppl.* 13(5):143.
4. Krivenko, M. P. 2014. Sravnitel'nyy analiz protsedur regressiionnogo analiza [Comparative analysis of regression analysis procedures]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):70–78.
5. Harris, E. K., and J. C. Boyd. 1995. *Statistical bases of reference values in laboratory medicine*. New York, NY: Marcel Dekker. 361 p.
6. McLachlan, G., and D. Peel. 2000. *Finite mixture models*. New York, NY: Wiley & Sons. 456 p.
7. Korolev, V. Yu. 2007. *EM-algoritm, ego modifikatsii i ikh primeneniye k zadache razdeleniya smesey veroyatnostnykh raspredeleniy. Teoreticheskiy obzor* [EM-algorithm. Modifications and their application to the separation of mixtures of probability distributions. Theoretical review]. Moscow: IPI RAN. 94 p.
8. McLachlan, G., and T. Krishnan. 2008. *The EM algorithm and extensions*. Hoboken, NJ: Wiley & Sons. 359 p.
9. Krivenko, M. P. 2011. *Prikladnye metody otsenivaniya raspredeleniya mnogomernykh dannykh maloy vyborki* [Applied methods for estimating the distribution of multi-dimensional data of small sample size]. Moscow: IPI RAN. 146 p.
10. Vasil'ev, V. G., and M. P. Krivenko. 2008. *Metody avtomatizirovannoy obrabotki tekstov* [Methods for automatic text processing]. M.: IPI RAN. 305 p.
11. Seidel, W., K. Mosler, and M. Alker. 2000. A cautionary note on likelihood ratio tests in mixture models. *Ann. Inst. Statist. Math.* 52(3):481–487.
12. Lehmann, E. L., and J. P. Romano. 2005. *Testing statistical hypotheses*. New York, NY: Springer. 792 p.
13. PSA Measurements. British Association of Urological Surgeons (BAUS). March 2014. Available at: <http://www.baus.org.uk> (accessed February 11, 2015).
14. Brosman, S. A. January 13, 2015 (updated). Age-specific PSA reference ranges. 2015. Available at: <http://emedicine.medscape.com/article/457394-overview> (accessed February 11, 2015).

Received February 19, 2015

Contributor

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

СОГЛАСОВАНИЕ ПРОГНОЗОВ ПРИ РЕШЕНИИ ЗАДАЧ ПРОГНОЗИРОВАНИЯ ИЕРАРХИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ*

М. М. Стенина¹, В. В. Стрижов²

Аннотация: Рассматривается задача одновременного прогнозирования набора временных рядов, объединенных в иерархическую многоуровневую структуру. Требуется, чтобы полученные прогнозы удовлетворяли физическим ограничениям и структуре иерархии. Предложен алгоритм согласования прогнозов иерархических временных рядов GTOp (Game-theoretically optimal reconciliation), гарантирующий неухудшение качества прогнозов после проведения процедуры согласования по сравнению с качеством прогнозов, полученных для каждого временного ряда независимо. Подход базируется на поиске равновесия Нэша в антагонистической игре заданного вида и сводит задачу согласования прогнозов к задаче оптимизации с ограничениями типа равенства и неравенства. Доказывается, что при выполнении ряда общих предположений о свойствах структуры иерархии, физических ограничений и функции потерь в игре существует равновесие Нэша в чистых стратегиях. Работа алгоритма демонстрируется на разных типах иерархических структур с использованием данных посуточной загруженности железнодорожных узлов.

Ключевые слова: иерархические временные ряды; согласование прогнозов временных рядов; антагонистическая игра; равновесие Нэша

DOI: 10.14357/19922264150209

1 Введение

Рассматривается задача одновременного прогнозирования набора временных рядов, связанных в иерархическую многоуровневую структуру, в которой временные ряды каждого следующего (более высокого) уровня формируются путем поэлементного суммирования некоторой части (возможно, всех) временных рядов предыдущего уровня. Используемые при решении этой задачи принцип равновесия и регрессионные методы широко обсуждаются в научной литературе, в том числе на страницах журнала «Информатика и её применения» [1, 2].

Задача прогнозирования иерархических временных рядов возникает в различных прикладных областях. В [3] описан конкурс, проведенный в 2012 г. на Kaggle [4], в котором одной из задач было прогнозирование иерархических временных рядов с требованием согласования прогнозов по иерархии. В работе [5] обсуждается задача прогнозирования туристической активности по регионам и целям поездов. В статье [6] решается задача прогнозирования потребительского спроса на различные группы товаров в ряде магазинов. В [7]

решается задача согласования прогнозов объемов железнодорожных перевозок различных типов грузов по ряду железнодорожных веток.

В связи с высоким интересом к задаче разными авторами предлагаются различные подходы к ее решению. Как правило, сперва получают каким-либо образом прогнозы всех (или некоторой части) временных рядов независимо друг от друга, а затем корректируют (согласуют) эти прогнозы.

Самыми простыми и самыми распространенными способами согласования являются нисходящий (top-down) и восходящий (bottom-up) подходы [8–15]. Нисходящий подход предполагает получение прогноза на верхнем уровне иерархии (агрегированный временной ряд), а затем деагрегирование этого прогноза на следующий (более низкий) уровень иерархии на основании долей, наблюдаемых в истории. Восходящий подход использует прогнозы временных рядов нижнего уровня иерархии (неагрегированных), из которых получает прогнозы рядов из верхних уровней путем агрегирования. Также встречаются подходы, комбинирующие нисходящий и восходящий.

Нет единой точки зрения на то, какой из этих подходов позволяет получать более точные про-

* Работа выполнена при финансовой поддержке РФФИ (проект 13-07-13139).

¹Московский физико-технический институт, Национальный исследовательский университет «Высшая школа экономики», mmedvednikova@gmail.com

²Вычислительный центр Российской академии наук им. А. А. Дородницына, strijov@ccas.com

гнозы. Наиболее ранние исследования проведены в работе [8], где авторы считают, что агрегированные данные содержат много ошибок и поэтому нисходящее прогнозирование дает более точные прогнозы. К таким же выводам приходят авторы работ [12, 13]. В [15] также утверждается, что агрегированные прогнозы более точны. С другой стороны, в [9, 10] обсуждается, что основные потери информации происходят при агрегировании и поэтому восходящий подход предпочтительнее. В [11] сравниваются оба подхода к согласованию прогнозов и утверждается, что восходящий предпочтительнее при выполнении некоторых условий на структуру иерархии и горизонт прогноза. В [14] исследуется смещение и устойчивость прогнозов, получаемых с помощью обоих подходов, и заключается, что восходящий надежнее, за исключением случаев с пропусками значений и выбросами на нижних уровнях иерархии.

Авторы статьи [5] обобщают нисходящий и восходящий подходы к согласованию иерархических прогнозов и предлагают оптимальное согласование с использованием регрессии, позволяющей согласовывать одновременно по всем уровням иерархии любой сложности. Однако предложенный ими алгоритм согласования предполагает, что полученные независимые прогнозы являются несмещенными оценками и что ошибки прогнозов временных рядов удовлетворяют структуре иерархии, т. е. являются согласованными. Оба эти требования являются достаточно строгими и сильно ограничивают применимость этого способа согласования.

Предложенный в [7] способ согласования прогнозов не требует несмещенности независимых прогнозов и согласованности ошибок. При этом, как продемонстрировано в статье, он не уступает методу из [5] по качеству согласованных прогнозов. Однако и этот метод имеет ряд недостатков. Для его использования необходимо оценивать погрешность независимых прогнозов, что не всегда удается корректно сделать. Метод не гарантирует, что качество согласованных прогнозов не будет уступать качеству независимых. И согласование происходит поэтапно по узлам иерархии, что не позволяет учесть сразу всю информацию о взаимосвязи между временными рядами.

В настоящей статье предлагается обобщение алгоритма согласования из [7] с сохранением его преимуществ и устранением недостатков. Предлагается алгоритм согласования прогнозов ГТОр, основанный на идеях из [16]. Алгоритм ГТОр не требует оценки погрешности независимых прогнозов временных рядов, не требует несмещенности независимых прогнозов и имеет теоретическое обоснование улучшения качества прогнозов

после проведения согласования. Задача согласования прогнозов рассматривается как поиск равновесия Нэша [17, 18] в антагонистической игре игрока, выбирающего согласованные прогнозы, с природой, которая выбирает действительные значения временных рядов, и сводится к решению оптимизационной задачи с ограничениями типа равенства и неравенства. Вид равновесия Нэша задает параметры оптимизационной задачи.

Работа предлагаемого алгоритма демонстрируется на данных о посуточном отправлении вагонов с 37 типами грузов с 98 веток Российских железных дорог (РЖД). Проведены эксперименты по согласованию прогнозов для различных типов иерархических структур и показано, что на практике действительно наблюдается улучшение качества прогнозов после проведения согласования алгоритмом ГТОр.

Статья включает следующие разделы. В разд. 2 вводятся необходимые обозначения, в разд. 3 формулируется задача согласования прогнозов иерархических временных рядов. Раздел 4 содержит необходимые определения и факты из теории игр, описание антагонистической игры, соответствующей задаче согласования прогнозов, и доказательство существования в этой игре равновесия Нэша. Раздел 5 описывает оптимизационную задачу, к которой сводится задача согласования прогнозов, а также преимущества и недостатки алгоритма. В разд. 6 приводятся примеры функций потерь, для которых применим алгоритм ГТОр. В разд. 7 приведены результаты экспериментов. В разд. 8 подводятся итоги и делаются выводы.

2 Обозначения

В этом разделе вводится система обозначений, которая будет использоваться в настоящей работе. Будем обозначать временной ряд через вектор \mathbf{x} , элементы временного ряда будем снабжать индексом t , $t = 1, \dots, T$, где T — длина истории.

$$\mathbf{x} = \{x_t\}_{t=1}^T.$$

Общее количество временных рядов во всей иерархии будем обозначать d . Далее для наглядности будем рассматривать иерархию рядов, изображенную на рис. 1. Она содержит один временной ряд на верхнем уровне и n рядов на нижнем. Для этой иерархии $d = 1 + n$.

Временные ряды нижнего уровня будут обозначаться $\mathbf{x}(i)$, $i = 1, \dots, n$, где n — число временных рядов на нижнем уровне. Временной ряд верхнего уровня обозначается как $\mathbf{x}(\cdot)$. Во избежание

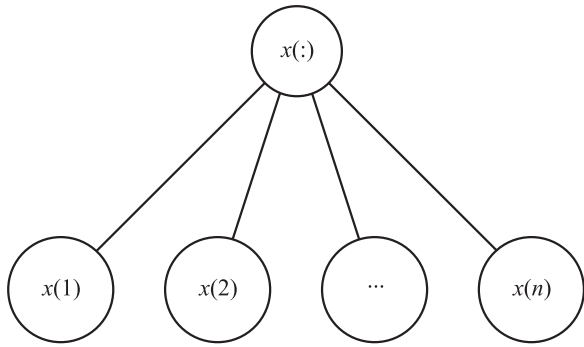


Рис. 1 Плоская двухуровневая иерархия

путаницы условимся использовать нижние индексы для обозначения отсчетов времени, а индексы в скобках использовать для обозначения положения временных рядов в структуре иерархии. Элементы временных рядов, составляющих плоскую двухуровневую иерархию, обозначаются соответственно $x_t(i)$, $i = 1, \dots, n$, $x_t(:)$. Их соотношение задается формулой

$$x_t(:) = \sum_{i=1}^n x_t(i), \quad t = 1, \dots, T. \quad (1)$$

Будем называть соотношение (1) **условием согласованности**. Прогнозы этих временных рядов будем обозначать «шляпками», опуская нижние индексы, чтобы избежать излишне громоздких обозначений. Прогнозироваться будет всегда $(T + 1)$ -е значение временного ряда $\hat{x}(i)$, $i = 1, \dots, n$, $\hat{x}(:)$. Согласованные прогнозы будут также обозначаться без нижних индексов: $\hat{y}(i)$, $i = 1, \dots, n$, $\hat{y}(:) = \sum_{i=1}^n \hat{y}(i)$.

Запишем все временные ряды в матрицу, каждая строка которой соответствует одному временному ряду. Для иерархии с рис. 1 эта матрица будет размера $(1 + n) \times T$ и выглядеть следующим образом:

$$X = \begin{pmatrix} x_1(:) & x_2(:) & \cdots & x_T(:) \\ x_1(1) & x_2(1) & \cdots & x_T(1) \\ \cdots & \cdots & \cdots & \cdots \\ x_1(n) & x_2(n) & \cdots & x_T(n) \end{pmatrix}.$$

Будем называть срезом иерархии в момент времени t столбец матрицы X , соответствующий моменту времени t . Для удобства записи введем векторы, соответствующие срезу иерархии в момент времени t , прогнозам и согласованным прогнозам. В этих векторах значения, соответствующие разным временным рядам, записаны в столбец, начиная с верхнего уровня иерархии и заканчивая нижним уровнем:

$$\chi_t = \begin{pmatrix} x_t(:) \\ x_t(1) \\ \vdots \\ x_t(n) \end{pmatrix}; \quad \hat{\chi}_t = \begin{pmatrix} \hat{x}(:) \\ \hat{x}(1) \\ \vdots \\ \hat{x}(n) \end{pmatrix}; \quad \hat{\varphi}_t = \begin{pmatrix} \hat{y}(:) \\ \hat{y}(1) \\ \vdots \\ \hat{y}(n) \end{pmatrix}.$$

Условие (1) для векторов χ_t и $\hat{\varphi}_t$ запишем, введя матрицу связей размером $1 \times (n + 1)$:

$$S = (-1 \ 1 \ \cdots \ 1).$$

Тогда условие согласованности запишется кратко

$$S\chi_t = 0; \quad S\hat{\varphi}_t = 0.$$

В случае, когда иерархия имеет более сложную структуру, чем на рис. 1, векторы χ_t , $\hat{\chi}_t$ и $\hat{\varphi}_t$ имеют размерность d , матрица X имеет ровно d строк, временные ряды в ней записываются от верхних уровней к нижним. А размерность матрицы связи S равна $c \times d$, где c — число узлов в графе иерархии или, другими словами, количество ограничений-равенств, наложенных на элементы срезов иерархии χ_t .

3 Задача согласования прогнозов иерархических временных рядов

Сформулируем задачу согласования прогнозов иерархических временных рядов. Для этого введем еще ряд необходимых обозначений.

Пусть дан набор из d временных рядов, значения которых записаны в матрицу X размера $d \times T$:

$$X = (\chi_1 \ \chi_2 \ \cdots \ \chi_T),$$

где каждый столбец χ_t соответствует срезу в момент времени t , а каждая строка x_i — одному временному ряду. Пусть структура иерархии задана матрицей связи S так, что для всех $t = 1, \dots, T$ выполнено условие согласованности:

$$S\chi_t = 0.$$

Пусть даны прогнозы значений $\hat{\chi}_t$ для всех временных рядов в момент времени $T + 1$ и задана функция суммарных потерь при прогнозировании иерархии

$$l_h(\hat{\chi}_T, \chi_{T+1}). \quad (2)$$

Определим множество

$$\mathcal{A} = \{\chi \in \mathbb{R}^d \mid S\chi = 0\}, \quad (3)$$

где χ — произвольный d -мерный вектор, а S — заданная матрица связи. Отметим, что все срезы χ_t , $t = 1, \dots, T$, лежат в множестве \mathcal{A} . Также в нем должны лежать согласованные прогнозы $\hat{\varphi}_t$.

В ряде задач прогнозы должны удовлетворять некоторым ограничениям, связанным с физической природой прогнозируемой величины. В связи с чем введем множество

$$\mathcal{B} = \{\chi \in \mathbb{R}^d \mid \chi(i) \in [A_i, B_i], \\ A_i, B_i \in [-\infty, +\infty], i = 1, \dots, d\}, \quad (4)$$

где χ — произвольный d -мерный вектор; A_i и B_i задают отрезок, в котором должна находиться i -я компонента этого вектора. Например, в случае $A_1 = \dots = A_d = 0, B_1 = \dots = B_d = +\infty$ вектор χ лежит в положительном октанте, а конечные значения $A_1, \dots, A_d, B_1, \dots, B_d$ могут задавать интервалы, в которых должны лежать прогнозы. Отсутствие каких-либо ограничений задается значениями $A_1 = \dots = A_d = -\infty, B_1 = \dots = B_d = +\infty$.

Введя все необходимые дополнительные обозначения, можно сформулировать задачу поиска согласованных прогнозов. Требуется найти вектор прогнозов $\hat{\varphi}$, удовлетворяющий следующим требованиям.

Согласованность: вектор прогнозов $\hat{\varphi}$ должен удовлетворять структуре иерархии, заданной матрицей связи S , т. е. $\hat{\varphi} \in \mathcal{A}$.

Ограничения: вектор прогнозов $\hat{\varphi}$ должен удовлетворять наложенным ограничениям, т. е. $\hat{\varphi} \in \mathcal{B}$.

Качество: общие потери при использовании согласованных прогнозов не должны превышать общие потери при использовании независимых прогнозов, т. е. $l_h(\hat{\varphi}, \chi_{T+1}) \leq l_h(\hat{\chi}, \chi_{T+1})$.

4 Задача согласования прогнозов как поиск равновесия в антагонистической игре

В этом разделе задача согласования прогнозов, сформулированная выше, рассматривается как антагонистическая игра. Такое представление не влияет на решение задачи согласования прогнозов и направлено лишь на достижение наглядности и интерпретируемости полученных результатов. В первой части раздела приводятся необходимые определения и факты из теории игр [17–19], во второй части вводится антагонистическая игра, соответствующая задаче согласования прогнозов, в третьей части формулируется и доказывается теорема о существовании в этой игре равновесия Нэша, которая является обобщением теоремы, доказанной в [16], и приводится следствие из этой теоремы, однозначно определяющее выбор оптимального вектора согласованных прогнозов.

4.1 Понятие антагонистической игры

Определение 1. Система

$$\Gamma = (\mathcal{M}, \mathcal{N}, L),$$

где \mathcal{M} и \mathcal{N} — непустые множества и функция $L: \mathcal{M} \times \mathcal{N} \rightarrow \mathbb{R}$, называется антагонистической игрой (игрой с нулевой суммой) в нормальной форме. Элементы $\mu \in \mathcal{M}$ и $\nu \in \mathcal{N}$ называются стратегиями игроков 1 и 2 соответственно, L — функцией потерь игрока 1. Потери игрока 2 полагаются равными $-L(\mu, \nu)$.

Определение 2. Говорят, что игра разыгрывается в чистых стратегиях, если оба игрока из имеющихся наборов действий \mathcal{M} и \mathcal{N} выбирают по одному действию μ и ν соответственно.

Определение 3. Введем на множествах стратегий \mathcal{M} и \mathcal{N} вероятностные распределения $p(\mu)$ и $q(\nu)$ соответственно:

$$\int_{\mathcal{M}} p(\mu) d\mu = 1; \quad \int_{\mathcal{N}} q(\nu) d\nu = 1.$$

Распределения $p(\mu)$ и $q(\nu)$ задают смешанные стратегии в игре Γ , если игрок 1 выбирает действие в соответствии с распределением $p(\mu)$ и игрок 2 выбирает действие в соответствии с распределением $q(\nu)$.

Чистые стратегии являются частным случаем смешанных. Поэтому далее будут рассматриваться смешанные стратегии, за исключением специально оговоренных моментов. Обозначать стратегии будем $p(\cdot)$ и $q(\cdot)$.

Математическое ожидание потерь игрока 1 при паре смешанных стратегий $p(\cdot)$ и $q(\cdot)$ обозначим через

$$\bar{L}(p(\cdot), q(\cdot)) = \int_{\mathcal{M}} \int_{\mathcal{N}} p(\mu)q(\nu)L(\mu, \nu) d\mu d\nu.$$

Игрок 1 преследует цель минимизировать эту величину при любых действиях игрока 2.

Определение 4. Пара стратегий $(p(\cdot), q(\cdot))$ называется равновесием Нэша в смешанных стратегиях в игре Γ , если для любых $p(\cdot)'$ и $q(\cdot)'$ выполнено неравенство:

$$\bar{L}(p(\cdot), q(\cdot)') \leq \bar{L}(p(\cdot), q(\cdot)) \leq \bar{L}(p(\cdot)', q(\cdot)).$$

При равновесии Нэша ни одному из игроков не выгодно отклоняться от равновесной стратегии, если второй продолжает придерживаться равновесной стратегии. При этом игрок 1 минимизирует свои потери в ситуации, когда игрок 2 действует наиболее выгодным для себя образом. Отметим

также, что равновесие Нэша является седловой точкой функции $\bar{L}(p(\cdot), q(\cdot))$.

Теорема 1. *В антагонистической игре равновесие Нэша существует тогда и только тогда, когда определена величина*

$$V = \min_{p(\cdot)'} \max_{q(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)') = \max_{q(\cdot)'} \min_{p(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)').$$

Величина V называется ценой игры. Доказательство этой теоремы в настоящей статье не приводится, при желании его можно найти в [17–19].

4.2 Антагонистическая игра, описывающая задачу согласования прогнозов

Вернемся к рассмотрению введенных в разд. 3 множеств \mathcal{A} (3) и \mathcal{B} (4). Напомним, что множество \mathcal{A} содержит все d -мерные векторы, удовлетворяющие структуре иерархии, заданной матрицей связи S . Множество \mathcal{B} содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам рассматриваемой задачи прогнозирования.

Будем рассматривать антагонистическую игру Γ , в которой игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$, которые одновременно удовлетворяют структуре иерархии, заданной матрицей связи S , и ограничениям-неравенствам, задающим множество \mathcal{B} , игрок 2 также выбирает вектор действительных значений χ_{T+1} , удовлетворяющих структуре иерархии и физическим ограничениям (можно считать игрока 2 природой).

Определим множества стратегий \mathcal{M} и \mathcal{N} игроков 1 и 2 как пересечение множеств $\mathcal{A} \cap \mathcal{B}$. Функцию потерь игрока 1 определим с помощью функции потерь по иерархии (2):

$$L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi}),$$

где вектор независимых прогнозов $\hat{\chi}$ считается заданным и не зависит от действий, выбираемых игроками. Считается, что $\hat{\chi} \in \mathcal{B}$. Такой выбор функции потерь игрока 1 связан с тем соображением, что при $L(\hat{\varphi}, \chi_{T+1}) = 0$ качество согласованных прогнозов не хуже, чем качество независимых прогнозов, а при $L(\hat{\varphi}, \chi_{T+1}) < 0$ и вовсе превосходит его.

Таким образом, получена антагонистическая игра

$$\Gamma = (\mathcal{A} \cap \mathcal{B}, \mathcal{A} \cap \mathcal{B}, l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})), \quad (5)$$

где игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$, а игрок 2 выбирает вектор

действительных значений элементов временных рядов χ_{T+1} из множества $\mathcal{A} \cap \mathcal{B}$. При этом первый игрок преследует цель минимизировать свои потери, выраженные функцией $L(\hat{\varphi}, \chi_{T+1})$, при любых действиях игрока 2, т. е. при любом векторе действительных значений χ_{T+1} . Эта цель достигается, если игрок 1 воспользуется стратегией, входящей в равновесие Нэша.

4.3 Существование равновесия Нэша

В этой части раздела будет показано, что при выполнении ряда естественных требований к множествам \mathcal{A} (3) и \mathcal{B} (4) и функции суммарных потерь при прогнозировании иерархии l_h (2) в антагонистической игре (5), описывающей задачу согласования прогнозов, существует равновесие Нэша в чистых стратегиях. Также будет показано, что соответствующее этому равновесию значение функции потерь игрока 1 неположительно, что гарантирует неухудшение качества прогнозов при переходе от независимых прогнозов к согласованным. Рассмотрим эти требования.

Определение 5. Множество $\mathcal{C} \subseteq \mathbb{R}^d$ называется выпуклым [20], если для любых $\chi_1 \in \mathcal{C}$ и $\chi_2 \in \mathcal{C}$ и любого $0 \leq \alpha \leq 1$ выполнено

$$\alpha\chi_1 + (1 - \alpha)\chi_2 \in \mathcal{C}.$$

Заметим, что множества \mathcal{A} (3) и \mathcal{B} (4) выпуклы и замкнуты.

Предположение 1. *Будем предполагать, что пересечение множеств \mathcal{A} и \mathcal{B} не пусто: $\mathcal{A} \cap \mathcal{B} \neq \emptyset$.*

Требование непустого пересечения этих множеств естественно, так как в противном случае неразрешима задача поиска вектора согласованных прогнозов $\hat{\varphi}$, который должен одновременно принадлежать обоим множествам. Также отметим, что множество $\mathcal{A} \cap \mathcal{B}$ является выпуклым и замкнутым как пересечение двух выпуклых и замкнутых множеств [20].

Предположение 2. *Будем считать, что функция суммарных потерь l_h (2) неотрицательна и равна нулю только при равенстве аргументов:*

$$l_h(\chi_{T+1}, \hat{\chi}) \geq 0 \text{ для всех } \chi_{T+1}, \hat{\chi};$$

$$l_h(\chi_{T+1}, \hat{\chi}) = 0 \Leftrightarrow \chi_{T+1} = \hat{\chi}.$$

Равенство аргументов $\chi_{T+1} = \hat{\chi}$ соответствует случаю, когда прогноз полностью совпадает с действительными значениями. В этом случае потери равны нулю. Во всех остальных случаях потери при прогнозе положительные.

Определение 6. Проекцией точки $\chi_0 \in \mathbb{R}^d$ на множество $C \subseteq \mathbb{R}^d$, инициированной функцией расстояния f , называется точка

$$\chi_{\text{proj}} = \arg \min_{\chi \in C} f(\chi, \chi_0).$$

Предположение 3. Пусть существует проекция точки из \mathbb{R}^d , соответствующей вектору независимых прогнозов $\hat{\chi}$, на выпуклое и замкнутое множество $A \cap B$, инициированная функцией суммарных потерь l_h :

$$\chi_{\text{proj}} = \arg \min_{\chi \in A \cap B} l_h(\chi, \hat{\chi}).$$

Предположение 4. Пусть $\chi_{\text{proj}} = \arg \min_{\chi \in A \cap B} l_h(\chi, \hat{\chi})$.

Будем предполагать, что для всех $\chi \in B$ и для всех $\psi \in A \cap B$ выполняется неравенство

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi).$$

Для пояснения этого требования рассмотрим частный случай, когда $\mathbb{R}^d = \mathbb{R}^3$, $A \subset \mathbb{R}^2$, $B \subset \mathbb{R}^3$, $A \cap B \subset \mathbb{R}^2$ и l_h — квадрат метрики Евклида (рис. 2). Точки ψ , χ , χ_{proj} образуют треугольник. Обозначим θ угол при вершине χ_{proj} и запишем теорему косинусов:

$$l_h(\psi, \chi) = l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi) - 2\sqrt{l_h(\psi, \chi_{\text{proj}})l_h(\chi_{\text{proj}}, \chi)} \cos \theta.$$

Поскольку χ_{proj} является проекцией, то угол θ не может быть острым. Он прямой, если проецируемая точка χ находится «над» множеством $A \cap B$, и тупой, если точка находится «в стороне». Таким образом получаем, что $\cos \theta \leq 0$, а значит, последнее слагаемое в теореме косинусов неотрицательное. Исключая его и заменяя знак равенства на знак нестрогого неравенства, получаем, что предположение 4 соответствует естественным свойствам проекции.

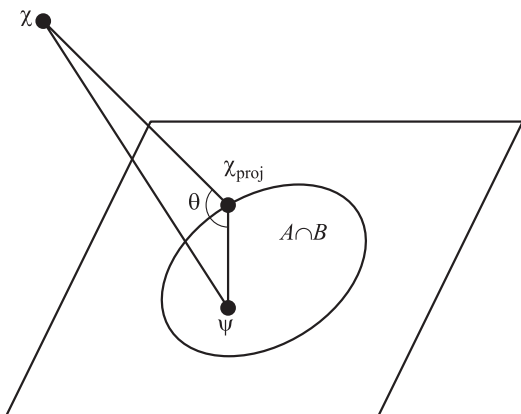


Рис. 2 Пояснение к предположению 4

Введя предположения 1–4, сформулируем теорему.

Теорема 2. Пусть выполнены предположения 1–4. Тогда пара стратегий $(\chi_{\text{proj}}, \chi_{\text{proj}})$ является равновесием Нэша в игре (5) и седловой точкой функции $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$. Цена игры при этом равна $V = -l_h(\chi_{\text{proj}}, \hat{\chi})$.

Доказательство. Найдем седловую точку функции $L(\hat{\varphi}, \chi_{T+1})$ в соответствии с определением 4 и теоремой 1. Найдем максимум этой функции по второму аргументу при $\hat{\varphi} = \chi_{\text{proj}}$. Для этого воспользуемся предположением 4:

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi).$$

Применяя неравенство к функции потерь L игрока 1 (подставляем $\psi = \chi_{T+1}$, $\chi = \hat{\chi}$), получаем:

$$\begin{aligned} L(\chi_{\text{proj}}, \chi_{T+1}) &= l_h(\chi_{T+1}, \chi_{\text{proj}}) - l_h(\chi_{T+1}, \hat{\chi}) \leq \\ &\leq l_h(\chi_{T+1}, \hat{\chi}) - l_h(\chi_{\text{proj}}, \hat{\chi}) - l_h(\chi_{T+1}, \hat{\chi}) = \\ &= -l_h(\chi_{\text{proj}}, \hat{\chi}) \end{aligned}$$

для всех $\chi_{T+1} \in A \cap B$. Заметим также, что из предположения 2 вытекает

$$L(\chi_{\text{proj}}, \chi_{\text{proj}}) = l_h(\chi_{\text{proj}}, \chi_{\text{proj}}) - l_h(\chi_{\text{proj}}, \hat{\chi}) = -l_h(\chi_{\text{proj}}, \hat{\chi}).$$

Приходим к выводу, что

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}})$$

для всех $\chi_{T+1} \in A \cap B$.

Следовательно, максимум по второму аргументу достигается при $\chi_{T+1} = \chi_{\text{proj}}$.

Минимум по первому аргументу при $\chi_{T+1} = \chi_{\text{proj}}$ находим, используя предположение 2, из соотношения

$$\begin{aligned} \arg \min_{\hat{\varphi} \in A \cap B} L(\hat{\varphi}, \chi_{\text{proj}}) &= \\ &= \arg \min_{\hat{\varphi} \in A \cap B} l_h(\chi_{\text{proj}}, \hat{\varphi}) - l_h(\chi_{\text{proj}}, \hat{\chi}). \end{aligned}$$

Второе слагаемое не зависит от $\hat{\varphi}$, а по предположению 2 функция суммарных потерь неотрицательна и обращается в ноль только при равенстве аргументов, поэтому получаем

$$\arg \min_{\hat{\varphi} \in A \cap B} L(\hat{\varphi}, \chi_{\text{proj}}) = \chi_{\text{proj}}.$$

Таким образом, получаем, что

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}}) \leq L(\hat{\varphi}, \chi_{\text{proj}}), \quad \chi_{T+1}, \hat{\varphi} \in A \cap B.$$

Следовательно, $(\chi_{\text{proj}}, \chi_{\text{proj}})$ — седловая точка функции L . И эта пара является равновесием Нэша в игре (5), и цена игры выражается как

$$\begin{aligned} V &= \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} \max_{\chi_{T+1} \in \mathcal{A}} L(\hat{\varphi}, \chi_{T+1}) = \\ &= \max_{\chi_{T+1} \in \mathcal{A}} \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} L(\hat{\varphi}, \chi_{T+1}) = \\ &= L(\chi_{\text{proj}}, \chi_{\text{proj}}) = -l_h(\chi_{\text{proj}}, \hat{\chi}). \end{aligned}$$

Следствие 1. *Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции вектора независимых прогнозов $\hat{\chi} \in \mathbb{R}^d$ на множество $\mathcal{A} \cap \mathcal{B}$, инициированной функцией суммарных потерь l_h , гарантирует значение функции суммарных потерь не большее, чем при использовании вектора независимых прогнозов $\hat{\chi}$. Доказательство.* По теореме 2 цена игры (5) равна значению функции потерь игрока 1 в точке $(\chi_{\text{proj}}, \chi_{\text{proj}})$ и неположительна в силу предположения 2

$$V = L(\chi_{\text{proj}}, \chi_{\text{proj}}) = -l_h(\chi_{\text{proj}}, \hat{\chi}) \leq 0.$$

А выбор рассматриваемой функции потерь игрока 1 $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$ был обусловлен тем, что ее знак совпадает со знаком изменения суммарных потерь при переходе от вектора независимых прогнозов $\hat{\chi}$ к вектору согласованных прогнозов $\hat{\varphi}$. Следовательно, при $\chi_{T+1} = \chi_{\text{proj}}$ суммарные потери при согласованных прогнозах меньше, чем при независимых.

Согласно определению 4 равновесия Нэша

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}}) \leq 0$$

для любых $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$. Поэтому при любом векторе действительных значений $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$ согласованные прогнозы $\hat{\varphi}$ оказываются предпочтительнее независимых прогнозов $\hat{\chi}$.

5 Алгоритм согласования прогнозов ГТОр

Согласно следствию 1 оптимальным выбором вектора согласованных прогнозов $\hat{\varphi}$ является проекция вектора независимых прогнозов $\hat{\chi}$ на множество $\mathcal{A} \cap \mathcal{B}$, инициированная функцией суммарных потерь l_h . Множество $\mathcal{A} \cap \mathcal{B}$ содержит векторы размерности d , удовлетворяющие структуре иерархии, так как множество \mathcal{A} задается ограничениями-равенствами, порожденными матрицей связи иерархии S . В то же время $\mathcal{A} \cap \mathcal{B}$ содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам из множества \mathcal{B} . Таким образом, задача поиска проекции — это оптимизационная задача с ограничениями типа равенства и неравенства:

$$\begin{cases} l_h(\chi, \hat{\chi}) \rightarrow \min_{\chi}, \\ \chi \in \mathcal{A} \text{ (ограничения-равенства)}; \\ \chi \in \mathcal{B} \text{ (ограничения-неравенства)}. \end{cases} \quad (6)$$

Алгоритм согласования прогнозов иерархических временных рядов ГТОр заключается в решении оптимизационной задачи (6). Достоинства этого алгоритма заключаются в том, что он требует от вектора независимых прогнозов $\hat{\chi}$ лишь принадлежности множеству \mathcal{B} , и не требует несмещенности независимых прогнозов, а следовательно, для получения независимых прогнозов можно использовать любой алгоритм прогнозирования. Также ГТОр не требует оценки погрешностей независимых прогнозов. Самое важное, что ГТОр обеспечивает неухудшение качества прогнозирования при замене независимых прогнозов на согласованные прогнозы. При этом на структуру иерархии, ограничения-неравенства на прогнозы и функцию суммарных потерь накладываются лишь общие ограничения, гарантирующие существование решения оптимизационной задачи (6). Еще одно достоинство алгоритма ГТОр в том, что он позволяет согласовывать прогнозы для иерархий любой сложности одновременно по всем уровням, учитывая все связи в иерархии и решая одну оптимизационную задачу. От сложности иерархии и количества временных рядов и уровней в оптимизационной задаче зависит число переменных и ограничений.

6 Дивергенция Брегмана

В этом разделе будет описано семейство функций двух переменных, удовлетворяющих предположениям 1–4, и приведен ряд примеров функций из этого семейства, которые можно использовать в задаче согласования прогнозов в качестве функции суммарных потерь l_h (2). Все эти функции двух переменных называются дивергенциями Брегмана [19, 21] и порождаются функциями одной переменной, обладающими следующими свойствами.

Определение 7. Функцией Лежандра [19] называется функция $F: \mathcal{B} \rightarrow \mathbb{R}$, которая удовлетворяет следующим условиям:

- $\mathcal{B} \subseteq \mathbb{R}^d$ — непустое множество, и внутренность \mathcal{B} выпукла;
- F — строго выпуклая функция с непрерывной первой производной на множестве \mathcal{B} ;
- если $\chi_1, \chi_2, \dots \in \mathcal{B}$ — последовательность, сходящаяся к граничной точке \mathcal{B} , то $\|\nabla F(\chi_n)\| \rightarrow \infty$ при $n \rightarrow \infty$.

Определение 8. Дивергенцией Брегмана, порожденной функцией Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, называется неотрицательная функция $D_F: \mathcal{B} \times \text{int}(\mathcal{B}) \rightarrow \mathbb{R}$, определенная как

$$D_F(\mathbf{u}, \mathbf{v}) = F(\mathbf{u}) - F(\mathbf{v}) - (\mathbf{u} - \mathbf{v})\nabla F(\mathbf{v}).$$

Свойства дивергенции Брегмана:

- для всех \mathbf{u} и \mathbf{v} выполнено $D_F(\mathbf{u}, \mathbf{v}) \geq 0$. Это следует из выпуклости функции F ;
- $D_F(\mathbf{u}, \mathbf{v})$ выпукла по первому аргументу \mathbf{u} , но необязательно выпукла по второму аргументу \mathbf{v} ;
- для любых $\alpha, \beta \in \mathbb{R}$ и любых функций Лежандра F_1 и F_2 выполнено

$$D_{\alpha F_1 + \beta F_2}(\mathbf{u}, \mathbf{v}) = \alpha D_{F_1}(\mathbf{u}, \mathbf{v}) + \beta D_{F_2}(\mathbf{u}, \mathbf{v}).$$

Определение 9. Пусть $F: \mathcal{B} \rightarrow \mathbb{R}$ — функция Лежандра и $\mathcal{A} \subset \mathbb{R}^d$ — замкнутое выпуклое множество, такое что $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Проекция Брегмана \mathbf{w}' точки $\mathbf{w} \in \text{int}(\mathcal{B})$ на множество \mathcal{A} — это

$$\mathbf{w}' = \arg \min_{\mathbf{u} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{u}, \mathbf{w}).$$

Теорема 3. Для всех функций Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, для всех замкнутых выпуклых множеств $\mathcal{A} \subset \mathbb{R}^d$, имеющих непустое пересечение $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, и для всех точек $\mathbf{w} \in \text{int}(\mathcal{B})$ проекция Брегмана точки \mathbf{w} на множество \mathcal{A} существует и единственна.

Доказательство этой теоремы приведено в [21].

Теорема 4. Пусть F — функция Лежандра. Для всех $\mathbf{w} \in \text{int}(\mathcal{B})$ и для всех замкнутых выпуклых множеств $\mathcal{A} \subseteq \mathbb{R}^d$ с непустым пересечением $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, если \mathbf{w}' — проекция Брегмана точки \mathbf{w} на множество \mathcal{A} ($\mathbf{w}' = \arg \min_{\mathbf{v} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{v}, \mathbf{w})$), то верно неравенство:

$$D_F(\mathbf{u}, \mathbf{w}) \geq D_F(\mathbf{u}, \mathbf{w}') + D_F(\mathbf{w}', \mathbf{w})$$

для всех $\mathbf{u} \in \mathcal{A}$.

Доказательство этого факта можно найти в [19].

Соотнесем перечисленные свойства дивергенции Брегмана и предположения 1–4. Определение 8 и свойство 1 дивергенции Брегмана обеспечивают выполнение предположения 2 о знаке функции суммарных потерь. В определении 9 и теореме 3 предполагается выпуклость и замкнутость множеств \mathcal{A} и \mathcal{B} и их непустое пересечение, как и в предположении 1. Теорема 4 гарантирует выполнение предположения 3 о выпуклости множеств и существовании и единственности проекции. Наконец, теорема 4 гарантирует выполнение предположения 4. Следовательно, для функций

суммарных потерь l_h (2), являющихся дивергенциями Брегмана, выполнены все условия теоремы 2. Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции Брегмана вектора независимых прогнозов $\hat{\chi} \in \mathcal{B}$, где множество \mathcal{B} определено по формуле (4) и является выпуклым и замкнутым, имеющим непустое пересечение с множеством \mathcal{A} (3), на множество $\mathcal{A} \cap \mathcal{B}$, включающее векторы прогнозов, удовлетворяющих структуре иерархии, гарантирует неухудшение качества прогнозов.

Следующие функции являются дивергенциями Брегмана и могут быть использованы в качестве функций суммарных потерь (2) при согласовании прогнозов.

Квадрат евклидова расстояния

$D_F(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|^2$ — канонический пример дивергенции Брегмана, порождается функцией $F(\mathbf{u}) = \|\mathbf{u}\|^2$.

Квадрат расстояния Махаланобуса

$D_F(\mathbf{u}, \mathbf{v}) = (1/2)(\mathbf{u} - \mathbf{v})^T Q(\mathbf{u} - \mathbf{v})$ — обобщение евклидова расстояния, порождается квадратичной формой $F(\mathbf{u}) = (1/2)\mathbf{u}^T Q\mathbf{u}$.

Обобщенная дивергенция Кульбака–Лейблера

$D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n u_i \log(u_i/v_i) - \sum_{i=1}^n u_i + \sum_{i=1}^n v_i$ порождается функцией $F(\mathbf{u}) = \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n u_i$.

Расстояние Itakura–Saito

$D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n ((u_i/v_i) - \log(u_i/v_i) - 1)$ порождается функцией $F(\mathbf{u}) = -\sum_{i=1}^n \log u_i$.

7 Эксперимент

В экспериментальной части рассматриваются данные о посуточной загруженности железнодорожных узлов РЖД. Из имеющихся данных были сформированы временные ряды, описывающие отправление 37 различных типов груза со станций 98 железнодорожных веток посуточно. Рассмотрены два вида иерархии.

В случае плоской двухуровневой иерархии решены задачи согласования прогнозов отправления грузов по отдельности и всех грузов в сумме для каждой ветки, а также решены задачи согласования прогнозов отправления груза с каждой ветки и суммарного отправления груза со всех веток для каждого типа груза.

В случае неплоской трехуровневой иерархии решена задача согласования прогнозов всех имеющихся временных рядов. Демонстрируется уменьшение значения функции суммарных потерь при переходе от независимых прогнозов к согласованным. В качестве функции суммарных потерь был

Вид записи базы данных железнодорожных перевозок

Дата погрузки	Станция отправления	Станция назначения	Количество вагонов	Код груза	Род вагона	Суммарный вес груза	Признак маршрутной отправки
2007-01-01	020108	932902	1	1	216	56	9

использован квадрат евклидового расстояния. Независимые прогнозы были получены алгоритмом Hist, описанным в работах [7, 22].

Экспериментальные данные. В эксперименте использованы данные о посуточной загруженности железнодорожных узлов РЖД с 1 января 2007 г. по 22 апреля 2008 г. В таблице приведен пример записи базы данных.

Коды станций представляют собой шестизначные числа. Станции, в коде которых две первые цифры совпадают, входят в одну железнодорожную ветку. Станций отправления — 1566, станций назначения — 1902, веток — 98. Код груза — натуральное число от 1 до 37; также имеются перевозки, где код груза не указан. Род вагона — натуральное число, в имеющихся данных 75 различных родов вагонов.

Иерархическая структура. Экспериментальные данные удовлетворяют структуре, изображенной на рис. 3.

Как видно из рисунка, иерархия не является плоской и содержит три уровня временных рядов. Временные ряды нижнего уровня этой иерархии

имеют два индекса, соответствующих номеру ветки и коду груза: $x(i, j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, где n — число веток, а m — количество грузов. На среднем уровне — два семейства временных рядов. Временные ряды, соответствующие суммарному отправлению всех грузов с каждой ветки, обозначаются $x(i, :)$, $i = 1, \dots, n$. Ряды среднего уровня, соответствующие суммарному отправлению со всех веток каждого из грузов, обозначаются $x(:, j)$, $j = 1, \dots, m$. Временной ряд верхнего уровня обозначается $x(:, :)$.

Условие согласованности для трехуровневой иерархии задается равенствами ($t = 1, \dots, T$):

$$\left. \begin{aligned} x_t(:, :) &= \sum_{i=1}^n x_t(i, :); \\ x_t(:, j) &= \sum_{i=1}^n x_t(i, j); \\ x_t(i, :) &= \sum_{j=1}^m x_t(i, j), \quad i = 1, \dots, n; \\ x_t(i, j) &= \sum_{i=1}^n x_t(i, j), \quad j = 1, \dots, m. \end{aligned} \right\} \quad (7)$$

Векторная запись срезов иерархии, независимых прогнозов и согласованных прогнозов имеет размерность $d = 1 + n + m + nm$ и выглядит следующим образом:

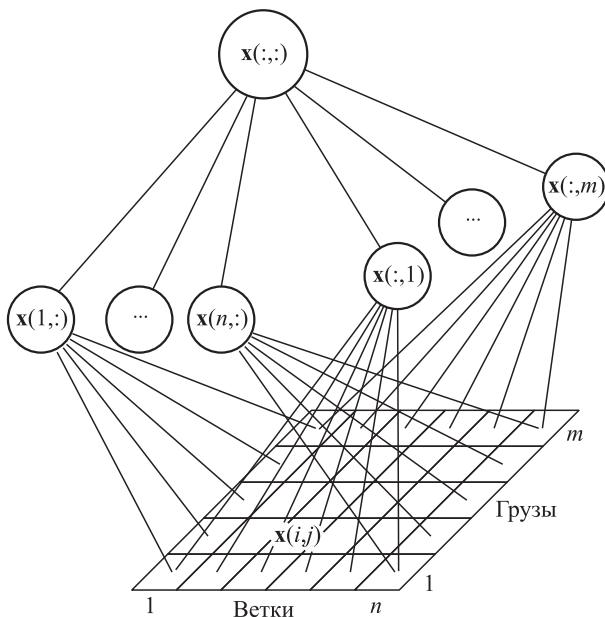


Рис. 3 Неплоская трехуровневая иерархия

$$\chi_t = \begin{pmatrix} x_t(:, :) \\ x_t(1, :) \\ \vdots \\ x_t(n, :) \\ x_t(:, 1) \\ \vdots \\ x_t(:, m) \\ x_t(1, 1) \\ \vdots \\ x_t(1, m) \\ \vdots \\ x_t(n, 1) \\ \vdots \\ x_t(n, m) \end{pmatrix}; \quad \hat{\chi} = \begin{pmatrix} \hat{x}(:, :) \\ \hat{x}(1, :) \\ \vdots \\ \hat{x}(n, :) \\ \hat{x}(:, 1) \\ \vdots \\ \hat{x}(:, m) \\ \hat{x}(1, 1) \\ \vdots \\ \hat{x}(1, m) \\ \vdots \\ \hat{x}(n, 1) \\ \vdots \\ \hat{x}(n, m) \end{pmatrix};$$

$$\hat{\varphi} = \begin{pmatrix} \hat{y}(:, :) \\ \hat{y}(1, :) \\ \vdots \\ \hat{y}(n, :) \\ \hat{y}(:, 1) \\ \vdots \\ \hat{y}(:, m) \\ \hat{y}(1, 1) \\ \vdots \\ \hat{y}(1, m) \\ \vdots \\ \hat{y}(n, 1) \\ \vdots \\ \hat{y}(n, m) \end{pmatrix}.$$

Матрица связей для условия (7) имеет размер $(2 + n + m) \times (1 + n + m + nm)$:

$$S = \begin{pmatrix} -1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 & \dots \\ -1 & 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 & \dots \\ 0 & -1 & \dots & 0 & 0 & \dots & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \dots & \vdots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 0 & \dots & 0 & -1 & \dots & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \dots & \vdots & \dots & \dots & \vdots & \vdots & \vdots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & -1 & 0 & 0 & \dots \\ \dots & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots \\ \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \vdots & \vdots & \dots \\ \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots \\ \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots \end{pmatrix}.$$

Требуется, чтобы все прогнозы были неотрицательны, поэтому множество \mathcal{B} (4) задается как

$$\mathcal{B} = \{\chi \in \mathbb{R}^d \mid \chi(i) \in [0, +\infty], i = 1, \dots, d\}.$$

В качестве функции суммарных потерь l_h (2) используется квадрат евклидова расстояния.

Оптимизационная задача. Задача, решенная для согласования прогнозов рассматриваемой иерархии, имеет вид:

$$\begin{cases} \|\chi - \hat{\chi}\|^2 \rightarrow \min_{\chi}, \\ S\chi = 0, \\ \chi \geq 0. \end{cases}$$

Результаты эксперимента. Для прогноза были использованы временные ряды, описывающие суммарный вес отправленных грузов разных типов по каждой ветке. Значения временных рядов нижнего уровня иерархии были отнормированы на отрезок $[0; 1]$ по формуле

$$x_t^{\text{norm}}(i, j) = \frac{x_t(i, j) - m(i, j)}{M(i, j) - m(i, j)},$$

где

$$m = \min_{t=1, \dots, T} x_t(i, j), \quad M = \max_{t=1, \dots, T} x_t(i, j).$$

Прогнозы были построены и согласованы для 100 последних точек истории. Для каждого отсчета времени строился вектор независимых прогнозов $\hat{\chi}$ и вектор согласованных прогнозов $\hat{\varphi}$. Для каждого вектора вычислялось значение функции суммарных потерь, затем вычислялись потери игрока 1 в игре (5), равные разности суммарных потерь при использовании согласованных прогнозов и суммарных потерь при использовании независимых прогнозов:

$$L_t = l_h(\hat{\varphi}, \chi_t) - l_h(\hat{\chi}, \chi_t), \quad t = T - 100 + 1, \dots, T.$$

Значения этой величины изображены на рис. 4. Теоретические выкладки подтверждаются на практике. Для плоских двухуровневых иерархий есть случаи, когда суммарные потери при переходе к согласованным прогнозам не изменяются, и случаи, когда суммарные потери уменьшаются. Для неплоской трехуровневой иерархии суммарные потери во всех контрольных точках уменьшаются.

8 Заключение

Предложен алгоритм ГТОр для согласования прогнозов иерархических временных рядов. Алгоритм не требует оценки погрешностей независимых прогнозов и не требует их несмещенности. Для любого набора независимых прогнозов временных рядов алгоритм согласования не ухудшает качество прогнозирования. Возможна работа с иерархическими структурами любой сложности. Все свойства алгоритма согласования прогнозов ГТОр подтверждаются на практике.

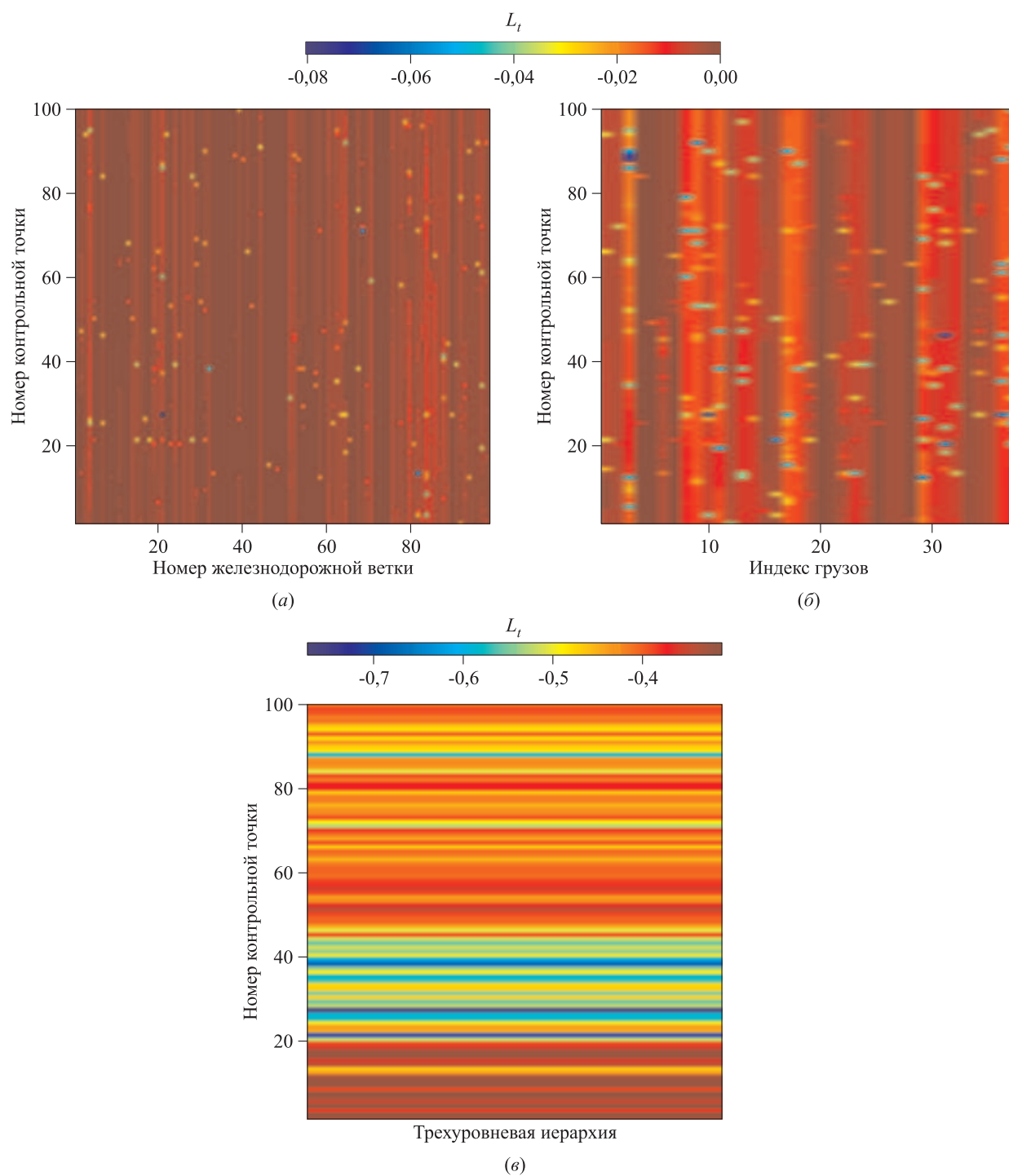


Рис. 4 Изменение суммарных потерь L_t : (а) для каждой ветки; (б) для каждого груза; (в) вся иерархия

Литература

1. Токмакова А. А., Стрижов В. В. Оценивание гиперпараметров линейных и регрессионных моделей при отборе шумовых и коррелирующих признаков // Информатика и её применения, 2012. Т. 6. Вып. 4. С. 66–75.
2. Васильев Н. С. Использование принципа равновесия для управления маршрутизацией в транспортных сетях // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 28–35.
3. Hong T., Pinson P., Fan S. Global energy forecasting competition 2012 // Int. J. Forecasting, 2014. Vol. 30. No. 2. P. 357–363.

4. Kaggle. <https://www.kaggle.com>.
5. Hyndman R. J., Ahmed R. A., Athanasopoulos G., Shang H. L. Optimal combination forecasts for hierarchical time series // *Comput. Stat. Data Anal.*, 2011. Vol. 55. No. 9. P. 2579–2589.
6. Кузнецов М. П., Мафусалов А. А., Животовский Н. К., Зайцев Е. Ю., Сунгуров Д. С. Сглаживающие алгоритмы прогнозирования // *Машинное обучение и анализ данных*, 2011. Т. 1. Вып. 1. С. 104–112.
7. Стенина М. М., Стрижов В. В. Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования // *Системы и средства информатики*, 2014. Т. 24. Вып. 2. С. 21–34.
8. Grunfeld Y., Griliches Z. Is aggregation necessarily bad? // *Rev. Econ. Stat.*, 1960. Vol. 42. No. 1. P. 1–13.
9. Orcutt G. H., Watts H. W., Edwards J. B. Data aggregation and information loss // *Am. Econ. Rev.*, 1968. Vol. 58. No. 4. P. 773–787.
10. Edwards J. B., Orcutt G. H. Should aggregation prior to estimation be the rule? // *Rev. Econ. Stat.*, 1969. Vol. 51. No. 4. P. 409–420.
11. Shlifer E., Wolff R. W. Aggregation and proration in forecasting // *Manage. Sci.*, 1979. Vol. 25. No. 6. P. 594–603.
12. Fogarty D. W., Blackstone J. H., Hoffman T. R. Production and inventory management. — 2nd ed. — Cincinnati, OH, USA: South-Western Publication Co., 1990. 880 p.
13. Narasimhan S. L., McLeavey D. W., Billington P. J. Production planning and inventory control. — 2nd ed. — Englewood Cliffs, NJ, USA: Prentice Hall, 1995. 716 p.
14. Schwarzkopf A. B., Tersine R. J., Morris J. S. Top-down versus bottom-up forecasting strategies // *Int. J. Prod. Res.*, 1998. Vol. 26. No. 11. P. 1833–1843.
15. Fliedner G. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation // *Comput. Oper. Res.*, 1999. Vol. 26. No. 10–11. P. 1133–1149.
16. Van Erven T., Cugliari J. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. 2013. <https://hal.inria.fr/hal-00920559>.
17. Петросян Л. А., Зенкевич Н. А., Семина Е. А. Теория игр. — М.: Университет, 1998. 301 с.
18. Меньшиков И. С. Лекции по теории игр и экономическому моделированию. — 2-е изд., испр. и доп. — М.: Контакт Плюс, 2010. 336 с.
19. Cesa-Bianchi N., Lugosi G. Prediction, learning, and games. — Cambridge: Cambridge University Press, 2006. Vol. 1. 403 p.
20. Boyd S., Vandenberghe L. Convex optimization. — Cambridge: Cambridge University Press, 2009. 732 p.
21. Bregman L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // *USSR Comput. Math. Math. Phys.*, 1967. Vol. 7. No. 3. P. 200–217.
22. Вальков А. С., Кожанов Е. М., Медведникова М. М., Хусаинов Ф. И. Непараметрическое прогнозирование загруженности системы железнодорожных узлов по историческим данным // *Машинное обучение и анализ данных*, 2012. Т. 1. Вып. 4. С. 448–465.

Поступила в редакцию 27.10.14

FORECASTS RECONCILIATION FOR HIERARCHICAL TIME SERIES FORECASTING PROBLEM

M. M. Stenina¹ and V. V. Strijov²

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

²Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The hierarchical time series forecasting problem is researched. Time series forecasts must satisfy the physical constraints and the hierarchical structure. In this paper, a new algorithm for hierarchical time series forecasts reconciliation is proposed. The algorithm is called GTOp (Game-theoretically optimal reconciliation). It guarantees that the quality of reconciled forecasts is not worse than the quality of self-dependent forecasts. This approach is based on Nash equilibrium search for the antagonistic game and turns the forecasts reconciliation problem into the optimization problem with equality and inequality constraints. It is proved that the Nash equilibrium in pure strategies exists in the game if some assumptions about the hierarchical structure, the physical constraints, and the loss function are satisfied. The algorithm performance is demonstrated for different types of hierarchical structures of time series.

Keywords: hierarchical time series; reconciliation of time series forecasts; antagonistic game; Nash equilibrium

DOI: 10.14357/19922264150209

Acknowledgments

The research was financially supported by the Russian Foundation for Basic Research (project 13-07-13139).

References

1. Tokmakova, A. A., and V. V. Strizhov. 2012. Otsenivanie giperparametrov lineynykh i regressionnykh modeley pri otbore shumovykh i korrelirovannykh priznakov [Estimation of linear model hyperparameters for noise or correlated feature selection problem]. *Informatika i ee Primeneniya — Inform. Appl.* 6(4):66–75.
2. Vasil'ev, N. S. 2014. Ispol'zovanie printsipa ravnovesiya dlya upravleniya marshrutizatsiy v transportnykh setyakh [Equilibrium principle application to routing control in packet data transmission networks]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):28–35.
3. Hong, T., P. Pinson, and S. Fan. 2014. Global energy forecasting competition 2012. *Int. J. Forecasting* 30(2):357–363.
4. Kaggle. Available at: <https://www.kaggle.com> (accessed May 20, 2015).
5. Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. 2011. Optimal combination forecasts for hierarchical time series. *Comput. Stat. Data Anal.* 55(9):2579–2589.
6. Kuznetsov, M. P., A. A. Mafusalov, N. K. Zhivotovskiy, E. Yu. Zaytsev, and D. S. Sungurov. 2011. Sglazhivayushchie algoritmy prognozirovaniya [Smoothing forecast algorithms]. *Mashinnoe obuchenie i analiz dannykh* [J. Machine Learning Data Anal.] 1(1):104–112.
7. Stenina, M. M., and V. V. Strizhov. 2014. Soglasovanie agregirovannykh i detalizirovannykh prognozov pri reshenii zadach neparametricheskogo prognozirovaniya [Reconciliation of aggregated and disaggregated time series forecasts in nonparametric forecasting problem]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(2):21–34.
8. Grunfeld, Y., and Z. Griliches. 1960. Is aggregation necessarily bad? *Rev. Econ. Stat.* 42(1):1–13.
9. Orcutt, G. H., H. W. Watts, and J. B. Edwards. 1968. Data aggregation and information loss. *Am. Econ. Rev.* 58(4):773–787.
10. Edwards, J. B., and G. H. Orcutt. 1969. Should aggregation prior to estimation be the rule? *Rev. Econ. Stat.* 51(4):409–420.
11. Shlifer, E., and R. W. Wolff. 1979. Aggregation and proration in forecasting. *Manage. Sci.* 25(6):594–603.
12. Fogarty, D. W., J. H. Blackstone, and T. R. Hoffman. 1990. *Production and inventory management*. 2nd ed. Cincinnati, OH: South-Western Publication Co. 880 p.
13. Narasimhan, S. L., D. W. McLeavey, and P. J. Billington. 1995. *Production planning and inventory control*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall. 716 p.
14. Schwarzkopf, A. B., R. J. Tersine, and J. S. Morris. 1998. Top-down versus bottom-up forecasting strategies. *Int. J. Prod. Res.* 26(11):1833–1843.
15. Fliedner, G. 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Comput. Oper. Res.* 26(10–11):1133–1149.
16. Van Erven, T., and J. Cugliari. 2013. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. Available at: <https://hal.inria.fr/hal-00920559> (accessed May 20, 2015).
17. Petrosyan, L. A., N. A. Zenkevich, and E. A. Semina. 1998. *Teoriya igr* [Games theory]. Moscow: Knizhnyy Dom Universitet. 301 p.
18. Men'shikov, I. S. 2010. *Lektsii po teorii igr i ekonomicheskomu modelirovaniyu* [Games theory and economics modeling lectures]. 2nd ed. Moscow: OOO Kontakt Plyus. 336 p.
19. Cesa-Bianchi, N., and G. Lugosi. 2006. *Prediction, learning, and games*. Cambridge: Cambridge University Press. Vol. 1. 403 p.
20. Boyd, S., and L. Vandenberghe. 2009. *Convex optimization*. Cambridge: Cambridge University Press. 732 p.
21. Bregman, L. M. 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* 7(3):200–217.
22. Val'kov, A. S., E. M. Kozhanov, M. M. Medvednikova, and F. I. Khusainov. 2012. Neparametricheskoe prognozirovanie zagruzhennosti sistemy zheleznodorozhnykh uzlov po istoricheskim dannym [Nonparametric forecasting of railroad stations occupancy according to historical data]. *Mashinnoe Obuchenie i Analiz Dannykh* [J. Machine Learning Data Anal.] 1(4):448–465.

Received October 27, 2014

Contributors

Stenina Mariya M. (b. 1991) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; mmedvednikova@gmail.com

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; strijov@gmail.com

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ МНОГОМЕРНОЙ ПЛОТНОСТИ С ПОМОЩЬЮ ВЕЙВЛЕТ-ОЦЕНОК ОДНОМЕРНЫХ ПРОЕКЦИЙ*

О. В. Шестаков¹

Аннотация: Рассматривается используемый в реконструктивной томографии метод обращения преобразования Радона для получения статистических оценок многомерных вероятностных плотностей. Этот метод использует нелинейные вейвлет-оценки одномерных проекций для построения оценки многомерной плотности. Нелинейные вейвлет-оценки обладают возможностью адаптации к локальным свойствам оцениваемой функции плотности и, следовательно, менее чувствительны к наличию сингулярных точек, чем линейные оценки. Еще одним важным практическим аспектом рассматриваемого метода является его параллельная структура, позволяющая значительно ускорить построение оценок на компьютерах, поддерживающих параллельные вычисления. Также показано, что при выполнении некоторых условий регулярности равномерное расстояние между построенной оценкой и истинной многомерной вероятностной плотностью стремится к нулю в среднем, и получены оценки скорости этой сходимости.

Ключевые слова: вейвлеты; многомерная плотность; преобразование Радона

DOI: 10.14357/19922264150210

1 Введение

Проблема непараметрического оценивания многомерной плотности вероятностного распределения возникает во многих областях статистического анализа данных, например в задачах классификации и распознавания изображений.

Существуют различные подходы к решению этой проблемы [1], такие как использование ядерных оценок, оценивание по ближайшим соседям, аппроксимация сплайнами и разложение в ортогональные ряды. В последние годы все более популярными становятся методы, основанные на вейвлет-разложении [2]. Преимуществом этих методов является адаптивность к локальной гладкости оцениваемой плотности. Возможные многомерные обобщения вейвлет-методов оценивания плотности рассмотрены в работах [3–6].

В данной работе рассматривается томографический подход, основанный на построении оценки многомерной плотности по одномерным оценкам проекций. Как отмечено в работе [7], данный подход не нуждается в использовании структурных предположений, таких как мультипликативность или аддитивность плотности. Кроме того, важным практическим аспектом такого метода является воз-

можность параллельного вычисления одномерных оценок.

Для рассматриваемого метода получены оценки скорости сходимости к нулю в среднем равномерного расстояния между построенной оценкой и оцениваемой многомерной плотностью.

2 Преобразование Радона

Преобразование Радона в d -мерном пространстве ставит в соответствие функции $f(x)$, $x \in \mathbb{R}^d$, одномерные проекции $R_\theta f(s)$:

$$R_\theta f(s) = \int_{s\theta=s} f(x) dx = \int_{\theta^\perp} f(s\theta + y) dy,$$

где $\theta \in S^{d-1}$, а S^{d-1} — единичная сфера в \mathbb{R}^d .

Зная $R_\theta f(s)$ для всех $\theta \in S^{d-1}$ и всех $s \in \mathbb{R}$, функцию $f(x)$ можно вычислить по следующей формуле [7]:

$$f(x) = \frac{1}{2(2\pi)^d} \int_{S^{d-1} - \infty}^{\infty} |\omega|^{d-1} \widehat{R_\theta f}(\omega) \exp\{i\omega(x\theta)\} d\omega d\theta, \quad (1)$$

* Работа выполнена при финансовой поддержке РФФИ (проект 15-07-02652).

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

где $\widehat{R_\theta f}(\omega)$ — преобразование Фурье функции $R_\theta f(s)$ по переменной s :

$$\widehat{R_\theta f}(\omega) = \int_{-\infty}^{\infty} R_\theta f(s) \exp\{-i\omega s\} ds.$$

Метод оценивания многомерной плотности основан на следующих соображениях. Легко видеть, что если X — d -мерный случайный вектор с плотностью $f(x)$, то одномерная случайная величина $X\theta$ имеет плотность $R_\theta f(s)$ [7]. Таким образом, построив оценки одномерных плотностей $R_\theta f(s)$ по выборкам $(X\theta)_j, j = 1, \dots, N$, можно получить оценку многомерной плотности $f(x)$, используя методы обращения преобразования Радона.

3 Вейвлет-оценки одномерных вероятностных плотностей

Вейвлет-базисы приобрели большую популярность в задачах непараметрического оценивания функций благодаря своей адаптивности к локальным особенностям оцениваемых функций. В работе [8] предложено использовать вейвлет-разложение для оценивания вероятностных плотностей.

Рассмотрим неоднородный ортонормированный вейвлет-базис $\{\phi_{j_0,k}(s) = 2^{j_0/2} \phi(2^{j_0} s - k)\}_{k \in \mathbb{Z}}$ и $\{\psi_{j,k}(s) = 2^{j/2} \psi(2^j s - k)\}_{k \in \mathbb{Z}, j \geq j_0}$, где $\phi(s)$ и $\psi(s)$ — масштабирующая и вейвлет-функции, порождающие некоторый кратномасштабный анализ [9]. Будем полагать, что функции $\phi(s)$ и $\psi(s)$ имеют компактный носитель и непрерывно дифференцируемы достаточное число раз (этому требованию удовлетворяют, например, вейвлеты Добеши [10]). Кроме того, будем полагать, что функция $\phi(s)$ нормализована следующим образом:

$$\int_{-\infty}^{\infty} \phi(s) ds = 1.$$

Вероятностную плотность $g(s)$ можно формально представить в виде ряда

$$g(s) = \sum_{k \in \mathbb{Z}} a_{j_0,k} \phi_{j_0,k}(s) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} b_{j,k} \psi_{j,k}(s), \quad (2)$$

где

$$a_{j_0,k} = \int_{-\infty}^{\infty} g(s) \phi_{j_0,k}(s) ds; \quad b_{j,k} = \int_{-\infty}^{\infty} g(s) \psi_{j,k}(s) ds.$$

Если Y_1, \dots, Y_N — выборка из независимых одинаково распределенных случайных величин с плотностью $g(s)$, то

$$\hat{a}_{j_0,k} = \frac{1}{N} \sum_{m=1}^N \phi_{j_0,k}(Y_m); \quad \hat{b}_{j,k} = \frac{1}{N} \sum_{m=1}^N \psi_{j,k}(Y_m)$$

являются несмещенными оценками коэффициентов $a_{j_0,k}$ и $b_{j,k}$ соответственно. Нелинейная оценка вероятностной плотности $g(s)$ имеет вид [8]:

$$g_N(s) = \sum_{k \in \mathbb{Z}} \hat{a}_{j_0,k} \phi_{j_0,k}(s) + \sum_{j=j_0}^J \sum_{k \in \mathbb{Z}} \hat{b}_{j,k} \mathbf{1}(\hat{b}_{j,k} > T_{j,N}) \psi_{j,k}(s), \quad (3)$$

где $T_{j,N}$ — «порог», а J — еще один сглаживающий параметр. Первая сумма в (3) представляет собой несмещенную оценку первой суммы в (2). Во второй сумме необходимо производить пороговую обработку и усечение до J . Параметры $T_{j,N}$ и J позволяют находить компромисс между смещением и дисперсией.

4 Оценка многомерной плотности и ее свойства

Для построения оценки многомерной плотности нельзя использовать непосредственно формулу (1) в силу некорректности задачи обращения преобразования Радона. Вместо этого используется регуляризованный вариант этой формулы, в которой $|\omega|$ умножается на функцию окна $W_\alpha(|\omega|)$, зависящую от сглаживающего параметра α . Часто в качестве $W_\alpha(|\omega|)$ выбирается функция $W_\alpha(|\omega|) = \exp(-\alpha^2 \omega^2 / 2)$, т.е. фактически осуществляется свертка проекций с нормальным распределением. В результате вместо $f(x)$ оценивается

$$f_\alpha(x) = \frac{1}{2(2\pi)^d} \int_{S^{d-1}} \int_{-\infty}^{\infty} |\omega|^{d-1} \exp\left(-\frac{\alpha^2 \omega^2}{2}\right) \times \widehat{R_\theta f}(\omega) \exp\{i\omega(x\theta)\} d\omega d\theta \quad (4)$$

и оценка многомерной плотности вычисляется по формуле:

$$f_N(x) = \frac{1}{2(2\pi)^d} \int_{S^{d-1}} \int_{-\infty}^{\infty} |\omega|^{d-1} \exp\left(-\frac{\alpha^2 \omega^2}{2}\right) \times \widehat{R_N f}(\theta, \omega) \exp\{i\omega(x \cdot \theta)\} d\omega d\theta, \quad (5)$$

где $R_N f(\theta, s)$ — вейвлет-оценки одномерных плотностей $R_\theta f(s)$, построенные по правилу (3). На практике можно вычислить оценки одномерных плотностей только для конечного числа направлений $\theta_1, \dots, \theta_M$. Для случая, когда $f(x)$ имеет конечный носитель, некоторые равномерные оценки

близости функции плотности, восстановленной по конечному числу проекций, к истинной функции плотности можно найти в работах [11, 12]. Здесь будет предполагаться, что M достаточно велико ($M \gg N$), чтобы можно было игнорировать ошибку, возникающую из-за использования конечного числа направлений. Докажем, что равномерное расстояние между $f_N(x)$ и $f_\alpha(x)$ стремится к нулю в среднем.

Пусть $W^r(\mathbb{R}^d)$ — пространство Соболева, т. е. пространство всех таких функций $f(x)$, для которых

$$\|f\|_r^2 = \int_{\mathbb{R}^d} (1 + |\omega|^2)^r |\widehat{f}(\omega)|^2 d\omega < \infty.$$

Обозначим через $W_0^r(U)$ пространство всех $f(x) \in W^r$, носитель которых содержится в некотором компактном множестве $U \subset \mathbb{R}^d$.

Теорема 1. Пусть $f(x) \in W_0^r(U)$, $r > 0$. Пусть $T_{j,N} = C_r \sqrt{j/N}$ ($C_r > 0$ — константа, зависящая от r [8, 13]), j_0 фиксировано, а $J = J(N)$ выбирается из условия $2^{J(N)} \approx (N/\log N)^{1/(2r+1)}$. Тогда

$$\mathbf{E} \sup_{x \in \mathbb{R}^d} |f_\alpha(x) - f_N(x)| \leq C_d \alpha^{1/2-d} N^{-(2r+d-1)/(4r+2d)}, \quad (6)$$

где C_d — некоторая положительная константа.

Доказательство. Используя формулы (4) и (5) и неравенство Коши–Буняковского, получаем:

$$\begin{aligned} & \mathbf{E} \sup_{x \in \mathbb{R}^d} |f_\alpha(x) - f_N(x)| = \\ & = \mathbf{E} \sup_{x \in \mathbb{R}^d} \frac{1}{2(2\pi)^d} \left| \int_{S^{d-1}} \int_{-\infty}^{\infty} |\omega|^{d-1} \exp\left(-\frac{\alpha^2 \omega^2}{2}\right) \times \right. \\ & \quad \left. \times \left(\widehat{R_\theta f}(\omega) - \widehat{R_N f}(\theta, \omega) \right) \exp\{i\omega(x\theta)\} d\omega d\theta \right| \leq \\ & \leq \frac{1}{2(2\pi)^d} \int_{S^{d-1}} \int_{-\infty}^{\infty} |\omega|^{d-1} \exp\left(-\frac{\alpha^2 \omega^2}{2}\right) \mathbf{E} \times \\ & \quad \times \left| \widehat{R_\theta f}(\omega) - \widehat{R_N f}(\theta, \omega) \right| d\omega d\theta \leq \\ & \leq \frac{1}{2(2\pi)^d} \int_{S^{d-1}} \left(\int_{-\infty}^{\infty} |\omega|^{2(d-1)} \exp(-\alpha^2 \omega^2) d\omega \right)^{1/2} \times \\ & \quad \times \left(\int_{-\infty}^{\infty} \left(\mathbf{E} \left| \widehat{R_\theta f}(\omega) - \widehat{R_N f}(\theta, \omega) \right|^2 d\omega \right)^{1/2} d\theta \leq \end{aligned}$$

$$\begin{aligned} & \leq \frac{1}{2(2\pi)^d} \int_{S^{d-1}} \left(\int_{-\infty}^{\infty} |\omega|^{2(d-1)} \exp(-\alpha^2 \omega^2) d\omega \right)^{1/2} \times \\ & \quad \times \left(\int_{-\infty}^{\infty} \mathbf{E} \left| \widehat{R_\theta f}(\omega) - \widehat{R_N f}(\theta, \omega) \right|^2 d\omega \right)^{1/2} d\theta. \end{aligned}$$

Поскольку $f(x) \in W_0^r(U)$, функция $R_\theta f(s) \in W^p(\mathbb{R})$ для любого $\theta \in S^{d-1}$, где $p = r + (d-1)/2$ [14]. В работе [8] показано, что

$$\begin{aligned} & \left(\int_{-\infty}^{\infty} \mathbf{E} \left| \widehat{R_\theta f}(\omega) - \widehat{R_N f}(\theta, \omega) \right|^2 d\omega \right)^{1/2} = \\ & = \left(\int_{-\infty}^{\infty} \mathbf{E} |R_\theta f(s) - R_N f(\theta, s)|^2 ds \right)^{1/2} \leq \\ & \leq C_p N^{p/(2p+1)}, \end{aligned}$$

где $C_p > 0$ — некоторая константа. Следовательно, существует константа $C_d > 0$ такая, что

$$\mathbf{E} \sup_{x \in \mathbb{R}^d} |f_\alpha(x) - f_N(x)| \leq C_d \alpha^{1/2-d} N^{-p/(2p+1)}.$$

Теорема доказана.

Функция $f_\alpha(x)$ аппроксимирует функцию $f(x)$, и чем меньше α , тем точнее эта аппроксимация. Если дополнительно предположить, что функция $f(x)$ равномерно регуляерна по Липшицу с показателем $1 \geq \gamma > 0$ на U , т. е. существует константа $C_L > 0$, такая что

$$\sup_{x \in U} |f(x) - f(x+y)| \leq C_L |y|^\gamma \quad \forall y \in \mathbb{R}^d, \quad (7)$$

то можно оценить скорость стремления к нулю в среднем равномерного расстояния между $f_N(x)$ и $f(x)$.

Теорема 2. Пусть выполнены условия теоремы 1 и условие регулярности (7). Тогда

$$\mathbf{E} \sup_{x \in \mathbb{R}^d} |f(x) - f_N(x)| \leq C_{d,\gamma} N^{-((2r+d-1)\gamma)/((4r+2d)(\gamma+d-1/2))}, \quad (8)$$

где $C_{d,\gamma}$ — некоторая положительная константа.

Доказательство. Имеем

$$\begin{aligned} \mathbf{E} \sup_{x \in \mathbb{R}^d} |f(x) - f_N(x)| & \leq \sup_{x \in \mathbb{R}^d} |f(x) - f_\alpha(x)| + \\ & + \mathbf{E} \sup_{x \in \mathbb{R}^d} |f_\alpha(x) - f_N(x)|. \end{aligned}$$

Далее

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d} |f(x) - f_\alpha(x)| \leq \\ & \leq \frac{1}{(2\pi\alpha^2)^{d/2}} \int_{\mathbb{R}^d} |f(x) - f(x+y)| \exp\left(-\frac{|y|^2}{2\alpha^2}\right) dy \leq \\ & \leq \frac{C_L}{(2\pi\alpha^2)^{d/2}} \int_{\mathbb{R}^d} |y|^\gamma \exp\left(-\frac{|y|^2}{2\alpha^2}\right) dy \leq C_{d,L} \alpha^\gamma, \quad (9) \end{aligned}$$

где $C_{d,L}$ — положительная константа. Выберем $\alpha = \alpha(N) = N^{-\beta}$, где

$$\beta = \frac{2r + d - 1}{(4r + d)(\gamma + d - 1/2)}.$$

При таком выборе β степени у N в оценках (6) и (9) равны. Следовательно, используя (6) и (9), получаем (8). Теорема доказана.

Литература

1. *Silverman B. W.* Density estimation for statistics and data analysis. — London: Chapman and Hall, 1986. 176 p.
2. *Vidacovic B.* Statistical modeling by wavelets. — New York, NY, USA: John Wiley and Sons, 1999. 408 p.
3. *Tribouley K.* Practical estimation of multivariate densities using wavelet methods // Stat. Neerl., 1995. Vol. 49. P. 41–62.
4. *Masry E.* Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series // Stoch. Proc. Appl., 1997. Vol. 67. P. 177–193.
5. *Dong J., Jiang R.* Multinomial probability estimation by wavelet thresholding // Communication in Statistics. Theory and Methods, 2009. Vol. 38. P. 1486–1507.
6. *Chesneau C., Dewan I., Doosti H.* Nonparametric estimation of a two dimensional continuous-discrete density function by wavelets // Stat. Methodol., 2014. Vol. 18. P. 64–78.
7. *O'Sullivan F., Pawitan Y.* Multidimensional density estimation by tomography // J. Roy. Stat. Soc. B, 1993. Vol. 55. No. 2. P. 509–521.
8. *Donoho D. L., Johnstone I. M., Kerkyacharian G., Picard D.* Density estimation by wavelet thresholding // Ann. Stat., 1996. Vol. 23. P. 508–539.
9. *Mallat S.* A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 851 p.
10. *Daubechies I.* Ten lectures on wavelets. — Philadelphia, PA, USA: SIAM, 1992. 357 p.
11. *Khalfin L. A., Klebanov L. B.* A solution of the computer tomography paradox and estimating the distances between the densities of measures with the same marginals // Ann. Probab., 1994. Vol. 22. P. 2235–2241.
12. *Шестаков О. В., Савенков Т. Ю.* Оценка расстояния между плотностями вероятностных мер, имеющих близкие проекции // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн., 2001. № 4. С. 44–46.
13. *Hall P., Patil P.* Formulae for mean integrated squared error of nonlinear wavelet-based density estimators // Ann. Stat., 1995. Vol. 23. No. 3. P. 905–928.
14. *Smith K. T., Solmon D. C., Wagner S. L.* Practical and mathematical aspects of the problem of reconstructing objects from radiographs // Bull. Amer. Math. Soc., 1977. Vol. 83. P. 1227–1270.

Поступила в редакцию 02.03.15

NONPARAMETRIC ESTIMATION OF MULTIDIMENSIONAL DENSITY WITH THE USE OF WAVELET ESTIMATES OF UNIVARIATE PROJECTIONS

O. V. Shestakov^{1,2}

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

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

Abstract: The paper explores the computerized tomography method of inverting the Radon transformation for obtaining statistical estimates of multidimensional probability densities. This method utilizes nonlinear wavelet estimators of univariate projections to construct the multidimensional density estimate. Nonlinear wavelet estimators possess the ability to adapt to the local properties of the estimated density function and, therefore, are less sensitive to the singular points than linear estimators. Another important practical feature of the considered

method is its parallel structure, which allows a considerable speedup of constructing the estimates on the computers supporting parallel processing. It is also proved that under some regularity conditions, the uniform distance between the constructed estimate and the true multidimensional probability density converges to zero in the mean, and some estimates of the rate of this convergence are obtained.

Keywords: wavelets; multidimensional density; Radon transformation

DOI: 10.14357/19922264150210

Acknowledgments

The research was financially supported by the Russian Foundation for Basic Research (project 15-07-02652).

References

1. Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall. 176 p.
2. Vidacovic, B. 1999. *Statistical modeling by wavelets*. New York, NY: John Wiley and Sons. 408 p.
3. Tribouley, K. 1995. Practical estimation of multivariate densities using wavelet methods. *Stat. Neerl.* 49:41–62.
4. Masry, E. 1997. Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stoch. Proc. Appl.* 67:177–193.
5. Dong, J., and R. Jiang. 2009. Multinomial probability estimation by wavelet thresholding. *Communication in Statistics. Theory and Methods* 38:1486–1507.
6. Chesneau, C., Dewan, I., and H. Doosti. 2014. Non-parametric estimation of a two dimensional continuous-discrete density function by wavelets. *Stat. Methodol.* 18:64–78.
7. O’Sullivan, F., and Y. Pawitan. 1993. Multidimensional density estimation by tomography. *J. Roy. Stat. Soc. B* 55(2):509–521.
8. Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard. 1996. Density estimation by wavelet thresholding. *Ann. Stat.* 23:508–539.
9. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 851 p.
10. Daubechies, I. 1992. *Ten lectures on wavelets*. Philadelphia, PA: SIAM. 357 p.
11. Khalfin, L. A., and L. B. Klebanov. 1994. A solution of the computer tomography paradox and estimating the distances between the densities of measures with the same marginals. *Ann. Probab.* 22:2235–2241.
12. Shestakov, O. V., and T. Yu. Savenkov. 2001. Otsenka rasstoyaniya mezhdru plotnostyami veroyatnostnykh mer, imeyushchikh blizkie proektsii [Estimating the distances between the densities of measures with ε -identical marginals]. *Vestn. Mosk. Un-ta. Ser. 15. Vychisl. matem. i kibern.* [Moscow University Computational Mathematics and Cybernetics] 4:44–46.
13. Hall, P., and P. Patil. 1995. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Stat.* 23(3):905–928.
14. Smith, K. T., D. C. Solmon, and S. L. Wagner. 1977. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bull. Amer. Math. Soc.* 83:1227–1270.

Received March 2, 2015

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

АССОЦИАТИВНЫЕ ПОРТРЕТЫ ПРЕДМЕТНОЙ ОБЛАСТИ — ИНСТРУМЕНТ АВТОМАТИЗИРОВАННОГО ПОСТРОЕНИЯ СИСТЕМ BIG DATA ДЛЯ ИЗВЛЕЧЕНИЯ ЗНАНИЙ: ТЕОРИЯ, МЕТОДИКА, ВИЗУАЛИЗАЦИЯ, ВОЗМОЖНОЕ ПРИМЕНЕНИЕ*

И. В. Галина¹, Е. Б. Козеренко², Ю. И. Морозова³, Н. В. Сомин⁴, М. М. Шарнин⁵

Аннотация: Представлена методика создания систем извлечения знаний, основанная на подходе, главным инструментом которого является автоматизированное формирование ассоциативного портрета предметной области (АППО) и построение семантического контекстного пространства (СКП). Идеология АППО базируется на дистрибутивной гипотезе, утверждающей, что семантически близкие (или связанные) лексемы имеют похожий контекст и, наоборот, при похожем контексте лексемы семантически близки. В применяемой модели используется расширенная гипотеза, включающая исследование сходства и различия в контекстах не только отдельных лексем, но и произвольных многолексемных фрагментов — значимых словосочетаний (ЗС). Приведены примеры реализованных проектов для различных предметных областей (ПО).

Ключевые слова: семантическое моделирование; ассоциативные связи; математическая статистика; дистрибутивная семантика; big data; автоматизированные системы извлечения знаний; электронные корпуса ЕЯ-текстов; семантический поиск; интеллектуальные интернет-технологии

DOI: 10.14357/19922264150211

1 Введение. Постановка проблемы

Создание лингвистической и предметной базы и разработка эффективной методики для систем извлечения и обработки знаний на сверхбольших объемах текста (big data), свободно распространяемых в Интернете, относятся к фундаментальным задачам компьютерной лингвистики. При этом актуально существенное снижение трудозатрат, увеличение скорости автоматизированной обработки с минимальным участием человека, повышение точности семантического поиска, качества результатов работы системы. Один из традиционных методов — составление вручную универсальных хорошо выверенных тезаурусов, который требует колоссальных трудовых и временных ресурсов, поэтому в последние годы ведется поиск новых пу-

тей семантического моделирования для создания лингвистической предметно-ориентированной базы знаний.

Одним из таких направлений является инструмент, названный авторами «ассоциативный портрет предметной области», на основе которого можно с достаточной надежностью решать большинство важнейших проблем, связанных с информационным поиском, извлечением структур знаний и построением классификаций [1–3]. Разработка данного инструмента основана на гипотезе о привлечении ассоциативных связей для определения значения, полный смысл которого выявляется с помощью контекстных окружений, что дает возможность автоматизации процесса разграничения значений и извлечения знаний из текстов.

Ассоциативный портрет — это совокупность наиболее характерных предметных и лингвистических знаний, свойственных определенной ПО.

* Работа выполнена при поддержке РФФИ (проект 13-07-00272).

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

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

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, yulia-iri@yandex.ru

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

⁵ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, mc@keywen.com

Под предметными знаниями понимаются присущие ПО термины, понятия, связанные различного рода ассоциативными связями. Лингвистические знания — это варианты словарного (лексического) выражения понятий. Суть предлагаемой методики заключается в анализе текстов естественного языка (ЕЯ), относящихся к определенной ПО, с выявлением ассоциативных связей и построении на этой основе ассоциативных портретов, характеризующих данную область. В качестве таких портретов предлагается СКП, служащее для представления как лингвистических, так и предметных знаний.

Создаваемая методика извлечения знаний на основе АППО строится на сочетании методов статистики, корпусной лингвистики, дистрибутивной семантики и машинного обучения и реализуется в комплексной технологии, которая предполагает разработку лингвистической компоненты — СКП, представляющего собой словарь ЗС ПО, элементы которого соединены ассоциативными связями. Таким образом, инструмент выделения АППО из текстов лег в основу технологии, позволяющей автоматически выявлять ЗС и ассоциативные связи из неструктурированных естественно-языковых текстов (ЕЯ-текстов), проходя через этапы предварительного обучения на текстах, в том числе взятых из различных интернет-ресурсов. Обработка больших массивов текстов (big data), постоянно пополняемых в интернет-среде, позволяет собирать необходимые статистические данные для формирования достаточно полной картины о ПО, представленной в виде СКП. Возможность проводить машинное обучение на большом количестве примеров придает системе определенную гибкость и улучшает результаты.

Отметим, что традиционное для дистрибутивной семантики использование диагностических смысловых различительных контекстов для разграничения дистрибутивных признаков исследуемых лексем в предлагаемой модели дополняется расширенным пониманием дистрибутивной гипотезы, при котором семантически близкими (или связанными) при сходстве контекстов признаются не только отдельные лексемы (термины), но и произвольные многолексемные фрагменты текста (ЗС). Предлагаемый подход [1–7] предусматривает исследование различных типов и источников контекста, а также различных методов выделения контекста и оценки силы ассоциативной связи по контекстным векторам; при этом исследуется как простая совместная встречаемость ЗС в тексте, так и совместная встречаемость ЗС в текстах в рамках заданных лексико-синтаксических шаблонов. Кроме того, учитывается порядок слов в лексических

последовательностях, что позволяет добиться более качественного выполнения вышеперечисленных задач, в том числе выявления ассоциативных связей словосочетаний, объектов и именованных сущностей. Объективные критерии статистической оценки, включенные в методику наряду с логико-лингвистическими методами и правилами дистрибутивной семантики, применяются для разрешения семантической неоднозначности лексем (терминов) и ЗС на основе вероятностных механизмов, что повышает скорость и эффективность семантического поиска, а также обеспечивает автоматизированное отнесение обрабатываемого текста к той или иной ПО. Некоторым побочным ответвлением указанной методики стала разработка средств визуализации АППО, которые позволяют точно отобразить координаты выявленных объектов в двумерном СКП и использовать эвристические методы для установления степени близости объектов [8]. Исследования по выявлению АППО проводятся на материале ЕЯ-текстов на русском (в основном) и английском (в меньшем объеме) языках.

Разработка и реализация на практике методики АППО и СКП предусматривает следующие этапы:

1. Создание и постоянное пополнение корпусов ЕЯ-текстов на основе обработки большого потока данных из Интернета (круглосуточным мониторингом на двух серверах).
2. Создание методов автоматизированного отнесения текстов к различным ПО.
3. Применение методов статистики, дистрибутивной семантики, математической лингвистики и машинного обучения на постоянно обновляемой выборке текстов для совершенствования технологии и создания автоматизированной системы выявления ЗС и их ассоциативных связей.
4. Обработка полученных данных и построение СКП различных ПО с их визуализацией.
5. Проектирование общей архитектуры системы, основанной на предлагаемой технологии.

Конечным итогом развития инструмента АППО будет создание комплекса программного обеспечения, отлаженного на обучающих выборках и решающего следующие задачи:

- поиск в интернет-среде корпусов текстов, связанных с заданной ПО;
- выделение из полученного электронного корпуса текстов терминологического материала и оценка значимости каждого термина на основе статистических критериев;

- выделение семантических связей (ассоциаций) из корпуса текстов с использованием методов машинного обучения на основе АППО, а также стандартных тезаурусов в качестве обучающей выборки;
- выделение терминологических словосочетаний и именованных сущностей из корпуса текстов на основе методов машинного обучения и кластерного анализа в пространстве контекстных векторов (СКП);
- демонстрация возможностей методики на примере визуализации ассоциативных портретов ЗС определенной ПО.

2 Современное состояние исследований

В настоящее время известно значительное число работ по автоматическому извлечению семантических связей из больших массивов текстов на ЕЯ [1–33].

Наиболее успешные подходы используют метод дистрибутивной семантики [1–3, 5–7, 10, 12–15, 18, 19, 25] и модели семантических векторных пространств (СВП). Дистрибутивная семантика занимается вычислением степени семантической близости между лингвистическими единицами на основании их контекстных окружений, или дистрибутивных признаков [14, 15, 18]. Основными сферами применения дистрибутивных моделей являются: разрешение лексической неоднозначности, информационный поиск, кластеризация документов, автоматическое формирование словарей (словарей семантических отношений, двуязычных словарей), создание семантических (визуальных) карт, определение тематики документа и др. Теоретические основы исследования дистрибуций восходят еще к методике З. Харриса. Похожие идеи выдвигали и основоположники структурной лингвистики Ф. де Соссюр и Л. Витгенштейн. Дистрибутивная семантика, как и все опирающиеся на нее подходы, базируется на гипотезе о том, что лингвистические элементы со схожей дистрибуцией имеют близкие значения [24, 26]. В основе всех современных вариантов дистрибутивного подхода лежат количественные оценки, которые характеризуют совместную встречаемость языковых единиц текста в контекстах определенной величины.

Для оценки степени связанности этих единиц вводится коэффициент «силы связи», который рассчитывается по некоторой формуле. Обычно в качестве вычислительного инструмента и способа представления моделей используется линейная алгебра.

Информация о дистрибуции лингвистических единиц представляется в виде многомерных векторов, а семантическая близость между лингвистическими единицами вычисляется как расстояние между векторами. Многомерные векторы образуют матрицу, где каждый вектор соответствует лингвистической единице (слову или словосочетанию), а каждое измерение вектора соответствует контексту (документу, параграфу, предложению, словосочетанию, слову). Для вычисления меры близости между векторами могут использоваться различные формулы: расстояние Минковского, расстояние Манхэттена, евклидово расстояние, расстояние Чебышева, скалярное произведение, косинусная мера. Наиболее популярной в настоящее время является косинусная мера. Вне зависимости от выбранной формулы расчета в ней обычно используются характеристики совместной встречаемости пар слов и одиночной встречаемости каждого из слов. Величина контекста, в рамках которого осуществляются подсчеты коэффициентов «силы связи», как показывают результаты исследований, позволяет наиболее вероятно устанавливать:

- при малых размерах контекста, ограниченного одним или двумя соседними словами, — контактные синтагматические связи словосочетаний;
- при размере 5–10 слов — дистантные синтагматические связи и парадигматические отношения;
- дальнейшее увеличение ширины контекста до 50–100 слов (размер предложения, сверхфразового единства, абзаца) — тематические связи между словами.

Тематические связи могут оказаться доминирующими, если принять размер контекста величиной с сам текст. Модели векторных пространств находят все более широкое применение в исследованиях, связанных с семантическими моделями ЕЯ, и имеют разнообразный спектр потенциальных и действующих приложений [22, 33]. Данная область в настоящее время является одной из наиболее актуальных. Следует отметить модели Word-Space Model [24] и Semantic Space Model [17]. В основе этих работ М. Sahlgren лежит пространство лексем и термов. Такое пространство базируется на распределении слов в корпусе текстов с целью представления их семантической связанности путем оценки пространственной близости. В настоящее время существуют разнообразные варианты подобных методов исследования и построенных моделей. Так, Себастьян Падо из университета Саарланд (Германия) и Мирелла Лапата из Шеффилдского университета

(Великобритания) рассматривают вопрос построения семантических пространств на основе традиционных векторных моделей с учетом синтаксических отношений. Семантические свойства слов представляются в виде частотной матрицы, каждый ряд которой соответствует уникальному целевому слову, а каждая колонка — лингвистическому контексту. Семантическая информация извлекается из текстов большого объема на основе анализа окружения слова. Слово рассматривается как точка в многомерном семантическом пространстве. На основе близости между точками семантического пространства вычисляется семантическое сходство между словами с использованием метрик. Анализ семантического сходства выполняется на основе статистических методов с расчетом частотности появления в тексте близких точек семантического пространства. В процессе анализа рассматриваются метрики Евклида, Жаккара, Кульбака—Лейблера и др. По результатам исследования делается вывод о том, что контекстное окружение играет важную роль в распознавании лексических отношений между словами.

Существует множество разновидностей моделей дистрибутивной семантики, которые различаются по следующим параметрам:

- тип контекста (размер контекста, правый или левый контекст, ранжирование);
- количественная оценка частоты встречаемости слова в данном контексте (абсолютная частота, энтропия, совместная информация и пр.);
- метод вычисления расстояния между векторами (косинусная мера, скалярное произведение, расстояние Минковского и пр.);
- метод уменьшения размерности матрицы (случайная проекция, сингулярное разложение и пр.).

Наиболее известными современными моделями дистрибутивной семантики являются латентный семантический анализ, разработанный для решения проблемы синонимии при информационном поиске [19], и модель языка как гиперпространства, разработанная в виде модели семантической памяти человека [13].

Концепция СВП впервые была реализована в информационно-поисковой системе SMART [9]. Идея СВП состоит в представлении каждого документа из коллекции текстов в виде точки в многомерном семантическом пространстве, которой соответствует вектор в векторном пространстве ЗС. Точки, расположенные ближе друг к другу в этом пространстве, считаются более близкими по смыслу. Пользовательский запрос рассматривается как

псевдодокумент и тоже представляется как точка в этом же пространстве. Документы сортируются в порядке возрастания расстояния, т. е. в порядке уменьшения семантической близости от запроса, и в таком виде предоставляются пользователю.

В настоящее время многие поисковики используют СВП для измерения степени близости запроса и найденных документов [23]. Varoni и Lenci [25] предложили обобщенную модель, названную «дистрибутивная память», которая является обобщением ранее известных моделей векторных пространств (vector spaces), семантических пространств (semantic spaces), пространств слов (word spaces), семантических моделей корпусной статистики (corpus-based semantic models) и дистрибутивных семантических моделей (distributional semantic models). Rapp в своих исследованиях [4] использовал контекстное векторное пространство для оценки семантической близости слов. Его система достигла результата 92,5% на тесте по выбору наиболее подходящего синонима из стандартного теста английского языка TOEFL (Test of English as Foreign Language), в то время как средний результат людей был 64,5%.

В разрабатываемом проекте используется более широкое понятие, выраженное термином «семантическое контекстное пространство», где точки пространства соответствуют контекстным векторам не отдельных слов (терминов), а ЗС [1, 6]. Под ЗС понимаются лексические последовательности, имеющие тенденцию к совместной встречаемости. В лингвистике для обозначения ЗС используется также термин «коллокация», впервые введенный в «Словаре лингвистических терминов» О. С. Ахмановой еще в 1966 г. Исследованиям коллокаций русского языка посвящен большой объем литературы (см., например, [11, 27]). В теоретической лингвистике под коллокациями понимают словосочетания из двух или более слов, которые обуславливают друг друга семантически и грамматически [20]. В корпусной лингвистике коллокациями называют статистически устойчивые словосочетания, причем они могут быть как фразеологизмами, так и свободными. Для выделения ЗС в компьютерной лингвистике используются различные статистические меры (названные мерами ассоциации или ассоциативной связанности — association measures (англ.)), вычисляющие силу связи между элементами в составе коллокации. В литературе упоминаются десятки способов расчета мер ассоциативной связанности; например, MI, *t*-score и log-likelihood методы вычисления.

Мера MI (mutual information), введенная в работе [12], сравнивает зависимые контекстно-связанные частоты с независимыми частотами слов

в тексте. Если значение MI превосходит определенное пороговое значение, то словосочетание считают статистически значимым. Мера MI вычисляется по следующей формуле:

$$MI = \log_2 \frac{f(n, c)N}{f(n)f(c)},$$

где n — первое слово словосочетания; c — второе слово словосочетания; $f(n, c)$ — частота совместной встречаемости двух слов; $f(n)$ и $f(c)$ — абсолютные частоты встречаемости каждого слова по отдельности; N — общее число словоупотреблений в корпусе.

Мера t -score также используется при ответе на вопрос, насколько неслучайным является сочетание двух или более слов в тексте. Для вычисления t -score используется следующая формула:

$$t\text{-score} = \frac{f(n, c) - f(n)f(c)/N}{\sqrt{f(n, c)}}.$$

Также достаточно часто применяется мера, известная под названием \log -likelihood, или логарифмическая функция правдоподобия, введенная в статье [10]. Для вычисления \log -likelihood применяется формула:

$$\log\text{-likelihood} = 2 \sum f(n, c) \log_2 \frac{f(n, c)N}{f(n)f(c)}.$$

Новым и чрезвычайно важным в предлагаемом авторами подходе — по сравнению с существующими методиками — является то, что изначально СКП представляет собой абстрактную модель, не связанную с конкретной ПО. Приложение разрабатываемой модели связано с автоматизированным выбором той или иной ПО и ее компонент: ЗС, ассоциативных связей и, соответственно, контекстов для их выделения. В результате строится система множественных ассоциативных связей, а затем формируется АППО. Разработчики СВП отмечают, что основная проблема этого подхода заключается в трудностях учета порядка слов, составляющих контексты. В рамках же данного проекта (СКП) эта проблема решается путем перехода от контекста слов к контексту ЗС. Отметим, что первоначально технология СВП развивалась для английского языка. Настоящий проект с использованием модели СКП уже на первом этапе его реализации работает как с русскими (основное внимание будет уделено все же именно русскому языку), так и с английскими ЕЯ-текстами. Также в дальнейшем возможно включение других языков. Таким образом, модель СКП является расширением моделей СВП.

3 Материал исследования

Работы над проектом по созданию системы для извлечения знаний, основанной на подходе, главным инструментом которого является автоматизированное формирование АППО и построение СКП, ведутся коллективом авторов на протяжении последних нескольких лет [1–3, 5–7, 16, 21, 28, 29, 33].

Данный проект реализуется на сверхбольших объемах данных (big data), свободно представленных в интернет-среде, что существенно повышает качество формируемых АППО. Материалом исследований служат свободно распространяемые сетевые русскоязычные ЕЯ-тексты, а также — для некоторых ПО — тексты на английском языке. В дальнейшем возможно расширение проекта на другие языки.

За данный период были разработаны методы итеративного формирования корпусов по определенной ПО, на основе которых созданы постоянно пополняемые из Интернета электронные корпуса текстов. Для решения этой задачи научным коллективом были арендованы 2 сервера, на которых запущено непрерывное скачивание больших массивов сетевых текстов (big data) и разбиение этого массива на ПО.

К настоящему времени обработаны следующие ПО:

- на русском языке — «Дистанционное зондирование сельскохозяйственных земель», «Психология делового общения»;
- «Бизнес-процессы предприятий»;
- «Социально-политический портрет регионов России» (в качестве обучающей выборки для этой тематики взято сужение темы до одного федерального региона — «Социально-политический портрет Татарстана», источниками сообщений для этой темы являются: официальные сайты государственных учреждений, партий; общероссийские и региональные средства массовой информации (СМИ); социальная сеть Вконтакте, микроблоги Твиттера и т. п.);
- «Мониторинг общественного мнения в социально-политической сфере», суженная до обучающей выборки «Протестная активность»;
- АНПА («Автономные необитаемые подводные аппараты») на русском и английском языке;
- «Computer Science» (на английском языке) и др.

По данным темам найдено и обработано от десятка тысяч до полутора миллионов документов на русском и английском языках, общим объемом около 160 ГБ. Так, по теме «Социально-политический портрет Татарстана» сформирован дайджест

текстов объемом в 20 ГБ, а по тематике «Протестная активность» — в 28 ГБ.

4 Методика построения ассоциативного портрета предметной области

4.1 Основа методики

Методика построения АППО опирается на современный уровень развития науки и на имеющийся у коллектива авторов большой задел в области теоретических исследований и создания действующих систем различных лингвистических процессоров и обработки текстовых знаний. Элементы предлагаемого подхода уже были частично апробированы при создании энциклопедии ключевых понятий KEYWEN (содержит 260 000 статей и 5 000 000 ключевых фраз, число которых постоянно растет) и в ряде работ других участников проекта.

Система KEYWEN представляет собой средство построения больших энциклопедий по материалам Интернета и на их основе составления рефератов и аналитических статей. Имеется опыт построения корпуса английских текстов из Интернета размером более 1 ТБ, проведены эксперименты по построению русских корпусов текстов для ряда ПО.

Авторский коллектив много лет ведет исследования в области аналитической обработки текстов и извлечения информационных объектов (именованных сущностей) и их связей. Разработан ряд прикладных систем, основанных на глубинном анализе текстов:

- «Криминал» и «Аналитик» — системы, извлекающие информацию из милицейских сводок для логико-аналитических решений, которые служат основой для проведения следственных действий;
- «Semantix» — многоязычный лингвистический процессор, обеспечивающий выделение объектов (сущностей) и их связей из текстов на английском и русском языке;
- «Антитеррор» — система, извлекающая из текстов информацию о террористах и террористических акциях с автоматическим формированием базы знаний;
- «Резюме» — система семантической обработки резюме и их отображения в формат сайта крупной рекрутинговой компании;
- «Памятники» — обеспечивает извлечение информации из текстов описания памятников культуры и формирования ролевых функций относящихся к ним лиц;

«ДИЕС» — диалоговая среда, обеспечивающая построение конкретных систем языкового взаимодействия с пользователем;

«ДЕКЛАР» — программная среда, включающая в себя систему представления знаний и язык программирования задач, связанных с обработкой знаний. В рамках этой среды авторами разработаны системы лексико-морфологического и синтактико-семантического анализа русских и английских текстов, которые используются при извлечении знаний.

Таким образом, разработка инструмента АППО как главного механизма построения системы извлечения знаний ведется с привлечением некоторых теоретических и практических достижений, унаследованных авторским коллективом от прошлых разработок, а также путем создания новых алгоритмов методики и программных средств реализации.

4.2 Развитие методов на основе статистической парадигмы

Подход, создаваемый на основе АППО, носит комплексный характер, так как использует методы корпусной лингвистики, статистические механизмы, сочетающие приемы дистрибутивной семантики и математической статистики (контекстные векторы, веса, вероятности и др.), а также некоторые логико-лингвистические методы и способы машинного обучения для построения АППО (базового понятия, на которое опирается СКП). Наиболее близки по духу к этому проекту методы известного британского исследователя Karen Spärk Jones [14, 15, 18].

Особенность предлагаемого подхода состоит в возможности создавать и использовать АППО автоматически. В связи с этим трудоемкость разработки АППО на несколько порядков ниже создания традиционных тезаурусов, что позволяет АППО стать реальным дополнением тезаурусов в новом поколении интеллектуальных интернет-технологий. При таком подходе достигается большая полнота охвата терминологии лишь при незначительном снижении точности. Тем самым преодолевается главная проблема — трудоемкости ручного труда лингвистов, значительно увеличивается скорость.

Данный проект создается в несколько этапов. На первом этапе был разработан макетный вариант алгоритма формирования АППО на материале различных ПО. Создан и успешно развивается основной компонент комплексной методики — модуль статистического анализа для алгоритма формирования АППО. Тематический поиск ЕЯ-текстов,

скачиваемых из Интернета, формирование больших корпусов этих текстов, разбиение их на ПО, выделение ЗС и создание словарей ЗС осуществляются в автоматизированном режиме и соединены в единый итеративный алгоритм, который по сути и является методикой получения АППО для любой ПО. Отдельные ее этапы будут уточняться и совершенствоваться в процессе дальнейшей работы.

На втором этапе существования проекта методика автоматического формирования АППО получила дальнейшее развитие в виде уточненных механизмов семантического поиска в пространстве ЕЯ-текстов из Интернета и детального улучшения алгоритмов методики, разработанных в первый год существования проекта, — за счет обучения модуля статистического анализа на тестовых выборках с последующей его апробацией на реальных текстовых потоках big data из различных ПО. Семантический поиск — первое звено в неразрывной цепочке методики формирования АППО, который осуществляется в виде постоянного интернет-серфинга, ведущегося как с помощью специально разработанных для данного проекта программных средств семантического поиска (частью которых стали механизмы поиска, функционирующие в собственной разработке научного коллектива — энциклопедии ключевых понятий KEYWEN [16, 21, 28]), так и с привлечением крупнейших поисковиков, каталогов библиотек, электронных магазинов и т. п. Круглосуточный мониторинг сети ведется на двух рабочих станциях (серверах), что обеспечивает обработку большого потока сетевых данных, из которых формируются и постоянно пополняются корпуса ЕЯ-текстов. Механизмы поиска были усовершенствованы на обучающих и тестовых выборках и апробированы при непосредственном интернет-серфинге для поиска и извлечения ЕЯ-текстов определенных ПО (подробнее об алгоритмах поиска см. ниже).

Коллективом авторов были проведены эксперименты на тестовых выборках и составлен план дальнейших экспериментов на весь срок выполнения проекта; на данном этапе продолжается разработка и улучшение модулей и алгоритмов программного обеспечения.

В настоящее время методика позволяет автоматически выделять ключевые слова (отдельные слова-лексемы, термины и ЗС) из ЕЯ-текстов, выявлять ассоциативные связи между ними, задавать иерархию терминов и ЗС методом кластеризации и строить словари ЗС различных ПО, а также — в качестве сопутствующего продукта — создавать визуальные карты ПО. На выходе алгоритма получается достаточно информативное «облако» ассоциативных связей для автоматически выделенных

лексем (терминов) и ЗС, из которых и формируются АППО. Собственно, полученный «пучок» ассоциативных связей определяет значение термина (ЗС) и его место в терминосистеме исследуемой ПО, и наоборот: выявленные ассоциативные связи позволяют отнести рассматриваемый термин и/или ЗС к той или иной тематике.

На втором этапе работа была сосредоточена на детализации алгоритма формирования АППО и завершилась созданием успешно функционирующего варианта. Ядро алгоритма — модуль статистического анализа, основанный на вычислительных методах дистрибутивной семантики; в частности, используется косинусная мера вычисления силы ассоциативной связи (ассоциативная мера) между контекстными векторами (компонентами вектора ЗС являются частоты совместной встречаемости данного ЗС с другими ЗС в одном и том же контексте). Применяв формулу вычисления косинусной меры между контекстными векторами, получаем коэффициенты семантической близости между рассматриваемыми ЗС.

Идеология АППО базируется на дистрибутивной гипотезе, утверждающей, что семантически близкие (или связанные) лексемы имеют похожий контекст и, наоборот, при похожем контексте лексемы семантически близки. В применяемой модели используется расширенная гипотеза, включающая исследование сходства и различия в контекстах не только отдельных лексем, но и произвольных многолексемных фрагментов — ЗС.

В качестве теоретической модели АППО в настоящем проекте принято СКП. Формально СКП определяется как граф $G = (V, E)$ с узлами v из V , представляющими значимые термины/словосочетания ПО, и дугами графа $(v_i, v_j, \text{Link}, w_{ij})$ из E , описывающими отношения/связи между словосочетаниями, где w_{ij} — это вес, выражающий силу связи, а Link — тип связи, определяемый типом лексико-синтаксического шаблона/конструкции, связывающей словосочетания. Содержательно СКП является словарем ЗС, элементы которого связаны ассоциативными связями. Любая АППО является конкретной реализацией СКП [1].

4.3 Процесс расширения семантической модели

Суть предлагаемого метода формирования АППО состоит в итерационном расширении первоначального словаря ЗС до полноценного АППО. Метод можно изложить в виде приведенного ниже алгоритма.

1. Выбор ключевых терминов, задающих ПО, осуществляется пользователем:
 - вручную (на первом этапе обучающей выборки или для исследования новых тем);
 - извлекается системой автоматически на основе частотных характеристик («весов» терминов);
 - из предыдущего АППО автоматически берется множество начальных ЗС (ключевых слов).
2. Семантический поиск по заданным ключевым словам и накопление базы релевантных интернет-текстов, из которых формируются корпуса ЕЯ-текстов различных ПО (big data). При отладке и совершенствовании поиска используются методы машинного обучения.
3. Автоматическое выделение терминов из текстов и составление частотного словаря ПО (методами статистики).
4. Формирование списка наиболее значимых понятий ПО из ключевых терминов и ЗС.
5. Разбиение текстов на сегменты или предложения.
6. Построение контекста каждого термина в виде набора содержащих его сегментов/предложений.
7. Построение контекстных векторов, т.е. для каждого значимого термина (или ЗС) ведется подсчет других значимых терминов/ЗС в его контексте и формируется вектор с полученной статистикой. Эти контекстные векторы определяют меру близости и ассоциативные связи между ключевыми словами.
8. Расчет косинусной меры близости контекстных векторов (компонентами вектора ЗС — ключевых слов данной ПО — являются частоты совместной встречаемости данного ЗС с другими ЗС в одном и том же контексте) и выбор наиболее сильных ассоциативных связей. Для вычисления косинусной меры применяется следующая формула:
9. Выбор кандидатов в состав ключевых терминов (для расширения первоначального частотного словаря ЗС) из числа полученных наиболее сильных ассоциаций у ключевых терминов, осуществляемый следующим образом:
 - проводится поиск кластера тесно связанных ключевых слов;
 - выстраивается иерархическая структура на множестве ЗС методом иерархической кластеризации;
 - в векторном пространстве рассчитывается центр кластера заданных ключевых слов (ЗС), относящихся к ПО;
 - определяются наиболее сильные термины-кандидаты для включения в состав ЗС из числа новых терминов, находящиеся ближе к центру кластера;
 - наиболее отдаленные от центра кластера ключевые слова (ЗС) становятся кандидатами на исключение из состава ЗС;
 - принимается итоговое решение по изменению состава ЗС путем добавления или удаления полученных терминов-кандидатов.
10. Пополнение первоначального состава ключевых терминов наиболее сильными кандидатами, автоматически выбранными системой, и переход к шагу 1.

Таков макетный алгоритм, на базе которого разработан статистический модуль, решающий задачу создания АППО. Отдельные этапы методики, включая механизмы семантического интернет-серфинга, будут развиваться и совершенствоваться на следующем этапе проекта.

Итак, применив формулу вычисления косинусной меры между контекстными векторами, получаем коэффициенты семантической близости между рассматриваемыми ЗС. В реальных приложениях матрицы, содержащие оценочные коэффициенты, содержат порядка миллиона столбцов (строк). Но благодаря тому, что подавляющее большинство элементов заполняется нулями, реально хранимая информация вполне обозрима. Для нахождения ассоциативных связей, которые могут войти в ассоциативный портрет, необходимо выбирать из этой матрицы пары терминов или ЗС с самыми большими коэффициентами семантической близости.

Приведем пример на основе следующего текстового фрагмента (выборка получена из текстов, свободно существующих в Интернете) из ПО АНПА [32].

Для анализа берется обучающий фрагмент из дайджеста текстов.

где x и y — векторы в пространстве признаков, которыми являются сами ЗС из данного СКП, а i — это индекс, пробегающий число признаков.

Фрагмент дайджеста текстов предметной области «Автономные обитаемые подводные аппараты»

Прошел испытания и поставлен на вооружение новый подводный аппарат отечественной разработки «Обзор-600», который поможет флоту в разведке и исследовании морского дна.

Необитаемый подводный аппарат (АНПА) Knifefish, предназначен для борьбы с морскими минами различных типов.

Гидролокатор может быть размещен на ТНПА или АНПА и работать в качестве вперёдсмотрящей навигационной системы, а также функционировать в режиме системы 3D-визуализации, на рабочих глубинах до 2500 м.

Двухлучевой эхолот JJ-Connect Fisherman 600 обеспечивает точность определения глубины и структуры дна.

Далее строятся контекстные векторы для ЗС «АНПА», «Обзор-600», «Knifefish», «Гидролокатор», «эхолот JJ-Connect Fisherman» и слов (ЗС), встречающихся в текстовом фрагменте более одного раза. В клетки таблицы записывается число предложений с соответствующей парой ЗС.

Матрица оценочных коэффициентов

	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉	c ₁₀	c ₁₁
c ₁	0	0	1	1	0	1	1	1	1	0	1
c ₂	0	0	0	0	0	1	1	1	1	1	0
c ₃	1	0	0	0	0	1	1	1	1	0	0
c ₄	1	0	0	0	0	0	0	0	0	0	1
c ₅	0	0	0	0	0	0	0	0	0	1	1
c ₆	1	1	1	0	0	0	2	2	2	1	0
c ₇	1	1	1	0	0	2	0	2	2	1	0
c ₈	1	1	1	0	0	2	2	0	2	1	0
c ₉	1	1	1	0	0	1	1	1	1	1	0
c ₁₀	0	1	0	0	1	1	1	1	1	0	1
c ₁₁	1	0	0	1	1	0	0	0	0	1	0

В таблице используются следующие сокращенные обозначения:

- c₁ = АНПА;
- c₂ = Обзор-600;
- c₃ = Knifefish;
- c₄ = гидролокатор;
- c₅ = эхолот JJ-Connect;
- c₆ = подводный аппарат;
- c₇ = подводный;
- c₈ = аппарат;
- c₉ = морского дна, морскими;
- c₁₀ = Fisherman;
- c₁₁ = глубинах, глубины.

Применив формулу вычисления косинусной меры между контекстными векторами, получим

следующие коэффициенты семантической близости между рассматриваемыми ЗС в порядке убывания:

- «Гидролокатор» и «эхолот JJ-Connect Fisherman» — 0,50;
- «Обзор-600» и «Knifefish» — 0,40;
- «АНПА» и «Обзор-600» — 0,34;
- «АНПА» и «Knifefish» — 0,34;
- «Knifefish» и «гидролокатор» — 0,32;
- «АНПА» и «гидролокатор» — 0,26;
- «АНПА» и «эхолот JJ-Connect Fisherman» — 0,26;
- «Обзор-600» и «эхолот JJ-Connect Fisherman» — 0,17.

Для нахождения ассоциативных связей, которые могут войти в АППО, следует выбрать из этой матрицы пары терминов с самыми большими коэффициентами семантической близости (косинусной меры). В данном случае это будут следующие пары:

- («Гидролокатор», «эхолот JJ-Connect Fisherman»);
- («Обзор-600», «Knifefish»);
- («АНПА», «Обзор-600»);
- («АНПА», «Knifefish»).

4.4 Использование машинного обучения

Обработка больших массивов текстов из Интернета позволяет собирать необходимые статистические данные для формирования достаточно полной картины о ПО, представленной в виде СКП. Возможность проводить машинное обучение на большом количестве примеров существенно улучшает результаты.

Машинное обучение при выявлении ключевых слов ПО и ассоциативных связей между ними строится по следующей схеме:

- (1) создание обучающей и тестовой выборок, в которых для каждого текста (документа) указан его класс;
- (2) подготовка текстов к использованию их для обучения: разбиение текстов на элементы — слова, ЗС, знаки препинания и т. д.; выявление «шума» в тексте и его отсеивание; для некоторых обучающих выборок делается лемматизация (приведение слов к нормальной форме);
- (3) преобразование каждого документа в вектор признаков;
- (4) создание словаря ЗС (классификатора) и дальнейшее обучение классификатора на полученном обучающем множестве;
- (5) проверка точности работы классификатора на тестовом массиве текстов.

Совершенствование алгоритма формирования АППО осуществляется, в частности, и методами машинного обучения, включая проведение экспериментов по построению АППО в удобной для разработки онтологий форме. В качестве обучающей темы для решения указанной задачи была выбрана ПО «Социально-политический портрет регионов России», а в качестве ее сужения до обучающей выборки — ограничение регионом Татарстана. При построении онтологии определенной ПО происходит выделение наиболее значимых именованных сущностей, формирующих классы и подклассы. Примеры работы алгоритма на конкретных ПО представлены в разд. 5.

Использование онтологии в информационной системе позволяет решать несколько различных задач:

- эталонные сущности служат образцами для сравнения и идентификации сущностей, извлеченных из публикаций о ПО, происходящих из различных информационных источников;
- сущности могут быть представлены пользователю в качестве предопределенных объектов для формирования запросов к интегрированному хранилищу данных;
- связи эталонных сущностей, представленные в онтологии, позволяют дополнять и обобщать информацию, содержащуюся в текстах. Например, если в тексте найдено упоминание человека, то можно отнести его имя к организации, где он работает, а затем обобщить до уровня власти и сферы деятельности.

Достоинствами автоматизированного способа создания онтологий являются низкие трудозатраты и — для рассматриваемой темы — учет «медиаактивности» персон и организаций, т.е. числа упоминаний в СМИ. К недостаткам можно отнести отсутствие полноты полученной онтологии и порою случайный характер выделяемых объектов. Полностью автоматизированный способ получения подобных онтологий пока невозможен, он еще требует доработки. При составлении первоначального набора ключевых слов, анализе материала, создании онтологий для разных обучающих выборок потребовалась некоторая ручная доводка (несколько итераций по результатам машинного обучения).

Для повышения полноты и точности результатов эта работа осуществлялась методом автоматического интернет-серфинга в результате круглосуточного функционирования двух серверов и методом краудсорсинга как силами основного авторского коллектива, так и с привлечением магистрантов

и студентов Московского технического университета связи и информатики [5]. В итоге произошло существенное расширение сформированных на начальном этапе тематических корпусов ЕЯ-текстов. Данная работа проводится по следующей схеме:

- система производит в Интернете мониторинг информации, имеющей отношение к какой-либо ПО;
- пользователь задает поисковый запрос в виде набора ключевых терминов (слов и словосочетаний);
- аналитическая система находит на интернет-сайтах предложения с заданными ключевыми терминами и фиксирует ссылки (URL — uniform resource locator) на соответствующие документы;
- из найденных предложений автоматически выделяются новые ЗС, что позволяет расширить исходный поисковый запрос и обеспечивает большую полноту извлеченной информации.

4.5 Развитие механизмов серфинга

Отметим, что совершенствование механизмов серфинга — необходимое условие для успешного извлечения знаний из ЕЯ-текстов сверхбольших объемов (big data); улучшение этих механизмов — отдельная подзадача [31]. Используемый алгоритм поиска частично унаследован от электронной системы — энциклопедии ключевых слов KEYWEN; этот алгоритм порождает в качестве промежуточного результата дайджесты текстов конкретной ПО и выглядит так. Поисковый запрос формируется на основе множества ЗС. Осуществляется перебор всех комбинаций ключевых терминов (единичные термины, пары терминов, тройки и т.д.) так, чтобы длина запроса не превышала определенной границы n . В результате поиска должно быть найдено «среднее» количество (порядка нескольких сотен) релевантных документов. Алгоритмы объединения имеют эмпирический характер — значение границы n определяется опытным путем. Затем с помощью известных поисковых систем (Google, Яндекс и пр.) поисковые запросы обрабатываются, и в результате получается множество M_1 текстовых документов. Это множество просматривается на предмет выделения URL-ссылок (в этот момент включаются в работу средства отсекающего дорвея, сгенерированных шумов и т.п.), по выбранным ссылкам формируется расширенное множество документов M_2 . Множество M_2 , в свою очередь, тоже содержит URL-ссылки, потому процесс расширения множества документов можно продолжать

и далее, несколькими итерациями. При этом необходимо проверять вновь найденные документы на наличие в них первоначальных ключевых терминов. В этом случае и последующие итерации вносят существенный вклад в пополнение множества искомых документов и списка ключевых слов. Далее документы из множества M_2 делятся на предложения (или на фрагменты, близкие по длине к обычным предложениям). В результате составляется база данных с записями вида: ⟨текст фрагмента⟩–⟨URL, указывающий на этот фрагмент⟩.

Такая база данных имеет значительный объем (до нескольких терабайт). Для ее хранения и осуществления поиска в ней недостаточно традиционных средств работы с базами данных, необходимо применять специальные программно-аппаратные технологии big data. По полученной базе данных производятся статистические подсчеты: сколько различных URL ссылается на одно и то же предложение. Отметим, что при таком подсчете сравнение URL обычно имеет смысл производить с точностью до доменов первого или второго уровня. На данном этапе приходится решать задачу отсеивания сайтов-двойников, дорвеев и т. д., с тем чтобы в выборку попали только так называемые «независимые» сайты. Таким образом, каждый фрагмент получает определенный рейтинг, равный числу его вхождений в независимые сайты. Из топа отсортированных по рейтингу фрагментов (путем определенной процедуры редактирования) составляется своеобразный дайджест текстов, содержащий рейтинговую информацию по данной ПО (уже без URL). Примеры дайджестов из обучающих выборок по Татарстану и по другой ПО — «Протестная активность» представлены в разд. 5.

Разрабатываемый авторским коллективом инструмент АППО, опираясь на обнаруженные ассоциативные связи, позволяет соотносить лексемы (термины) и ЗС с элементами более высокого уровня — подкатегориями, категориями (или «рубриками»).

Это соотношение происходит в соответствии с выявленными первичными и/или вторичными признаками; рассчитываются вероятности вхождения словосочетаний в документы, связанные с рубриками. По этим вероятностям при помощи специальной формулы рассчитываются веса признаков, связанных с рубриками. Вес признака, связанного с категорией, может быть рассчитан различными методами, например по вероятности вхождения признака (термина, словосочетания) в документы, на которые ссылается категория. Следующий возможный метод — это расчет веса через вероятность нахождения признака в контексте основных при-

знаков категории. Под контекстом здесь понимается множество всех документов и их компонент, содержащих основные признаки рубрики.

Например, каждому ключевому слову может быть присвоен вес, который рассчитывается с учетом частоты встречаемости ключевого слова в текстах определенной ПО. Выделенные ключевые слова служат основой для построения классификаторов. Категории в классификаторах представлены в виде кортежа, где вес определяет силу ассоциативной связи: ⟨Первичное ключевое слово, Категория, Вес⟩.

Соотнесение ключевого слова с категорией осуществляется по следующей формуле:

$$\begin{aligned} \text{вес(ключевое слово)} &= \text{вес(категория-}i\text{)} = \\ &= \log \frac{p(\text{ключевое слово} \& \text{ категория-}i\text{)}}{p(\text{ключевое слово})p(\text{категория-}i\text{)}}. \end{aligned}$$

Здесь $p(\text{ключевое слово} \& \text{ категория-}i\text{)}$ есть вероятность совместного проявления двух факторов: ключевое слово существует в документе и документ принадлежит категории i . Если ключевое слово и категория статистически независимы, то вероятность их соотнесения определяется как произведение $p(\text{ключевое слово})p(\text{категория-}i\text{)}$. Если они не независимы, статистика отражает тенденцию к соотнесению ключевого слова с определенной категорией, в таком случае вероятность $p(\text{ключевое слово} \& \text{ категория-}i\text{)}$ будет больше, чем $p(\text{ключевое слово})p(\text{категория-}i\text{)}$. Соотношение между $p(\text{ключевое слово} \& \text{ категория-}i\text{)}$ и $p(\text{ключевое слово})p(\text{категория-}i\text{)}$ является мерой степени статистической зависимости между ключевым словом и категорией i . Логарифм этого соотношения есть объем информации о наличии ключевого слова, полученной при наблюдении документа категории i . Так как уравнение симметрично, это также объем информации о принадлежности документа к категории i , получаемой при наблюдении в нем ключевого слова (принцип взаимности). Приведенная формула позволяет вычислить вес ключевого слова, связанного с данной категорией, если известна вероятность наличия этого ключевого слова в документах категории.

Веса признаков, рассчитанных разными методами, должны по возможности совпадать. Для этого был предложен следующий метод согласования весов.

Допустим, с помощью двух различных методов найдены две группы признаков T_1 и T_2 , связанных с некоторой категорией. Например, T_1 — это признаки, входящие в тестовую выборку, на которую ссылается категория (это основные термины и их меньше), а T_2 — это признаки, связанные с кате-

гориями ассоциативными связями (т. е. это вторичные признаки) и выделенные по интернет-текстам (их больше).

Веса рассчитываются независимо для T_1 и T_2 двумя различными алгоритмами. Но так как множество T_1 входит в T_2 , то веса для T_1 будут рассчитаны обоими алгоритмами и необходимо подбирать коэффициент для весов из T_2 , при котором эти веса максимально приближаются к весам из T_1 . В дальнейшем этот коэффициент используется для подсчета весов вторичных признаков.

5 Примеры реализации инструмента ассоциативного портрета предметной области

Приведем некоторые примеры работы методики на обучающих выборках из различных ПО.

5.1 Предметная область «Социально-политический портрет Татарстана»

При построении онтологии данной ПО наиболее значимыми являются именованные сущности следующих классов:

- географические названия наиболее крупных населенных пунктов региона;

- персоны, исполняющие должности в административном аппарате региона, и их должности;
- наиболее важные организации региона;
- наиболее важные события в социально-политической жизни региона;
- сферы деятельности в регионе.

Данные классы важны при описании социально-политической жизни всех регионов Российской Федерации. В случае Татарстана большое значение имеет разработка полезных ископаемых (нефть, уголь, газ), а также определенные виды хозяйственной деятельности в этом регионе, в связи с чем в онтологию добавился класс «месторождения», а среди «организаций» были выделены подклассы «нефтехимия», «машиностроение», «агрокомплекс», «образование».

Основные классы сущностей, их атрибуты и взаимосвязи представлены на рис. 1.

Для создания дайджеста текстов — обучающей выборки по ПО «Социально-политический портрет Татарстана» был использован поисковый запрос, состоящий из двух составляющих:

- (1) территориальные образования Татарстана;
- (2) слова, характерные для текстов о социально-политической жизни.

Итогом работы системы семантического серфинга стал дайджест из предложений, удовлетво-

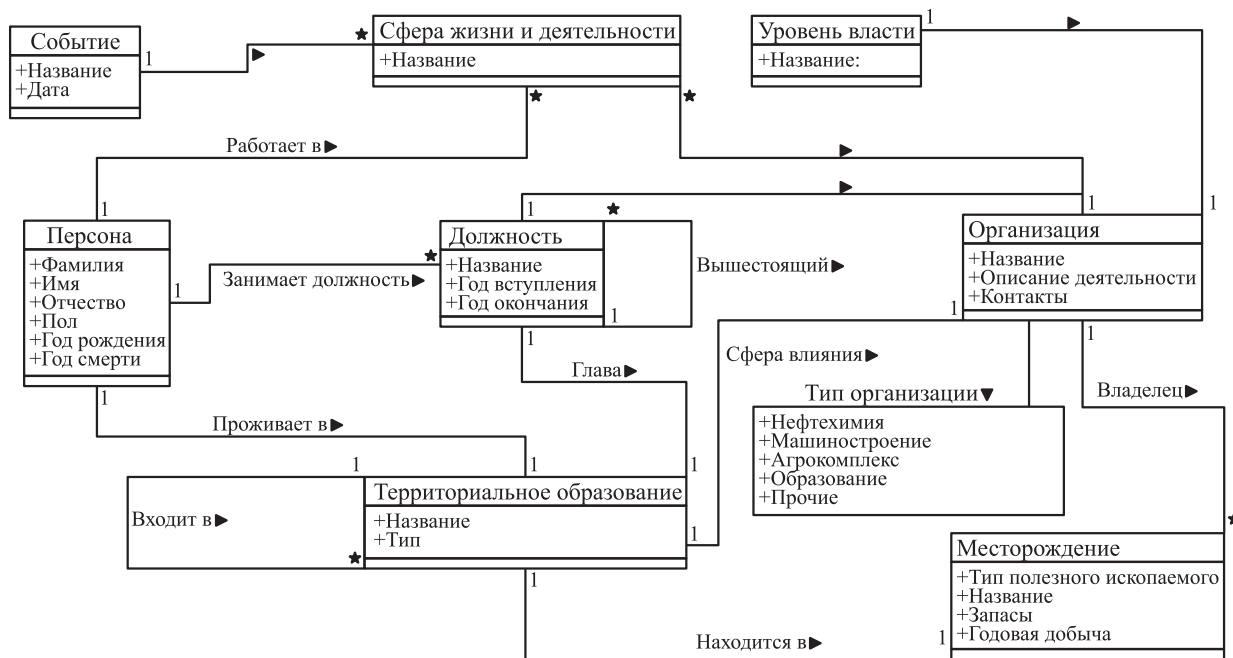


Рис. 1 Семантическая структура ПО «Социально-политический портрет Татарстана»

ряющих поисковому запросу, объемом около 40 000 предложений.

В результате дальнейшей работы системных модулей (анализа дайджеста) были получены списки должностных лиц и организаций Татарстана. Так как речь шла о подготовке обучающей выборки, впоследствии эти списки подверглись ручному редактированию, чтобы исключить объекты, не относящиеся к изучаемой ПО.

5.2 Предметная область «Протестная активность»

Приведем пример поискового запроса из выделенных ключевых слов (ЗС) для формирования дайджеста ПО «Протестная активность» с помощью автоматической системы семантического серфинга.

Поисковый запрос

(Активисты OR Акции Протеста OR Властвующая Элита OR Внутриэлитный Раскол OR Госпереворот OR Государственный переворот OR Давление На Власть OR Демонстрации OR Евромайдан OR Захват Власти OR Конфликт Элит OR Кризис Легитимной Власти OR Майдан OR Массовый Протест OR Митинги OR Недовольство OR Общественное Мнение OR Оппозиция OR Пикетирование OR Пикеты OR Политический Протест OR Протестное Движение OR Протестное Поведение OR Радикализация Протеста OR Смена Власти OR Собрания OR Состояние Тревожности OR Уровень Организованности OR Ценностные Конфликты OR Шестивия OR Эскалация Конфликта)

На основе запроса была получена обучающая текстовая выборка и построена онтология данной ПО. Фрагмент верхнего уровня онтологии ПО «Протестная активность»:

Объекты: события (акции протеста), люди (активисты, политики), организации (партии, оппозиция, Евромайдан), источники информации (агентства, веб-сайты), время, место.

Акции протеста: митинги, демонстрации, пикетирования, кампании гражданского неповиновения, забастовки, собрания, шествия.

6 Визуализация ассоциативного портрета предметной области

Отображение многомерных векторов на плоскость является удобным средством визуализации семантических связей, в результате чего образуются визуальные карты ЗС. На таких картах расстояние между ЗС тем меньше, чем больше сила их ассоциативной связи, что позволяет выделять сильно связанные и близкие по смыслу ЗС.

В ходе работы над проектом были разработаны методы создания визуальных карт для различных ПО с применением технологии «multidimensional scaling» (многомерного масштабирования). Визуальные карты ПО строятся на основе сочетания методик дистрибутивной семантики и математической статистики. Все они базируются на положении дистрибутивной семантики о том, что семантическое сходство между двумя словами/терминами/ЗС может быть смоделировано как функция степени перекрытия их языковых контекстов. Уменьшение размеров и масштаба оригинального семантического пространства дает возможность представить на обучающей выборке расстояния между терминами на карте ПО. Семантическое векторное пространство, измеренное методами статистики, позволяет визуализировать сверхбольшие объемы данных: расстояния между терминами зависят от распределения контекстных векторов, компоненты которых, в свою очередь, зависят от распределения большого количества не видимых на карте ЗС. Таким образом, относительно небольшое число видимых терминов (ЗС), отражающихся на карте, может зависеть от распределения в области знаний миллионов невидимых словосочетаний. Подобная карта распределения ЗС в СКП дает наглядное представление о важнейших ПО, их границах, взаимопроникновении, о терминах и ЗС и ассоциативных связях между ними. Пример такой визуальной карты представлен на рис. 2.

Пространственное положение терминов на карте определяется их семантической связью. Это расстояние постоянно изменяется в зависимости от количества обработанной информации. Так, в данном примере термин Friends («Друзья») является внешним термином по отношению к рассматриваемой карте ПО. Наблюдение за движением внешнего термина по отношению к «внутренним» ЗС также оказывается полезным для изучения семантических связей, их изменений (сдвигов) внутри огибающей кривой.

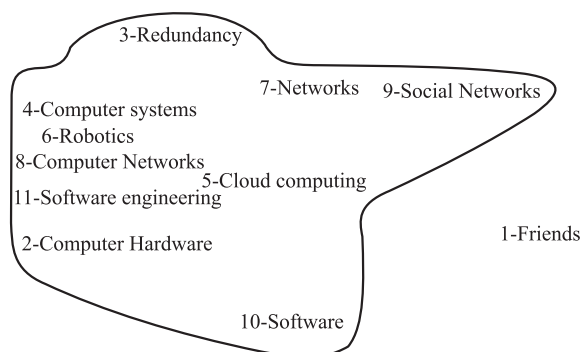


Рис. 2 Визуальная карта СКП

Метод итерационного накопления словарей ЗС дает возможность создать сначала компактную карту ПО; агрегация данных из сверхбольших массивов ЕЯ-текстов позволяет постепенно расширить и уточнить эту карту. Внедрение новой информации в корпус текстов может существенно изменить расстояния и зависимости в целом между различными терминами на карте, что делает видимым влияние новой информации. Это также позволяет выявить, обработать и быстро идентифицировать пропущенные ранее термины. Таким образом, карты ПО, построенные на основе пространственно-векторной семантической модели, становятся эффективным инструментом мониторинга для измерения расстояния и определения (визуализации) различий данных (терминов и ЗС) в сверхбольших текстовых корпусах.

7 Заключение

Инструмент автоматизированного формирования АППО — перспективное направление для создания систем извлечения знаний. Кроме того, этот инструмент подходит для решения ряда других задач.

Комплексный подход к автоматической обработке ЕЯ-текстов на основе сочетания методов статистики, корпусной лингвистики и дистрибутивной семантики с целью выявления ЗС и их ассоциативных связей для формирования СКП реализуется в виде итеративной методики формирования АППО, является новаторским и актуальным для исследований на материале русского языка. Применяемый подход использует объективные оценочные статистические критерии корпусной лингвистики (контекстные векторы, веса и др.), методы дистрибутивной семантики, дополненные математическими (вероятностными) и логико-лингвистическими методами. Проект реализуется на сверхбольших объемах сетевых русскоязычных и англоязычных ЕЯ-текстах (big data).

Полученные результаты могут быть использованы для решения задач автоматизированного семантического поиска на сверхбольших объемах данных, извлечения знаний из неструктурированных ЕЯ-текстов и автоматизированного формирования интерактивных предметно-ориентированных энциклопедий. Особенно перспективной областью применения АППО является задача построения онтологии ПО с последующим построением схем баз данных и знаний.

Новые результаты, полученные в ходе выполнения проекта, позволяют успешно решать проблему установления семантических соответствий (ас-

социативных связей) между разнотипными (терминальными и нетерминальными) языковыми объектами: лексемами, словосочетаниями, предложениями, сверхфразовыми единствами. Терминальная единица — лексема — термин определенной ПО; нетерминальные — ЗС, предложение, абзац, статья.

Статистические оценочные критерии, включенные в методику наряду с логико-лингвистическими методами и правилами дистрибутивной семантики, позволяют разрешать семантическую неоднозначность лексем (терминов) и ЗС на основе вероятностных механизмов, что повышает скорость и эффективность семантического поиска и обеспечивает автоматизированное отнесение обрабатываемого текста к той или иной ПО.

Построение АППО и СКП с использованием ассоциативных связей между выделенными терминами (ЗС), а также соотнесения терминов с категориями, рассчитанных по формуле косинусной меры, привело к следующим результатам:

- разработан комплекс методов классификации текстов, использующих методы статистики, дистрибутивной семантики, корпусной лингвистики, аппарат ассоциативных связей и некоторые другие, позволяющий классифицировать тексты, изначально не содержащие термины и ЗС (коллокации) из обучающей выборки;
- создано программное обеспечение, реализующее предложенные методы для отбора релевантных текстов по теме ПО;
- сформированы объемные электронные корпуса соотнесенных с определенной ПО ЕЯ-текстов с помощью автоматизированного разбиения текстов на ПО;
- созданы корпуса словарей терминов ЗС и предметных словарей, накопленных путем обработки больших объемов данных из Интернета;
- разработаны методы создания визуальных карт для различных ПО с применением технологии «multidimensional scaling» (многомерного масштабирования);
- получил дальнейшее развитие метод семантического поиска ЕЯ-текстов с целью направленного извлечения из сетевых текстов (big data) энциклопедической информации.

В настоящее время авторский коллектив продолжает работу по развитию методики АППО и созданию программных средств ее реализации.

Литература

1. Шарнин М. М., Сомин Н. В., Кузнецов И. П., Морозова Ю. И., Галина И. В., Козеренко Е. Б. Статистические механизмы формирования ассоциативных портретов предметных областей на основе естественно-языковых текстов больших объемов для систем извлечения знаний // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 92–99.
2. Charnine M., Petrov A., Kuznetsov I. Association-based identification of Internet User interests // WORLDCOMP'13, ICAI 2013: 2013 Conference (International) on Artificial Intelligence Proceedings. — Las Vegas: CSREA Press, 2013. Vol. I. P. 77–81.
3. Козеренко Е. Б. Интегральное моделирование языковых структур в лингвистических процессорах систем обработки знаний и машинного перевода // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 89–98.
4. Rapp R. Word sense discovery based on sense descriptor dissimilarity // 9th MT Summit Proceedings. — New Orleans, LA, USA, 2003. P. 315–322.
5. Charnine M., Protasov V. Optimal automated method for collaborative development of university curricula // WORLDCOMP'13, ICAI 2013: 2013 Conference (International) on Artificial Intelligence Proceedings. — Las Vegas: CSREA Press, 2013. Vol. I. P. 96–100.
6. Морозова Ю. И. Построение семантических векторных пространств различных предметных областей // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 90–93.
7. Charnine M., Somin N., Nikolaev V. Conceptual text generation based on key phrases // WORLDCOMP'14, ICAI 2014: 2014 Conference (International) on Artificial Intelligence Proceedings. — Las Vegas: CSREA Press, 2014. Vol. II. P. 639–643.
8. Bacon E., Hagel G., Charnine M., Foggie R., Kirk B., Schagayev I., Kravtsov G. WEDUCA: Web-enhanced design of university curricula // 2013 Conference (International) on Frontiers in Education: Computer Science and Computer Engineering (FECS'13) Proceedings. — Las Vegas: CSREA Press, 2013. P. 288–294.
9. The SMART retrieval system: Experiments in automatic document processing / Ed. G. M. Salton. — Prentice-Hall, 1971. 556 p.
10. Dunning T. Accurate methods for the statistics of surprise and coincidence // Comput. Linguist., 1993. Vol. 19. No. 1. P. 61–74.
11. Борисова Е. Г. Коллокации. Что это такое и как их изучать. — 2-е изд., стер. — М.: Филология, 1995. 49 с.
12. Church K., Hanks P. Word association norms, mutual information, and lexicography // Comput. Linguist., 1996. Vol. 16. No. 1. P. 22–29.
13. Lund K., Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence // Behav. Res. Meth. Inst. C., 1996. Vol. 28. No. 2. P. 203–208.
14. Readings in information retrieval / Eds. J. K. Spärck, P. Willett. — San Francisco, CA, USA: Morgan Kaufmann, 1997. 594 p.
15. Spärck Jones, K. A statistical interpretation of term specificity and its application in retrieval // J. Documentation, MCB University Press, 2004. Vol. 60. No. 5. P. 493–502.
16. Charnine M. M., Kuznetsov I. P., Kozerenko E. B. Semantic navigator for Internet search // MLMTA'05: Conference (International) on Machine Learning Proceedings. — Las Vegas: CSREA Press, 2005. P. 60–68.
17. Sahlgren M. Towards pertinent evaluation methodologies for word-space models // LREC 2006: 5th Conference (International) on Language Resources and Evaluation Proceedings. — Genoa, Italy, 2006. P. 821–824.
18. Spärck Jones, K. Statistics and retrieval: past and future // Conference (International) on Computing: Theory and Applications. Platinum Jubilee Conference of the Indian Statistical Institute. — Kolkata, India: IEEE, 2007.
19. Landauer Th. K., McNamara D. S., Dennis S., Kintsch W. Handbook of latent semantic analysis. — Mahwah, NJ, USA: Lawrence Erlbaum, 2007. 544 p.
20. Иорданская Л. Н., Мельчук И. А. Смысл и сочетаемость в словаре. — М.: Языки славянских культур, 2007. 672 с.
21. Charnine M., Charnine V. Keywen category structure. — Wordclay, USA, 2008. 60 p.
22. Lenci A. Distributional semantics in linguistic and cognitive research // Rivista di Linguistica, 2008. Vol. 1. P. 1–30.
23. Manning C., Raghavan P., Schütze H. Introduction to information retrieval. — Cambridge: Cambridge University Press, 2008. 581 p.
24. Sahlgren M. The distributional hypothesis. From context to meaning // Distributional Models of the Lexicon in Linguistics and Cognitive Science: Special issue of the Italian J. Linguistics: Rivista di Linguistica, 2008. Vol. 20. No. 1. P. 33–53.
25. Baroni M., Lenci A. Distributional memory: A general framework for corpus-based semantics // Comput. Linguist., 2010. Vol. 36. No. 4. P. 673–721.
26. Turney P. D., Pantel P. From frequency to meaning: Vector space models of semantics // JAIR, 2010. Vol. 37. P. 141–188.
27. Захаров В. П., Хохлова М. В. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог» (2010). — М.: РГГУ, 2010. Вып. 9(16). С. 147–143.
28. Kuznetsov I. P., Charnine M. M., Kozerenko E. B., Somin N. V., Nikolayev V. G., Matskevich A. G. Intelligent tools for the semantic Internet navigator design // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. XIV Всеросс. науч. конф. RCDL'2012. — Переславль-Залесский: Университет города Переславля, 2012. С. 274–283.
29. Kuznetsov I. P., Kozerenko E. B., Charnin M. M. Technological peculiarity of knowledge extraction for logical analytical systems // WORLDCOMP'12, ICAI'12:

- 2012 Conference (International) on Artificial Intelligence Proceedings. — Las Vegas: CSREA Press, 2012. Vol. II. P. 762–768.
30. Schumann A. Towards the automated enrichment of multilingual terminology databases with knowledge-rich contexts // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог» (2012). — М.: РГГУ, 2012. Вып. 11(18). Т. 1. С. 559–567.
31. Шарнин М. М., Кузнецов И. П. Особенности семантического поиска информационных объектов на основе технологии баз знаний // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 47–56.
32. Борисов Т. Н., Бронецкий А. Е., Клименко С. В., Рыков В. В., Шарнин М. М. Автономные необитаемые подводные аппараты: автоматическое формирование ассоциативного портрета предметной области // Ситуационные центры и информационно-аналитические системы класса 4i для задач мониторинга и безопасности (SC-IAS4i-VRТеггo2013): Тр. II Междунар. науч. конф. — Протвино: ИФТИ, 2013. С. 38–43.
33. Zolotarev O., Charnine M., Matskevich A. Conceptual business process structuring by extracting knowledge from natural language texts // WORLDCOMP'14, ICAI 2014: 2014 Conference (International) on Artificial Intelligence Proceedings. — Las Vegas: CSREA Press, 2014. Vol. I. P. 82–87.

Поступила в редакцию 21.04.15

ASSOCIATIVE PORTRAITS OF SUBJECT AREAS AS A TOOL FOR AUTOMATED CONSTRUCTION OF BIG DATA SYSTEMS FOR KNOWLEDGE EXTRACTION: THEORY, METHODS, VISUALIZATION, AND APPLICATION

I. V. Galina, E. B. Kozerenko, Yu. I. Morozova, N. V. Somin, and M. M. Charnine

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

Abstract: The paper presents the technique of developing systems for extraction of knowledge which employs the approach of automated association portrait of a subject area (APSA) formation and building a semantic context space (SCS). The ideology of the APSA is based on the distributional hypothesis claiming that semantically equal (or related) lexemes have a similar context and, *vice versa*, in a similar context, the lexemes are semantically close. The model uses an extended hypothesis that consists in the investigation of similarities and differences in contexts not only of individual words, but of arbitrary multilexeme fragments of meaningful word-combinations. The examples of implemented projects for different subject domains are given.

Keywords: semantic modeling; associations; mathematical statistics; distributive semantics; big data; automated extraction of knowledge; digital natural language text corpora; semantic search; intelligent Internet technology

DOI: 10.14357/19922264150211

Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (project 13-07-00272).

References

1. Charnine, M. M., N. V. Somin, I. P. Kuznetsov, Yu. I. Morozova, I. V. Galina, and E. B. Kozerenko. 2013. Statisticheskie mekhanizmy formirovaniya assotsiativnykh portretov predmetnykh oblastey na osnove estestvenno-yazykovykh tekstov bols'shikh ob'emov dlya system izvlecheniya znaniy [Statistical mechanisms of the subject domains associative portraits formation on the basis of big natural language texts for the systems of knowledge extraction]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):92–99.
2. Charnine, M., A. Petrov, and I. Kuznezov. 2013. Association-based identification of Internet user interests. *ICAI'13: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 77–81.
3. Kozerenko, E. B. 2014. Integral'noe modelirovanie yazykovykh struktur v lingvisticheskikh protessorakh sistem obrabotki znaniy i mashinnogo perevoda [Integrated modeling of language structures for linguistic processors of knowledge management and machine translation systems]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):89–98.

4. Rapp, R. 2003. Word sense discovery based on sense descriptor dissimilarity. *Conference (International) 9th MT Summit Proceedings*. New Orleans, LA. 315–322.
5. Charnine, M., and V. Protasov. 2013. Optimal automated method for collaborative development of university curricula. *ICAI'13: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 96–100.
6. Morozova, Yu. I. 2013. Postroenie semanticheskikh vektornykh prostranstv razlichnykh predmetnykh oblastey [Semantic vector spaces for different knowledge domains]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):90–93.
7. Charnine, M., N. Somin, and V. Nikolaev. 2014. Conceptual text generation based on key phrases. *ICAI'14: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 639–643.
8. Bacon, E., G. Hagel, M. Charnine, R. Foggie, B. Kirk, I. Schagaev, and G. Kravtsov. 2013. WEDUCA: Web-enhanced design of university curricula. *FECS'13 Conference (International) on Frontiers in Education: Computer Science and Computer Engineering Proceedings*. Las Vegas, USA: CSREA Press. 288–294.
9. Salton, G. M., ed. 1971. *The SMART retrieval system: Experiments in automatic document processing*. Prentice-Hall. 556 p.
10. Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Comput. Linguist.* 19(1):61–74.
11. Borisova, E. G. 1995. *Kollokatsii. Chto eto takoe i kak ikh izuchat'* [Collocations. What are they, and how are they to be studied]. 2nd ed. Moscow: Filologiya. 49 p.
12. Church, K., and P. Hanks. 1996. Word association norms, mutual information, and lexicography. *Comput. Linguist.* 16(1):22–29.
13. Lund, K., and C. Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Meth. Ins. C.* 28(2):203–208.
14. Spärck Jones, K., and P. Willett, eds. 1997. *Readings in information retrieval*. San Francisco, CA: Morgan Kaufmann. 594 p.
15. Spärck Jones, K. 2004. A statistical interpretation of term specificity and its application in retrieval. *J. Documentation, MCB University Press.* 60(5):493–502.
16. Charnine, M. M., I. P. Kuznetsov, and E. B. Kozerenko. 2005. Semantic navigator for Internet search. *MLMTA'05 Conference (International) on Machine Learning Proceedings*. Las Vegas, USA: CSREA Press. 60–68.
17. Sahlgren, M. 2006. Towards pertinent evaluation methodologies for word-space models. *LREC 2006: 5th Conference (International) on Language Resources and Evaluation Proceedings*. Genoa, Italy. 821–824.
18. Spärck Jones, K. 2007. Statistics and retrieval: Past and future. *Conference (International) on Computing: Theory and Applications Proceedings. Platinum Jubilee Conference of the Indian Statistical Institute*. Kolkata, India: IEEE.
19. Landauer, Th. K., D. S. McNamara, S. Dennis, and W. Kintsch. 2007. *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum. 544 p.
20. Iordanskaya, L. N., and I. A. Melchuk. 2007. *Smysl i sochetamost' v slovare* [Meaning and combinability in the dictionary]. Moscow: Slavonic Cultures Languages. 672 p.
21. Charnine, M., and V. Charnine. 2008. *Keywen category structure*. Wordclay, USA. 60 p.
22. Lenci, A. 2008. Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica* 2:1–30.
23. Manning, C., P. Raghavan, and H. Schütze. 2008. *Introduction to information retrieval*. Cambridge: Cambridge University Press. 581 p.
24. Sahlgren, M. 2008. The distributional hypothesis. From context to meaning. *Distributional Models of the Lexicon in Linguistics and Cognitive Science: Special issue of the Italian J. Linguistics: Rivista di Linguistica* 20(1):33–53.
25. Baroni, M., and A. Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Comput. Linguist.* 36(4):673–721.
26. Turney, P. D., and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. *JAIR* 37:141–188.
27. Zakharov, V. P., and M. V. Khokhlova. 2010. Analiz effektivnosti statisticheskikh metodov vyyavleniya kollokatsiy v tekstakh na Russkom yazyke [The analysis of statistical methods for the discovery of collocations in the Russian language texts]. *Computational Linguistics and Intelligent Technologies. Dialog'10: Conference (International) Proceedings*. Moscow: Russian State University for Humanities. 9(16):147–143.
28. Kuznetsov, I. P., M. M. Charnine, E. B. Kozerenko, N. V. Somin, V. G. Nikolayev, and A. G. Matskevich. 2012. Intelligent tools for the semantic Internet navigator design. *RCDL'2012 Conference (International) on Digital Libraries Proceedings*. Pereslavl-Zalesski, Russia. 274–283.
29. Kuznetsov, I. P., E. B. Kozerenko, and M. M. Charnin. 2012. Technological peculiarity of knowledge extraction for logical analytical systems. *ICAI'12: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 762–768.
30. Schumann, A. 2012. Towards the automated enrichment of multilingual terminology databases with knowledge-rich contexts. *Computational Linguistics and Intelligent Technologies. Dialog'12: Conference (International) Proceedings*. Moscow: Russian State University for Humanities. 11(18):559–567.
31. Sharnin, M. M., and I. P. Kuznetsov. 2012. Osobennosti semanticheskogo poiska informatsionnykh ob"ektov na osnove tekhnologii baz znaniy [Semantic search of natural language information on the basis of knowledge base technology]. *Informatika i ee Primeneniya — Inform. Appl.* 6(2):47–56.
32. Borisov, T. N., A. E. Bronetski, S. V. Klimentko, V. V. Rykov, and M. M. Charnine. 2013. Avtonomnye neobitaemye podvodnye apparaty: Avtomaticheskoe formirovanie assotsiativnogo portreta predmetnoy oblasti [Autonomous uninhabited submarine apparatus: Auto-

- matic creation of the subject area associative portrait]. *SC-IAS4i-VRTerro2013 Conference (International) on Situational Centers and Information Analytical Systems of 4i Class for the Tasks of Monitoring and Security Proceedings*. Protvino: IFTI. 38–43.
33. Zolotarev, O., M. Charnine, and A. Matskevich. 2014. Conceptual business process structuring by extracting knowledge from natural language texts. *ICAI'14: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 82–87.

Received April 21, 2015

Contributors

Galina Irina V. (b. 1967) — 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; irn_gl@mail.ru

Kozerenko Elena B. (b. 1959) — Candidate of Science (PhD) in linguistics, 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; kozerenko@mail.ru

Morozova Yulia I. (b. 1984) — 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; yulia-ipi@yandex.ru

Somin Nicolay V. (b. 1947) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow, 119333, Russian Federation; somin@post.ru

Charnine Mikhail M. (b. 1959) — 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; mc@keywen.com

ИНДИКАТОРЫ ТЕМАТИЧЕСКИХ ВЗАИМОСВЯЗЕЙ ОТРАСЛЕЙ НАУКИ И ИНФОРМАЦИОННО-КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В НАЧАЛЕ XXI ВЕКА*

В. А. Минин¹, И. М. Зацман², В. А. Хавансков³, С. К. Шубников⁴

Аннотация: Представлены результаты экспериментальных вычислений индикаторов тематических взаимосвязей науки и информационно-компьютерных технологий. Вычисленные значения индикаторов получены с помощью макета аналитической информационной системы (АИС), который был создан в рамках проекта Российского гуманитарного научного фонда (РГНФ) «Информационная система мониторинга и оценивания инновационно-технологического потенциала направлений фундаментальных научных исследований». Макет позволяет вычислять значения индикаторов взаимосвязей отраслей науки и направлений научных исследований (ННИ) с заданным видом технологий. В экспериментальных вычислениях в качестве исходной информации использовались официальные данные Роспатента об изобретениях по классу G06 Международной патентной классификации (МПК) (Обработка данных; Вычисление; Счет), опубликованные в 2000–2012 гг. Исходные данные для проведения расчетов не являются числовой информацией, а представляют собой полные тексты описаний изобретений на естественном языке (ЕЯ). Поэтому до начала экспериментальных вычислений индикаторов выполнялось извлечение из полных текстов изобретений информации о научных публикациях, цитируемых в описаниях изобретений, и определялось число публикаций по каждой отрасли науки и ННИ. Полученная числовая информация является исходной для вычислений значений индикаторов и дает возможность экспертам определять степень интенсивности переноса знаний из отраслей науки и ННИ в сферу технологий и оценивать их с помощью количественных индикаторов.

Ключевые слова: взаимосвязи науки и технологий; информационно-компьютерные технологии; обработка текста изобретения; регулярные выражения; рубрицирование; расчет значений индикаторов

DOI: 10.14357/19922264150212

1 Введение

Данная работа содержит описание итоговых результатов по проекту РГНФ (грант № 12-02-12019в) «Информационная система мониторинга и оценивания инновационно-технологического потенциала направлений фундаментальных научных исследований».

Методологические проблемы мониторинга, включая вопросы определения значений индикаторов тематических взаимосвязей науки и технологий, изложены в работах [1–6]. В [6] приводится описание индикаторов, предназначенных для количественного оценивания интенсивности процессов передачи знаний от науки к технологиям. В ка-

честве степени интенсивности процессов передачи знаний берется число научных публикаций, цитируемых экспертами в отчетах о патентном поиске или авторами изобретений в их описаниях. Определение степени интенсивности являлось целью проекта. В проекте использовались описания изобретений, опубликованных Федеральной службой по интеллектуальной собственности (Роспатент) за период с 2000 по 2012 гг. и относящихся к классу G06 (Обработка данных; Вычисление; Счет) МПК.

В текстовых описаниях отобранных изобретений были найдены сделанные авторами ссылки на научные публикации, библиографические данные которых позволяют отнести их к определенным отраслям науки и ННИ. Таким образом, с помощью

* Работа выполнена в Институте проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук при частичной поддержке РГНФ (грант № 12-02-12019в).

¹ Российский фонд фундаментальных исследований, minin@rflbr.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

ссылки цитирования устанавливалась связь между технологической сферой (в виде совокупности индексов МПК) и цитируемыми результатами научных исследований (для кодирования результатов использовались Государственный рубрикатор научно-технической информации (ГРНТИ) и рубрикатор Российского фонда фундаментальных исследований (РФФИ)).

Использование информационных ресурсов Роспатента было обусловлено тем, что они представлены в электронном виде и доступны для автоматизированной обработки. В соответствии со Страсбургским соглашением от 24 марта 1971 г. о МПК компетентные органы стран-участниц Союза по МПК при классифицировании патентных документов должны указывать «полные индексы МПК, присвоенные изобретению, описанному в документе» (ст. 4, п. 3). Это означает, что публикуемые в Роспатенте сведения о выданных патентах на изобретения содержат индексы МПК, которые можно использовать для описания тематики исследуемых групп технологий. Одновременно имеются полнотекстовые описания изобретений, представляющие собой неструктурированные тексты, в которых при изложении сути изобретения авторы нередко ссылаются на публикации в научных изданиях. Таким образом, библиографические ссылки на публикации в описаниях изобретений, привязанные к одной или нескольким рубрикам ННИ, можно сопоставить с индексами МПК, описывающими технологическую сферу, к которой относится изобретение, и таким образом косвенно оценить интенсивность процессов передачи знаний от науки к технологиям.

В процессе анализа изобретений объем отобранных полнотекстовых описаний может достигать нескольких сотен тысяч. Например, в работе [7] описывается процесс обработки массива из 656 695 патентов на изобретения, выданных Патентным ведомством США. Для установления взаимосвязей между индексами МПК и кодами рубрик ННИ из описаний изобретений в процессе обработки были выделены 1 147 160 ссылок на цитируемые публикации (ссылки на патенты не рассматривались). Затем из них для дальнейшей обработки были отобраны только те ссылки на журнальные статьи, для которых удалось идентифицировать название журнала и соотнести его с нормативным списком названий журналов, в котором каждому названию присвоена одна или несколько рубрик ННИ. При таких ограничениях для анализа было отобрано 106 636 ссылок, т.е. менее 10% от выделенных ссылок на цитируемые публикации.

При реализации подобной методологии для анализа отечественных описаний изобретений анало-

гичные зарубежные решения трудно адаптировать в силу ряда причин, подробно рассмотренных в работах [8, 9], а именно:

- отсутствие в «Административном регламенте исполнения Роспатентом приема заявок на изобретение, их рассмотрения и экспертизы» требований именно к структурированному представлению ссылок на цитируемые публикации (см. п. 10.11(12) Регламента [10]);
- отсутствие в опубликованных электронных версиях полнотекстовых описаний изобретений групп меток, выделяющих ссылки на цитируемые публикации согласно рекомендациям стандарта Всемирной организации интеллектуальной собственности (ВОИС) ST.14 [11];
- отсутствие или неполнота списка нормализованных и сокращенных названий журналов, используемых в ссылках на цитируемые публикации, в системах подготовки электронных патентных заявок.

Таким образом, при исследовании тематических взаимосвязей технологий и ННИ возникает задача анализа десятков и сотен тысяч полнотекстовых описаний изобретений и поиска в их текстах на ЕЯ ссылок на публикации с последующей их структуризацией и привязкой ссылок к рубрикам ННИ. Как следствие, возникает задача автоматизации данного процесса. При этом необходимо учитывать, что по своему содержанию само библиографическое описание является структурированным информационным объектом, состоящим из нескольких полей, который может размещаться в любом месте неструктурированного текста описания изобретения, а разные поля библиографической информации могут быть, в общем случае, на разных языках.

Целью данной работы является описание результатов завершеного проекта РГНФ и тех его задач, результаты решения которых в основном и определяют точность вычисленных значений индикаторов тематических взаимосвязей науки и технологий, а именно:

- поиск и выделение ссылок на цитируемые научные публикации в тексте изобретений на ЕЯ;
- рубрицирование выделенных ссылок по заданным классификаторам ННИ.

2 Поиск ссылок и их рубрикация

Всего для экспериментальных расчетов с использованием макета АИС [9] было отобрано 6665 изобретений, опубликованных Роспатентом

в 2000–2012 гг. и относящихся к классу G06 МПК. В полнотекстовых описаниях изобретений было выделено и верифицировано 8847 ссылок на цитируемые научные публикации. Архитектура АИС, ее описание и технология проведения расчетов подробно представлены в работах [9, 12]. Полнота и точность вычисленных значений индикаторов зависят от точности выделения ссылок в тексте описания изобретения и полноты перечня найденных ссылок (считается как доля от общего их числа, которые присутствуют в описании изобретения).

Построение шаблона поиска ссылки (в шаблонах используется язык регулярных выражений [13]; подробнее о построении шаблонов см. [14–17]) опирается, в основном, на требование «Административного регламента исполнения Роспатентом приема заявок на изобретение, их рассмотрения и экспертизы», содержащееся в п. 10.11(12) [10]. Это требование звучит следующим образом: «Библиографические данные источников информации указываются таким образом, чтобы источник информации мог быть по ним обнаружен». Как показывает анализ описаний изобретений, данное положение Административного регламента трактуется авторами достаточно широко, от классического представления ссылок на цитируемые публикации в научных статьях (опирающегося на стандарт представления библиографических данных [18]) до произвольной формы упоминания цитируемой публикации.

Например, в патенте 2144211 авторы упоминают источник следующим образом: «...как он [способ — прим. авт.] описан в главе 6 и главе 12 книги под названием «Адаптивная обработка сигналов» Бернарда Уидроу и Сэмьюэла Стернса, опубликованной издательством «Прентис Холл», копирайт 1985 г.».

Таким образом, создать шаблоны поиска ссылок на цитируемые публикации, которые покрывали бы все возможные варианты представления ссылок на цитируемые публикации, практически невозможно. Поэтому создаваемые шаблоны ориентированы, как правило, на поиск ссылок на публикации, приближенных к требованиям стандарта на представление библиографических данных публикаций, или нестандартных, но чаще используемых видов представления ссылок.

Ввиду того, что и сам стандарт допускает различные варианты представления библиографических данных для разных видов публикаций (книги, статьи, доклады на конференциях и пр.), возникла необходимость в разработке целой коллекции шаблонов $\{R\}$, каждый из которых нацелен на поиск заданного вида публикации в некоторой стандартной форме представления библиографических данных

или нестандартной, но частотной форме. В общем виде структура этих шаблонов опирается на обобщенную структуру ссылки цитирования публикации, которая может быть представлена в следующем виде: [автор $\{S_1\}$] [название публикации] [$\{S_2\}$ название источника] $\{S_3\}$ атрибуты публикации.

Наличие квадратных скобок говорит о необязательности присутствия данного элемента структуры в реальной ссылке на цитируемую публикацию.

В процессе поиска ссылок на публикацию в описании изобретений к его тексту применяется коллекция $\{R\}$ шаблонов поиска ссылок. Использование классов регулярных выражений обеспечивает возможность получения непересекающихся фрагментов текста, содержащих признаки ссылки на публикацию, для отдельно взятого шаблона поиска из коллекции $\{R\}$. Но в то же время разные шаблоны коллекции могут формировать пересекающиеся фрагменты текста. Следовательно, после применения к тексту всей коллекции $\{R\}$ необходимо выполнить процедуру интеграции выделенных каждым шаблоном фрагментов текста (см. [17, рис. 1]).

Каждый шаблон из коллекции $\{R\}$ имеет видовую направленность, но заранее вид публикации неизвестен, вследствие чего при его применении возможны следующие варианты выделения текста:

- (а) выделенный текст полностью соответствует ссылке на цитируемую публикацию в тексте описания изобретения;
- (б) выделенный текст содержит часть ссылки на цитируемую публикацию в тексте описания;
- (в) выделенный текст превышает ссылку на цитируемую публикацию в тексте описания и содержит часть содержательного текста описания;
- (г) выделенный текст не содержит ссылки на цитируемую публикацию, но соответствует шаблону поиска.

Для того чтобы определить качество шаблона для поиска ссылок соответствующего вида, был сформирован тестовый массив. Каждый элемент этого массива содержит неструктурированный текст описания изобретения, внутри которого находится одна ссылка на цитируемую публикацию определенного вида. Таблица 1 включает примеры ссылок на публикации в этих текстах.

Всего в тестовом массиве представлено 277 ссылок на публикации разных видов. Для каждой ссылки на публикацию указан вид публикации (статья, книга и пр.) и тип структуры ссылки. Распределение ссылок на публикации по видам и типам структуры даны в табл. 2 и 3 соответственно.

Таблица 1 Примеры из тестового массива

Номер патента	Текст ссылки на публикацию	Вид публикации
2144274	J. B. Postel. Simple mail transfer protokol. August 1982, Information Sciences Institute, University of Southern California, RFC 821	Стандарт (RFC, ГОСТ, ОСТ и пр.)
2144785	Журнал «Science», 1998. No. 5. Vol. 280. P. 1723	Статья в журнале или сборнике
2145466	Рекомендация V.110 (1988) «Голубой книги» Международного консультативного комитета по телефонии и телеграфии (ССИТТ)	Стандарт (RFC, ГОСТ, ОСТ и пр.)
2146840	Баранов С. И., Скляров В. А. Цифровые устройства на программируемых БИС с матричной структурой. — М.: Радио и связь, 1986. С. 43	Книга
2148274	Горбань А. Н. Обучение нейронных сетей. М.: СП Параграф, 1990	Книга
2149450	Вопросы проектирования радиоэлектронной аппаратуры. Опыт, результаты, проблемы. — Таллин: ЭстНИИИТИ, 1989. С. 87–90. Рис. 6, б	Книга
2149455	Омельченко В. В. Теоретические основы классификации нечетких ситуаций при испытаниях сложных технических комплексов. — М.: МО РФ, 1998. С. 328–351	Книга
2150140	Шустер Г. Детерминированный хаос. — М.: Мир, 1988. С. 33, 38	Книга

Таблица 2 Распределение ссылок по типам публикаций

Тип публикации	Количество
Книга	100
Статья в журнале или сборнике	71
Стандарт (ИСО, ГОСТ, ОСТ и пр.)	31
Веб-публикация	29
Материалы конференции	25
Отчет	16
Статья в энциклопедии	5
Всего	277

Таблица 3 Распределение по типам структур ссылок на публикации

Тип структуры текста ссылки	Количество
Авторы(ФИО)/Название/Источник/Атрибуты	67
Название/Источник/Атрибуты	64
Авторы(ИОФ)/Название/Источник/Атрибуты	59
Произвольный текст	29
Источник	18
Источник/Атрибуты	16
Название/Авторы(ФИО)/Источник/Атрибуты	9
Название/Авторы(ИОФ)/Источник/Атрибуты	4
Источник/Название/Атрибуты	5
Авторы(ИОФ)/Источник/Атрибуты	3
Авторы(ИОФ)/Название/Атрибуты	2
Авторы(ФИО)/Название/Атрибуты	1
Всего	277

¹ Исходный тестовый массив в течение времени может пополняться новыми образцами ссылок на цитируемые публикации. Поэтому при создании тестового задания для чистоты эксперимента и последующего корректного анализа создается копия исходного массива, существующего на момент создания тестового задания.

Для получения сравнительных количественных характеристик качества шаблонов поиска в рамках макета АИС была разработана отдельная подсистема, в которой реализована следующая методика анализа качества поиска ссылок на публикации: для каждого шаблона поиска ссылок создается задание на его тестирование (рис. 1).

В задании указывается имя тестируемого шаблона и формируется тестовый массив текстов изобретений на основании исходного¹, содержащих ссылки (табл. 1 содержит их примеры). В тестовом массиве помимо выделенной ссылки на публикацию в описании изобретения указывается стартовая позиция искомой ссылки на публикацию в тексте описания изобретения и длина текста ссылки в символах. Далее запускается программа поиска, которая для каждого исследуемого описания изобретения составляет список фрагментов текста, который данный шаблон поиска выделил как ссылки с указанием стартовой позиции фрагмента в тексте и его длины.

В результате формируется массив фрагментов, в котором путем сравнения стартовых позиций и длин имеющейся искомой ссылки и выделенного в процессе поиска фрагмента текста получают количественные характеристики тестируемого шаблона поиска ссылки на публикацию, которые определяются следующим образом.

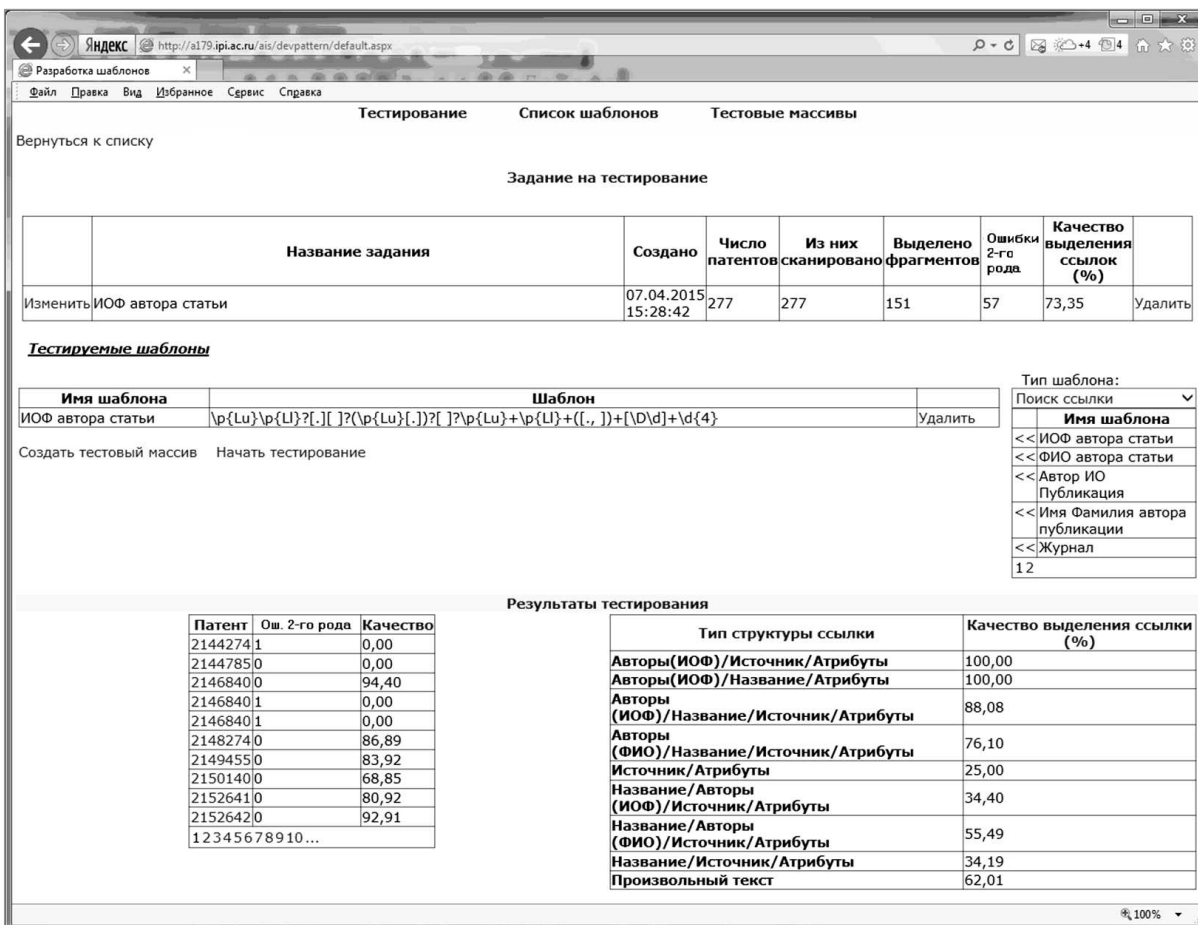


Рис. 1 Пример задания на тестирование шаблона поиска ссылок

Если выделенный фрагмент текста полностью находится за пределами искомой ссылки на публикацию, то считается, что это ложное выделение (так называемая ошибка 2-го рода). В противном случае вычисляется процент покрытия выделенным фрагментом текста искомого текста ссылки на публикацию; иными словами, определяется точность выделения ссылки цитирования.

Рисунок 1 иллюстрирует полученные результаты тестирования. Кроме того, дается список найденных фрагментов текста, который дает возможность качественной оценки работы шаблона для каждого случая его срабатывания (рис. 2).

На основе этой оценки предлагаются рекомендации по доработке (уточнению) существующего шаблона и/или созданию нового шаблона поиска. Проверять полноту и точность каждого шаблона с использованием тестового массива, можно получить ряд его характеристик:

1. Точность выделения текста ссылки на цитируемую публикацию.

2. Полноту поиска ссылок как отношение числа найденных фрагментов текста с ненулевой точностью выделения к общему числу искомых ссылок на публикации, существующих в тестовом массиве.
3. Коэффициент ложных выделений как отношение числа найденных фрагментов текста с нулевой точностью выделения к общему числу ссылок на публикацию в тестовом массиве.

Если первые две характеристики относятся к точности и полноте поиска с помощью анализируемого шаблона, то третья характеристика представляет собой частотность ошибок второго рода, порождаемых этим шаблоном.

Возможно получение и еще одной характеристики, определяющей приоритет использования анализируемого шаблона для поиска ссылок на публикации. Для этого в тестовом массиве указаны одновременно тип структуры ссылки и вид цитируемой публикации (книга, статья, доклад и пр.). Имея сводную таблицу результатов тестирования

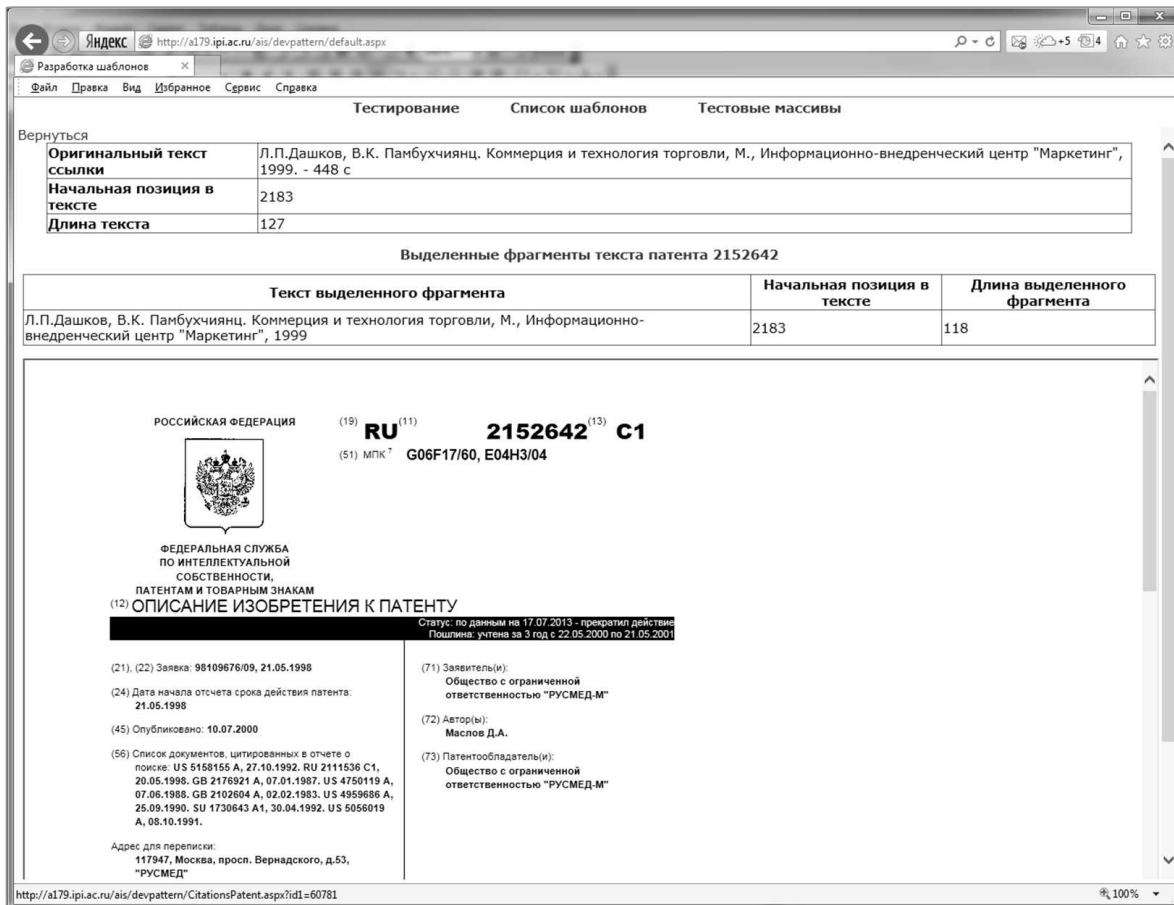


Рис. 2 Данные о выделенном фрагменте текста из описания изобретения, на которое был выдан патент 2152642

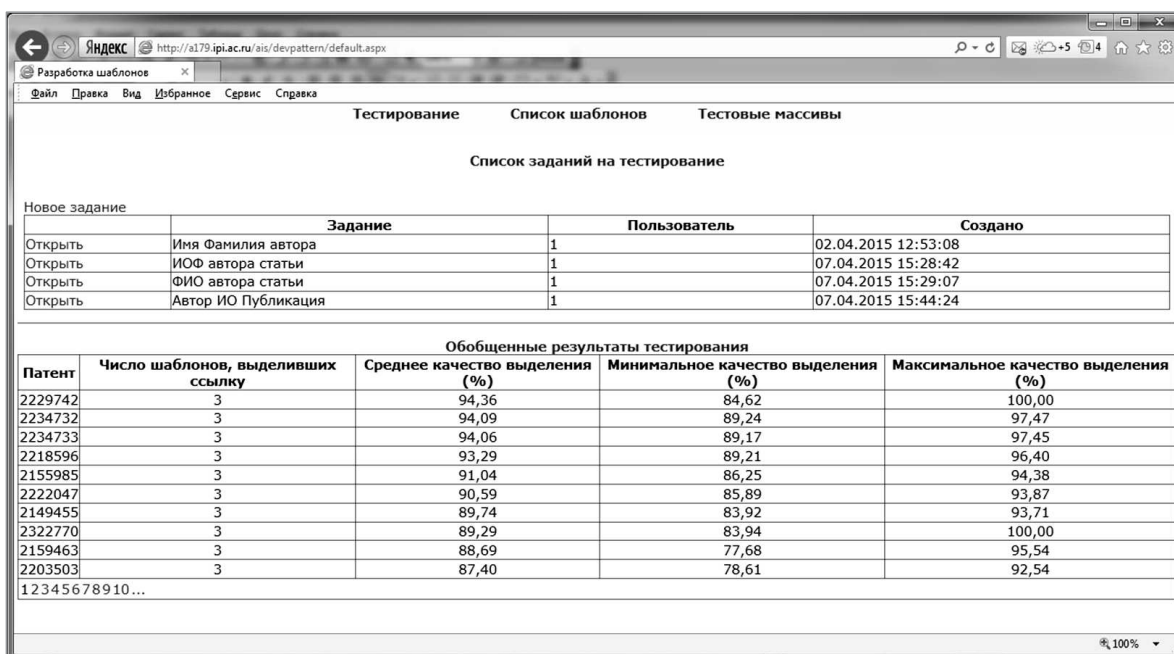


Рис. 3 Результаты выполнения тестовых заданий на тестирование шаблонов

для всех разработанных шаблонов (рис. 3), можно выбрать те шаблоны, которые дают наилучшие результаты поиска ссылок. Это позволяет сократить число используемых шаблонов поиска, выделив базовые, и, как следствие, увеличить производительность АИС на основании результатов тестирования.

Используемые в АИС методы определения ННИ, которые взаимосвязаны с заданными технологическими областями, предполагают оперирование классификаторами научной и научно-технической информации (ГРНТИ, РФФИ и пр.). Найденные и структурированные ссылки на цитируемые публикации в процессе работы АИС соотносятся с теми или иными рубриками используемых классификаторов. В процессе вычисления значений индикаторов с помощью макета АИС предполагалось, что рубрики цитируемых публикаций совпадают с рубриками изданий, в которых они были опубликованы. Это предположение, с одной стороны, существенно упростило макетирование. С другой стороны, оно снизило точность вычисленных значений индикаторов. Однако, так как основной задачей макета являлась демонстрация реализуемости методологии вычисления индикаторов взаимосвязей науки и технологий, то для этапа макетирования это предположение было допустимым.

Естественно, что в промышленном варианте АИС рубрики публикаций и изданий в общем случае совпадать не будут и должны использоваться именно рубрики публикаций.

Отметим, что существующие информационно-библиотечные системы предоставляют возможность использования именно рубрик публикаций. Например, такая возможность есть в электронных каталогах Государственной публичной научно-технической библиотеки или Всероссийского института научной и технической информации Российской академии наук [19, 20].

3 Индикаторы взаимосвязей науки и технологий

В проведенном эксперименте с использованием макета АИС вычислялись значения следующих двух индикаторов:

- (1) матрица корреляций между индексами МПК (для класса G06 и его подклассов) и рубриками ННИ ГРНТИ;
- (2) распределение времени отклика на статью (от момента ее публикации до момента публикации патента на изобретение, где она цитируется).

Подробное описание матрицы корреляций представлено в работе [9]. Частотности связей между индексами МПК и рубриками ГРНТИ, вычисленные для всего массива изобретений по информационно-компьютерным технологиям, запатентованных в РФ в период 2000–2012 гг., показаны в ячейках матрицы (табл. 4). Для компактности представления все рубрики ННИ представлены только самым верхним уровнем ГРНТИ, технологии — подклассами класса G06 МПК.

Таблица 5 отображает распределения цитируемых публикаций и изобретений для всего класса G06. Таблица 6 содержит названия индексов ГРНТИ.

В строках матрицы (см. табл. 4) приведены первые девять ННИ в классификации ГРНТИ в порядке убывания числа связей с индексами МПК по всему классу G06 (см. последний столбец таблицы). Эти частотности связей между индексами МПК и рубриками ГРНТИ были вычислены впервые в отечественной научно-технической сфере. Они определены в процессе обработки всех статей, цитируемых в описаниях изобретений по классу G06, независимо от того, кем цитируется статья: экспертами, что обозначается в описании изобретения меткой 56, и/или авторами изобретений.

Таблица 4 Частотности связей между индексами МПК и рубриками ГРНТИ, %

Код рубрики ГРНТИ	Подкласс								Класс G06
	G06E	G06F	G06G	G06K	G06M	G06N	G06Q	G06T	
50.00.00	0	9,58	0,27	4,85	0	0,62	1,01	2,05	18,38
28.00.00	0	8,40	0,11	4,10	0	0,59	1,05	1,35	15,60
47.00.00	0,10	5,16	0,41	3,86	0	0,17	0	0,50	10,20
45.00.00	0	4,88	0,26	3,49	0	0,19	0	0,35	9,17
20.00.00	0	4,41	0,02	3,25	0	0,12	0	0,05	7,85
30.00.00	0	4,19	0,00	3,23	0	0,11	0	0,04	7,57
29.00.00	0	3,79	0,02	3,17	0	0,01	0	0	6,99
84.00.00	0	3,50	0	3,11	0	0	0,01	0	6,62
27.00.00	0	2,08	0	0,71	0	0,52	1,01	1,30	5,62
Остальные рубрики	0,00	8,69	0,23	1,39	0,01	1,16	0,15	0,37	12,00

Таблица 5 Распределение изобретений и статей по подклассам класса G06

Индекс МПК	Название подкласса	Число изобретений	Число статей	Число статей на одно изобретение
G06C	Механические цифровые вычислительные машины	7	0	0
G06D	Гидравлические и пневматические цифровые вычислительные устройства	1	0	0
G06E	Оптические вычислительные устройства	52	8	0,15
G06F	Обработка цифровых данных с помощью электронных устройств	3415	107	0,03
G06G	Аналоговые вычислительные машины. . .	228	14	0,06
G06J	Гибридные вычислительные устройства	3	0	0
G06K	Распознавание, представление и воспроизведение данных; манипулирование носителями информации; носители информации	681	64	0,09
G06M	Счетчики; способы и устройства для подсчета предметов, не отнесенные к другим подклассам	12	0	0
G06N	Компьютерные системы, основанные на специфических вычислительных моделях	107	8	0,07
G06Q	Системы обработки данных или способы, специально предназначенные для административных, коммерческих (. . .) целей	417	5	0,01
G06T	Обработка или генерация данных изображения. . .	320	43	0,13
Всего по классу G06		5243	249	0,05

Таблица 6 Названия рубрик ГРНТИ из матрицы корреляций

Код ГРНТИ	Название рубрики
20.00.00	Информатика
27.00.00	Математика
28.00.00	Кибернетика
29.00.00	Физика
30.00.00	Механика
45.00.00	Электротехника
47.00.00	Электроника. Радиотехника
50.00.00	Автоматика. Вычислительная техника
84.00.00	Стандартизация

В рамках проведенного эксперимента вычислялись значения еще одного индикатора — распределения времени отклика на статьи (рис. 4). В процессе вычисления этого индикатора для каждой пары «индекс МПК — рубрика ГРНТИ» было определено время отклика как разность между годом публикации патента на изобретение и годом публикации статьи, на которую есть ссылка в описании этого изобретения. Отдельно отмечались статьи, цитируемые экспертами в отчетах о патентном поиске. Затем было построено распределение времени отклика с учетом авторства ссылок на статьи (экспер-

ты включили ссылку в отчет о патентном поиске или авторы изобретения в его полное описание).

Экспериментальные данные позволяют утверждать, что в патентах на изобретения, опубликованные в период 2000–2012 гг., эксперты в отчетах о поиске и авторы изобретений по информационным технологиям наиболее часто цитировали статьи, опубликованные за 10, 20 и 30 лет до публикации патентов на эти изобретения.

4 Заключение

Разработанный макет АИС и технология его применения впервые в отечественной научно-технической сфере дали возможность выявить количественные взаимосвязи отраслей науки и НИИ с заданным видом технологий. С помощью макета были вычислены значения индикатора тематических взаимосвязей информационных технологий, относящихся к классу G06 МПК, с рубриками ГРНТИ. Вычисленные значения показывают, что наиболее часто в изобретениях по информационно-компьютерным технологиям цитируются статьи по автоматике, вычислительной технике, кибернетике, электронике, радиотехнике, электротехнике

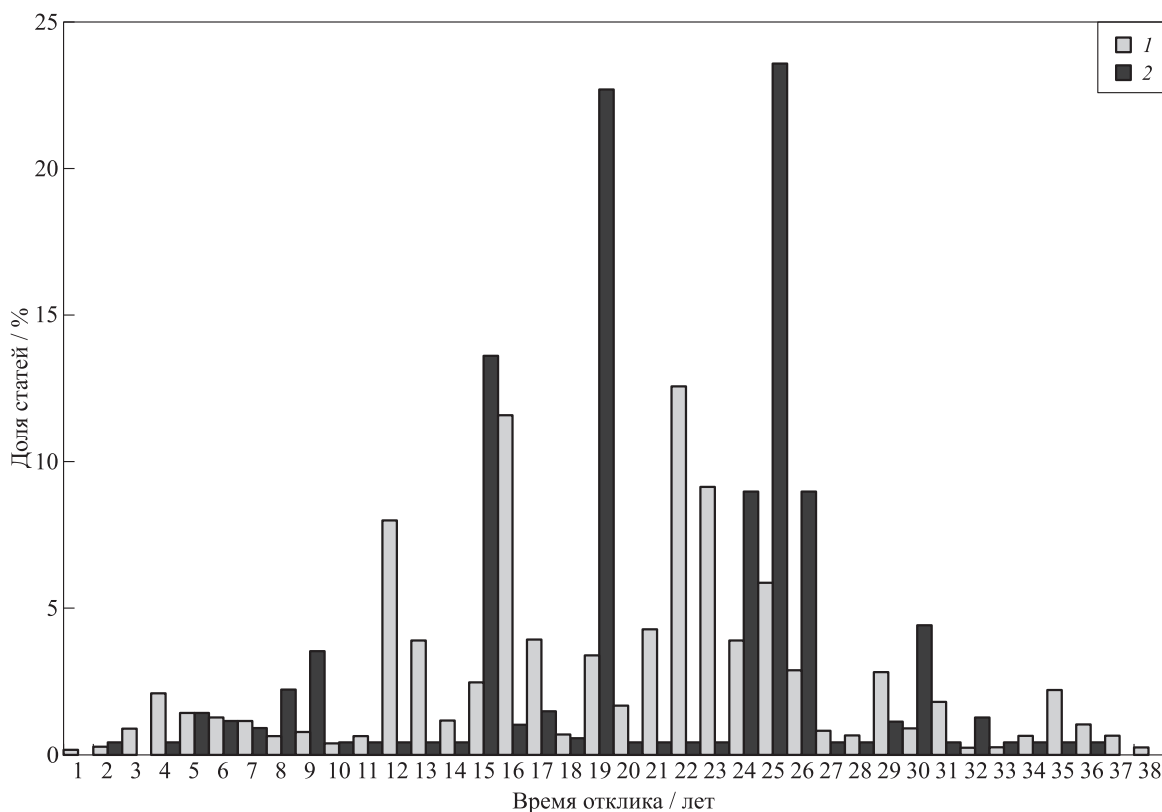


Рис. 4 Распределение времени между публикацией статьи и патента по классу G06: 1 — по полным описаниям изобретений; 2 — по спискам документов, цитируемых в отчетах об информационном поиске

и информатике (см. табл. 4). Таким образом, использование рубрик ГРНТИ дает в первую очередь прикладной разрез тематических взаимосвязей. Для получения более полной картины взаимосвязей отраслей науки с технологиями необходимо также использовать рубрики фундаментальных наук, например классификатор РФФИ.

Экспериментальные расчеты, проведенные с помощью макета АИС, позволяют сделать вывод о реализуемости методологии определения взаимосвязей отраслей науки и технологий с использованием отечественных патентных информационных ресурсов. В процессе проведения эксперимента была проведена оценка качества шаблонов поиска, используемых функциональными подсистемами макета АИС. Были получены численные оценки точности и полноты поиска ссылок на цитируемые публикации в полнотекстовых описаниях изобретений.

Выявлены наиболее сложные технологические операции поиска ссылок, рубрицирования публикаций в автоматическом режиме и определены подходы к повышению точности и полноты выделения ссылок и рубрицирования. В макете АИС предусмотрена возможность для экспертного уточнения

результатов автоматической рубрикации, выполняемой сейчас с использованием рубрик изданий, а не публикаций. Это позволяет уже сегодня использовать макет АИС для вычисления значений индикаторов взаимосвязей отраслей науки и НИИ с любым заданным видом технологий на ретроспективе в 12–15 лет и использовать ретроспективные данные для прогноза изменения значений этих индикаторов в краткосрочной перспективе.

Литература

1. Schmoch U. Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. No. 1. P. 193–211.
2. 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.
3. 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.
4. Архипова М. Ю., Зацман И. М., Шульга С. Ю. Индикаторы патентной активности в сфере информаци-

- онно-коммуникационных технологий и методика их вычисления // Экономика, статистика и информатика. Вестник УМО, 2010. № 4. С. 93–104.
5. 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.
 6. 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.
 7. 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.
 8. 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: Ун-т города Переславля, 2007. С. 35–44.
 9. 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.
 10. Административный регламент исполнения Роспатентом приема заявок на изобретение, их рассмотрения и экспертизы. — М.: ФИПС, 2008. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian.laws/order_minobr/administrative.regulations/test.8.
 11. Стандарт ВОИС ST.14. Рекомендации по включению ссылок, цитируемых в патентных документах. http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03_14_01.pdf.
 12. 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.
 13. Регулярные выражения в .NET Framework // MSDN. Библиотека. <http://msdn.microsoft.com/ru-ru/library/hs600312.aspx>.
 14. Васильев А., Козлов Д., Самусев С., Шамина О. Извлечение метаинформации и библиографических ссылок из текстов русскоязычных научных статей // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. IX Всеросс. науч. конф. RCDL'2007. — Переславль: Ун-т города Переславля, 2007. С. 175–184.
 15. Васильев А., Козлов Д., Самусев С., Шамина О. Создание электронной библиотеки русскоязычных научных статей // Сб. работ стипендиатов гранта «Интернет-математика 2007». — Екатеринбург: Уральский ун-т, 2007. С. 37–45.
 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. ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. <http://lib.usfeu.ru/index.php/gost-7-1-2003>.
 19. Сбойчаков К. О. Распределение ключевых слов по рубрикам ГРНТИ в базе данных Электронного каталога ГПНТБ России // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: Тр. XI Междунар. конф. «Крым 2004». — М., 2004. <http://www.gpntb.ru/win/inter-events/crimea2004/292.pdf>.
 20. Гиляревский Р. С., Шапкин А. В., Белоозеров В. Н. Рубрикатор как инструмент информационной навигации. — С.-Петербург: Профессия, 2008. 352 с.

Поступила в редакцию 21.04.15

INDICATORS FOR THEMATIC LINKAGES BETWEEN SCIENCE AND INFORMATION AND COMPUTER TECHNOLOGIES AT THE BEGINNING OF THE XXI CENTURY

V. A. Minin¹, I. M. Zatsman², V. A. Havanskov², and S. K. Shubnikov²

¹Russian Foundation for Basic Research, 32A Leninsky Prosp., Moscow 119991, Russian Federation

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

Abstract: Outcomes of experimental evaluation of thematic linkages between science and information and computer technologies (ICT) are presented. The indicator values for the linkages are calculated by the testbed of an analytical information system that was created within the project of the Russian Foundation for Humanities “Information system for monitoring and evaluating innovative and technological potential of the fields of basic research.” Texts of inventions on the class G06 (Data processing; Calculations; Account) of the International patent classification were used. These texts, which were published in 2000–2012 by Rospatent, are full-text descriptions of inventions in a natural language. Prior to experimental calculation of indicator values for the linkages, automated extraction of information on the cited scientific publications was retrieved from full-text descriptions. A number of publications was determined for each field of basic research. Obtained numerical information was used for quantitative evaluation of thematic science–ICT linkages and gave the possibility to define an intensity of knowledge transfer from science to ICT sphere and estimate the linkages by quantitative indicators.

Keywords: science–technology linkages; information and communication technologies; processing of invention text; regular expressions; classifying; evaluation of indicator values

DOI: 10.14357/19922264150212

Acknowledgments

The research was performed at the Institute of Informatics Problems of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences and was partially supported by the Russian Foundation of Humanities (grant No. 12-02-12019B).

References

- Schmoch, U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* 26(1):193–211.
- Zatsman, I. M., and G. F. Verevkin. 2006. Informatsionnyy monitoring sfery nauki v zadachakh programmno-tselevogo upravleniya [Information monitoring in sphere of science and problems of program-oriented management]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:164–189.
- Zatsman, I. M., and O. S. Kozhunova. 2007. Semanticheskiy slovar' sistemy informatsionnogo monitoringa v sfere nauki: Zadachi i funktsii [The semantic dictionary of system for information monitoring in science sphere: Tasks and functions]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 17:124–141.
- Arhipova, M. Yu., I. M. Zatsman, and S. Yu. Shul'ga. 2010. Indikatory patentnoy aktivnosti v sfere informatsionno-kommunikatsionnykh tekhnologiy 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.
- Zatsman, I. M., and A. A. Durnovo. 2011. Modelirovaniye 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.
- Minin, V. A., I. M. Zatsman, M. G. Kruzhkov, and T. P. Norekhan. 2013. Metodologicheskie osnovy sozdaniya informatsionnykh sistem dlya vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologiy [Methodological basis for the creation of information systems for the calculation of indicators of thematic linkages between science and technology]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):70–81.
- Verbeek, A., K. Debackere, M. Luwel, P. Andries, E. Zimmermann, and D. Deleus. 2002. Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(3):399–420.
- Zatsman, I. M., and S. K. Shubnikov. 2007. Printsipy obrabotki informatsionnykh resursov dlya otsenki innova-

- tsionnogo 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 kolleksii: 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': Pereslavl University. 35–44.
9. Minin, V.A., I.M. Zatsman, V.A. Havanskov, and S.K. Shubnikov. 2014. Indikatory tematicheskikh 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.
 10. Administrativnyy reglament ispolneniya Rospatentom priema zayavok na izobretenie, ikh rassmotreniya i ekspertizy [Administrative regulations of execution by Rospatent of demands acceptance for the invention, their considerations and examination]. Available at: http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/order_minobr/administrative_regulations/test_8/ (accessed April 17, 2015).
 11. Standart VOIS ST.14 “Rekomendatsii po vkluycheniyu ssylok, tsitiruemykh v patentnykh dokumentakh” [WIPO Standard ST.14 “Recommendations for the Inclusion of References Cited in Patent Documents”]. Available at: <http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03.14.01.pdf> (accessed April 17, 2015).
 12. Minin, V.A., I.M. Zatsman, V.A. Havanskov, and S.K. Shubnikov. 2013. Arkhitekturnye resheniya dlya sistem vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologii [Information system conceptual decisions for assessment of linkages between science and technologies]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):260–283.
 13. Regulyarnye vyrazheniya v .NET Framework [Regular expressions within .NET Framework]. Available at: <http://msdn.microsoft.com/ru-ru/library/hs600312.aspx> (accessed April 17, 2015).
 14. Vasil'ev, A., D. Kozlov, S. Samusev, and O. Shamina. 2007. Izvlechenie metainformatsii i bibliograficheskikh ssylok iz tekstov russkoyazychnykh nauchnykh statey [Extraction of metainformation and bibliographic references from texts of Russian language scientific articles]. *Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kolleksii: 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': Pereslavl University. 175–184.
 15. Vasil'ev, A., D. Kozlov, S. Samusev, and O. Shamina. 2007. Sozdanie elektronnoy biblioteki russkoyazychnykh nauchnykh statey [Creation of digital library of Russian language scientific articles]. *Sb. rabot stipendiatov granta “Internet-matematika 2007”* [Collection of works of scholars of a grant “Internet mathematics 2007”]. Ekaterinburg: Ural University. 37–45.
 16. Zatsman, I.M., V.A. Havanskov, and S.K. Shubnikov. 2013. Metod izvlecheniya bibliograficheskoy informatsii iz polnotekstovykh opisaniy izobreteniy [Method of bibliographic information extraction from full-text descriptions of inventions]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):52–65.
 17. Havanskov, V.A., and S.K. Shubnikov. 2013. Poisk i rubritirovaniye 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 kolleksii: 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: Joint Institute for Nuclear Research. 165–173.
 18. GOST 7.1-2003. Bibliograficheskaya zapis'. Bibliograficheskoe opisaniye. Obshchie trebovaniya i pravila sostavleniya [Bibliographic record. Bibliographic description. General requirements and drawing-up rules]. Available at: <http://lib.usfeu.ru/index.php/gost-7-1-2003> (accessed April 17, 2015).
 19. Sboychakov, K. O. 2004. Raspredeleniye klyuchevykh slov po rubrikam GRNTI v baze dannykh Elektronnoy kataloga GPNTB Rossii [Distribution of keywords on SCSTI headings in a database of the Electronic catalog of State Public Scientific Technical Library of Russia]. *Biblioteki i informatsionnye resursy v sovremennoy mire nauki, kul'tury, obrazovaniya i biznesa: 11-ya Mezhdunar. konf. “Krym 2004”* [Libraries and Information Resources in the Modern World of Science, Culture, Education, and Business: 11th Conference (International) “Crimea 2004”]. Moscow. Available at: <http://www.gpntb.ru/win/inter-events/crimea2004/292.pdf> (accessed May 27, 2015).
 20. Gilyarevskiy, R. S., A. V. Shapkin, and V. N. Beloozerov. 2008. *Rubrikator kak instrument informatsionnoy navigatsii* [Subject authority as instrument of information navigation]. St. Petersburg: Professiya. 352 p.

Received April 21, 2015

Contributors

Minin Vladimir A. (b. 1941) — Doctor of Science in physics and mathematics, adviser, Russian Foundation for Basic Research, 32A Leninsky Prosp., Moscow 119991, Russian Federation; minin@rfrb.ru

Zatsman Igor M. (b. 1952) — Doctor of Sciences 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@a170.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@a170.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

Беляев Константин Павлович (р. 1955) — доктор физико-математических наук, ведущий научный сотрудник Института океанологии им. П. П. Ширшова Российской академии наук

Березин Сергей Васильевич (р. 1986) — старший научный сотрудник Института прикладной математики и механики Санкт-Петербургского политехнического университета Петра Великого

Бронштейн Ефим Михайлович (р. 1946) — доктор физико-математических наук, профессор Уфимского государственного авиационного технического университета

Вагапова Диана Маратовна (р. 1987) — соискатель Уфимского государственного авиационного технического университета

Васильев Николай Семенович (р. 1952) — доктор физико-математических наук, профессор Московского государственного технического университета имени Н. Э. Баумана

Галина Ирина Владимировна (р. 1967) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зяц Олег Иванович (р. 1952) — кандидат физико-математических наук, доцент Института прикладной математики и механики Санкт-Петербургского политехнического университета Петра Великого

Козеренко Елена Борисовна (р. 1959) — кандидат филологических наук, заведующая лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, заведующий сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Мейханаджян Лусине Акобовна (р. 1990) — аспирант Российского университета дружбы народов

Милованова Татьяна Александровна (р. 1977) — кандидат физико-математических наук, старший преподаватель Российского университета дружбы народов

Минин Владимир Александрович (р. 1941) — доктор физико-математических наук, советник Российского фонда фундаментальных исследований

Морозова Юлия Игоревна (р. 1984) — научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Российского университета дружбы народов

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

Сомин Николай Владимирович (р. 1947) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Стенина Мария Михайловна (р. 1991) — студентка Московского физико-технического института

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Тучкова Наталия Павловна (р. 1955) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Хавансков Валерий Александрович (р. 1950) — научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шарнин Михаил Михайлович (р. 1959) — кандидат технических наук, старший научный сотрудник Ин-

ститута проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

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. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций, а также не содержит сведений, запрещенных к опубликованию в открытой печати.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или 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>