

Информатика и её применения

Том 13 Выпуск 3 Год 2019

СОДЕРЖАНИЕ

Методы выявления «слабых» признаков нарушений информационной безопасности	
<i>А. А. Грушо, Н. А. Грушо, Е. Е. Тимонина</i>	3
Об асимптотике кластерного коэффициента конфигурационного графа с неизвестным распределением степеней вершин	
<i>Ю. Л. Павлов</i>	9
Об оценках скорости сходимости для некоторых моделей массового обслуживания с неполно заданными интенсивностями	
<i>А. И. Зейфман, Я. А. Сатин, К. М. Киселева</i>	14
Нетранзитивные триплеты непрерывных случайных величин и их приложения	
<i>А. В. Лебедев</i>	20
Априорное обобщенное гамма-распределение в байесовских моделях баланса	
<i>А. А. Кудрявцев</i>	27
Гибридные модели экстремального градиентного бустинга для восстановления пропущенных значений в данных об осадках	
<i>А. К. Горшенин, О. П. Мартынов</i>	34
Управление выходом стохастической дифференциальной системы по квадратичному критерию. III. Анализ свойств оптимального управления	
<i>А. В. Босов, А. И. Стефанович</i>	41
О решении проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением	
<i>П. В. Шнурков, Н. А. Вахтанов</i>	50
Оценка уровня значимости критерия Шуирманна для проверки гипотезы биоэквивалентности при наличии пропущенных данных	
<i>Т. В. Захарова, А. А. Тархов</i>	58
Формализация метода ранжирования альтернатив для процесса группового принятия решений при анализе социальных сетей	
<i>А. А. Гайдамака, Н. В. Чухно, О. В. Чухно, К. Е. Самуйлов, С. Я. Шоргин</i>	63
Вычисление показателей оптимальных по CC-VaR портфелей на рынках опционов	
<i>Г. А. Агасандян</i>	72
Интерфейсы третьего порядка в информатике	
<i>И. М. Зацман</i>	82
Архитектура системы нейронного машинного перевода	
<i>В. А. Нуриев</i>	90

Информатика и её применения

Том 13 Выпуск 3 Год 2019

СОДЕРЖАНИЕ

Методика поиска имплицитных логико-семантических отношений в тексте А. А. Гончаров, О. Ю. Инькова	97
Персональный когнитивный ассистент: концепция и принципы работы И. В. Смирнов, А. И. Панов, А. А. Скрынник, Е. В. Чистова	105
Применение рекуррентных нейронных сетей для прогнозирования моментов конечных нормальных смесей А. К. Горшенин, В. Ю. Кузьмин	114
Методы моделирования и визуального представления конфликта в малом коллективе экспертов, решающих проблемы (обзор) С. Б. Румовская, И. А. Кириков	122
Разработка метода формирования признакового пространства и модели для оценки и прогнозирования антропогенного влияния на окружающую среду (на примере лесного фонда нефтедобывающего региона) В. В. Бурлуцкий, А. В. Якимчук, А. В. Мельников, А. Л. Царегородцев, С. В. Волошин	131
Научный результат как информационный объект в контексте системы управления научными сервисами А. П. Сучков	137
Об авторах	145
Правила подготовки рукописей	148
Requirements for manuscripts	151

МЕТОДЫ ВЫЯВЛЕНИЯ «СЛАБЫХ» ПРИЗНАКОВ НАРУШЕНИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ*

А. А. Грушо¹, Н. А. Грушо², Е. Е. Тимонина³

Аннотация: Предложен новый подход к выявлению «слабых» признаков нарушений информационной безопасности. Исходной информацией для выявления «слабых» признаков нарушений информационной безопасности инсайдером-злоумышленником служат наблюдаемые потенциальные цели инсайдера-злоумышленника. Появление новой ценной информации, в которой заинтересован инсайдер-злоумышленник, вызовет поведенческую реакцию инсайдера-злоумышленника в каких-то информационных пространствах. Методы поиска таких реакций в различных информационных пространствах являются целью данной работы. Построена вероятностная модель реакции инсайдера-злоумышленника в случае повторного появления цели. Показано, что использование многих информационных пространств существенно повышает возможности выявления влияния цели на поведение инсайдера-злоумышленника.

Ключевые слова: информационная безопасность; информационные пространства; поведенческие признаки нарушителя информационной безопасности

DOI: 10.14357/19922264190301

1 Введение

Цифровая экономика подразумевает массовое использование вычислительной техники в бизнесе. Предполагается, что через сеть (например, интернет) можно наблюдать за поведением пользователей и связывать это поведение с действиями нарушителя информационной безопасности.

Проблема поиска инсайдера-злоумышленника (далее — инсайдера) актуальна и исследовалась с различных сторон многими учеными. Надо отметить наиболее серьезные исследования в этой области. В рамках программы CERT (Computer Emergency Response Team) в 2002 г. был организован центр изучения угроз инсайдера [1]. Работы CERT носят, в основном, обзорный и методический характер. Согласно этим исследованиям отмечено следующее:

- почти всегда для инсайдера характерны специфические особенности поведения. В частности, отмечается наличие стресса, инициированного внешними причинами или осознанием противоправности своих действий [2];
- для атаки инсайдером создаются скрытые пути доступа к информации, которые изобретаются

и реализуются заранее с использованием специфики информационной системы [3].

Как правило, действия инсайдера в информационной системе связано с появлением где-либо аномалий. В 2010 г. DARPA (Defense Advanced Research Projects Agency) начала программу ADAMS (Anomaly Detection At Multiple Scales) [4] выявления аномалий в информационных системах. В рамках этой программы реализован проект PRODIGAL (PROactive Discovery of Insider threats using Graph Analysis and Learning) [5]. Главной целью проекта PRODIGAL ставилась разработка подходов к обнаружению «слабых сигналов», которые характерны для хорошо замаскированных действий инсайдеров. Предполагается, что «слабые сигналы» присутствуют в действиях и функциях пользователей компьютерных систем.

Среди целей инсайдеров выделим следующие:

- хищение информации;
- нарушение целостности информации и ресурсов.

Один из выводов проекта PRODIGAL состоял в том, что для выявления инсайдеров требуется дополнительная информация, не связанная с действиями пользователей в компьютерных системах,

* Работа частично поддержана РФФИ (проект 18-07-00274).

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

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

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

т. е. для выявления инсайдера необходимо искать различные дополнительные источники данных поведенческого характера.

В работах [6, 7] построена модель независимого исследования нескольких информационных пространств и предложены формальные методы объединения результатов этих исследований на основе понятия запрета вероятностной меры [8].

Из работы [5] следует, что многообразии возможностей для действий инсайдера порождает неопределенность в выборе средств анализа данных. Отметим, что возможны изощренные пути доступа к ценной информации, которые не выявляются с помощью логов компьютерных систем [3].

В основном поиск аномалий ведется статистическими методами в условиях множества малых выборок. При использовании этих методов следует учитывать различные виды шума и возникающих отсюда ошибок в принятии решений. В работе [9] показано, что при обработке большого числа малых выборок либо часто возникают ложные тревоги, либо существенно увеличивается вероятность потери правильного решения.

В данной работе предложен подход, отличный от подходов, указанных выше. Исходной информацией для выявления инсайдера служат его наблюдаемые потенциальные цели. Допустим, что квалификация инсайдера достаточно высока для того, чтобы сделать доступ к ценной информации «невидимым» для контролеров — офицеров службы информационной безопасности. Однако допустим, что новая ценная информация, представляющая интерес для инсайдера, появляется достаточно регулярно. Появление новой ценной информации, в которой заинтересован инсайдер, каким-то неизвестным образом становится ему известной, и это вызовет его поведенческую реакцию в каких-то информационных пространствах. Методы поиска таких реакций в различных информационных пространствах являются целью данной работы.

Обозначим через Θ появление цели для инсайдера в компьютерной системе. Например, Θ — это очередное появление новой ценной информации. Допустим, что Θ в обновленном виде повторяется, т. е. если инсайдер «добывает» ценную информацию в компьютерной системе, то он делает это неоднократно (и неизвестным образом).

Выделим моменты появления Θ и обозначим это множество моментов (+). События Θ позволяют выделить «разумные» сегменты наблюдаемых событий в процессах ожидаемой реакции инсайдера в выбранном информационном пространстве. Множество этих интервалов также будем обозначать (+). Если цели Θ отражаются в наблюдаемом процессе в выбранном информационном простран-

стве, то возникают некоторые повторы реакций в выделенных сегментах (+). Эти повторы могут появляться нерегулярно, поэтому их можно отнести к «слабым влияниям».

Обозначим через (–) данные в наблюдаемом процессе в выбранном информационном пространстве вне интервалов (+). Отсюда появляется возможность сравнения и выявления неслучайных повторов искомым поведенческих реакций. Если появлений Θ набирается достаточно много, то выявление влияния Θ на наблюдаемый процесс в выбранном информационном пространстве может быть хорошо заметно.

Решение задачи выявления влияния Θ на наблюдаемый процесс в выбранном информационном пространстве может вестись несколькими способами:

- логическими методами [10] — выделение знаний из фактов;
- статистическими методами [11] — выделение вкраплений, используя данные (–) для формирования понятия «обычное поведение».

В отличие от [11], в поиске влияния Θ порядок интервалов оказывается несущественным.

Работая статистическими методами, необходимо иметь в виду, что возможно случайное появление признаков, которые можно было бы считать влиянием. Поэтому надо выявлять различия между появлением искомым признаков (+) и (–).

2 Математическая модель

Построим математическую модель влияния появления целей для инсайдера на аспекты его поведения. Пусть в бизнес-процессе организации и в порождаемых им информационных технологиях возникают события Θ , которые предположительно интересны для инсайдера. Предположим, что рассматривается информационное пространство, в котором могут проявиться действия или изменения поведения инсайдера при появлениях Θ . Пусть время t дискретно и офицер безопасности может синхронизировать появление Θ и наблюдаемый в информационном пространстве процесс событий $\xi^+(t)$, $t = 1, 2, \dots$, который может отражать реакцию инсайдера на Θ . Предположим, что офицер безопасности знает допустимые времена задержки, которые происходят между событиями Θ и возможным влиянием этих событий на поведение инсайдера. Эти выделенные сегменты обозначим (+). Пусть имеется возможность наблюдать независимый аналогичный процесс $\xi^-(t)$, $t = 1, 2, \dots$, без событий Θ и вне временных сегментов.

Определим универсальный язык L значений наблюдаемого процесса в информационном пространстве следующим образом. Поскольку целью исследования являются действия инсайдера и их характеристики, то алфавит A в языке L — это пары (действия, значения характеристик при выполнении действий). Выражения языка L состоят из одной или нескольких букв алфавита A .

Пример. Пусть наблюдаются звонки пользователей. Тогда алфавит A состоит из пар (звонок, номер). Выделенные (+) пары в моменты Θ не означают, что именно данный звонок надо рассматривать как признак влияния. Однако выделение нескольких выражений после очередного появления Θ может нести информацию о влиянии.

В качестве основной исследуемой характеристики будем изучать превышение повторов некоторых букв или выражений в множестве (+) над аналогичными характеристиками, типичными для множества (–) для пользователя. С этой целью необходимо исследовать возможности случайных повторов в множестве (–). Будем считать, что буквы алфавита или выражения из выделенного фиксированного множества появляются равновероятно и независимо друг от друга (распределения можно оценивать из практики).

Воспользуемся сделанным ранее замечанием, что, в отличие от модели [10], в данной задаче представляют интерес только повторы букв или выражений. Тогда моделью для множества (–) служит классическая задача размещения частиц по ячейкам [12]. Пусть все рассматриваемые выражения имеют фиксированную длину h . Предположим, что в множестве всех выражений длины h в алфавите A определена равновероятная мера. Эти выражения будем рассматривать как уникальные имена ячеек, тогда последовательность $\xi^-(t)$, $t = 1, 2, \dots$, будет формально распределять различные частицы по этим ячейкам.

Задача о выявлении повторов сводится к существованию ячеек, число частиц в которых превышает некоторый порог, определяемый из модели (–). Пороги заполнения ячеек можно получить из данной модели (–).

Рассмотрим вероятностную схему размещения n частиц по N ячейкам [13]. Обозначим η_i число частиц в ячейке с номером i , $i = 1, \dots, N$. Пусть $\eta_{(1)}, \dots, \eta_{(N)}$ — вариационный ряд, построенный по набору (η_1, \dots, η_N) , где $\eta_{(1)}$ — минимальное число частиц в ячейке, а $\eta_{(N)}$ — максимальное число. Обозначим

$$\alpha = \frac{n}{N}; \quad p_k = \frac{\alpha^k}{k!} e^{-\alpha}, \quad k = 0, 1, \dots$$

Предельное поведение $\eta_{(N)}$ описывается, например, следующей теоремой.

Теорема 1 [13]. Если $n, N \rightarrow \infty$, $\alpha / \ln N \rightarrow 0$ и $r = r(\alpha, N)$ выбрано так, что $r > 0$ и $Np_r \rightarrow \lambda$, где λ — положительная постоянная, то

$$\begin{aligned} P \{ \eta_{(N)} = r - 1 \} &\rightarrow e^{-\lambda}; \\ P \{ \eta_{(N)} = r \} &\rightarrow 1 - e^{-\lambda}. \end{aligned}$$

Вернемся к примеру. Пусть у подозреваемого инсайдера 50 телефонных номеров. Рассмотрим $h = 1$, $N = 50$, $\sqrt{N} \approx 7,07$, а n — сумма длин сегментов (+). Согласно [13], в асимптотике при $n, N \rightarrow \infty$ и $n = \sqrt{N}$ возможны случайные повторения. Поэтому очевидно, что для такого короткого объема данных (+) поиск повторений не имеет смысла.

Однако можно рассмотреть повторы выражений, состоящих из трех упорядоченных звонков (которые могут быть разнесены между другими звонками), попавших в отдельные сегменты (+). В данном случае асимптотика работает достаточно уверенно, а именно: $N = 50^3 = 125 \cdot 10^3$, $\sqrt{N} \approx 353,55$.

Предположим, что множество сегментов (+) состоит из 5 сегментов длиной 10 каждый и все подпоследовательности длиной 3 независимы и равновероятны. В каждом сегменте содержится $n = 720$ подпоследовательностей длиной 3. Тогда необходимо оценить возможности появления повторов в аналогичном наборе сегментов (–). Выбранные параметры показывают, что каждый сегмент приблизительно удовлетворяет условиям теоремы 1 для $r = 2$ при $\lambda \approx 2$. Это означает, что в одном сегменте возможны повторы, но не более двух раз для одного знака. Поэтому в данных (–) целесообразно рассматривать возможность повторов трехчленных выражений по одному в каждом из трех сегментов, т. е. целесообразно рассматривать 3 сегмента B , C и D , длиной 10 букв каждый, так как в любых двух сегментах возможны указанные повторения выражений (по одному в каждом).

Выделим по одному фиксированному месту в каждом из сегментов C и D . Вероятность того, что на этих местах расположены одинаковые фиксированные буквы, равна $1/N^2$. Вероятность того, что первая буква сегмента B совпадает с буквой в выделенных местах сегментов C и D , равна $1/N^2$, а вероятность того, что на первом месте сегмента B находится буква, не совпадающая с буквой в выделенных местах сегментов C и D , равна $1 - 1/N^2$. Вероятность того, что ни одна буква из сегмента B не совпадет с буквой в выделенных местах сегментов C и D , равна $(1 - 1/N^2)^{720}$. Тогда вероятность хотя бы одного совпадения буквы

в сегменте B и в фиксированных местах сегментов C и D равна $1 - (1 - 1/N^2)^{720}$. Отсюда получаем среднее число повторов букв одновременно в B , C и D : $720^2 \left(1 - (1 - 1/N^2)^{720}\right)$.

Оценку искомой вероятности получаем из неравенства Маркова [14] и полученной формулы для математического ожидания. Отсюда получаем оценку искомой вероятности:

$$\frac{720^3}{(125 \cdot 10^3)^2} \cong 0,024.$$

Аналогично получаем оценку вероятности встретить такой повтор в трех из пяти сегментов:

$$\binom{5}{3} \cdot 0,024 = 0,24.$$

Этот результат можно считать подтверждением наличия влияния.

3 Случай многих информационных пространств

Различные информационные пространства могут в совокупности повышать уверенность в наличии влияния Θ на поведение инсайдера. Для простоты рассмотрим следующую схему.

Пусть определены два информационных пространства IS_1 и IS_2 и мощность каждого из них равна N . Пусть по-прежнему существует возможность синхронизации событий Θ с последовательностью значений во всех информационных пространствах, т. е. в предположении дискретного времени значения рассматриваемых процессов во всех пространствах описываются векторами \bar{x}_n , $n = 0, 1, \dots$, где $\bar{x}_n = (\xi_n^{(1)}, \xi_n^{(2)})$, $\xi_n^{(i)} \in IS_i$, $i = 1, 2$. Опять сведем задачу к поиску повторов и для этого используем вероятностную схему размещения частиц по ячейкам [13].

Одновременное влияние в информационных пространствах IS_1 и IS_2 описывается увеличением числа повторов по каждой координате в m сегментах (+) длины n последовательностей \bar{x}_t , $t = 0, 1, \dots$. Если n и N удовлетворяют теореме 1 при $r = 2$, то возможны случайные пары повторов по каждой координате вектора \bar{x}_t . Поэтому необходимо рассматривать тройные повторы по каждой координате. Выберем произвольным образом одно из информационных пространств. Число троек возможных сегментов в этом пространстве равно $\binom{m}{3}$.

В схеме (–) вероятность того, что данная тройка содержит один и тот же элемент на фиксированных местах, равна $N \cdot 1/N^3$. Тогда математическое ожидание числа повторов длины 3 во множестве сегментов выбранного информационного пространства равно $\binom{m}{3} \frac{n^3}{N^2}$.

Пусть $n = \sqrt{N}$, тогда указанное математическое ожидание имеет порядок $\binom{m}{3} / \sqrt{N}$.

Используя неравенство Маркова [14], получим оценку вероятности появления тройки повторов в различных сегментах одного информационного пространства. Независимость координат вектора \bar{x}_t позволяет оценить вероятность случайного возникновения влияния в информационных пространствах IS_1 и IS_2 следующим образом: $\binom{m}{3}^2 / N$.

4 Заключение

В работе рассмотрена задача выявления инсайдера по изменению его поведения при появлении цели его преступной деятельности. Предположим, что инсайдер обладает временем и интеллектом, чтобы скрыть от офицеров информационной безопасности, откуда и как он получает информацию о цели преступной деятельности. В связи с этим необходимо искать «следы» его интереса к цели в различных информационных пространствах, которые доступны наблюдению офицеру информационной безопасности.

Построена вероятностная модель реакции инсайдера в случае повторного появления цели. Показано, что использование многих информационных пространств существенно повышает возможности выявления влияния цели на поведение инсайдера.

Отметим, что асимптотические результаты используются в основном как «путеводная звезда», позволяющая правильно сформировать условия поиска влияния.

Литература

1. Cappelli D., Moore A., Trzeciak R. The CERT guide to insider threats: How to prevent, detect, and respond to information technology crimes (theft, sabotage, fraud). — Addison-Wesley Professional, 2012. 430 p.
2. Band S. R., Cappelli D., Fischer L. F., Moore A. P., Shaw E. D., Trzeciak R. F. Comparing insider IT sabotage and espionage: A model-based analysis. CMU/SEI-2006-TR-026. — Software Engineering Institute, Carnegie Mellon University, 2006. 108 p. <http://>

- resources.sei.cmu.edu/library/asset-view.cfm?AssetID=8163.
3. Тимонина Е. Е. Анализ угроз скрытых каналов и методы построения гарантированно защищенных распределенных автоматизированных систем: Дисс. . . . докт. техн. наук. — М.: РГГУ, 2004. 204 с.
 4. Anomaly Detection at Multiple Scales (ADAMS). 2011. <https://info.publicintelligence.net/DARPA-ADAMS.pdf>.
 5. Senator T. E., Goldberg H. G., Memory A., et al. Detecting insider threats in a real corporate database of computer usage activity // 19th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings. — New York, NY, USA: ACM, 2013. P. 1393–1401.
 6. Грушо А. А., Грушо Н. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. Параметризация в прикладных задачах поиска эмпирических причин // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 62–66.
 7. Грушо А. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. Модель множества информационных пространств в задаче поиска инсайдера // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 65–69.
 8. Грушо А., Тимонина Е. Запреты в дискретных вероятностно-статистических задачах // Дискретная математика, 2011. Т. 23. Вып. 2. С. 53–58.
 9. Axelsson S. The base-rate fallacy and its implications for the difficulty of intrusion detection // 6th Conference on Computer and Communications Security Proceedings. — New York, NY, USA: ACM, 1999. P. 1–7.
 10. Финн В. К. Искусственный интеллект: Методология, применения, философия. — М.: Красанд, 2011. 448 с.
 11. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ приложения / Пер. с англ. — М.: Мир, 1981. 694 с. (Tukey J. W. Exploratory data analysis. — Addison Wesley Pub. Co., Inc., 1977. 711 p.)
 12. Феллер В. Введение в теорию вероятностей и ее приложения / Пер. с англ. — М.: Мир, 1967. Т. 1. 499 с. (Feller W. An introduction to probability theory and its applications. — 2nd ed. — New York, NY, USA: John Wiley and Sons, Inc., 1950. Vol. 1. 520 p.)
 13. Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. — М.: Наука, 1976. 224 с.
 14. Ширяев А. Н. Вероятность. — В 2-х кн. — 3-е изд., перераб. и доп. — М.: МЦНМО, 2004. 521 с.

Поступила в редакцию 07.07.19

METHODS OF IDENTIFICATION OF “WEAK” SIGNS OF VIOLATIONS OF INFORMATION SECURITY

A. A. Grusho, N. A. Grusho, and E. E. Timonina

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: New approach of identification of “weak” signs of violations of information security is suggested. Initial information for identification of “weak” signs of violations of information security by the insider-malefactor are the observed potential purposes of the insider-malefactor. Emergence of new valuable information, in which the insider-malefactor is interested, will cause behavioral reaction of the insider-malefactor in some information spaces. Methods of searching of such reactions in various information spaces are the purpose of this work. The probability model of a reaction of an insider-malefactor in case of repeated emergence of a purpose is constructed. It is shown that usage of many information spaces significantly increases possibilities of identification of influence of a purpose on behavior of an insider-malefactor.

Keywords: information security; information spaces; behavioral signs of a violator of information security

DOI: 10.14357/19922264190301

Acknowledgments

The paper was partially supported by the Russian Foundation for Basic Research (project 18-07-00274-a).

References

1. Cappelli, D., A. Moore, and R. Trzeciak. 2012. *The CERT guide to insider threats: How to prevent, detect, and respond to information technology crimes (theft, sabotage, fraud)*. Addison-Wesley Professional. 430 p.
2. Band, S., D. Cappelli, L. Fischer, A. Moore, E. Shaw, and R. Trzeciak. 2006. Comparing insider IT sabotage and espionage: A model-based analysis. CMU/SEI-2006-TR-026. Software Engineering Institute, Carnegie Mellon University. 108 p. Available at: <http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=8163> (accessed June 20, 2019).
3. Timonina, E. E. 2004. Analiz ugroz skrytykh kanalov i metody postroeniya garantirovanno zashchishchennykh raspredelennykh avtomatizirovannykh sistem [The analysis of threats of covert channels and methods of creation

- of guaranteed protected distributed automated systems]. Moscow: Russian State University for the Humanities. D.Sc. Diss. 204 p.
4. Anomaly Detection at Multiple Scales (ADAMS). 2011. Available at: <https://info.publicintelligence.net/DARPA-ADAMS.pdf> (accessed June 20, 2019).
 5. Senator, T. E., H. G. Goldberg, A. Memory, *et al.* 2013. Detecting insider threats in a real corporate database of computer usage activity. *19th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings*. New York, NY: ACM. 1393–1401.
 6. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2018. Parametrizatsiya v prikladnykh zadachakh poiska empiricheskikh prichin [Parametrization in applied problems of search of the empirical reasons]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):62–66.
 7. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):65–69.
 8. Grusho, A., and E. Timonina. 2011. Prohibitions in discrete probabilistic statistical problems. *Discrete Math. Appl.* 21(3):275–281.
 9. Axelsson, S. 1999. The base-rate fallacy and its implications for the difficulty of intrusion detection. *6th ACM Conference on Computer and Communications Proceedings*. New York, NY: ACM. 1–7.
 10. Finn, V. K. 2011. *Iskusstvennyy intellekt: metodologiya, primeneniya, filosofiya* [Artificial intelligence: Methodology, applications, philosophy]. Moscow: KRASAND. 448 p.
 11. Tukey, J. W. 1977. *Exploratory data analysis*. Addison Wesley Pub. Co., Inc. 711 p.
 12. Feller, W. 1950. An introduction to probability theory and its applications. 2nd ed. New York, NY: John Wiley and Sons, Inc. Vol. 1. 520 p.
 13. Kolchin, V. F., B. A. Sevastyanov, and V. P. Chistyakov. 1978. *Random allocations*. Washington, DC: V. H. Winston & Sons. 270 p.
 14. Shiryaev, A. N. 2004. *Veroyatnost'* [Probability]. Moscow: MTsNMO. 521 p.

Received July 7, 2019

Contributors

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

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

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

ОБ АСИМПТОТИКЕ КЛАСТЕРНОГО КОЭФФИЦИЕНТА КОНФИГУРАЦИОННОГО ГРАФА С НЕИЗВЕСТНЫМ РАСПРЕДЕЛЕНИЕМ СТЕПЕНЕЙ ВЕРШИН*

Ю. Л. Павлов¹

Аннотация: Рассматриваются конфигурационные графы со случайными одинаково распределенными степенями вершин. Степень каждой вершины равна числу выходящих из нее различных полуребер. Граф строится путем попарного равновероятного соединения полуребер друг с другом для образования ребер. Такой граф допускает петли и кратные ребра. Конфигурационные графы широко используются для моделирования сложных сетей коммуникаций, включая интернет, социальные, транспортные, телефонные сети. Распределение степеней вершин может быть неизвестным. Предполагается только, что оно имеет конечное математическое ожидание и либо дисперсия тоже конечна, либо выполнены некоторые достаточно слабые ограничения на асимптотическое поведение хвоста распределения. Обсуждаются понятия кластерного коэффициента и его свойства в таких графах. При стремлении к бесконечности числе вершин доказана предельная теорема для кластерного коэффициента. Найдены условия, при которых этот коэффициент неограниченно возрастает.

Ключевые слова: случайные графы; конфигурационные графы; кластерный коэффициент; предельные теоремы

DOI: 10.14357/19922264190302

1 Введение

Случайные графы широко используются для моделирования сложных сетей коммуникаций, таких как интернет, системы мобильной связи, социальные, транспортные, электрические сети и др. Наиболее полный обзор таких моделей можно найти в книге [1]. Известно, что, хотя сети и имеют различную природу, они обладают общими свойствами. Одним из важнейших свойств является эмпирически установленный факт, что степени вершин в графах, моделирующих сети, можно считать независимыми одинаково распределенными случайными величинами. При этом число вершин, степень которых не меньше k , при достаточно больших k пропорционально $k^{-\tau}$, где $\tau > 0$. Поэтому распределение случайной величины ξ , равной степени любой вершины, можно представить следующим образом:

$$P\{\xi \geq k\} = \frac{h(k)}{k^\tau}, \quad k = 1, 2, \dots, \quad (1)$$

где $h(k)$ — медленно меняющаяся функция. В настоящее время существуют два основных класса случайных графов, удовлетворяющих усло-

вию (1). Это графы предпочтительного присоединения, идея которых была изложена в [2], и конфигурационные графы, впервые предложенные в [3].

Будем рассматривать конфигурационные графы с N вершинами, степени которых ξ_1, \dots, ξ_N являются независимыми случайными величинами с распределением (1). Степень каждой вершины равна числу выходящих из нее полуребер, т. е. ребер, для которых смежные вершины еще не определены. Все полуребра занумерованы в произвольном порядке. Сумма степеней вершин любого графа должна быть четной, поэтому в случае нечетной суммы в граф добавляется дополнительная вершина единичной степени. В статье [4] отмечалось, что эта дополнительная вершина, вместе с инцидентным ей ребром, не влияет на асимптотические свойства графа при $N \rightarrow \infty$. Поэтому далее не будем учитывать дополнительную вершину в случае ее появления.

Построение конфигурационного графа со случайными степенями вершин завершается путем попарного равновероятного соединения полуребер друг с другом для образования ребер. Поскольку такое соединение проводится без ограничений, в графе возможно появление петель и кратных ре-

*Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания Карельского научного центра Российской академии наук (Институт прикладных математических исследований КарНЦ РАН).

¹Институт прикладных математических исследований Федерального исследовательского центра «Карельский научный центр Российской академии наук», pavlov@krc.karelia.ru

бер. Ранее был доказан ряд предельных теорем о степенной структуре конфигурационных графов, описывающих асимптотическое поведение числа вершин заданной степени и максимальной степени вершин (см. [5, 6] и библиографию в них). Значительное внимание было уделено условным графам при условии, что число ребер известно. Такие случайные графы полезны при моделировании сетей, в которых можно оценить число связей. Но и для сетей без ограничений разумно использовать результаты об условных графах, усредняя их по известным распределениям числа ребер. В последние годы при изучении графов, степени которых имеют распределение вида (1), внимание исследователей привлечено не только к свойствам степенной структуры, но и к другим числовым характеристикам, более подробно описывающим особенности сетей (см., например, [7]). Одной из важнейших таких характеристик служит кластерный коэффициент. В следующем разделе дается определение этого коэффициента и обсуждаются некоторые его свойства, а в последнем разделе доказывается теорема о предельном поведении кластерного коэффициента при стремлении к бесконечности числа вершин графа.

2 Кластерный коэффициент

Для того чтобы дать определение кластерного коэффициента, воспользуемся терминологией и обозначениями из [1]. Рассмотрим граф $G = G(V, E)$ с N вершинами, образующими множество V , и с множеством ребер E . Пусть (ij) означает ребро из E , соединяющее вершины $i, j \in V$. Обозначим через i, j, t три разных вершины графа. Если $(ij), (jt), (ti) \in E$, то будем говорить, что вершины i, j, t образуют треугольник в G . Введем сумму

$$\begin{aligned} W_G &= \sum_{1 \leq i, j, t \leq N} I\{(ij), (jt) \in E\} = \\ &= 2 \sum_{1 \leq i < j < t \leq N} I\{(ij), (jt) \in E\}, \end{aligned}$$

где $I\{A\}$ — индикатор события A . Пусть также

$$\begin{aligned} \Delta_G &= \sum_{1 \leq i, j, t \leq N} I\{(ij), (jt), (ti) \in E\} = \\ &= 6 \sum_{1 \leq i < j < t \leq N} I\{(ij), (jt), (ti) \in E\}. \end{aligned}$$

Кластерным коэффициентом C_G графа G называется величина

$$C_G = \frac{\Delta_G}{W_G}. \quad (2)$$

Таким образом, C_G равен отношению утроенного числа различных треугольников графа к числу различных пар смежных ребер. Поэтому коэффициент (2) можно интерпретировать как вероятность того, что если у произвольной вершины i имеются две смежные с ней вершины j и t , то существует также и ребро (jt) . Заметим, что ребро (jt) может быть кратным, тогда вершины i, j, t при фиксированных ребрах (ij) и (it) образуют несколько треугольников, число которых равно кратности ребра (jt) .

Обозначим

$$p_k = P\{\xi_i = k\}, \quad k = 1, 2, \dots, \quad i = 1, \dots, N. \quad (3)$$

Далее будем считать, что распределение степени любой вершины имеет конечное математическое ожидание

$$m = \sum_{s=1}^{k_{\max}} sp_s, \quad (4)$$

где k_{\max} — максимально возможная степень вершины.

Пусть произвольные вершины i и j ($i \neq j$) соединены ребром (ij) , а в случае нескольких таких ребер рассмотрим одно из них. Если при построении графа из вершины j выходит $k + 1$ полуредер, то одно из них необходимо для формирования ребра (ij) , а остальные k полуредер могут быть использованы произвольно. Обозначим q_k вероятность того, что степень вершины j равна k без учета ребра (ij) . В [7] показано, что

$$q_k = \frac{(k + 1)p_{k+1}}{m}, \quad k = 0, 1, \dots, k_{\max} - 1. \quad (5)$$

Рассмотрим случай, когда кроме вершины j существует еще и другая смежная с i вершина t . Очевидно, что если вершины j и t имеют степени $k_j + 1$ и $k_t + 1$ соответственно, то число способов образования ребра (jt) равно $k_j k_t$. Поэтому в [7] доказано, что для рассматриваемых конфигурационных графов кластерный коэффициент (2) имеет вид:

$$C_G = \frac{\left(\sum_{k=1}^{k_{\max}-1} k q_k\right)^2}{Nm}. \quad (6)$$

В статье [4] утверждается, что медленно меняющаяся функция $h(k)$ в распределении (1) не влияет на асимптотические свойства графа при $N \rightarrow \infty$, и поэтому предложено использовать простейший случай $h(k) = 1$. Тогда из (1) и (3) следует, что $k_{\max} = \infty$ и

$$p_k = \frac{1}{k^\tau} - \frac{1}{(k + 1)^\tau}, \quad k = 1, 2, \dots, \quad \tau > 0. \quad (7)$$

Отсюда следует, что при $k \rightarrow \infty$

$$p_k \sim \frac{\tau}{k^{\tau+1}}. \quad (8)$$

Известно [1], что для большинства реальных сетей $\tau \in (1, 2)$. Однако в некоторых случаях, например при моделировании лесных пожаров [8], оказалось, что $\tau > 2$.

В [5] рассматривалось обобщение этой модели, в котором предполагалось, что распределение степеней вершин конфигурационного графа неизвестно и имеется только следующее ограничение на поведение p_k при $k \rightarrow \infty$:

$$p_k \sim \frac{d}{k^g (\ln k)^h}, \quad (9)$$

где $d > 0$, $g > 1$, $h \geq 0$. Понятно, что распределение (7), в силу (8), удовлетворяет условию (9) при $d = \tau$, $g = \tau + 1$ и $h = 0$. В [5] приведены также и другие примеры распределений вида (7), приводящие к (9). К числу таких примеров относятся и модели с распределением (1) степеней вершин, в котором параметр τ является случайной величиной. Заметим еще, что если $g > 3$ или $g = 3$, $h > 1$, то из (9) следует, что распределение ξ имеет конечную дисперсию.

В следующем разделе будет доказана теорема о предельном поведении кластерного коэффициента (6) в случайном конфигурационном графе с неизвестным распределением степеней вершин, имеющим конечную дисперсию либо удовлетворяющим условию (9).

3 Предельная теорема

Обозначим $\xi_{(N)}$ максимальную степень вершины в графе:

$$\xi_{(N)} = \max_{1 \leq s \leq N} \xi_i.$$

Целью данной статьи является получение следующего результата.

Теорема. Пусть $N \rightarrow \infty$. Тогда с вероятностью, сколь угодно близкой к единице, справедливы следующие утверждения.

1. Если дисперсия ξ конечна, то

$$C_G = O\left(\frac{1}{N}\right).$$

2. Если $g = 3$, $h = 1$, то

$$C_G \sim \frac{(d \ln \ln N)^2}{m^3 N}.$$

3. Если $g = 3$, $h < 1$, то

$$C_G \sim \left(\frac{d}{1-h} \left(\frac{\ln N}{2}\right)^{1-h}\right)^2 \frac{1}{m^3 N}.$$

4. Если $1 < g < 3$, то существует постоянная B такая, что $N^{1/(g-1)}/B \leq \xi_{(N)} \leq BN^{1/(g-1)}$ и при условии $\xi_{(N)} = uN^{1/(g-1)}$, $0 < u < \infty$, имеет место соотношение:

$$C_G \sim \left(\frac{du^{3-g}}{3-g}\right)^2 \left(\frac{g-1}{\ln N}\right)^{2h} \frac{N^{(\tau-3g)/(g-1)}}{m^3}.$$

Замечание. Нетрудно видеть, что при выполнении условия 4 теоремы, если $g > 7/3$ или $g = 7/3$, $h > 0$, то $C_G \rightarrow 0$, если $g = 7/3$, $h = 0$, то C_G стремится к зависящей от u константе, а если $g < 7/3$, то $C_G \rightarrow \infty$. Такое явление объясняется тем, что при малых значениях g пара смежных ребер (ij) и (it) , как упоминалось выше, может образовывать значительное число треугольников при наличии кратных ребер вида (jt) . Это свойство было описано в [7], где рассматривались модели, в которых $p_k = 1/(k^\tau \zeta(\tau))$, $k = 1, 2, \dots$, $\tau > 1$, а $\zeta(\tau)$ — значение дзета-функции Римана в точке τ .

Доказательство теоремы. Пусть распределение ξ имеет конечную дисперсию. Тогда из (5) следует, что ряд $\sum_{k=1}^{\infty} kq_k$ сходится и из (6) непосредственно вытекает первое утверждение теоремы.

Нетрудно видеть, что

$$P\{\xi_{(N)} < x\} = (1 - P\{\xi \geq x\})^N. \quad (10)$$

Пусть $g = 3$, $h \leq 1$. Тогда из (9) находим, что при $x \rightarrow \infty$

$$P\{\xi \geq x\} = d(1 + o(1)) \sum_{k \geq x} \frac{1}{k^3 (\ln k)^h} = d(1 + o(1)) \int_x^{\infty} \frac{dy}{y^3 (\ln y)^h}. \quad (11)$$

Из (11) следует, что для любого сколь угодно малого $\delta > 0$

$$d \int_x^{\infty} \frac{dy}{y^{3+\delta}} < P\{\xi \geq x\} < d \int_x^{\infty} \frac{dy}{y^3}. \quad (12)$$

Положим $x = \sqrt{N/z}$, где $0 < z < \infty$. Тогда для любого фиксированного $z > 0$ из (12) находим, что

$$P\left\{\xi \geq \sqrt{\frac{N}{z}}\right\} < \frac{dz}{2N}. \quad (13)$$

Если же $x = (zN)^{1/(2+\delta)}$, то

$$P \left\{ \xi \geq (zN)^{1/(2+\delta)} \right\} > \frac{d}{(2+\delta)zN}. \quad (14)$$

Из (10), (13) и (14) следует, что вероятность $P \left\{ (\varepsilon N)^{1/(2+\delta)} < \xi_{(N)} < \sqrt{N/\varepsilon} \right\}$ может быть сделана сколь угодно близкой к единице выбором достаточно малых положительных ε и δ . Поэтому достаточно рассмотреть множество графов, в которых $\xi_{(N)} = u\sqrt{N}$, где $(\varepsilon/N^{\delta/2})^{1/(2+\delta)} < u < 1/\sqrt{\varepsilon}$. Тогда из (4), (5) и (9) получаем, что при выполнении условия 2 теоремы и при достаточно больших натуральных D

$$\sum_{k=1}^{u\sqrt{N}} kq_k = C + \frac{d(1+o(1))}{m} \sum_{k=D}^{u\sqrt{N}-1} \frac{1}{k \ln k}, \quad (15)$$

где C — некоторая положительная постоянная. Заменяя суммирование интегрированием, нетрудно получить, что

$$\sum_{k=1}^{u\sqrt{N}-1} kq_k \sim \frac{d}{m} \int_D^{u\sqrt{N}} \frac{dy}{y \ln y} \sim \frac{d \ln \ln N}{m}.$$

Отсюда и из (6) следует второе утверждение теоремы.

Если $g = 3$ и $h < 1$, то, рассуждая аналогично, приходим к выводу, что

$$\sum_{k=1}^{u\sqrt{N}-1} kq_k = \frac{d}{m(1-h)} \left(\frac{\ln N}{2} \right)^{1-h} (1+o(1)).$$

Третье утверждение теоремы вытекает отсюда и из (6).

Пусть $1 < g < 3$. Аналогично предыдущему случаю нетрудно показать, с помощью (10), что вероятность $P \left\{ (\varepsilon N)^{1/(g+\delta-1)} < \xi_{(N)} < (N/\varepsilon)^{1/(g-1)} \right\}$ можно сделать сколь угодно близкой к единице, выбрав достаточно малые $\varepsilon, \delta > 0$. Если $\xi_{(N)} = uN^{1/(g-1)}$, $(\varepsilon/N^{\delta/(g-1)})^{1/(g+\delta-1)} < u < 1/\varepsilon^{1/(g-1)}$, то по аналогии с (15) видим, что

$$\begin{aligned} \sum_{k=1}^{uN^{1/(g-1)}} kq_k &= \\ &= C + \frac{d(1+o(1))}{m} \sum_D^{uN^{1/(g-1)}-1} \frac{1}{k^{g-2}(\ln k)^h}. \end{aligned} \quad (16)$$

Заметим, что при $N \rightarrow \infty$

$$\begin{aligned} \sum_D^{uN^{1/(g-1)}-1} \frac{1}{k^{g-2}(\ln k)^h} &= \\ &= (1+o(1)) \int_D^{uN^{1/(g-1)}} \frac{dy}{y^{g-2}(\ln y)^h}. \end{aligned} \quad (17)$$

Рассматривая последний интеграл как функцию верхнего предела и используя правило Лопиталья, получаем, что при $x \rightarrow \infty$

$$\int_D^x \frac{dy}{y^{g-2}(\ln y)^h} \sim \frac{x^{3-g}}{(3-g)(\ln x)^h}.$$

Отсюда и из (16), (17) следует, что

$$\begin{aligned} \sum_{k=1}^{uN^{1/(g-1)}-1} kq_k &= \\ &= \frac{du^{3-g}N^{(3-g)/(g-1)}}{m(3-g)} \left(\frac{g-1}{\ln N} \right)^h (1+o(1)), \end{aligned}$$

и из (6) получаем последнее утверждение теоремы.

Литература

1. Hofstad R. Random graphs and complex networks. — Cambridge: Cambridge University Press, 2017. Vol. 1. 337 p.
2. Barabasi A.-L., Albert R. Emergence of scaling in random networks // Science, 1999. Vol. 286. P. 509–512.
3. Bollobas B. A. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // Eur. J. Combin., 1980. Vol. 1. P. 311–316.
4. Reittu H., Norros I. On the power-law random graph model of massive data networks // Perform. Evaluation, 2004. Vol. 55. P. 3–23.
5. Павлов Ю.Л. Условные конфигурационные графы со случайным параметром степенного распределения степеней // Математический сборник, 2018. Т. 209. Вып. 2. С. 120–137.
6. Павлов Ю.Л., Челюкова И. А. Об асимптотике степенной структуры конфигурационных графов с ограничениями на число ребер // Дискретная математика, 2018. Т. 30. Вып. 1. С. 77–94.
7. Newman M. E. J. The structure and function of complex networks // SIAM Review, 2003. Vol. 45. Iss. 3. P. 3–23.
8. Leri M., Pavlov Y. Forest fire models on configuration random graphs // Fund. Inform., 2016. Vol. 145. Iss. 3. P. 313–322.

Поступила в редакцию 09.01.19

ON THE ASYMPTOTICS OF CLUSTERING COEFFICIENT IN A CONFIGURATION GRAPH WITH UNKNOWN DISTRIBUTION OF VERTEX DEGREES

Yu. L. Pavlov

Institute of Applied Mathematical Research, Karelian Research Centre of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Karelia, Russian Federation

Abstract: The author considers configuration graphs with vertex degrees being independent identically distributed random variables. The degree of each vertex equals to the number of incident half-edges that are numbered in an arbitrary order. The graph is constructed by joining each half-edge to another equiprobably to form edges. Configuration graphs are widely used for modeling of complex communication networks such as the Internet, social, transport, telephone networks. The distribution of vertex degrees can be unknown. It is only assumed that this distribution either has a finite variance or that some sufficient weak constraints on the asymptotic behavior of the tail are satisfied. The notion of clustering coefficient and its properties in such graphs are discussed. The author proves the limit theorem for the clustering coefficient with the number of vertices tending to infinity. The conditions under which this coefficient increases indefinitely are found.

Keywords: random graphs; configuration graphs; clustering coefficient; limit theorems

DOI: 10.14357/19922264190302

Acknowledgments

Financial support for research was carried out from the federal budget for the implementation of the state assignment of the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research, KarRC RAS).

References

1. Hofstad, R. 2017. *Random graphs and complex networks*. Cambridge: Cambridge University Press. Vol. 1. 337 p.
2. Barabasi, A.-L., and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.
3. Bollobas, B. A. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* 1:311–316.
4. Reittu, H., and I. Norros. 2004. On the power-law random graph model of massive data networks. *Perform. Evaluation* 55:3–23.
5. Pavlov, Yu. L. 2018. Conditional configuration graphs with discrete power-law distribution of vertex degrees. *Sb. Math.* 209(2):258–275.
6. Pavlov, Yu. L., and I. A. Cheplyukova. 2018. On asymptotics of degree structure of configuration graphs with restrictions on number of edges. *Discrete Math.* 28(2):58–79.
7. Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Rev.* 45(3):3–23.
8. Leri, M., and Y. Pavlov. 2016. Forest fire models on configuration random graphs. *Fund. Inform.* 145(3):313–322.

Received January 9, 2019

Contributor

Pavlov Yuri L. (b. 1949) — Doctor of Science in physics and mathematics, principal scientist, Institute of Applied Mathematical Research, Karelian Research Centre of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Karelia, Russian Federation; pavlov@krc.karelia.ru

ОБ ОЦЕНКАХ СКОРОСТИ СХОДИМОСТИ ДЛЯ НЕКОТОРЫХ МОДЕЛЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕПОЛНО ЗАДАНЫМИ ИНТЕНСИВНОСТЯМИ*

А. И. Зейфман¹, Я. А. Сатин², К. М. Киселева³

Аннотация: Рассматриваются некоторые модели массового обслуживания с неполно заданными интенсивностями. Авторы изучают систему $M_t/M_t/S$ с произвольным числом серверов S и систему $M_t/M_t/S/S$ (модель Эрланга) с интенсивностями, удовлетворяющими соответствующим условиям. Для получения оценок скорости сходимости используется понятие логарифмической нормы операторной функции и связанные с ней оценки нормы матрицы Коши.

Ключевые слова: система массового обслуживания; неполно заданные интенсивности; скорость сходимости; эргодичность; логарифмическая норма; $M_t/M_t/S$; $M_t/M_t/S/S$

DOI: 10.14357/19922264190303

1 Введение

В работе изучаются простые нестационарные системы обслуживания $M_t/M_t/S/S$ и $M_t/M_t/S$ в ситуации, когда интенсивности поступления и обслуживания требований являются 1-периодическими функциями времени и известны не сами эти функции, а их «средние» значения, т. е. числа

$$\lambda^* = \int_0^1 \lambda(t) dt; \quad \mu^* = \int_0^1 \mu(t) dt.$$

Такого типа модели (в которых известны не сами интенсивности поступления и обслуживания требований, а какие-либо их характеристики) достаточно часто встречаются в современной литературе (см., например, [1–6]).

Оценки скорости сходимости для рассматриваемых моделей в стационарной и нестационарной ситуациях изучались во многих работах, (см., например, [7–17] и приведенную в этих работах библиографию).

Сами эти модели достаточно стандартны, они характеризуются числом серверов S , а также интенсивностями поступления и обслуживания требований $\lambda(t)$ и $\mu(t)$ соответственно.

Число требований $X(t)$ описывается неоднородным процессом рождения и гибели (ПРГ):

– для системы $M_t/M_t/S/S$ с конечным пространством состояний $0, 1, \dots, S$ и интенсивностями $\lambda_k(t) = \lambda(t)$ и $\mu_k(t) = k\mu(t)$;

– для системы $M_t/M_t/S$ со счетным пространством состояний $0, 1, \dots, S, \dots$ и интенсивностями $\lambda_k(t) = \lambda(t)$ и $\mu_k(t) = \mu(t) \min(k, S)$.

Функции $\lambda(t)$ и $\mu(t)$ предполагаются неотрицательными, 1-периодическими и интегрируемыми на $[0, 1]$.

Обозначим через

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

$$i, j \geq 0 \quad 0 \leq s \leq t,$$

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

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

Будем обозначать через $\|\bullet\|$ l_1 -норму вектора и матрицы, т. е. $\|\mathbf{x}\| = \sum |x_i|$, а $\|B\| = \max_j \sum_i |b_{ij}|$ при $B = (b_{ij})_{i,j=0}^S$, а через Ω — множество всех стохастических векторов, т. е. множество векторов с неотрицательными координатами и единичной l_1 -нормой.

Через $E(t, k) = E \{X(t) | X(0) = k\}$ будем далее обозначать математическое ожидание процесса (среднее число требований) в момент t при условии, что в нулевой момент времени он находится в состоянии k .

* Исследование выполнено за счет гранта Российского научного фонда (проект 19-11-00020).

¹ Вологодский государственный университет; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Вологодский научный центр Российской академии наук, a_zeifman@mail.ru

² Вологодский государственный университет, yacovi@mail.ru

³ Вологодский государственный университет, ksushakiseleva@mail.ru

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

2 Оценки для системы $M_t/M_t/S/S$

Прежде всего отметим, что слабая эргодичность $X(t)$ эквивалентна условию $\lambda^* + \mu^* > 0$, это вытекает, например, из следствия 1 [15].

Сформулируем получаемые оценки отдельно для каждого из двух случаев.

Теорема 1. Пусть $\mu^* > 0$. Тогда $X(t)$ имеет предельный 1-периодический режим

$$\pi(t) = (\pi_0(t), \pi_1(t), \dots, \pi_S(t))^T$$

и соответствующее предельное среднее

$$\phi(t) = \sum_{k=0}^S k \pi_k(t)$$

и при любом $t \geq 0$ выполняются следующие оценки скорости сходимости:

$$\|\mathbf{p}^*(t) - \pi(t)\| \leq 8S e^{-\mu^*(t-1)} \quad (1)$$

при любых начальных условиях;

$$|E(t, k) - \phi(t)| \leq 8S^2 e^{-\mu^*(t-1)} \quad (2)$$

при любом k . При $\lambda^* > 0$ вместо (1) и (2) справедлива оценка:

$$\|\mathbf{p}^*(t) - \pi(t)\| \leq 8S e^{-\lambda^*(t-1)/S}. \quad (3)$$

Для доказательства теоремы 1 рассмотрим положительные числа d_i и положим

$$d = \inf_{i \geq 1} d_i = 1; \quad W = \inf_{i \geq 1} \frac{d_i}{i};$$

$$g_i = \sum_{n=1}^i d_n; \quad G = \sum_{i=1}^S d_i.$$

Рассмотрим теперь выражения:

$$\alpha_k(t) = \lambda_k(t) + \mu_{k+1}(t) -$$

$$- \frac{d_{k+1}}{d_k} \lambda_{k+1}(t) - \frac{d_{k-1}}{d_k} \mu_k(t), \quad k \geq 0;$$

$$\beta_*(t) = \inf_{k \geq 0} \alpha_k(t). \quad (4)$$

Тогда, применяя соответствующее преобразование прямой системы Колмогорова, из следствия 1 [15] получаем такие оценки:

$$\|\mathbf{p}^*(t) - \pi(t)\| \leq \frac{8G}{d} e^{-\int_0^t \beta_*(\tau) d\tau};$$

$$|E(t, k) - \phi(t)| \leq \frac{8G}{W} e^{-\int_0^t \beta_*(\tau) d\tau}.$$

Положим вначале все $d_i = 1$, тогда все $\alpha_k(t) = \beta_* = \mu(t)$, и с учетом неравенства

$$e^{-\int_0^t \mu(\tau) d\tau} \leq e^{-\int_0^{[t]} \mu(\tau) d\tau} \leq e^{-\mu^*(t-1)}$$

получаем оценки (1) и (2).

Теперь, следуя ходу рассуждений теоремы 93 из [13], положим

$$d_1 = 1; \quad d_{k+1} = \frac{S-1}{S} d_k < 1.$$

Тогда получим все $\alpha_k(t) > \lambda(t)/S$, откуда $\beta_* \geq \lambda(t)/S$. При этом $G \leq S$, а $e^{-\int_0^t \lambda(\tau) d\tau} \leq e^{-\lambda^*(t-1)}$ и, значит, справедливо неравенство (3).

3 Оценки для системы $M_t/M_t/S$

Отметим, что для наличия слабой эргодичности достаточно, чтобы выполнялось условие $\lambda^* < S\mu^*$.

Рассмотрим треугольную матрицу D :

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

и пространства последовательностей l_{1D} :

$$l_{1D} = \{\mathbf{z} = (p_1, p_2, \dots)^T : \|\mathbf{z}\|_{1D} \equiv \|D\mathbf{z}\| < \infty\}$$

и l_{1E} :

$$l_{1E} = \{\mathbf{z} = (p_1, p_2, \dots)^T : \|\mathbf{z}\|_{1E} \equiv \sum k|p_k| < \infty\}.$$

Выберем число $\delta \in (1, S/(S-1))$. Тогда, как показано в [13, параграф 4.2], в (4) получается

$$\beta_*(t) \geq (S\mu(t) - \delta\lambda(t)) (1 - \delta^{-1}).$$

А теперь, применяя теорему 1 и следствие 1 из работы [17], получаем общие оценки:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1D} \leq$$

$$\leq 4e^{-\int_0^t \beta_*(\tau) d\tau} \sum_{i \geq 1} g_i |p_i^*(0) - p_i^{**}(0)|;$$

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1E} = |\phi(t) - E(t, k)| \leq$$

$$\leq \frac{4}{W} e^{-\int_0^t \beta_*(\tau) d\tau} \sum_{i \geq 1} g_i |p_i^*(0) - p_i^{**}(0)|$$

при любом $t \geq 0$ и любых начальных условиях $\mathbf{p}^*(0)$ и $\mathbf{p}^{**}(0)$.

Далее, имеем в 1-периодическом случае

$$\begin{aligned} e^{-\int_0^t (S\mu(\tau) - \delta\lambda(\tau))(1 - \delta^{-1}) d\tau} &\leq \\ \leq e^{-\int_0^t (S\mu(\tau) - \delta\lambda(\tau))(1 - \delta^{-1}) d\tau} e^{\int_0^t \delta\lambda(\tau)(1 - \delta^{-1}) d\tau} &\leq \\ \leq e^{\lambda^*(\delta - 1) - (S\mu^* - \delta\lambda^*)(1 - \delta^{-1})(t - 1)}. \end{aligned}$$

Теорема 2. Пусть $\lambda^* < S\mu^*$. Тогда $X(t)$ имеет предельный 1-периодический режим $\pi(t) = (\pi_0(t), \pi_1(t), \dots)^T$ и соответствующее предельное среднее $\phi(t) = \sum_{k \geq 0} k\pi_k(t)$ и выполняются следующие оценки скорости сходимости:

$$\begin{aligned} \|\mathbf{p}^*(t) - \pi(t)\|_{1D} &\leq 4e^{\lambda^*(\delta - 1) - (S\mu^* - \delta\lambda^*)(1 - \delta^{-1})(t - 1)} \times \\ &\times \sum_{i \geq 1} g_i |p_i^*(0) - \pi_i(0)|; \\ |\phi(t) - E(t, k)| &\leq \\ &\leq \frac{4g_k}{W} e^{\lambda^*(\delta - 1) - (S\mu^* - \delta\lambda^*)(1 - \delta^{-1})(t - 1)}. \end{aligned}$$

4 Примеры

Рассмотрим модели систем обслуживания $M_t/M_t/S/S$ и $M_t/M_t/S$ с интенсивностями $\mu = 1$, $\lambda(t) = \lambda(1 + M \sin 2\pi\omega t)$ и различными вариациями «амплитуды» M и «частоты» ω . Выбор $S = 100$ в первом примере обусловлен возможностью непосредственного вычисления предельных характеристик этой системы. Для модели $M_t/M_t/S$ построение предельного среднего проводится с помощью усеченных процессами с конечным пространством состояний в соответствии с методикой [17], здесь для простоты взято $S = 10$. На приведенных рис. 1–6 показано влияние амплитуды и частоты интенсивности поступления требований на предельные характеристики процесса, описывающего число требований в системе. Интересно отметить, что «двойное среднее», т. е. величина $E = \int_0^1 \phi(t) dt$, по крайней мере в пределах получаемой точности, оказывается не зависящей от этих характеристик.

Для построения всех интересующих величин задача Коши для прямой системы Колмогорова с начальным условием, соответствующим ситуации $X(0) = 0$, решается на $[0, t^*]$, а затем нужная величина с точностью 10^{-3} строится на отрезке $[t^*, t^* + 1]$. Во втором примере предварительно выбирается нужная размерность «усеченного» процесса N .

Пример 1. Система $M_t/M_t/S/S$, $S = 100$, $\lambda = 10$. Получаем $t^* = 15$. Рассмотрены случаи:

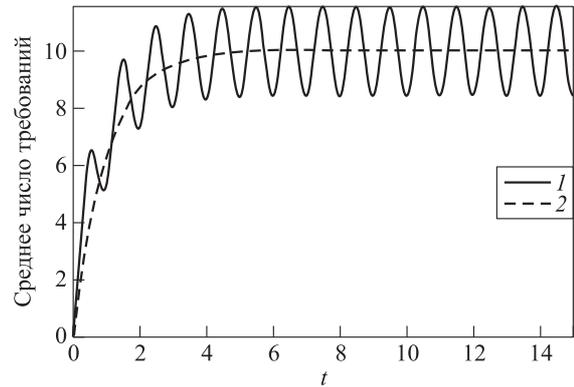


Рис. 1 Среднее число требований в системе $E(t, 0)$ на отрезке $[0, 15]$: 1 — пример I, случай 1; 2 — соответствующий однородный процесс

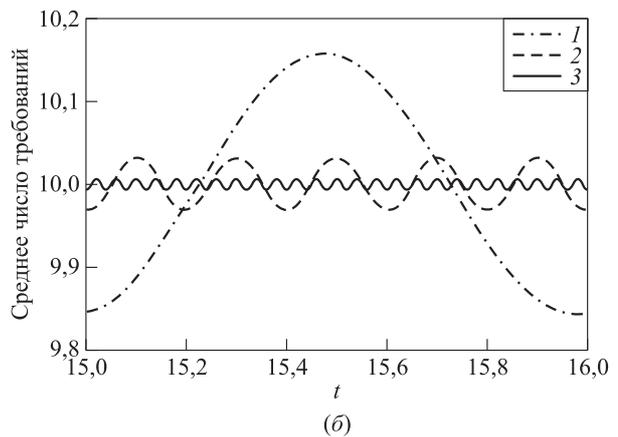
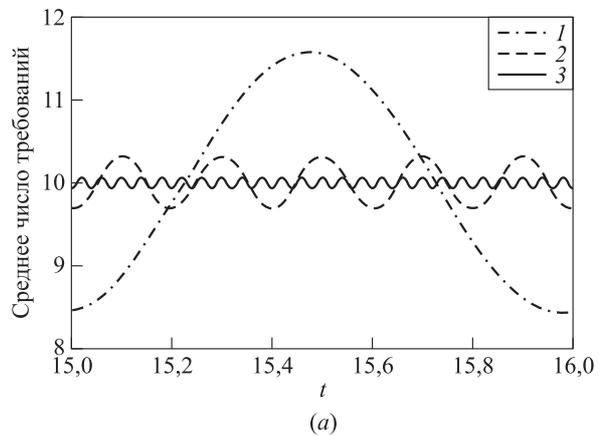


Рис. 2 Среднее число требований в системе $E(t, 0)$ на отрезке $[15, 16]$, пример I, амплитуды $M = 1$ (а) и $0,1$ (б): 1 — $w = 1$; 2 — 5; 3 — $w = 25$

- (1) $M = 1, \omega = 1$;
- (2) $M = 1, \omega = 5$;
- (3) $M = 1, \omega = 25$;

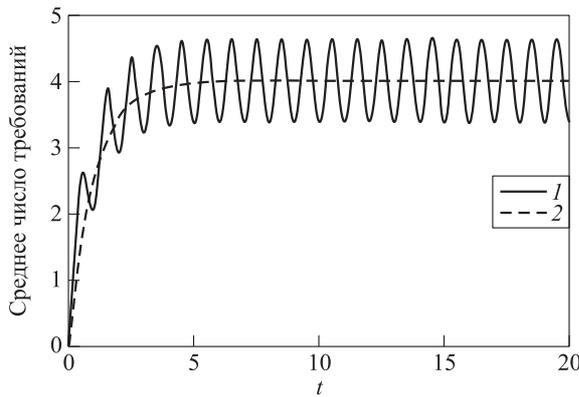


Рис. 3 Среднее число требований в системе $E(t, 0)$ на отрезке $[0, 20]$: 1 — пример II, случай 1; 2 — соответствующий однородный процесс

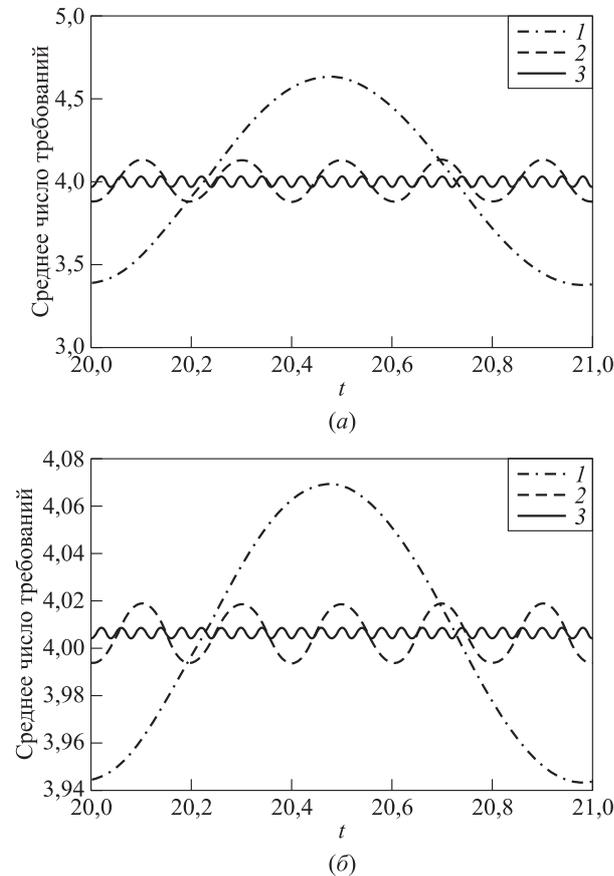


Рис. 4 Среднее число требований в системе $E(t, 0)$ на отрезке $[20, 21]$, пример II, амплитуды $M = 1$ (а) и $0,1$ (б): 1 — $\omega = 1$; 2 — $\omega = 5$; 3 — $\omega = 25$

- (4) $M = 0,1, \omega = 1$;
- (5) $M = 0,1, \omega = 5$;
- (6) $M = 0,1, \omega = 25$.

При этом $E \approx \int_0^1 \phi(t) dt = 10,000$.

Пример II. Система $M_t/M_t/S, S = 10, \lambda = 4$. Здесь получается $N = 100, t^* = 20$. Рассмотрены случаи:

- (1) $M = 1, \omega = 1$;
- (2) $M = 1, \omega = 5$;
- (3) $M = 1, \omega = 25$;
- (4) $M = 0,1, \omega = 1$;
- (5) $M = 0,1, \omega = 5$;
- (6) $M = 0,1, \omega = 25$.

При этом $E \approx \int_0^1 \phi(t) dt = 4,006$.

Литература

1. Guo Y., Wang Z. Stability of Markovian jump systems with generally uncertain transition rates // J. Frankl. Inst., 2013. Vol. 350. Iss. 9. P. 2826–2836.
2. Crawford F. W., Minin V. N., Suchard M. A. Estimation for general birth-death processes // J. Am. Stat. Assoc., 2014. Vol. 109. Iss. 506. P. 730–747.
3. Dong J., Whitt W. Stochastic grey-box modeling of queueing systems: Fitting birth-and-death processes to data // Queueing Syst., 2015. Vol. 79. P. 391–426.
4. Zhu D. M., Ching W. K., Guu S. M. Sufficient conditions for the ergodicity of fuzzy Markov chains // Fuzzy Set. Syst., 2016. Vol. 304. P. 82–93.
5. Cruz F. R. B., Quinino R. C., Ho L. L. Bayesian estimation of traffic intensity based on queue length in a multi-server $M/M/s$ queue // Commun. Stat. Simulat., 2017. Vol. 46. P. 7319–7331.
6. Ho L. S. T., Xu J., Crawford F. W., Minin V. N., Suchard M. A. Birth/birth–death processes and their computable transition probabilities with biological applications // J. Math. Biol., 2018. Vol. 76. P. 911–944.
7. Зейфман А. И. Некоторые свойства системы с потерями в случае переменных интенсивностей // Автоматика и телемеханика, 1989. Вып. 1. С. 107–113.
8. Kijima M. On the largest negative eigenvalue of the infinitesimal generator associated with $M/M/n/n$ queues // Oper. Res. Lett., 1990. Vol. 9. P. 59–64.
9. Zeifman A. I. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // Stoch. Proc. Appl., 1995. Vol. 59. P. 157–173.
10. Fricker C., Robert P., Tibi D. On the rate of convergence of Erlang’s model // J. Appl. Probab., 1999. Vol. 36. P. 1167–1184.
11. Voit M. A note of the rate of convergence to equilibrium for Erlang’s model in the subcritical case // J. Appl. Probab., 2000. Vol. 37. P. 918–923.
12. Zeifman A., Leorato S., Orsingher E., Satin Ya., Shilova G. Some universal limits for nonhomogeneous birth and death processes // Queueing Syst., 2006. Vol. 52. P. 139–151.
13. Зейфман А. И., Бенинг В. Е., Соколов И. А. Марковские цепи и модели с непрерывным временем. — М.: ЭЛЕКС-КМ, 2008. 168 с.

14. Van Doorn E. A., Zeifman A. I. On the speed of convergence to stationarity of the Erlang loss system // *Queueing Syst.*, 2009. Vol. 63. P. 241–252.
15. Зейфман А. И. О нестационарной модели Эрланга // *Автоматика и телемеханика*, 2009. Вып. 12. С. 71–80.
16. Van Doorn E. A., Zeifman A. I., Panfilova T. L. Bounds and asymptotics for the rate of convergence of birth–death processes // *Theor. Probab. Appl.*, 2010. Vol. 54. P. 97–113.
17. Zeifman A., Satin Ya., Korolev V., Shorgin S. On truncations for weakly ergodic inhomogeneous birth and death processes // *Int. J. Appl. Math. Comp.*, 2014. Vol. 24. P. 503–518.

Поступила в редакцию 06.06.19

ON THE BOUNDS OF THE RATE OF CONVERGENCE FOR SOME QUEUEING MODELS WITH INCOMPLETELY DEFINED INTENSITIES

A. I. Zeifman^{1,2,3}, Y. A. Satin¹, and K. M. Kiseleva¹

¹Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

³Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation

Abstract: The authors consider some queueing systems with incompletely defined 1-periodical intensities under corresponding conditions. The authors deal with $M_t/M_t/S$ queue for any number of servers S and $M_t/M_t/S/S$ (the Erlang model). Estimates of the rate of convergence in weakly ergodic situation are obtained by applying the method of the logarithmic norm of the operator of a linear function. The examples with exact given values of intensities and different variations of amplitude and frequency are considered, ergodicity conditions and estimates of the rate of convergence are obtained for each model, and plots of the effect of intensities’ amplitude and frequency of incoming requirements on the limiting characteristics of the process are constructed. The authors use the general algorithm to build graphs, it is associated with solving the Cauchy problem for the forward Kolmogorov system on the corresponding interval, which has already been used by the authors in previous papers.

Keywords: queueing systems; incompletely defined intensities; rate of convergence; ergodicity; logarithmic norm; $M_t/M_t/S$ queue; $M_t/M_t/S/S$ queue

DOI: 10.14357/19922264190303

Acknowledgments

This work was financially supported by the Russian Science Foundation (grant No 19-11-00020).

References

1. Guo, Y., and Z. Wang. 2013. Stability of Markovian jump systems with generally uncertain transition rates. *J. Frankl. Inst.* 350(9):2826–2836.
2. Crawford, F. W., V. N. Minin, and M. A. Suchard. 2014. Estimation for birth-death processes. *J. Am. Stat. Assoc.* 109(506):730–747.
3. Dong J., and W. Whitt. 2015. Stochastic grey-box modeling of queueing systems: Fitting birth-and-death processes to data. *Queueing Syst.* 79:391–426.
4. Zhu, D. M., W. K. Ching, and S. M. Guu. 2016. Sufficient conditions for the ergodicity of fuzzy Markov chains. *Fuzzy Set. Syst.* 304:82–93.
5. Cruz, F. R. B., R. C. Quinino, and L. L. Ho. 2017. Bayesian estimation of traffic intensity based on queue length in a multi-server $M/M/s$ queue. *Commun. Stat. Simulat.* 46:7319–7331.
6. Ho, L. S. T., J. Xu, F. W. Crawford, V. N. Minin, and M. A. Suchard. 2018. Birth/birth-death processes and their computable transition probabilities with biological applications. *J. Math. Biol.* 76:911–944.
7. Zeifman, A. I. 1989. Some properties of a system with losses in the case of variable rates. *Automat. Rem. Contr.* 50(1):82–87.
8. Kijima, M. 1990. On the largest negative eigenvalue of the infinitesimal generator associated with $M/M/n/n$ queues. *Oper. Res. Lett.* 9:59–64.

9. Zeifman, A. I. 1995. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stoch. Proc. Appl.* 59:157–173.
10. Fricker C., P. Robert, and D. Tibi. 1999. On the rate of convergence of Erlang’s model. *J. Appl. Probab.* 36:1167–1184.
11. Voit, M. 2000. A note of the rate of convergence to equilibrium for Erlang’s model in the subcritical case. *J. Appl. Probab.* 37:918–923.
12. Zeifman, A., S. Leorato, E. Orsingher, Ya. Satin, and G. Shilova. 2006. Some universal limits for nonhomogeneous birth and death processes. *Queueing Syst.* 52:139–151.
13. Zeifman, A. I., V. E. Bening, and I. A. Sokolov. 2008. *Markovskie tsepi i modeli s nepreryvnyim vremenem* [Markov chains and models with continuous time]. Moscow: ELEKS-KM Publ. 168 p.
14. Van Doorn, E. A., and A. I. Zeifman. 2009. On the speed of convergence to stationarity of the Erlang loss system. *Queueing Syst.* 63:241–252.
15. Zeifman, A. I. 2009. On the nonstationary Erlang loss model. *Automat. Rem. Contr.* 70(12):2003–2012.
16. Van Doorn, E. A., A. I. Zeifman, and T. L. Panfilova. 2010. Bounds and asymptotics for the rate of convergence of birth–death processes. *Theor. Probab. Appl.* 54:97–113.
17. Zeifman, A., Ya. Satin, V. Korolev, and S. Shorgin. 2014. On truncations for weakly ergodic inhomogeneous birth and death processes. *Int. J. Appl. Math. Comp.* 24:503–518.

Received June 6, 2019

Contributors

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of Department, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; principal scientist, Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation; a_zeifman@mail.ru

Satin Yacov A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University, 15 Lenin Str., Vologda 160000; yacovi@mail.ru

Kiseleva Ksenia M. (b. 1992) — Candidate of Science (PhD) in physics and mathematics, scientist, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; ksushakiseleva@mail.ru

НЕТРАНЗИТИВНЫЕ ТРИПЛЕТЫ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН И ИХ ПРИЛОЖЕНИЯ

А. В. Лебедев¹

Аннотация: Изучается явление нетранзитивности отношения стохастического предшествования для трех независимых случайных величин с распределениями из некоторых классов непрерывных распределений. Первоначально этот вопрос был поставлен в связи с задачей из теории прочности. При парных сравнениях железных брусков с трех заводов может сложиться парадоксальная ситуация, когда бруски с первого завода «хуже» брусков со второго завода, бруски со второго «хуже» брусков с третьего, а бруски с третьего «хуже» брусков с первого. В дальнейшем тема нетранзитивности стала популярной на примере так называемых нетранзитивных (игральных) костей, однако это привело к ее сужению на дискретные случайные величины с конечным множеством значений. В работе показано, что для смесей нормальных и показательных распределений нетранзитивность возможна в широком диапазоне параметров. Указаны характерные особенности взаимного расположения графиков функций распределения в этих случаях.

Ключевые слова: нетранзитивность; нетранзитивные кости; стохастическое предшествование; непрерывные распределения; смеси распределений

DOI: 10.14357/19922264190304

1 Введение

В теории и практике различные отношения превосходства между объектами часто обладают свойством транзитивности: если A превосходит B и B превосходит C , то A превосходит C . Это имеет место, например, если каждый объект можно характеризовать действительным числом и сравнение объектов сводится к сравнению этих чисел.

Однако бывает и по-другому. Например, в классической игре «камень, ножницы, бумага» «камень» побеждает «ножницы», «ножницы» побеждают «бумагу», но при этом «бумага» побеждает «камень», а не наоборот. Таким образом, отношение между стратегиями в этой игре нетранзитивно и не существует оптимальной чистой стратегии.

Различным аспектам и многочисленным примерам нетранзитивности отношений превосходства в природе, технике и обществе посвящен ряд научных и научно-популярных статей А. Н. Поддякова, в частности [1–4]. Отмечается, что нетранзитивность часто имеет контринтуитивный характер и может затруднять принятие правильных решений.

Оказывается, даже в шахматах существуют нетранзитивные наборы позиций. Это значит, что каждую позицию нельзя описать каким-то одним числом (функцией от позиции), по которому она лучше или хуже других, и, как следствие, основанные на этой идее шахматные программы имеют ограниченные возможности.

Далее будем изучать явление нетранзитивности в вероятностно-статистическом аспекте. Рассмотрим три независимые случайные величины (триплет) X , Y и Z , такие что

$$\mathbf{P}(X < Y) > \frac{1}{2} \quad (1)$$

и

$$\mathbf{P}(Y < Z) > \frac{1}{2}. \quad (2)$$

Можно было бы ожидать, что из (1) и (2) следует

$$\mathbf{P}(X < Z) > \frac{1}{2},$$

однако существуют примеры, когда это не так, а напротив,

$$\mathbf{P}(Z < X) > \frac{1}{2}.$$

Таким образом, отношение (1) между случайными величинами нетранзитивно и может идти по кругу, как в игре «камень, ножницы, бумага».

Отношение (1) называют *стохастическим предшествованием* (stochastic precedence). В работах [5,6] оно применялось в задачах статистического анализа. В работе [7] оно использовалось в задачах ранжирования и называлось доминированием по вероятности. Отметим также недавнюю работу [8], где

¹Московский государственный университет им. М. В. Ломоносова, механико-математический факультет, кафедра теории вероятностей, avlebed@yandex.ru

рассматривались различные стохастические порядки, в том числе стохастическое предшествование.

Далее для определенности под нетранзитивностью будем понимать ситуацию, когда

$$p_{XYZ} = \min \{P(X < Y), P(Y < Z), P(Z < X)\} > \frac{1}{2}.$$

Первой попыткой изучения этого явления стало исследование С. Трыбулы, начатое совместно с Г. Штейнгаузом [9, 10]. Было показано, что

$$\max_{X,Y,Z} p_{XYZ} = \frac{\sqrt{5} - 1}{2} \approx 0,618$$

и максимум достигается, например, на триплете

$$\left. \begin{aligned} X &= \begin{cases} 1 \text{ с вероятностью } p; \\ 4 \text{ с вероятностью } 1 - p; \end{cases} \\ Y &= 2; \\ Z &= \begin{cases} 0 \text{ с вероятностью } 1 - p; \\ 3 \text{ с вероятностью } p, \end{cases} \end{aligned} \right\} \quad (3)$$

где

$$p = \frac{\sqrt{5} - 1}{2},$$

тогда

$$P(X < Y) = P(Y < Z) = P(Z < X) = \frac{\sqrt{5} - 1}{2}.$$

Далее в работах З. Усыскина [11] и С. Трыбулы [12] был рассмотрен случай n независимых случайных величин, получены некоторые оценки максимума вероятностей, однако явное выражение было лишь недавно выведено И. И. Богдановым [13]:

$$\begin{aligned} \max_{X_1, \dots, X_n} \min \{ &P(X_1 < X_2), \dots, \\ &\dots, P(X_{n-1} < X_n), P(X_n < X_1) \} = \\ &= 1 - \left(4 \cos^2 \frac{\pi}{n+2} \right)^{-1}, \quad n \geq 3. \end{aligned}$$

В качестве приложения в [10] речь шла о прочности материалов. Пусть в лаборатории сравнивают попарно на прочность железные бруски одинакового размера и формы, помещая их в одну рамку (тиски) и прилагая к ним одинаковую силу путем закручивания винта, пока один из брусков не сломается. Предположим, что бруски производятся на трех разных заводах (которые дают разное распределение прочности) и сравниваются бруски с первого и второго, второго и третьего, первого и третьего заводов. Тогда теоретически может сложиться

парадоксальная ситуация, что бруски с первого завода «хуже» (т. е. чаще ломаются раньше) брусков со второго завода, бруски со второго «хуже» брусков с третьего, а бруски с третьего «хуже» брусков с первого.

Понятно, что конкретно в приведенном примере можно поступить иначе: не сравнивать прочности брусков друг с другом, а измерять их явно, затем выбирать лучший завод по какой-то важной на практике числовой характеристике распределения (по которой они достаточно заметно различаются). Однако в приложениях бывают и ситуации, когда какие-то величины просто не измеряются в явном виде, и нет другого способа выявить соотношения объектов или их совокупностей, кроме парных сравнений.

Например, в биологии речь может идти о парных взаимодействиях животных в борьбе за пищу, территорию, размножение или доминирование в группе. Аналогичным образом при сравнительном анализе работы алгоритмов на различных наборах данных или соревнованиях компьютерных программ могут быть победители в парах, но не быть лучшего в группе из трех или более объектов. Для людей речь может идти как о соревнованиях, так и о сравнительных оценках различных товаров и услуг, голосовании и др. Проблема нетранзитивности предпочтений здесь была известна с XVIII в. (*парадокс Кондорсе*).

Разумеется, человеческий интеллект может найти какой-то выход из подобной ситуации, однако в случае искусственного интеллекта (компьютерных программ) проявления нетранзитивности могут привести машину в тупик, если не будут предусмотрены программистом, что необходимо учитывать в разработках.

Тема нетранзитивности стохастического предшествования приобрела популярность на примере так называемых *нетранзитивных костей* (nontransitive, intransitive dice). Имеются в виду наборы игральных костей, на грани которых нанесены числа таким образом, чтобы создать нетранзитивные отношения соответствующих случайных величин.

Нетранзитивные кости были популяризованы М. Гарднером [14, 15], им посвящена обширная литература, например работы [16–19]. К сожалению, при этом произошло, во-первых, сужение темы до дискретных случайных величин с конечным числом значений, во-вторых, сложилось не слишком серьезное отношение со стороны ученых, зачастую относящих эту тему к игровой, развлекательной математике.

Однако изначально С. Трыбулой была затронута вполне серьезная теоретико-вероятностная проблема с возможными приложениями на практике.

Прочность материалов, конечно, описывается непрерывным распределением, как и многие другие характеристики объектов в природе, технике и обществе. И при попарных сравнениях объектов из нескольких разных совокупностей (категорий) теоретически может возникать нетранзитивность. Таким образом, вполне закономерно задаться вопросом, в каких классах непрерывных распределений она бывает, а в каких нет.

В [10, теорема 2] было показано, что если распределения случайных величин X, Y и Z относятся к одному сдвиговому семейству, то нетранзитивности быть не может. Понятно, что этот вывод также распространяется на масштабные семейства распределений на положительной или отрицательной полупрямой (путем логарифмирования).

В частности, прочность материалов часто описывается распределением Вейбулла

$$F(x) = 1 - \exp \left\{ - \left(\frac{x-a}{b} \right)^\alpha \right\}, \quad x \geq a, \quad a, b, \alpha > 0,$$

предложенным еще в книге [20]. Таким образом, сразу можно сделать вывод, что если случайные величины (прочности железных брусков с разных заводов) имеют распределение Вейбулла с постоянными a и α и разными b или с постоянными b и α и разными a , то нетранзитивность невозможна.

В работе автора [21] выведены некоторые новые критерии, когда нетранзитивности быть не может. Показано, что в случае распределения Вейбулла с постоянными a и b и разными α нетранзитивность также невозможна. Рассмотрен класс распределений с полиномиальной плотностью на единичном отрезке, где возможность нетранзитивности зависит от степени многочлена.

Далее будет приведена одна общая схема построения нетранзитивных триплетов и содержательные примеры в классах конечных смесей нормальных и показательных распределений.

Конечные смеси распределений возникают в различных приложениях, проблемам их статистического анализа посвящена обширная литература. Из недавних работ отметим [22–27], где смеси нормальных распределений рассматривались в [23–27], гамма-распределений — в [22, 24, 27], показательных распределений — в [22].

2 Когда нетранзитивность возможна

Проведем обобщение базовой схемы (3), заменив 0, 1, 2, 3 и 4 на независимые непрерывные случайные величины $\xi_0, \xi_1, \xi_2, \xi_3$ и ξ_4 , такие что для

вероятностей $p_{kl} = \mathbf{P}(\xi_k < \xi_l)$ верно $p_{kl} > 1/2$ при $k < l$ и эти вероятности убывают по k и возрастают по l .

Положим

$$\begin{aligned} X &= \begin{cases} \xi_1 & \text{с вероятностью } r, \\ \xi_4 & \text{с вероятностью } 1-r, \end{cases} \quad 0 \leq r \leq 1; \\ Y &= \xi_2; \\ Z &= \begin{cases} \xi_0 & \text{с вероятностью } 1-s, \\ \xi_3 & \text{с вероятностью } s, \end{cases} \quad 0 \leq s \leq 1. \end{aligned}$$

Предположим, что p_{kl} фиксированы, а r и s можно варьировать. Тогда

$$\begin{aligned} p_{XY}(r) &= \mathbf{P}(X < Y) = rp_{12} + (1-r)p_{42}; \\ p_{YZ}(s) &= \mathbf{P}(Y < Z) = (1-s)p_{20} + sp_{23}; \\ p_{ZX}(r, s) &= \mathbf{P}(Z < X) = r(1-s)p_{01} + \\ &\quad + (1-r)(1-s)p_{04} + rsp_{31} + (1-r)sp_{34}. \end{aligned}$$

Обозначим также

$$\begin{aligned} p_{XYZ}(r, s) &= \min \{p_{XY}(r), p_{YZ}(s), p_{ZX}(r, s)\}; \\ p_{XYZ}^{\max} &= \max_{0 \leq r, s \leq 1} p_{XYZ}(r, s). \end{aligned}$$

Максимизация $p_{XYZ}(r, s)$ в общем случае ведет к множеству вариантов и громоздким выражениям, поэтому далее рассмотрим случай, когда p_{kl} зависят только от $l-k$, обозначим $p_{l-k} = p_{kl}$. Тогда

$$\begin{aligned} p_{01} &= p_{12} = p_{23} = p_{34} = p_1; \\ p_{42} &= p_{20} = p_{31} = p_{-2} = 1 - p_2; \\ p_{04} &= p_4; \\ \frac{1}{2} &< p_1 < p_2 < p_4 \leq 1, \end{aligned}$$

в силу симметрии можно положить $r = s$ и формулы упрощаются до

$$\left. \begin{aligned} p_{XY}(r) &= rp_1 + (1-r)(1-p_2); \\ p_{ZX}(r) &= 2r(1-r)p_1 + (1-r)^2 p_4 + \\ &\quad + r^2(1-p_2); \\ p_{XYZ}(r) &= \min \{p_{XY}(r), p_{ZX}(r)\}. \end{aligned} \right\} \quad (4)$$

Замечание 1. $p_{XY}(r)$ линейно возрастает от $1-p_2 < 1/2$ до $p_1 > 1/2$, а $p_{ZX}(r)$ — квадратичная функция, принимающая значения между $p_4 > p_1 > 1/2$ ($r=0$) и $1-p_2 < 1/2$ ($r=1$), а значит, не имеющая локальных экстремумов на $(0, 1)$ и убывающая от $p_4 > 1/2$ до $1-p_2 < 1/2$.

Теорема 1. Нетранзитивность в модели (4) возможна тогда и только тогда, когда

$$p_{ZX} \left(\frac{p_2 - 1/2}{p_1 + p_2 - 1} \right) > \frac{1}{2}.$$

Доказательство. Функция $p_{XY}(r)$ возрастает и достигает значения $1/2$ в точке

$$r^* = \frac{p_2 - 1/2}{p_1 + p_2 - 1}.$$

Необходимым и достаточным условием наличия нетранзитивности является $p_{ZX}(r^*) > 1/2$. Действительно, тогда в некоторой правой окрестности r^* верно $p_{XY}(r) > 1/2$ и $p_{ZX}(r) > 1/2$. Если же $p_{ZX}(r^*) \leq 1/2$, то далее $p_{ZX}(r) < 1/2$, поскольку $p_{ZX}(r)$ убывает. \square

Теорема 2. Максимум $p_{XYZ}(r)$ достигается в точке $0 < r_{\max} < 1$, являющейся единственным на $(0, 1)$ корнем уравнения

$$Ar^2 + Br + C = 0,$$

где

$$\begin{aligned} A &= 2p_1 + p_2 - p_4 - 1; \\ B &= 2p_4 + p_2 - p_1 - 1; \\ C &= 1 - p_2 - p_4. \end{aligned}$$

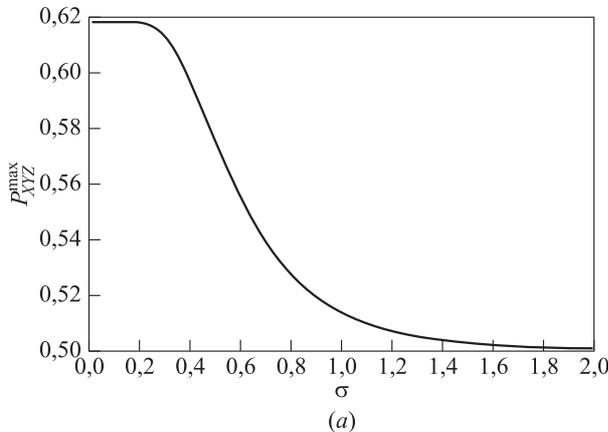
Доказательство. В силу замечания 1 максимум минимум функций $p_{XY}(r)$ и $p_{ZX}(r)$ достигается в единственной точке их равенства:

$$\begin{aligned} rp_1 + (1-r)(1-p_2) &= \\ &= 2r(1-r)p_1 + (1-r)^2p_4 + r^2(1-p_2), \end{aligned}$$

откуда и получаем уравнение. \square

Пример 1. Пусть ξ_k имеют нормальное распределение со средними k и дисперсией σ^2 , тогда

$$p_m = \Phi\left(\frac{m}{\sigma\sqrt{2}}\right), \quad m = 1, 2, 4.$$



На рис. 1, а представлен график p_{XYZ}^{\max} , построенный в соответствии с расчетами по теореме 2. Фактически, нетранзитивность наблюдается при любых $\sigma > 0$. При изменении σ от 0 до $+\infty$ вероятность r_{\max} пробегает от $(\sqrt{5} - 1)/2$ до $2/3$ и p_{XYZ}^{\max} от $(\sqrt{5} - 1)/2$ до $1/2$.

В частности, при $\sigma = 1$ получаем $r_{\max} \approx 0,638$, $p_{XYZ}^{\max} \approx 0,514$. На рис. 2, а для этого случая представлены графики функций распределения X , Y и Z .

Пример 2. Пусть ξ_k имеют показательное распределение со средними μ^k , $\mu > 1$, тогда

$$p_m = 1 - \frac{1}{1 + \mu^m}, \quad m = 1, 2, 4.$$

На рис. 1, б представлен график p_{XYZ}^{\max} . Фактически, нетранзитивность наблюдается при любых $\mu > 1$. При изменении μ от 1 до $+\infty$ вероятность r_{\max} пробегает от $2/3$ до $(\sqrt{5} - 1)/2$ и p_{XYZ}^{\max} от $1/2$ до $(\sqrt{5} - 1)/2$.

В частности, при $\mu = 3$ получаем $r_{\max} \approx 0,639$, $p_{XYZ}^{\max} \approx 0,515$. На рис. 2, б для этого случая представлены графики функций распределения X , Y и Z .

Замечание 2. Обозначим $d_m = p_m - 1/2$, $m = 1, 2, 4$. В обоих примерах имеют место соотношения $d_4 \sim \sim 4d_1$, $d_2 \sim 2d_1$ при $d_1 \rightarrow 0$ (когда $\sigma \rightarrow \infty$ в примере 1 или $\mu \rightarrow 1$ в примере 2). По теореме 2 отсюда следует $r_{\max} \rightarrow 2/3$ и $p_{XYZ}^{\max} \rightarrow 1/2$.

Замечание 3. В обоих примерах на рис. 2 видно, что значения функций распределения одинаковым образом меняют свое взаимное расположение (четыре раза), графики F_X и F_Z имеют две точки пересечения, а график F_Y пересекается с каждым из них по одному разу. Аналогичное явление было отмечено и в [21]. Можно предположить, что такая картина характерна при нетранзитивности.

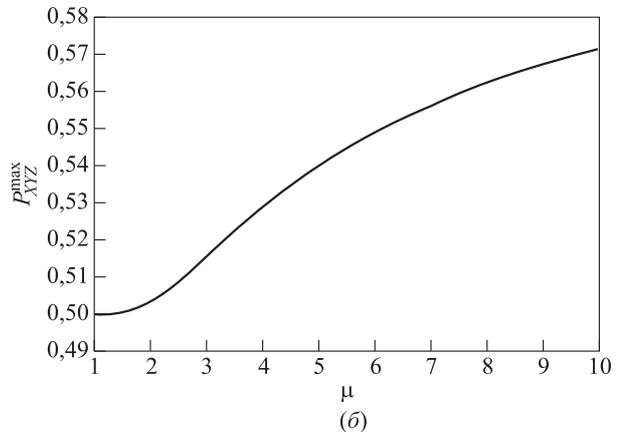


Рис. 1 График p_{XYZ}^{\max} к примерам 1 (а) и 2 (б)

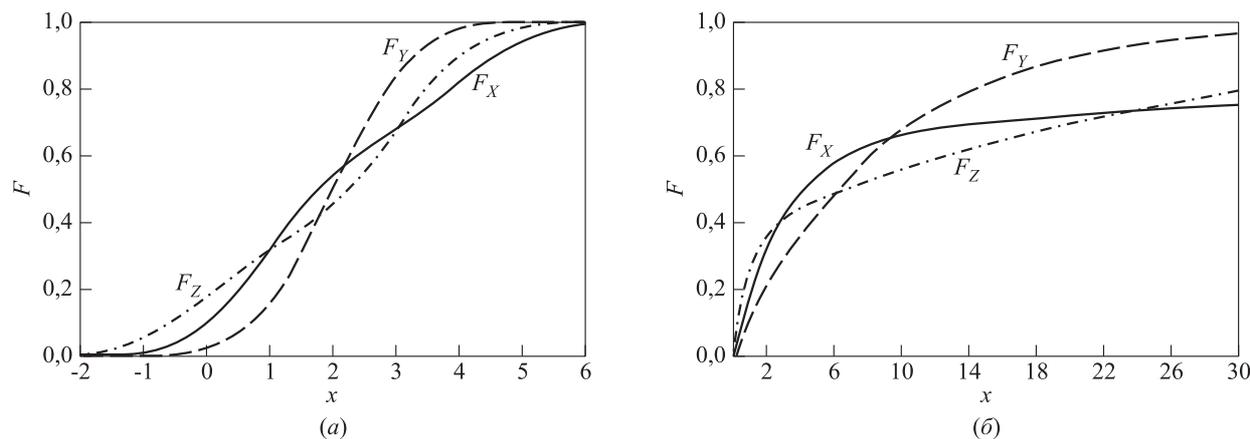


Рис. 2 Графики функций распределения к примерам 1 (а) и 2 (б)

3 Заключение

В работе показано, что для смесей нормальных и показательных распределений нетранзитивность возможна в широком диапазоне параметров. Учитывая инвариантность относительно любого непрерывного строго возрастающего преобразования (применяемого ко всем случайным величинам) этот вывод можно распространить на смеси логнормальных распределений, распределений Вейбулла, Парето и др.

Возвращаясь к примеру С. Трыбулы по прочности брусков, если допустить, что хотя бы на двух из трех заводов имеются не менее двух цехов или смен, выпускающих бруски с заметно разным распределением прочности (Вейбулла), нетранзитивность теоретически исключать нельзя.

Можно сделать вывод, что появлению нетранзитивности могут способствовать внутренние неоднородности совокупностей, из которых производятся выборки, что и проявляется статистически как смешивание классических распределений, имеющих различные параметры.

Литература

1. Поддьяков А. Н. Непереходность (нетранзитивность) отношений превосходства и принятие решений // Психология. Ж. ВШЭ, 2006. № 3. С. 88–111.
2. Пермгорский М. С., Поддьяков А. Н. Отношение превосходства между объектами и нетранзитивность их предпочтений человеком // Вопросы психологии, 2014. № 2. С. 3–14.
3. Поддьяков А. Н. Нетранзитивность — кладезь для избретателей // Троицкий вариант, 21.11.2017. № 242.
4. Poddiakov A. Intransitive machines // arXiv.org, 2018. ArXiv:1809.03869 [math.NO]. 11 p.

5. Arcones M. A., Kvam P. H., Samaniego F. J. Nonparametric estimation of a distribution subject to a stochastic precedence constraint // J. Am. Stat. Assoc., 2002. Vol. 97. No. 457. P. 170–182.
6. Boland P. J., Singh H., Cukic B. The stochastic precedence ordering with applications in sampling and testing // J. Appl. Probab., 2004. Vol. 41. No. 1. P. 73–82.
7. Шахнов И. Ф. Задачи ранжирования интервальных величин при многокритериальном анализе сложных систем // Изв. РАН. ТиСУ, 2008. № 1. С. 37–44.
8. Лепский А. Е. Стохастическое и нечеткое упорядочивание методом минимальных преобразований // Автоматика и телемеханика, 2017. № 1. С. 59–79.
9. Steinhaus H., Trybula S. On a paradox in applied probabilities // В. Acad. Pol. Sci., 1959. Vol. 7. P. 67–69.
10. Trybula S. On the paradox of three random variables // Zastos. Matem., 1961. Vol. 5. No. 4. P. 321–332.
11. Usyskin Z. Max–min probabilities in the voting paradox // Ann. Math. Stat., 1964. Vol. 35. No. 2. P. 857–862.
12. Trybula S. On the paradox of n random variables // Zastos. Matem., 1965. Vol. 8. No. 2. P. 143–156.
13. Богданов И. И. Нетранзитивные рулетки // Матем. просв., 2010. Сер. 3. Вып. 14. С. 240–255.
14. Gardner M. The paradox of the nontransitive dice and the elusive principle of indifference // Sci. Am., 1970. Vol. 223. No. 6. P. 110–114.
15. Gardner M. On the paradoxical situations that arise from nontransitive relations // Sci. Am., 1974. Vol. 231. No. 6. P. 120–125.
16. Savage R. The paradox of nontransitive dice // Am. Math. Mon., 1994. Vol. 101. No. 5. P. 429–436.
17. Bozoki S. Nontransitive dice sets releasing the Paley tournament for solving Shütte’s tournament problem // Miskolc Math. Notes, 2014. Vol. 15. No. 1. P. 39–50.
18. Conrey B., Gabbard J., Grant K., Liu A., Morrison K. E. Intransitive dice // Math. Mag., 2016. Vol. 89. P. 133–143.
19. Buhler I., Graham R., Hales A. Maximally nontransitive dice // Am. Math. Mon., 2018. Vol. 125. No. 5. P. 387–399.

20. Weibull W. A statistical theory of the strength of materials. — Stockholm: Generalstabens litografiska anstalts förlag, 1939. 45 p.
21. Лебедев А. В. Проблема нетранзитивности для трех независимых случайных величин // Автоматика и телемеханика, 2019. № 6. С. 91–103.
22. Батракова Д. А., Королев В. Ю., Шоргин С. Я. Новый метод вероятностно-статистического анализа информационных потоков в телекоммуникационных сетях // Информатика и её применения, 2007. Т. 1. Вып. 1. С. 40–53.
23. Кривенко М. П. Расщепление смеси вероятностных распределений на две составляющие // Информатика и её применения, 2008. Т. 2. Вып. 4. С. 48–56.
24. Бенинг В. Е., Горшенин А. К., Королев В. Ю. Асимптотически оптимальный критерий проверки гипотез о числе компонент смеси вероятностных распределений // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 4–16.
25. Горшенин А. К. Об устойчивости сдвиговых смесей нормальных законов по отношению к изменениям смешивающего распределения // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 22–28.
26. Королев В. Ю., Горшенин А. К., Гулев С. К., Беляев К. П. Статистическое моделирование турбулентных потоков тепла между океаном и атмосферой с помощью метода скользящего разделения конечных нормальных смесей // Информатика и её применения, 2015. Т. 9. Вып. 4. С. 3–13.
27. Горшенин А. К. Зашумление данных конечными смесями нормальных и гамма-распределений с применением к задаче округления наблюдений // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 28–34.

Поступила в редакцию 03.10.18

NONTRANSITIVE TRIPLETS OF CONTINUOUS RANDOM VARIABLES AND THEIR APPLICATIONS

A. V. Lebedev

Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, Main Building, 1 Leninskiye Gory, Moscow 119991, Russian Federation

Abstract: The phenomenon of nontransitivity of the stochastic precedence relation for three independent random variables with distributions from some classes of continuous distributions is studied. Initially, this question was posed in connection with the application in strength theory. With paired comparisons of iron bars from three factories, a paradoxical situation may arise when the bars from the first factory are “worse” than the bars from the second factory, the bars from the second factory are “worse” than the bars from the third factory, and the bars from the third factory are “worse” than the bars from the first factory. Further, the nontransitivity topic gained popularity for the example of the so-called nontransitive dice; however, this led to its narrowing down to discrete random variables with finite sets of values. The paper presents that for mixtures of normal and exponential distributions, nontransitivity is possible in a wide range of parameters. Specific features of the mutual arrangement of the graphs of the distribution functions in these cases are indicated.

Keywords: nontransitivity; nontransitive dice; stochastic precedence; continuous distributions; mixtures of distributions

DOI: 10.14357/19922264190304

References

1. Poddiakov, A. N. 2006. Neperekhodnost' (netranzitivnost') otnosheniy prevoskhodstva i prinyatie resheniy [Intransitive character of superiority relations and decision-making]. *Psikhologiya. Zh. VShE* [Psychology. J. HSE] 3(3):88–111.
2. Permogorskiy, M. S., and A. N. Poddiakov. 2014. Otnoshenie prevoskhodstva mezhdub'ektami i netranzitivnost' ikh predpochteniy chelovekom [The relation of superiority between objects and the nontransitivity of their preferences by human]. *Voprosy psikhologii* [Psychology issues] 2:3–14.
3. Poddiakov, A. November 21, 2017. Netranzitivnost' — kladez' dlya izobretateley [Nontransitivity is a treasure for inventors]. *Troitskiy variant* [Troitsk Variant] 242.
4. Poddiakov, A. 2018. Intransitive machines. Available at: <https://arxiv.org/abs/1809.03869> (accessed September 8, 2018).
5. Arcones, M. A., P. H. Kvam, and F. J. Samaniego. 2002. Nonparametric estimation of a distribution subject to a stochastic precedence constraint. *J. Am. Stat. Assoc.* 97(457):170–182.
6. Boland, P. J., H. Singh H., and B. Cukic. 2004. The stochastic precedence ordering with applications in sampling and testing. *J. Appl. Probab.* 41(1):73–82.

7. Shakhnov, I. F. 2008. A problem of ranking interval objects in a multicriteria analysis of complex systems. *J. Comput. Sys. Sc. Int.* 47(1):33–39.
8. Lepskiy, A. E. 2017. Stochastic and fuzzy ordering with the method of minimal transformations. *Automat. Rem. Contr.* 78(1):50–66.
9. Steinhaus, H., and S. Trybula. 1959. On a paradox in applied probabilities. *B. Acad. Pol. Sci.* 7:67–69.
10. Trybula, S. 1961. On the paradox of three random variables. *Zastos. Matem.* 5(4):321–332.
11. Usyskin, Z. 1964. Max–min probabilities in the voting paradox. *Ann. Math. Stat.* 35(2):857–862.
12. Trybula, S. 1965. On the paradox of n random variables. *Zastos. Matem.* 8(2):143–156.
13. Bogdanov, I. I. 2010. Netranzitivnye ruletki [Nontransitive roulettes]. *Mat. Pros.* [Math. Enlight.] 14:240–255.
14. Gardner, M. 1970. The paradox of the nontransitive dice and the elusive principle of indifference. *Sci. Am.* 223(6):110–114.
15. Gardner, M. 1974. On the paradoxical situations that arise from nontransitive relations. *Sci. Am.* 231(6):120–125.
16. Savage, R. 1994. The paradox of nontransitive dice. *Am. Math. Mon.* 101(5):429–436.
17. Bozoki, S. 2014. Nontransitive dice sets releasing the Pauley tournament for solving Shütte’s tournament problem. *Miskolc Math. Notes* 15(1):39–50.
18. Conrey, B., J. Gabbard, K. Grant, A. Liu, and K. E. Morrison. 2016. Intransitive dice. *Math. Mag.* 89:133–143.
19. Buhler, I., R. Graham, and A. Hales. 2018. Maximally nontransitive dice. *Am. Math. Mon.* 125(5):387–399.
20. Weibull, W. 1939. *A statistical theory of the strength of materials*. Stockholm: Generalstabens litografiska anstalts förlag. 45 p.
21. Lebedev, A. V. 2019. Nontransitivity problem for three continuous random variables. *Automat. Rem. Contr.* 80(6): 1025–1036.
22. Batrakova, D. A., V. Yu. Korolev, and S. Ya. Shorgin. 2007. Novyy metod veroyatnostno-statisticheskogo analiza informatsionnykh potokov v telekommunikatsionnykh setyakh [A new method for the probabilistic and statistical analysis of information flows in telecommunication networks]. *Informatika i ee Primeneniya — Inform. Appl.* 1(1):40–53.
23. Krivenko, M. P. 2008. Rasshcheplenie smesi veroyatnostnykh raspredeleniy na dve sostavlyayushchie [Splitting of distribution mixture in two components] *Informatika i ee Primeneniya — Inform. Appl.* 2(4):48–56.
24. Bening, B. E., A. K. Gorshenin, and V. Yu. Korolev. 2011. Asimptoticheski optimal’nyy kriteriy proverki gipotez o chisle komponent smesi veroyatnostnykh raspredeleniy [An asymptotically optimal test for the number of components of a mixture of probability distributions]. *Informatika i ee Primeneniya — Inform. Appl.* 5(3):4–16.
25. Gorshenin, A. K. 2012. Ob ustoychivosti sdvigoverykh smesey normal’nykh zakonov po otnosheniyu k izmeneniyam smeshivayushchego raspredeleniya [On stability of normal location mixtures with respect to variations in mixing distribution]. *Informatika i ee Primeneniya — Inform. Appl.* 6(2):22–28.
26. Korolev, V. Yu., A. K. Gorshenin, S. K. Gulev, and K. P. Belyaev. 2015. Statisticheskoe modelirovanie turbulentnykh potokov tepla mezhdru okeanom i atmosferoy s pomoshch’yu metoda skol’zyashchego razdeleniya konechnykh normal’nykh smesey [Statistical modeling of air–sea turbulent heat fluxes by the method of moving separation of finite normal mixtures]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):3–13.
27. Gorshenin, A. K. 2018. Zashumlenie dannykh konechnymi smesyami normal’nykh i gamma-raspredeleniy s primeneniem k zadache okrugleniya nablyudeniy [Data noising by finite normal and gamma mixtures with application to the problem of rounded observations]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):28–34.

Received October 3, 2018

Contributor

Lebedev Alexey V. (b. 1971) — Doctor of Science in physics and mathematics, associate professor, Department of Probability Theory, Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, Main Building, 1 Leninskiye Gory, Moscow 119991, Russian Federation; avlebed@yandex.ru

АПРИОРНОЕ ОБОБЩЕННОЕ ГАММА-РАСПРЕДЕЛЕНИЕ В БАЙЕСОВСКИХ МОДЕЛЯХ БАЛАНСА*

А. А. Кудрявцев¹

Аннотация: Работа посвящена исследованию байесовских моделей баланса, предполагающих разбиение параметров системы на два класса: способствующих функционированию позитивных факторов и препятствующих функционированию негативных факторов. Рассматривается индекс баланса, определяемый как отношение негативного фактора к позитивному. Изучается постановка задачи, заключающаяся в нахождении основных вероятностных характеристик (плотность, функция распределения, моменты) индекса баланса факторов, имеющих априорные обобщенные гамма-распределения с параметрами формы одного знака. Результаты формулируются в терминах гамма-экспоненциальной функции. Приводится ряд новых свойств последней. Показано, что приводимые утверждения легко переформулируются для масштабных смесей обобщенных гамма-распределений, имеющих параметры формы разных знаков. Полученные результаты могут найти широкое применение в моделях, использующих для описания процессов и явлений распределения с положительным неограниченным носителем.

Ключевые слова: байесовский подход; обобщенное гамма-распределение; гамма-экспоненциальная функция; модели баланса; смешанные распределения

DOI: 10.14357/19922264190305

1 Введение

Байесовский подход к моделям баланса в терминах теории массового обслуживания был впервые предложен С. Я. Шоргиным в работе [1] и распространен на задачи повышения надежности сложных информационно-телекоммуникационных систем в [2]. Подробное описание байесовских моделей баланса и сферы их применения в разнообразных областях знания было дано в статье [3]. Остановимся на нем вкратце.

Всевозможные аспекты современной жизни усложнились настолько, что вынуждают прибегать к использованию разного рода индексов и рейтингов для определения эффективности функционирования исследуемой системы. Для построения таких показателей целесообразно условно разделять факторы, оказывающие влияние на систему, на способствующие функционированию (р-факторы) и препятствующие функционированию (п-факторы). Разделение параметров на негативные и позитивные факторы свойственно моделям из разных областей науки [3] от теории массового обслуживания до моделирования чрезвычайных ситуаций.

Вполне естественно, что функционирование исследуемой системы в итоге зависит не столько от значений п- и р-факторов, сколько от их соотношения — индекса баланса. При этом во избежа-

ние чрезмерных затрат на преодоление негативного влияния и с целью недопущения недооценки негативного воздействия следует добиваться сбалансированности системы, т. е. стремиться приблизить к единице индекс баланса.

С течением времени п- и р-факторы, а следовательно и индекс баланса, претерпевают изменения, связанные с неустойчивостью и неопределенностью среды, в которой происходит функционирование. Это дает предпосылки для рассмотрения факторов и индексов как случайных величин. Однако глобальные изменения окружающей среды происходят достаточно редко, поэтому законы, влияющие на значения факторов, можно считать неизменными. Из этого следует, что распределения рассматриваемых случайных величин можно полагать заданными априорно, что в свою очередь обуславливает применение к моделям баланса байесовского метода.

Поскольку в байесовских моделях баланса оба смешиваемых распределения в большинстве случаев имеют положительные носители, в качестве априорных имеет смысл рассматривать распределения гамма-типа, зарекомендовавшие себя за последние столетия в качестве адекватных моделей многих реальных процессов и явлений. Стоит также отметить, что вопрос об отыскании характеристик отношения двух факторов (индекса балан-

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

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

са) в данном случае эквивалентен задаче нахождения характеристик произведения двух величин из гамма-класса, что существенно расширяет область применимости рассматриваемых результатов.

2 Гамма-экспоненциальная функция

Приведем ряд вспомогательных инструментов.

Определение 1. Назовем функцию вида

$$\text{Ge}_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta),$$

$$x \in \mathbb{R}, \quad 0 \leq \alpha < 1, \quad \beta > 0, \quad (1)$$

гамма-экспоненциальной функцией [4].

Функция (1) обобщает на случай $\beta \neq 1$ преобразование, введенное Леруа [5] для исследования производящих функций специального вида. Кроме того, функцию (1) можно рассматривать (при некоторых допущениях) как частный случай функции Сривастава–Томовски [6], обобщающей функцию Миттаг-Леффлера [7].

Лемма 1. Пусть $\alpha, \theta > 0$, а $r, u, v \neq 0$ имеют один знак. Тогда

$$\int_0^{\infty} y^{r-1} e^{-(y/\alpha)^u - (y/\theta)^v} dy =$$

$$= \begin{cases} \frac{\theta^r}{|v|} \text{Ge}_{u/v, r/v} \left(- \left(\frac{\theta}{\alpha} \right)^u \right), & |v| > |u|; \\ \frac{\alpha^r}{|u|} \text{Ge}_{v/u, r/u} \left(- \left(\frac{\alpha}{\theta} \right)^v \right), & |u| > |v|; \\ \frac{\Gamma(r/u)}{|u|(\alpha^{-u} + \theta^{-u})^{r/u}}, & u = v. \end{cases}$$

Доказательство. Рассмотрим случай $r > 0$, $u > 0$ и $v > 0$. Предположим, что $v > u$. Тогда, используя теорему Лебега о предельном переходе, получаем:

$$\int_0^{\infty} y^{r-1} e^{-(y/\alpha)^u - (y/\theta)^v} dy =$$

$$= \frac{\alpha^r}{u} \int_0^{\infty} t^{r/u-1} e^{-t - (\alpha/\theta)^v t^{v/u}} dt =$$

$$= \frac{\alpha^r}{u} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^{\infty} t^{r/u+k-1} e^{-(\alpha/\theta)^v t^{v/u}} dt =$$

$$= \frac{\theta^r}{v} \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{uk}}{\alpha^{uk} k!} \Gamma \left(\frac{r+uk}{v} \right).$$

Случай $u > v$ рассматривается аналогично. Случай $u = v$ напрямую следует из определения гамма-функции. Так же аналогично получаются результаты для случая $r < 0$, $u < 0$ и $v < 0$. Лемма доказана.

Замечание 1. Несмотря на то что при $uv < 0$ и $r \neq 0$ интеграл $\int_0^{\infty} y^{r-1} e^{-(y/\alpha)^u - (y/\theta)^v} dy$, очевидно, сходится, непосредственное его вычисление связано с принципиальными трудностями.

Определение 2. Интегральной гамма-экспоненциальной функцией [8] назовем функцию

$$\text{Gi}(r, s, t; x) = \frac{1}{t\Gamma(s)\Gamma(t)} \int_0^x \text{Ge}_{r, tr+s} \left(-z^{1/t} \right) dz,$$

где $0 \leq r < 1$, $s, t > 0$.

3 Масштабные смеси обобщенных гамма-распределений

Будем обозначать обобщенное гамма-распределение, или распределение Стейси [9], с плотностью

$$f(x) = \frac{|v|x^{vq-1} e^{-(x/\theta)^v}}{\theta^{vq} \Gamma(q)},$$

$$v \neq 0, \quad q > 0, \quad \theta > 0, \quad x > 0,$$

через GG (v, q, θ) .

Класс обобщенных гамма-распределений достаточно широк и включает экспоненциальное распределение [10] ($v = 1, q = 1$); χ^2 -распределение [10] ($v = 1, q = k/2, k \in \mathbb{N}, \theta = 2$); масштабированное χ^2 -распределение [11] ($v = 1, q = k/2, k \in \mathbb{N}$); распределение Эрланга [10] ($v = 1, q \in \mathbb{N}$); гамма-распределение [10] ($v = 1$); полунормальное распределение [12], или распределение максимума процесса броуновского движения [13] ($v = 2, q = 1/2$); распределение Рэлея [10] ($v = 2, q = 1$); распределение Максвелла–Больцмана [10] ($v = 2, q = 3/2$); χ -распределение [10] ($v = 2, q = k/2, k \in \mathbb{N}, \theta = \sqrt{2}$); масштабированное χ -распределение [14] ($v = 2, q = k/2, k \in \mathbb{N}$); m-распределение Накагами [15] ($v = 2$); распределение Вильсона–Хильфerti [16] ($v = 3$); распределение Вейбулла–Гнеденко [10] ($v > 0, q = 1$); обобщенное распределение Вейбулла [17] ($v > 0, q \in \mathbb{N}$); псевдовейбулловское распределение [18] ($v > 0, q = 1 + 1/v$); распределение Пирсона третьего и пятого

типов [10] ($|v| = 1$); распределение Леви с нулевым смещением [10] ($v = -1, q = 1/2$); обратное экспоненциальное распределение [12] ($v = -1, q = 1$); обратное χ^2 -распределение [14] ($v = -1, q = k/2, k \in \mathbb{N}, \theta = 1/2$); масштабированное обратное χ^2 -распределение [14] ($v = -1, q = k/2, k \in \mathbb{N}$); обратное гамма-распределение [14] ($v = -1$); обратное распределение Рэлея [19] ($v = -2, q = 1$); обратное χ -распределение [11] ($v = -2, q = k/2, k \in \mathbb{N}, \theta = 1/\sqrt{2}$); масштабированное обратное χ -распределение [11] ($v = -2, q = k/2, k \in \mathbb{N}$); распределение Фреше [20] ($v < 0, q = 1$) и др.

Замечание 2. Если случайная величина ξ имеет распределение $\text{GG}(v, q, \theta)$, то случайная величина $1/\xi$ имеет распределение $\text{GG}(-v, q, 1/\theta)$. Поэтому приведенные ниже результаты для вероятностных характеристик отношения двух независимых случайных величин, имеющих обобщенное гамма-распределение с параметрами формы одного знака, несложно переформулировать для масштабных смесей обобщенных гамма-распределений, имеющих параметры формы разных знаков, с соответствующей заменой параметров.

Следующее утверждение непосредственно вытекает из леммы 1.

Теорема 1. Пусть независимые случайные величины λ и μ имеют соответственно распределения $\text{GG}(v, q, \theta)$ и $\text{GG}(u, p, \alpha)$, причем $uv > 0$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность:

$$f_\rho(x) = \begin{cases} \frac{|v| \alpha^{vq} x^{vq-1}}{\theta^{vq} \Gamma(p) \Gamma(q)} \text{Ge}_{v/u, vq/u+p} \left(- \left(\frac{\alpha x}{\theta} \right)^v \right), & |u| > |v|; \\ \frac{|u| \theta^{up} x^{-up-1}}{\alpha^{up} \Gamma(p) \Gamma(q)} \text{Ge}_{u/v, up/v+q} \left(- \left(\frac{\alpha x}{\theta} \right)^{-u} \right), & |v| > |u|; \\ \frac{|v| (\alpha/\theta)^{vq} x^{vq-1}}{B(p, q) (1 + (\alpha x/\theta)^v)^{p+q}}, & u = v. \end{cases}$$

Заметим, что при $u = v$ случайная величина ρ имеет обобщенное бета-распределение второго рода [21] $\text{GB2}(v, \theta/\alpha, q, p)$, частные случаи которого представляют распределение Бурра [22], или распределение Сингх–Маддала [23], $\text{GB2}(|v|, \theta/\alpha, 1, p)$; распределение Дагума [24] $\text{GB2}(|v|, \theta/\alpha, q, 1)$; распределение Ломакса [25] $\text{GB2}(1, \theta/\alpha, 1, p)$; F-распределение Фишера–Снедекора [10] $\text{GB2}(1, \theta/\alpha, \alpha/2, \theta/2)$ и др. Легко показать, что при $v < 0$ распределение $\text{GB2}(v, \theta/\alpha, q, p)$

совпадает с $\text{GB2}(|v|, \theta/\alpha, p, q)$. Таким образом, теорема 1 вместе с замечанием 2 дают возможность выражать характеристики смесей прямых и обратных гамма-распределений, а также распределений Вейбулла и Фреше с равными параметрами формы в терминах хорошо изученных распределений.

Несмотря на то что плотность отношения двух независимых случайных величин, имеющих обобщенное гамма-распределение, выражается в терминах специальной гамма-экспоненциальной функции, моментные характеристики ρ вычисляются достаточно просто.

Теорема 2. Пусть независимые случайные величины λ и μ имеют соответственно распределения $\text{GG}(v, q, \theta)$ и $\text{GG}(u, p, \alpha)$, причем $uv > 0$. Тогда для их отношения $\rho = \lambda/\mu$ при всех $z \in \mathbb{R}$ имеет место соотношение:

$$E\rho^z = \frac{(\theta/\alpha)^z \Gamma(q + z/v) \Gamma(p - z/u)}{\Gamma(q) \Gamma(p)}, \quad q + \frac{z}{v} > 0, \quad p - \frac{z}{u} > 0.$$

Доказательство непосредственно вытекает из независимости случайных величин λ и μ , замечания 2 и того факта, что

$$E\lambda^z = \frac{\theta^z \Gamma(q + z/v)}{\Gamma(q)}, \quad q + \frac{z}{v} > 0.$$

Теорема 2 дает возможность сформулировать следующее свойство гамма-экспоненциальной функции.

Следствие 1. Для всех $z \in \mathbb{R}$ при $0 < v/u < 1, p - z/u > 0, q + z/v > 0, c > 0$ выполняется соотношение:

$$\int_0^\infty x^{z+vq-1} \text{Ge}_{v/u, vq/u+p}(-cx^v) dx = \frac{\Gamma(q + z/v) \Gamma(p - z/u)}{|v| c^{q+z/v}},$$

в частности

$$\int_0^\infty x^{q-1} \text{Ge}_{\alpha, \alpha q+p}(-cx) dx = \frac{\Gamma(q) \Gamma(p)}{c^q}.$$

Перейдем к рассмотрению результата, касающегося описания функции распределения отношения двух независимых случайных величин, имеющих обобщенное гамма-распределение.

Теорема 3. Пусть независимые случайные величины λ и μ имеют соответственно распределения $\text{GG}(v, q, \theta)$

и $GG(u, p, \alpha)$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет функцию распределения:

$$F_\rho(x) = \begin{cases} Gi\left(\frac{v}{u}, p, q; \left(\frac{\alpha x}{\theta}\right)^{vq}\right), & u > v > 0; \\ 1 - Gi\left(\frac{v}{u}, p, q; \left(\frac{\alpha x}{\theta}\right)^{vq}\right), & u < v < 0; \\ 1 - Gi\left(\frac{u}{v}, q, p; \left(\frac{\alpha x}{\theta}\right)^{-up}\right), & v > u > 0; \\ Gi\left(\frac{u}{v}, q, p; \left(\frac{\alpha x}{\theta}\right)^{-up}\right), & v < u < 0. \end{cases}$$

Доказательство. Пусть $u > v > 0$. Имеем при $x > 0$:

$$\begin{aligned} F_\rho(x) &= \int_0^x \frac{v(\alpha/\theta)^{vq} y^{vq-1}}{\Gamma(p)\Gamma(q)} Ge_{v/u, vq/u+p}\left(-\left(\frac{\alpha y}{\theta}\right)^v\right) dy = \\ &= \frac{v(\alpha/\theta)^{vq}}{\Gamma(p)\Gamma(q)} \int_0^x \sum_{k=0}^{\infty} \frac{(-1)^k y^{vk+vg-1}}{(\theta/\alpha)^{vk} k!} \times \\ &\quad \times \Gamma\left(\frac{vk + vq + up}{u}\right) dy = \frac{1}{q\Gamma(p)\Gamma(q)} \times \\ &\quad \times \int_0^{(\alpha x/\theta)^{vq}} \sum_{k=0}^{\infty} \frac{(-y^{1/q})^k}{k!} \Gamma\left(\frac{vk + vq + up}{u}\right) dy = \\ &= \frac{1}{q\Gamma(p)\Gamma(q)} \int_0^{(\alpha x/\theta)^{vq}} Ge_{v/u, vq/u+p}\left(-y^{1/q}\right) dy. \end{aligned}$$

Пусть $u < v < 0$. Имеем при $x > 0$:

$$\begin{aligned} F_\rho(x) &= - \int_0^x \frac{v(\alpha/\theta)^{vq} y^{vq-1}}{\Gamma(p)\Gamma(q)} \times \\ &\quad \times Ge_{v/u, vq/u+p}\left(-\left(\frac{\alpha y}{\theta}\right)^v\right) dy = \\ &= - \frac{v(\alpha/\theta)^{vq}}{\Gamma(p)\Gamma(q)} \int_0^x \sum_{k=0}^{\infty} \frac{(-1)^k y^{vk+vg-1}}{(\theta/\alpha)^{vk} k!} \times \\ &\quad \times \Gamma\left(\frac{vk + vq + up}{u}\right) dy = \frac{1}{q\Gamma(p)\Gamma(q)} \times \\ &\quad \times \int_0^{(\alpha x/\theta)^{vq}} \sum_{k=0}^{\infty} \frac{(-y^{1/q})^k}{k!} \Gamma\left(\frac{vk + vq + up}{u}\right) dy = \\ &= 1 - \frac{1}{q\Gamma(p)\Gamma(q)} \int_0^{(\alpha x/\theta)^{vq}} Ge_{v/u, vq/u+p}\left(-y^{1/q}\right) dy. \end{aligned}$$

Случаи $v > u > 0$ и $v < u < 0$ рассматриваются аналогично. Теорема доказана.

Следствие 2. Для интегральной гамма-экспоненциальной функции справедливо соотношение:

$$\lim_{x \rightarrow \infty} Gi(r, s, t; x) = 1.$$

Заметим, что при $q = 1$ и $p = 1$ функция распределения отношения двух независимых случайных величин, имеющих обобщенное гамма-распределение, может быть представлена через гамма-экспоненциальную функцию. Это дает возможность исследовать масштабные смеси распределений Вейбулла и Фреше без использования интегральной гамма-экспоненциальной функции.

Следствие 3. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет функцию распределения:

$$F_\rho(x) = \begin{cases} \frac{v\alpha^{vq} x^{vq}}{u\theta^{vq}\Gamma(q)} Ge_{v/u, vq/u} \left(-\left(\frac{\alpha x}{\theta}\right)^v\right), & u > v > 0, p = 1; \\ 1 - \frac{1}{\Gamma(p)} Ge_{v/u, p} \left(-\left(\frac{\alpha x}{\theta}\right)^v\right), & u < v < 0, q = 1; \\ 1 - \frac{1}{\Gamma(q)} Ge_{u/v, q} \left(-\left(\frac{\alpha x}{\theta}\right)^{-u}\right), & v < u < 0, p = 1; \\ \frac{u\theta^{up} x^{-up}}{v\alpha^{up}\Gamma(p)} Ge_{u/v, up/v} \left(-\left(\frac{\alpha x}{\theta}\right)^{-u}\right), & v < u < 0, q = 1. \end{cases}$$

Доказательство. Пусть $u > v > 0$. Имеем при $x > 0$:

$$\begin{aligned} F_\rho(x) &= \frac{v(\alpha/\theta)^{vq}}{\Gamma(p)\Gamma(q)} \times \\ &\quad \times \int_0^x \sum_{k=0}^{\infty} \frac{(-1)^k y^{vk+vg-1}}{(\theta/\alpha)^{vk} k!} \Gamma\left(\frac{vk + vq + up}{u}\right) dy = \\ &= \frac{(\alpha/\theta)^{vq}}{\Gamma(p)\Gamma(q)} \sum_{k=0}^{\infty} \frac{(-1)^k x^{vk+vg}}{(\theta/\alpha)^{vk} (k+q)k!} \Gamma\left(\frac{vk + vq + up}{u}\right). \end{aligned}$$

При $p = 1$ получаем:

$$F_\rho(x) = \frac{v(\alpha/\theta)^{vq}}{u\Gamma(q)} \sum_{k=0}^{\infty} \frac{(-1)^k x^{vk+vg}}{(\theta/\alpha)^{vk} k!} \Gamma\left(\frac{vk + vq}{u}\right).$$

При $q = 1$ получаем

$$F_{\rho}(x) = -\frac{1}{\Gamma(p)} \times \sum_{k=0}^{\infty} \frac{(-1)^{k+1} x^{v(k+1)}}{(\theta/\alpha)^{v(k+1)} (k+1)!} \Gamma\left(\frac{v(k+1) + up}{u}\right) = 1 - \frac{1}{\Gamma(p)} \sum_{k=0}^{\infty} \frac{(-1)^k x^{vk}}{(\theta/\alpha)^{vk} k!} \Gamma\left(\frac{vk + up}{u}\right).$$

Случай $v < u < 0$ рассматривается аналогично. Следствие доказано.

Замечание 3. Следствие 3 можно переформулировать в терминах гамма-экспоненциальной функции как свойство инвариантности относительно интегрального преобразования при $\alpha, c, p, q > 0$:

$$\int_0^x y^{\alpha q - 1} \text{Ge}_{\alpha, \alpha q + 1}(-cy^{\alpha}) dy = x^{\alpha q} \text{Ge}_{\alpha, \alpha q}(-cx^{\alpha});$$

$$\int_0^x c\alpha y^{-\alpha - 1} \text{Ge}_{\alpha, \alpha + p}(-cy^{-\alpha}) dy = \Gamma(p) - \text{Ge}_{\alpha, p}(-cx^{-\alpha}).$$

4 Заключение

Приведенные результаты дают возможность исследовать модели, в которых используются отношения и произведения случайных величин, имеющих обобщенное гамма-распределение, в терминах гамма-экспоненциальной функции, свойства которой во многом унаследованы от показательной функции.

Автор выражает признательность В. Ю. Королеву, Ю. Н. Недоливко, С. И. Палионной, В. В. Саенко, А. И. Титовой, Ю. С. Хохлову и С. Я. Шоргину за содействие при подготовке публикации.

Литература

1. Шоргин С. Я. О байесовских моделях массового обслуживания // II Научная сессия Института проблем информатики РАН: Тезисы докладов. — М.: ИПИ РАН, 2005. С. 120–121.
2. Кудрявцев А. А., Соколов И. А., Шоргин С. Я. Байесовская рекуррентная модель роста надежности: равномерное распределение параметров // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 55–59.

3. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
4. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
5. Le Roy É. Sur les séries divergentes et les fonctions définies par un développement de Taylor // Annales de la faculté des sciences de Toulouse 2e série, 1900. Vol. 2. No. 3. P. 317–384.
6. Srivastava H. M., Tomovski Ž. Fractional calculus with an integral operator containing a generalized Mittag-Leffler function in the kernel // Appl. Math. Comput., 2009. Vol. 211. P. 198–210.
7. Gorenflo R., Kilbas A. A., Mainardi F., Rogosin S. V. Mittag-Leffler functions, related topics and applications. — Berlin–Heidelberg: Springer-Verlag, 2014. 443 p.
8. Кудрявцев А. А., Палионная С. И., Шоргин В. С. Априорное обобщенное распределение Фреше в байесовских моделях баланса // Системы и средства информатики, 2019. Т. 29. № 2. С. 39–45.
9. Stacy E. W. A generalization of the gamma distribution // Ann. Math. Stat., 1962. Vol. 33. P. 1187–1192.
10. Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф. Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985. 640 с.
11. Lee P. M. Bayesian statistics: An introduction. — 4th ed. — Chichester, U.K.: John Wiley & Sons, 2012. 462 p.
12. Kleiber C., Kotz S. Statistical size distributions in economics and actuarial sciences. — New York, NY, USA: John Wiley & Sons, 2003. 332 p.
13. Круглов В. М. Случайные процессы. Ч. 1. Основы общей теории. — 2-е изд., перераб. и доп. — М.: Юрайт, 2016. 276 с.
14. Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B. Bayesian data analysis. — 3rd ed. — CRC Press, 2014. 639 p.
15. Nakagami M. The m-distribution, a general formula of intensity of rapid fading // Statistical Methods in Radio Wave Propagation: Symposium Proceedings / Ed. W. C. Homan. — New York, NY, USA: Pergamon Press, 1960. P. 3–36.
16. Wilson E. B., Hilferty M. M. The distribution of chi-square // P. Natl. Acad. Sci. USA, 1931. Vol. 17. No. 12. P. 684–688.
17. Lai C.-D. Generalized Weibull distributions. — New York, NY, USA: Springer, 2014. 128 p.
18. Vodá V. G. New models in durability tool-testing: Pseudo-Weibull distribution // Kybernetika, 1989. Vol. 25. No. 3. P. 209–215.
19. Panwar M. S., Sudhir B. A., Bundel R., Tomer S. K. Parameter estimation of inverse Rayleigh distribution under

- competing risk model for masked data // J. Inst. Sci. Technol., 2015. Vol. 20. No. 2. P. 122–127.
20. Fréchet M. Sur la loi de probabilité de l'écart maximum // Ann. Soc. Polonaise Math., 1927. Vol. 6. P. 93–116.
21. McDonald J. B. Some generalized functions for the size distribution of income // Econometrica, 1984. Vol. 52. No. 3. P. 647–665.
22. Burr I. W. Cumulative frequency functions // Ann. Math. Stat., 1942. Vol. 13. Iss. 2. P. 215–232.
23. Singh S. K., Maddala G. S. A function for size distribution of incomes // Econometrica, 1976. Vol. 44. No. 5. P. 963–970.
24. Dagum C. A new model of personal income-distribution-specification and estimation // Econ. Appl., 1977. Vol. 30. No. 3. P. 413–437.
25. Lomax K. S. Business failures: Another example of the analysis of failure data // J. Am. Stat. Assoc., 1954. Vol. 49. No. 268. P. 847–852.

Поступила в редакцию 07.06.19

A PRIORI GENERALIZED GAMMA DISTRIBUTION IN BAYESIAN BALANCE MODELS

A. A. Kudryavtsev

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

Abstract: The work is devoted to the study of Bayesian balance models, involving the division of the system parameters into two classes: supporting system functioning positive factors and interfering with the functioning negative factors. The balance index, defined as the ratio of the negative factor to the positive factor, is considered. The formulation of the problem, which consists in finding the main probabilistic characteristics (density, distribution function, and moments) of the balance index of factors having *a priori* generalized gamma distribution with the parameters of the form of one sign, is studied. The results are formulated in terms of the gamma-exponential function. A number of new properties of the latter are given. It is shown that the given statements are easily reformulated for large-scale mixtures of generalized gamma distributions with parameters of the form of different signs. The obtained results can be widely used in models, which describe the processes and phenomena using distributions with a positive unlimited support.

Keywords: Bayesian approach; generalized gamma distribution; gamma-exponential function; balance models; mixed distributions

DOI: 10.14357/19922264190305

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

References

- Shorgin, S. Ya. 2005. O bayesovskikh modelyakh massovogo obsluzhivaniya [On Bayesian queuing models]. *II Nauchnaya sessiya Instituta problem informatiki RAN: Tezisy dokladov* [2nd Scientific Session of the Institute of Informatics Problems of the Russian Academy of Sciences: Abstracts]. Moscow: IPI RAN. 120–121.
- Kudryavtsev, A. A., I. A. Sokolov, and S. Ya. Shorgin. 2013. Bayesovskaya rekurrentnaya model' rosta nadezhnosti: ravnomernoe raspredelenie parametrov [Bayesian recurrent model of reliability growth: Uniform distribution of parameters]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):55–59.
- Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
- Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
- Le Roy, É. 1900. Sur les séries divergentes et les fonctions définies par un développement de Taylor. *Annales de la faculté des sciences de Toulouse 2e série* 2(3): 317–384.
- Srivastava, H. M., and Ž. Tomovski. 2009. Fractional calculus with an integral operator containing a generalized Mittag-Leffler function in the kernel. *Appl. Math. Comput.* 211:198–210.
- Gorenlo, R., A. A. Kilbas, F. Mainardi, and S. V. Rogosin. 2014. *Mittag-Leffler functions, related topics and applications*. Berlin – Heidelberg: Springer-Verlag. 443 p.

8. Kudryavtsev, A. A., S. I. Palionnaia, and V. S. Shorgin. 2019. Apriornoe obobshchennoe raspredelenie Freshe v bayesovskikh modelyakh balansa [*A priori* generalized Frechet distribution in Bayesian balance models]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):39–45.
9. Stacy, E. W. 1962. A generalization of the gamma distribution. *Ann. Math. Stat.* 33:1187–1192.
10. Korolyuk, V. S., N. I. Portenko, A. V. Skorokhod, and A. F. Turbin. 1985. *Spravochnik po teorii veroyatnostey i matematicheskoy statistike* [Handbook of probability theory and mathematical statistics]. Moscow: Nauka. 640 p.
11. Lee, P. M. 2012. *Bayesian statistics: An introduction*. 4th ed. Chichester, U.K.: John Wiley & Sons. 462 p.
12. Kleiber, C., and S. Kotz. 2003. *Statistical size distributions in economics and actuarial sciences*. New York, NY: John Wiley & Sons. 332 p.
13. Kruglov, V. M. 2016. *Sluchaynye protsessy. Ch. 1. Osnovy obshchey teorii* [Stochastic processes. Part 1. Bases of general theory]. 2nd ed. Moscow: Yurayt. 276 p.
14. Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. 3rd ed. CRC Press. 639 p.
15. Nakagami, M. 1960. The m-distribution, a general formula of intensity of rapid fading. *Statistical Methods in Radio Wave Propagation: Symposium Proceedings*. Ed. W. C. Hoffman. New York, NY: Pergamon Press. 3–36.
16. Wilson, E. B., and M. M. Hilferty. 1931. The distribution of chi-square. *P. Natl. Acad. Sci. USA* 17(12):684–688.
17. Lai, C.-D. 2014. *Generalized Weibull distributions*. New York, NY: Springer. 128 p.
18. Voda, V. G. 1989. New models in durability tool-testing: Pseudo-Weibull distribution. *Kybernetika* 25(3):209–215.
19. Panwar, M. S., B. A. Sudhir, R. Bundel, and S. K. Tomer. 2015. Parameter estimation of inverse Rayleigh distribution under competing risk model for masked data. *J. Inst. Sci. Technol.* 20(2):122–127.
20. Fréchet, M. 1927. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polonaise Math.* 6:93–116.
21. McDonald, J. B. 1984. Some generalized functions for the size distribution of income. *Econometrica* 52(3):647–665.
22. Burr, I. W. 1942. Cumulative frequency functions. *Ann. Math. Stat.* 13(2):215–232.
23. Singh, S. K., and G. S. Maddala. 1976. A function for size distribution of incomes. *Econometrica* 44(5):963–970.
24. Dagum, C. 1977. A new model of personal income-distribution-specification and estimation. *Econ. Appl.* 30(3):413–437.
25. Lomax, K. S. 1954. Business failures: Another example of the analysis of failure data. *J. Am. Stat. Assoc.* 49(268):847–852.

Received June 7, 2019

Contributor

Kudryavtsev Alexey A. (b. 1978) — Candidate of Sciences (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

ГИБРИДНЫЕ МОДЕЛИ ЭКСТРЕМАЛЬНОГО ГРАДИЕНТНОГО БУСТИНГА ДЛЯ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ В ДАННЫХ ОБ ОСАДКАХ*

А. К. Горшенин¹, О. П. Мартынов²

Аннотация: Проведено сравнение классического метода экстремального градиентного бустинга, реализованного во фреймворке XGBoost (eXtreme Gradient Boosting, экстремальный градиентный бустинг) и категориальной модификации CatBoost (Categorical Boosting, категориальный бустинг), которая достаточно редко встречается в научных исследованиях. Предложены некоторые гибридные модели классификации-регрессии для повышения точности заполнения пропусков в реальных данных на примере 14 станций в Германии. Достигнутая точность в задачах классификации составила до 92% при весьма умеренных значениях ошибок прогнозов в метрике RMSE (Root Mean-Square Error, среднеквадратичная ошибка). Гибридные методы превзошли по качеству предсказания простые модели классификации и регрессии. Развиваемые подходы могут быть успешно использованы как для непосредственного анализа метеорологических данных методами машинного обучения, так и для улучшения качества предсказания на основе физических моделей атмосферных процессов.

Ключевые слова: заполнение пропусков; осадки; классификация; регрессия; градиентный бустинг; XGBoost; CatBoost

DOI: 10.14357/19922264190306

1 Введение

Повышение эффективности алгоритмов машинного обучения привело к росту их востребованности как в задачах анализа результатов физических моделей предсказания погоды с целью получения более точного прогноза, так и в качестве самостоятельных инструментов исследования пространственно-временных метеорологических рядов, полученных со спутников и метеостанций. Такие наблюдения в больших объемах поступают с огромного числа датчиков и зачастую содержат пропуски, которые могут существенным образом повлиять на качество обучения методов или изменить решения статистических моделей анализа различных метеорологических явлений, например экстремальных осадков [1]. Поэтому весьма важной оказывается задача корректного заполнения пропусков в данных.

В настоящей статье развивается подход к обработке пропущенных значений для объемов осадков на основе популярного алгоритма машинного

обучения, называемого градиентным бустингом над деревьями решений [2]. Это семейство методов, включая и его наиболее часто применяемую модификацию XGBoost [3], широко используется для решения задач классификации и регрессии в значительном спектре прикладных областей. Например, можно упомянуть работы по предсказанию биологической активности лекарств [4], кредитному скоррингу [5], прогнозированию кризисов на финансовых рынках [6], обнаружению дефектов в ветровых турбинах [7], ветровой энергетике и солнечной радиации [8, 9].

Предложенная компанией Яндекс модификация метода градиентного бустинга, получившая название CatBoost [10], несмотря на ее применение в реальных продуктах, в научных статьях используется достаточно редко. В качестве примеров можно упомянуть протеомные исследования [11] и «умные» сети электроснабжения [12]. В данной статье восполняется этот пробел для метеорологических данных, для которых традиционно применяется классический экстремальный гради-

*Постановка задачи и анализ полученных результатов в данной статье проведены А. К. Горшениным, чьи исследования поддержаны РНФ (проект 18-71-00156). Разработка и программная реализация методов анализа пропущенных значений выполнены О. П. Мартыновым.

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, agorshenin@frcsc.ru

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, martynov.oleg.mipt@gmail.com

ентный бустинг [13, 14]. Будет проведено сравнение XGBoost и CatBoost, а также предложены некоторые гибридные модели для повышения точности заполнения пропусков в реальных данных на примере 14 станций в Германии. Кроме того, для извлечения признаков будет использована новая библиотека `tsfresh` [15] для языка программирования Python.

2 Постановка задачи

Исходные данные представляют собой наборы признаков $\{\hat{x}_t | t \in T\}$ и соответствующие им измерения $\{y_t | t \in T\}$ за некоторый временной интервал T . На первом этапе наблюдения $\{y_t | t \in T\}$ случайным образом разделяются на тренировочную $\{y_t | t \in T_{\text{train}}\}$ и тестовую $\{y_t | t \in T_{\text{test}}\}$ части, причем

$$T_{\text{train}} \cup T_{\text{test}} = T, \quad T_{\text{train}} \cap T_{\text{test}} = \emptyset.$$

На втором этапе без увеличения объема исходных наблюдений с помощью ряда специальных процедур (см. статью [15], где также приведено описание пакета на языке Python) формируются дополнительные признаки, которые в дальнейшем будут использованы при обучении моделей. Таким образом, проводится дополнение исходных признаков $\{\hat{x}_t\}$ наборами функций $\{f_i(\cdot)\}$ и $\{g_i(\cdot)\}$ для получения расширенного набора $\{\hat{x}_t^e\}$:

$$\{\hat{x}_t^e\} = \{\hat{x}_t\} \cup \{f_i(y_{t-1}, y_{t-2}, \dots, y_{t-k+1})\} \cup \{g_i(y_{t+1}, y_{t+2}, \dots, y_{t+l-1})\}, \quad (1)$$

который и будет использован для обучения моделей. Отметим, что в формуле (1) величины k и l также могут рассматриваться в качестве свободных параметров. Таким образом, для обучения используются расширенные наборы признаков $\{\hat{x}_t^e | t \in T_{\text{train}}\}$ и соответствующие им измерения $\{y_t | t \in T_{\text{train}}\}$, а в качестве выходных данных выступают прогнозы $\{y_t^{\text{pred}} | t \in T_{\text{test}}\}$.

Для оценивания корректности работы методов в данной статье будут использоваться несколько метрик. Для задач классификации используется величина АСС, определяемая выражением:

$$\text{ACC} = \text{TPR} + \text{TNR}, \quad (2)$$

где TPR (True Positive Rate) — доля верно угаданных положительных классов, а TNR

(True Negative Rate) — доля верно угаданных отрицательных классов (подробнее они будут описаны в подразд. 4.1 с привязкой к наличию или отсутствию осадков в конкретный день). Также для сравнения бинарных классификаторов используется площадь под ROC-кривой (ROC AUC — area under ROC (receiver operating characteristic) curve) [16], которая описывает точность решения модели как вероятность принадлежности к определенному классу и задается выражением:

$$\text{ROC AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx, \quad (3)$$

где FPR (False Positive Rate) — доля неверно угаданных положительных классов. Она позволяет корректнее оценивать точность в случае, когда один из классов является доминирующим, что вполне соответствует исходным данным об осадках, в которых достаточно много нулевых значений. Кроме того, используется и стандартная для непрерывных данных метрика RMSE:

$$\text{RMSE} = \sqrt{|T|^{-1} \sum_{t \in T} (y_t - y_t^{\text{pred}})^2}. \quad (4)$$

3 Подготовка данных и программная реализация методов машинного обучения

В качестве исходных данных использованы точные метеорологические наблюдения за различные периоды времени, собранные на 14 станциях в восьми городах Германии¹. Анализ в настоящей работе проводится для объемов осадков, а такие показатели, как средняя температура, точка росы и средняя скорость ветра, включены в модели в качестве дополнительных признаков. Предварительно проведена простая регуляризация исходных данных — отсечение аномальных выбросов, связанных с пропущенными значениями (в имеющихся в распоряжении рядах они заполнены большими величинами), на уровне 95% выборочного квантиля. Отметим, что такое преобразование не повлекло изменений очевидно корректных значений в тестовых наборах, следовательно, не повлияло на качество работы методов.

Для возможности верификации разрабатываемых методов заполнения пропусков использовались интервалы наблюдений, в которых данные

¹Открытая база NNDC (NOAA's National Data Centers) Climate Data (<https://www7.ncdc.noaa.gov/CDO/cdo>) Национального управления океанических и атмосферных исследований (NOAA — National Oceanic and Atmospheric Administration), США.

были полны. Для искусственного внесения пропусков исходный интервал разбивался на непересекающиеся отрезки на некотором расстоянии друг от друга, затем в каждый из них случайным образом помещалось единственное пропущенное значение:

$$\dots \underbrace{y_{t+1}, \dots, y_{t+g}}_{0 \text{ пропусков}}, \underbrace{y_{t+g+1}, \dots, y_{t+g+s}}_{1 \text{ пропуск}}, \dots, \underbrace{y_{t+g+s+1}, \dots, y_{t+2g+s+1}}_{0 \text{ пропусков}}, \dots$$

где $g = \max\{k, l\}$ (величины k и l определяются из формулы (1)), а параметр s задается следующим выражением:

$$s = \frac{\{\text{длина интервала}\} - g}{\{\text{число пропусков}\}} - g.$$

В целях повышения качества предсказания рассматриваемых моделей из тренировочной и тестовой частей данных были исключены значения, для которых нельзя было выбрать в точности k предыдущих и l последующих значений для корректного извлечения дополнительных признаков (см. разд. 2).

Предварительная статистическая обработка, а также результаты статьи [17] демонстрируют, что распределения суточных объемов осадков хорошо аппроксимируются с помощью гамма-распределения. Этот факт будет использован в некоторых моделях, в частности в XGBoost_Gamma.

Программные инструменты анализа данных реализованы на языке Python с использованием библиотек XGBoost и CatBoost для соответствующих моделей градиентного бустинга, tsfresh для извлечения дополнительных признаков, а также классических инструментов NumPy, Pandas, SciPy

и scikit-learn. Все модели обучались для различных конфигураций параметров k и l (1), итоговые результаты приведены для $k \equiv l \equiv 5$, для которых получена наибольшая точность из всех рассмотренных комбинаций. Всего для рассмотренных станций были апробированы порядка 8500 различных конфигураций.

4 Классические модели градиентного бустинга

В задачах прогнозирования традиционно выделяются два возможных направления: определение попадания предсказываемого наблюдения в некоторую группу (классификация) и установление точной величины неизвестного значения (регрессия). В этом разделе рассмотрим применение алгоритмов XGBoost и CatBoost для решения каждой из них.

4.1 Классификация

Поставим в соответствие измерениям осадков $\{y_t\}$ последовательность $\{c_t = \mathcal{I}(y_t \geq \varepsilon)\}$ для некоторого наперед заданного значения ε (например, $\varepsilon = 0,005$), где $\mathcal{I}(\cdot)$ обозначает индикатор соответствующего множества. Так формируется разбиение исходных данных на два класса — наличие и отсутствие осадков. Первый класс $\{c_t = 1\}$ будем называть положительным, а второй $\{c_t = 0\}$ — отрицательным. Модель классификации на вход получает значения \hat{x}_t^e , а в качестве результата выдает вероятность принадлежности наблюдений к положительному классу $\{p_t^{\text{pred}} \equiv \mathbb{P}(c_t = 1)\}$. Результирующий класс определяется как $\{c_t^{\text{pred}} = \mathcal{I}(p_t^{\text{pred}} \geq 0,5)\}$.

Таблица 1 Сравнение моделей классификации

Город	Станция	Лучшая модель	ACC	ROC AUC
Берлин	93850	XGBoost_Logistic	86,11%	95,64%
Берлин	103810	CatBoost_Logistic	80,56%	82,19%
Доберлуг	94900	CatBoost_Logistic	83,33%	88,1%
Доберлуг	104900	CatBoost_Logistic	86,11%	84,23%
Хольцдорф	104760	XGBoost_Logistic	61,11%	72,1%
Линденберг	93930	CatBoost_Logistic	75%	76,92%
Линденберг	103930	CatBoost_Logistic	86,11%	86,55%
Нойруппин	92700	XGBoost_Logistic	72,22%	75,6%
Нойруппин	102700	CatBoost_Logistic	69,44%	66,9%
Потсдам	93790	XGBoost_Logistic	63,89%	68,75%
Потсдам	103790	XGBoost_Logistic	66,7%	78,41%
Визенбург	103680	XGBoost_Logistic	66,7%	70,63%
Виттенберг	94740	XGBoost_Logistic	77,78%	74,6%
Виттенберг	104740	XGBoost_Logistic	69,44%	73,96%

Таблица 2 Усредненные результаты предсказания моделей классификации

Модель	Метрика	
	ACC	ROC AUC
XGBoost_Logistic	72,22%	78,26%
CatBoost_Logistic	71,03%	75,96%

Для решения задачи классификации были выбраны модели градиентного бустинга XGBoost_Logistic и CatBoost_Logistic с целевой функцией логистической регрессии — для них требуется меньший объем данных для обучения. В табл. 1 приведены результаты для лучшей из рассмотренных конфигураций по каждой станции в метриках ACC (2) и ROC AUC (3).

В табл. 2 приведены усредненные сразу по всем вариантам параметров значения точности для каждой модели.

В данной задаче модель XGBoost_Logistic продемонстрировала в среднем несколько более высокие (порядка 1%–3%) показатели точности по сравнению с CatBoost_Logistic. Однако, как видно из табл. 1, для шести из 14 станций более успешным оказалось применение категориального бустинга.

4.2 Регрессия

В данном случае в качестве входных параметров используются величины $\{\hat{x}_t^e\}$, а на выходе формируется набор $\{y_t^{\text{pred}}\}$ (см. разд. 2). При этом результирующий класс имеет следующий вид: $\{c_t^{\text{pred}} = \mathcal{I}(y_t^{\text{pred}} \geq \varepsilon)\}$. В качестве моделей регрессии используются XGBoost и CatBoost с различными целевыми функциями. В XGBoost_RMSE и CatBoost_RMSE используется RMSE (4), CatBoost_MAE — средняя абсолютная

ошибка (MAE), а в XGBoost_Gamma — гамма-регрессия с логарифмической связью. В табл. 3 приведены результаты лучшей модели для каждой станции в метриках ACC (2) и RMSE (4).

5 Гибридные модели

Сравнение результатов точности описанных в предыдущем разделе моделей показывает, что модели регрессии гораздо хуже справляются с задачей классификации. В этом разделе предложены гибридные модели, которые сочетают преимущества обоих подходов.

XGBoost_Logistic+Mean. Простую комбинированную модель можно получить, совместив выход классификатора $\{c_t^{\text{cls}}\}$ и среднее значение объемов осадков $m = |T|^{-1} \sum_{t \in T} y_t$. Тогда выход комбинированной модели определим как $\{y_t^{\text{pred}} = c_t^{\text{cls}} m\}$.

XGBoost_Logistic+RMSE. Здесь вместо среднего значения m будет использоваться выход простого регрессора $\{y_t^{\text{reg}}\}$. Тогда выход новой модели определим как $\{y_t^{\text{pred}} = c_t^{\text{cls}} y_t^{\text{reg}}\}$.

XGBoost_Logistic+RMSE+Sigmoid. Пусть теперь выход классификатора $\{p_t^{\text{cls}}\}$ определяет вероятность принадлежности к положительному классу. Введем функцию связи вида

$$s(x) = \left(1 + e^{-\alpha(x-\beta)}\right)^{-1},$$

где α и β — некоторые заданные действительные коэффициенты. Она позволяет использовать вместо бинарного решения $\{c_t^{\text{cls}}\}$ набор

Таблица 3 Сравнение моделей регрессии

Город	Станция	Лучшая модель	ACC	RMSE
Берлин	93850	XGBoost_Gamma	83,33%	0,0631
Берлин	103810	XGBoost_Gamma	69,44%	0,1111
Доберлуг	94900	XGBoost_Gamma	86,11%	0,075
Доберлуг	104900	CatBoost_MAE	83,33%	0,0876
Хольцдорф	104760	XGBoost_Gamma	75%	0,1027
Линденберг	93930	XGBoost_Gamma	75%	0,0721
Линденберг	103930	XGBoost_Gamma	80,56%	0,0912
Нойруппин	92700	XGBoost_Gamma	66,7%	0,061
Нойруппин	102700	CatBoost_MAE	66,7%	0,0643
Потсдам	93790	XGBoost_Gamma	66,7%	0,0910
Потсдам	103790	XGBoost_Gamma	66,7%	0,0886
Визенбург	103680	XGBoost_Gamma	61,11%	0,1767
Виттенберг	94740	CatBoost_MAE	72,22%	0,0591
Виттенберг	104740	XGBoost_RMSE	77,78%	0,0734

Таблица 4 Сравнение результатов гибридных моделей

Город	Станция	Лучшая модель	ACC	RMSE
Берлин	93850	XGBoost_Logistic+RMSE+Sigmoid	91,67%	0,0588
Берлин	103810	XGBoost_Logistic+RMSE	75%	0,1012
Нойруппин	92700	XGBoost_Logistic+RMSE	72,22%	0,0556
Нойруппин	102700	XGBoost_Logistic+RMSE+Sigmoid	69,44%	0,0593
Потсдам	93790	XGBoost_Logistic+RMSE+Sigmoid	69,44%	0,0697
Потсдам	103790	XGBoost_Logistic+RMSE+Sigmoid	69,44%	0,0777
Визенбург	103680	XGBoost_Logistic+RMSE	66,7%	0,1418
Виттенберг	94740	XGBoost_Logistic+RMSE+Sigmoid	80,56%	0,052

непрерывных значений, а кроме того, предоставляет возможность гибкого подбора коэффициентов, наиболее подходящих для конкретных данных. Выход такой модели определим как $\{y_t^{\text{pred}} = s(p_t^{\text{cls}}) y_t^{\text{reg}}\}$. Отметим, что для рассматриваемых в статье данных наилучшие результаты были получены для конфигураций с $\alpha = 10$ и $\beta = 0,45$.

В табл. 4 приведены станции, для которых были улучшены результаты чисто регрессионных моделей (см. табл. 3). Отметим, что этого удалось достичь для 8 из 14 станций, причем увеличение точности классификации составило от 2% до 8% в зависимости от местоположения. Таким образом, использование гибридных моделей оказывается вполне оправданным с точки зрения более точной обработки данных и корректного заполнения пропусков в них.

В табл. 5 приведены усредненные сразу по всем вариантам параметров значения точности как для чисто регрессионных, так и для гибридных моделей в порядке возрастания величины ACC (2). Кроме того, для сравнения представлена простая модель заполнения средним значением, обозначенная как Mean.

Отметим, что и величины ошибок RMSE для непрерывных значений также остаются весьма умеренными.

Таблица 5 Усредненные результаты точности моделей регрессии, включая гибридные

Модель	Метрика	
	ACC	RMSE
Mean	45,24%	0,0801
CatBoost_RMSE	47,02%	0,0756
XGBoost_RMSE	58,73%	0,0759
CatBoost_MAE	60,91%	0,0804
XGBoost_Gamma	70,83%	0,0877
XGBoost_Logistic+Mean	72,22%	0,0808
XGBoost_Logistic+RMSE	72,42%	0,0787
XGBoost_Logistic+RMSE+Sigmoid	74,4%	0,0763

6 Заключение

В работе продемонстрирована возможность высокоточного заполнения пропусков в данных с использованием гибридных моделей градиентного бустинга. Достигнутая точность в задачах классификации составила до 92% при весьма умеренных значениях ошибок прогнозов в метрике RMSE. Гибридные методы превосходили по качеству предсказания как простые модели классификации, так и регрессии.

Было проведено сравнение методов библиотек XGBoost и CatBoost, которое показало, что в большинстве случаев для работы с осадками более перспективным представляется использование экстремального градиентного бустинга, однако для отдельных рядов результаты могут быть улучшены за счет применения категориальных моделей. Развиваемые подходы могут быть успешно использованы как для непосредственного анализа метеорологических данных методами машинного обучения, так и для улучшения качества предсказания на основе физических моделей атмосферных процессов.

Авторы выражают особую признательность профессору В. Ю. Королеву за полезные обсуждения в рамках совместных исследований метеорологических явлений.

Литература

1. Горшенин А. К., Королев В. Ю. Определение экстремальности объемов осадков на основе модифицированного метода превышения порогового значения // Информатика и её применения, 2018. Т. 12. Вып. 4. С. 16–24.
2. Friedman J. H. Greedy function approximation: A gradient boosting machine // Ann. Stat., 2001. Vol. 29. Iss. 5. P. 1189–1232.
3. Chen T., Guestrin C. XGBoost: A scalable tree boosting system // 22nd ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings. — San Francisco, CA, USA, 2016. P. 785–794.

4. *Mustapha I. B., Saeed F.* Bioactive molecule prediction using extreme gradient boosting // *Molecules*, 2016. Vol. 21. Iss. 8. Art. No. 983.
5. *Xia Y., Liu C., Li Y., Liu N.* A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring // *Expert Syst. Appl.*, 2017. Vol. 78. P. 225–241.
6. *Chatzis S. P., Siakoulis V., Petropoulos A., Stavroulakis E., Vlachogiannakis N.* Forecasting stock market crisis events using deep and statistical machine learning techniques // *Expert Syst. Appl.*, 2018. Vol. 112. P. 353–371.
7. *Zhang D., Qian L., Mao B., Huang C., Huang B., Si Y.* A data-driven design for fault detection of wind turbines using random forests and XGboost // *IEEE Access*, 2018. Vol. 6. P. 21020–21031.
8. *Aler R., Galvan I. M., Ruiz-Arias J. A., Gueymard C. A.* Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting // *Sol. Energy*, 2017. Vol. 150. P. 558–569.
9. *Torres-Barran A., Alonso A., Dorronsoro J. R.* Regression tree ensembles for wind energy and solar radiation prediction // *Neurocomputing*, 2018. Vol. 326. P. 151–160.
10. *Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A.* CatBoost: Unbiased boosting with categorical features // *Adv. Neur. In.*, 2018. Vol. 31. P. 6638–6648.
11. *Ivanov M. V., Levitsky L. I., Bubis J. A., Gorshkov M. V.* Scavager: A versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting // *Proteomics*, 2019. Vol. 19. Iss. 3. Art. No. 1800280.
12. *Punmiya R., Choe S.* Energy theft detection using gradient boosting theft detector with feature boost engineering-based preprocessing // *IEEE T. Smart Grid*, 2019. Vol. 10. Iss. 2. P. 2326–2329.
13. *Korner P., Kronenberg R., Genzel S., Bernhofer C.* Introducing Gradient Boosting as a universal gap filling tool for meteorological time series // *Meteorol. Z.*, 2018. Vol. 27. Iss. 5. P. 369–376.
14. *Fan J., Wang X., Wu L., Zhou H., Zhang F., Yu X., Lu X., Xiang Y.* Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China // *Energ. Convers. Manage.*, 2018. Vol. 164. P. 102–111.
15. *Christ M., Braun N., Neuffer J., Kempa-Liehr A. W.* Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh — a Python package) // *Neurocomputing*, 2018. Vol. 307. P. 72–77.
16. *Huang J., Ling C. X.* Using AUC and accuracy in evaluating learning algorithms // *IEEE T. Knowl. Data En.*, 2005. Vol. 17. Iss. 3. P. 299–310.
17. *Gorshenin A. K., Korolev V. Yu.* Scale mixtures of Frechet distributions as asymptotic approximations of extreme precipitation // *J. Math. Sci.*, 2018. Vol. 234. Iss. 6. P. 886–903.

Поступила в редакцию 08.07.19

HYBRID EXTREME GRADIENT BOOSTING MODELS TO IMPUTE THE MISSING DATA IN PRECIPITATION RECORDS

A. K. Gorshenin^{1,2} and O. P. Martynov²

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

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

Abstract: The article compares the classical method of extreme gradient boosting implemented in the XGBoost (eXtreme Gradient Boosting) framework with the new modification CatBoost (Categorical Boosting), which is rarely involved in scientific researches. Some hybrid classification-regression models are proposed to improve the accuracy of imputation in missing values in real data using 14 meteorological stations in Germany. The achieved accuracy of the classification is up to 92% and the root-mean-square errors are quite moderate. The hybrid methods outperformed both simple classification and regression models in prediction accuracy. The proposed approaches can be successfully used for meteorological data analysis by machine learning methods as well as for improving the forecasting accuracy in physical models of atmospheric processes.

Keywords: data imputation; precipitation; classification; regression; gradient boosting; XGBoost; CatBoost

DOI: 10.14357/19922264190306

Acknowledgments

The problems formulation and analysis of results through the paper were performed by A. K. Gorshenin whose research was supported by the Russian Science Foundation (project 18-71-00156). Machine learning algorithms for imputation of missing values were implemented by BSc student O. P. Martynov.

References

- Gorshenin, A. K., and V. Yu. Korolev. 2018. Opredelenie ekstremal'nosti ob'emov osadkov na osnove modifitsirovannogo metoda prevysheniya porogovogo znacheniya [Determining the extremes of precipitation volumes based on a modified "Peaks over Threshold"]. *Informatika i ee Primeniya — Inform. Appl.* 12(4):16–24.
- Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29(5):1189–1232.
- Chen, T., and C. Guestrin. 2016. XGBoost: A scalable tree boosting system. *22nd ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings*. San Francisco, CA. 785–794.
- Mustapha, I. B., and F. Saeed. 2016. Bioactive molecule prediction using extreme gradient boosting. *Molecules* 21(8):983.
- Xia, Y., C. Liu, Y. Li, and N. Liu. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst. Appl.* 78:225–241.
- Chatzis, S. P., V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis. 2018. Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Syst. Appl.* 112:353–371.
- Zhang, D., L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si. 2018. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access* 6:21020–21031.
- Aler, R., I. M. Galvan, J. A. Ruiz-Arias, and C. A. Gueymard. 2017. Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. *Sol. Energy* 150:558–569.
- Torres-Barran, A., A. Alonso, and J. R. Dorronsoro. 2018. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing* 326:151–160.
- Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. 2018. CatBoost: Unbiased boosting with categorical features. *Adv. Neur. In.* 31:6638–6648.
- Ivanov, M. V., L. I. Levitsky, J. A. Bubis, and M. V. Gorshkov. 2019. Scavager: A versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting. *Proteomics* 19(3):1800280.
- Punmiya, R., and S. Choe. 2019. Energy theft detection using gradient boosting theft detector with feature boost engineering-based preprocessing. *IEEE T. Smart Grid* 10(2):2326–2329.
- Korner, P., R. Kronenberg, S. Genzel, and C. Bernhofer. 2018. Introducing Gradient Boosting as a universal gap filling tool for meteorological time series. *Meteorol. Z.* 27(5):369–376.
- Fan, J., X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, and Y. Xiang. 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energ. Convers. Manage.* 164:102–111.
- Christ, M., N. Braun, J. Neuffer, and A. W. Kempa-Liehr. 2018. Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh — a Python package). *Neurocomputing* 307:72–77.
- Huang, J., and C. X. Ling. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE T. Knowl. Data En.* 17(3):299–310.
- Gorshenin, A. K., and V. Yu. Korolev. 2018. Scale mixtures of Frechet distributions as asymptotic approximations of extreme precipitation. *J. Math. Sci.* 234(6):886–903.

Received July 8, 2019

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilova Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991, Russian Federation; agorshenin@frccsc.ru

Martynov Oleg P. (b. 1996) — BSc student, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991, Russian Federation; martynov.oleg.mipt@gmail.com

УПРАВЛЕНИЕ ВЫХОДОМ СТОХАСТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ СИСТЕМЫ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ. III. АНАЛИЗ СВОЙСТВ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ*

А. В. Босов¹, А. И. Стефанович²

Аннотация: Продолжено исследование задачи оптимального управления для диффузионного процесса Ито и линейного управляемого выхода с квадратичным критерием качества. Изучаются свойства оптимального решения, определяемого функцией Беллмана вида $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$, коэффициенты $\beta_t(y)$ и $\gamma_t(y)$ которой описываются линейными уравнениями в частных производных параболического типа. Для данных коэффициентов определяются альтернативные эквивалентные описания в форме стохастических дифференциальных уравнений и теоретико-вероятностного представления их решений, известного как уравнение А. Н. Колмогорова. Показано, что полученное дифференциальное представление эквивалентно интегральной формуле Фейнмана–Каца. В перспективе полученное описание коэффициентов и, как следствие, решение исходной задачи управления могут использоваться для реализации альтернативного численного метода их расчета как результата имитационного моделирования решения стохастического дифференциального уравнения.

Ключевые слова: стохастическое дифференциальное уравнение; оптимальное управление; функция Беллмана; линейные уравнения параболического типа; уравнение А. Н. Колмогорова; формула Фейнмана–Каца

DOI: 10.14357/19922264190307

1 Введение

В работах [1, 2] получены оптимальное и приближенное численные решения в задаче управления линейным выходом стохастической дифференциальной системы по квадратичному критерию качества, краткая выдержка полученного результата приведена в разд. 2 статьи. Оптимизируемая динамическая система описывается двумя уравнениями: нелинейным стохастическим дифференциальным уравнением Ито для состояния и линейным уравнением для управляемого выхода, цель оптимизации — квадратичным функционалом. Аналитическое решение определяет функция Беллмана вида

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

коэффициенты α_t , $\beta_t(y)$ и $\gamma_t(y)$ которой представлены решениями определенных дифференциальных уравнений (обыкновенного Риккати и двух в частных производных линейных параболического типа). Приближенное решение осуществляется

численным методом расчета указанных коэффициентов. При этом применение сеточных методов для параболического уравнения осложняется тем обстоятельством, что метод динамического программирования дает только начальные и не дает граничных условий для уравнений, определяющих $\beta_t(y)$, $\gamma_t(y)$, точнее, отсутствуют основания для их обоснованного выбора. Из-за этого граничные условия выбираются довольно волюнтаристски, обсуждается устойчивость решения в отношении этих граничных условий. Так, при вычислении $\beta_t(y)$ рассматривались и сравнивались два традиционных варианта граничных условий — в задаче Дирихле:

$$\beta_t(y) = 0,$$

т. е. условие поглощения, и в задаче Неймана:

$$\frac{\partial \beta_t(y)}{\partial y} = 0,$$

т. е. условие отражения, явные и неявные схемы с разными вариантами дискретизации.

* Работа выполнена при частичной поддержке РФФИ (проект 19-07-00187-А).

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

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

Получение удовлетворительного результата не сделало менее актуальной задачу поиска альтернативного метода для приближенного вычисления коэффициентов функции Беллмана в рассматриваемой задаче, которой посвящена данная статья (здесь и далее под коэффициентами подразумевается пара $\beta_t(y)$, $\gamma_t(y)$, так как эффективных вариантов решения уравнения Риккати для α_t известно предостаточно). Поскольку обсуждаемые коэффициенты описываются линейными уравнениями в частных производных второго порядка параболического типа, то перспективным представлялось опираться в исследовании этой задачи на известные факты, ассоциирующие такие уравнения с уравнениями стохастическими. Видимо, самая известная ассоциация такого рода — это уравнение А. Н. Колмогорова [3], которое и было использовано в качестве инструмента для реализации альтернативного метода приближенного вычисления коэффициентов $\beta_t(y)$ и $\gamma_t(y)$. Точнее, использовалось утверждение, которое определяет связь между решением задачи Коши для уравнения А. Н. Колмогорова, с одной стороны, и теоретико-вероятностным представлением ее терминального условия, с другой. Адаптированная для рассматриваемой задачи оптимизации формулировка этого результата приведена в разд. 3 статьи, а ее применение к уравнениям для $\beta_t(y)$ и $\gamma_t(y)$ — в разд. 4.

Связь дифференциальных уравнений в частных производных и стохастических изучается давно, существенных результатов в этой области получено множество. Наиболее развита тематика так называемых обратных стохастических уравнений, начало которой положено в работах [4, 5], некоторые итоги имеются в [6, 7], а актуальные исследования проводятся и в настоящее время. Надо отметить, что и приложении этой тематики именно к численному решению параболических уравнений также успешно реализовано [8]. Так что предложенный в статье подход можно считать исследованным и апробированным. Другое дело, что предложенный в статье порядок изложения результата на основе уравнения А. Н. Колмогорова позволяет не только получить альтернативную численную процедуру, но качественно описать свойства рассматриваемых коэффициентов, так как полученное эквивалентное описание коэффициентов — это стохастические дифференциальные уравнения и соотношение, определяющее связь между математическим ожиданием их решения и искомыми коэффициентами. Соответственно, такое представление служит основой для численного решения рассматриваемых параболических уравнений вероятностными методами, т. е. методом имитационного моделирования решения стохастического дифференциаль-

ного уравнения. Кроме того, полученное дифференциальное представление имеет известный интегральный аналог — формулу Фейнмана—Каца [9], является частным случаем классического варианта этой формулы. Обсуждению данного факта посвящен разд. 5.

2 Оптимальное управление Выходом

Аналитическое решение задачи оптимизации, исследуемое далее, получено в [1] и кратко описывается следующими положениями.

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

$$dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t, \quad y_0 = Y, \quad (1)$$

где v_t — стандартный винеровский процесс; Y — случайная величина с конечным вторым моментом, функции A_t ; Σ_t удовлетворяют условиям Ито, обеспечивающим существование единственного решения [10].

С состоянием y_t линейно связан выход z_t :

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z, \quad (2)$$

где w_t — не зависящий от v_t , Y и Z стандартный винеровский процесс; Z — случайная величина с конечным вторым моментом; u_t — допустимое управление. Функции a_t , b_t , c_t и σ_t предполагаются ограниченными, процесс управления — допустимым неупреждающим [10], что обеспечивает существование решения уравнения (2) для любого допустимого управления.

Используется целевой функционал следующего вида:

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

$$U_0^T = \{u_t, 0 \leq t \leq T\}, \quad (3)$$

где S_t , G_t и H_t — неотрицательные ограниченные функции.

Оптимальное управление, доставляющее минимум функционалу $J(U_0^T)$,

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

решение задачи поиска u_t^* обеспечивается представлением функции Беллмана $V_t(y, z)$ в виде:

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

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

$$\begin{aligned} \frac{\partial \alpha_t}{\partial t} + 2\alpha_t \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) + \\ + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) g_t^2 + G_t - \\ - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t^2 = 0, \quad \alpha_T = S_T g_T^2 + G_T; \end{aligned} \quad (4)$$

$$\begin{aligned} \frac{\partial \beta_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t(y)}{\partial y^2} + A_t(y) \frac{\partial \beta_t(y)}{\partial y} + \\ + 2\alpha_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ + \beta_t(y) \left(b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t \right) - \\ - 2 \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t y - \\ - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t \beta_t(y) = 0, \\ \beta_T(y) = -2S_T s_T g_T y; \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{\partial \gamma_t(y)}{\partial t} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \gamma_t(y)}{\partial y^2} + \sigma_t^2 \alpha_t + A_t(y) \frac{\partial \gamma_t(y)}{\partial y} + \\ + \beta_t(y) \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 - \\ - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2(y) = 0, \\ \gamma_T(y) = S_T s_T^2 y^2. \end{aligned} \quad (6)$$

Уравнение (4) является уравнением Риккати и имеет единственное неотрицательное решение для всех $0 \leq t \leq T$, так как предполагается $S_t h_t^2 + H_t > 0$. Уравнения (5) и (6) представляют собой линейные уравнения в частных производных второго порядка, относятся к параболическому типу, поскольку $\Sigma_t^2(y) > 0$. Численное решение этих уравнений традиционными сеточными методами [2] обеспечивает возможность получения приближенного решения рассматриваемой задачи оптимизации. Однако с реализацией такого численного решения возникают проблемы как вычислительного характера, так и алгоритмического, связанные с неограниченной областью допустимых значений аргумента y и, как следствие, отсутствием граничного условия, необходимого сеточному алгоритму.

3 Уравнение А. Н. Колмогорова

Помимо сложностей реализации приближенного решения полученное выражение для функции

Беллмана представляется целесообразным сопроводить разъяснениями в отношении физического (содержательного) смысла коэффициентов $\beta_t(y)$ и $\gamma_t(y)$, т. е. ответить на вопрос, в чем смысл уравнений (5) и (6), а также предложить эффективный способ их приближенного решения. Отметим, что основную ценность составляет, конечно, определение коэффициента $\beta_t(y)$, поскольку именно он требуется для реализации оптимального управления. Но и коэффициент $\gamma_t(y)$ заслуживает внимания в целях анализа качества управления, динамики изменения целевой функции $J(U_0^t)$. В качестве такого способа видится использование имитационного моделирования для стохастического дифференциального уравнения, ассоциированного с решаемым параболическим уравнением (5) или (6). Для реализации такой ассоциации будет использован следующий результат [3].

Рассмотрим процесс $X_\tau \in \mathbb{R}^n$, описываемый стохастической дифференциальной системой

$$dX_\tau = m_\tau(X_\tau) dt + \sigma_\tau(X_\tau) dW_\tau,$$

где W_τ — винеровский процесс подходящей размерности. Отметим, что здесь в постановке требуется рассматривать многомерный случай, с учетом скалярного y_t в (1) можно ограничиться $X_\tau \in \mathbb{R}^2$, что для дальнейшего уже несущественно. Кроме того, отметим, что формулируется результат в существенно более частном виде, чем дано в [3]. Результат общего вида не ограничивается случаем непрерывной гауссовской меры dW_τ и приводит к интегродифференциальному уравнению, здесь достаточно частного случая дифференциального уравнения. Решение X_τ будем рассматривать на интервале $t \leq \tau \leq T$ с начальным условием $X_t = X$. Подчеркивая это, обозначим такое решение $X_\tau(t, X)$. Далее рассмотрим его в правой точке горизонта управления $X_T(t, X)$, интерпретируя как функцию переменных (t, X) . Пусть $f(X)$ — некоторая дважды непрерывно дифференцируемая и вместе со своими частными производными первого и второго порядка равномерно ограниченная функция, $F(t, X) = E\{f(X_T(t, X))\}$. Тогда функция $F(t, X)$ дважды непрерывно дифференцируема по X , дифференцируема по t и удовлетворяет уравнению:

$$\begin{aligned} \frac{\partial F(t, x)}{\partial t} + \sum_{i=1}^n m_{t_i} \frac{\partial F(t, X)}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^n \sigma_{t_{ij}}^2 \frac{\partial^2 F(t, X)}{\partial X_i \partial X_j} = 0, \\ \lim_{t \rightarrow T} F(t, X) = f(X). \end{aligned} \quad (7)$$

Уравнение (7) известно как уравнение А. Н. Колмогорова, а данное утверждение определяет связь

между решением задачи Коши с терминальным условием для уравнения с частными производными (7), с одной стороны, и теоретико-вероятностным представлением $F(t, X) = E\{f(X_T(t, X))\}$ этого решения, с другой. Кроме того, этот результат может служить основой для приближенного решения (7) методом имитационного моделирования (моделирование решения $X_T(t, X)$ стохастического дифференциального уравнения и метод Монте-Карло для оценки $F(t, X)$).

Уравнение (7) имеет очевидное сходство с уравнениями (5) и (6). Соответственно, представляется обоснованным, используя (7), прояснить смысл решений (5) и (6) и предложить метод их приближенного решения.

4 Стохастическое представление коэффициентов функции Беллмана

Требуемый результат получается в итоге несложных манипуляций, приводящих уравнения (5) и (6) к виду (7) уравнения А. Н. Колмогорова. Сначала приведем выкладки для $\beta_t(y)$, для чего уравнение (5) запишем в виде:

$$\frac{\partial \beta_t(y)}{\partial t} + A_t(y) \frac{\partial \beta_t(y)}{\partial y} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t(y)}{\partial y^2} + M_t y + N_t \beta_t(y) = 0, \quad \beta_T(y) = -2S_T S_T g_T y, \quad (8)$$

где

$$M_t = 2 \left(\alpha_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) - \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t \right);$$

$$N_t = b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t.$$

Введем новую функцию $\beta_t^{(0)}(y)$ следующим равенством:

$$\beta_t(y) = \beta_t^{(0)}(y) \exp \left\{ - \int_0^t N_\tau d\tau \right\} = \beta_t^{(0)} \exp \{-i_t\}, \quad (9)$$

где

$$i_t = \int_0^t N_\tau d\tau.$$

Замена переменных (корректная, с учетом сделанных в рамках модели (1)–(3) предположений,

гарантирующих ограниченность подынтегрального выражения) в (8) с учетом

$$\frac{\partial i_t}{\partial t} = N_t;$$

$$\frac{\partial \beta_t(y)}{\partial t} = \frac{\beta_t^{(0)}(y)}{\partial t} \exp \{-i_t\} - \beta_t^{(0)} \exp \{-i_t\} N_t;$$

$$\frac{\partial \beta_t(y)}{\partial y} = \frac{\beta_t^{(0)}(y)}{\partial y} \exp \{-i_t\};$$

$$\frac{\partial^2 \beta_t(y)}{\partial y^2} = \frac{\partial^2 \beta_t^{(0)}(y)}{\partial y^2} \exp \{-i_t\}$$

и дополнительного обозначения

$$I_t = \exp \{-i_t\}$$

дает:

$$\frac{\partial \beta_t^{(0)}(y)}{\partial t} I_t + A_t(y) \frac{\partial \beta_t^{(0)}(y)}{\partial y} I_t + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t^{(0)}(y)}{\partial y^2} I_t + M_t y + N_t \beta_t^{(0)}(y) I_t - \beta_t^{(0)}(y) I_t N_t = 0,$$

откуда с учетом $I_t^{-1} = \exp \{i_t\}$ получаем:

$$\frac{\partial \beta_t^{(0)}(y)}{\partial t} + A_t(y) \frac{\partial \beta_t^{(0)}(y)}{\partial y} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \beta_t^{(0)}(y)}{\partial y^2} + M_t I_t^{-1} y = 0, \quad \beta_T^{(0)}(y) = -2S_T S_T g_T I_T^{-1} y. \quad (10)$$

Далее введем еще одну функцию $\beta_t^{(1)}(y, y^{(1)})$, обозначив:

$$\beta_t^{(1)}(y, y^{(1)}) = \beta_t^{(0)}(y) + y^{(1)}. \quad (11)$$

Это функция времени и двух фазовых переменных, «старой» y и «новой/дополнительной» $y^{(1)}$, частные производные которой связаны с производными $\beta_t^{(0)}(y)$ следующим образом:

– производная по времени

$$\frac{\partial \beta_t^{(1)}(y, y^{(1)})}{\partial t} = \frac{\partial \beta_t^{(0)}(y)}{\partial t};$$

– градиент

$$\Delta \beta_t^{(1)} = \left(\frac{\partial \beta_t^{(1)}(y, y^{(1)})}{\partial y} \right) = \left(\frac{\partial \beta_t^{(0)}(y)}{\partial y} \right);$$

– гессиан

$$\nabla\beta_t^{(1)} = \begin{pmatrix} \frac{\partial^2\beta_t^{(1)}(y, y^{(1)})}{(\partial y)^2} & \frac{\partial^2\beta_t^{(1)}(y, y^{(1)})}{\partial y \partial y^{(1)}} \\ \frac{\partial^2\beta_t^{(1)}(y, y^{(1)})}{\partial y^{(1)} \partial y} & \frac{\partial^2\beta_t^{(1)}(y, y^{(1)})}{\partial (y^{(1)})^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2\beta_t^{(0)}(y)}{(\partial y)^2} & 0 \\ 0 & 0 \end{pmatrix}.$$

С учетом сделанных замечаний уравнение (10) записывается в виде:

$$\frac{\partial\beta_t^{(1)}(y, y^{(1)})}{\partial t} + \begin{pmatrix} A_t(y) \\ M_t I_t^{-1} y \end{pmatrix}^T \Delta\beta_t^{(1)} + \frac{1}{2} \begin{pmatrix} \Sigma_t^2(y) \\ 0 \end{pmatrix}^T \nabla\beta_t^{(1)} \begin{pmatrix} \Sigma_t^2(y) \\ 0 \end{pmatrix} = 0 \quad (12)$$

с терминальным условием

$$\beta_T^{(1)}(y, y^{(1)}) = -2S_T S_T g_T I_T^{-1} y + y^{(1)}.$$

Полученное уравнение (12) – это уравнение А. Н. Колмогорова (7), которое порождается случайным процессом

$$X_\tau = \begin{pmatrix} x_\tau \\ y_\tau^{(1)} \end{pmatrix}, \quad t \leq \tau \leq T,$$

заданным следующей стохастической дифференциальной системой:

$$\left. \begin{aligned} dx_\tau &= A_\tau(x_\tau) d\tau + \Sigma_\tau(x_\tau) dv_\tau, \\ t &\leq \tau \leq T, \quad x_t = y_t; \\ dy_\tau^{(1)} &= M_\tau I_\tau^{-1} x_\tau d\tau, \quad y_t^{(1)} = \int_0^t M_\tau I_\tau^{-1} y_\tau d\tau. \end{aligned} \right\} \quad (13)$$

Полученная система очевидным образом эквивалентна исходной стохастической системе: ее состояние y_t описывается таким же уравнением, что и первое уравнение (13) для x_τ , где в записи использован тот же самый винеровский процесс v_τ , плюс представляет некоторый среднеквадратичный (с.к.) интеграл этого состояния (второе уравнение).

Теперь, полагая в уравнении (7)

$$X = \begin{pmatrix} y \\ y^{(1)} \end{pmatrix},$$

используя вместо обозначения $X_T(t, X)$ пару $x_T(t, y, y^{(1)})$ и $y_T^{(1)}(t, y, y^{(1)})$ и подставляя

$$F(t, X) = \beta_t^{(1)}(y, y^{(1)});$$

$$\begin{aligned} f(x_T(t, X)) &= \\ &= \beta_T^{(1)}(x_T(t, y, y^{(1)}), y_T^{(1)}(t, y, y^{(1)})) = \\ &= -2S_T S_T g_T I_T^{-1} x_T(t, y, y^{(1)}) + y_T^{(1)}(t, y, y^{(1)}), \end{aligned}$$

получаем:

$$\begin{aligned} F(t, X) &= E\{f(X_T(t, X))\} \Rightarrow \beta_t^{(1)}(y, y^{(1)}) = \\ &= E\{-2S_T S_T g_T I_T^{-1} y_T(t, y, y^{(1)}) + y_T^{(1)}(t, y, y^{(1)})\}. \end{aligned}$$

Далее заметим, что согласно (13) величина $x_T(t, y, y^{(1)})$ не зависит от $y^{(1)}$, поэтому далее ее следует обозначать $x_T(t, y)$. Значение $y_T^{(1)}$ можно представить в виде:

$$y_T^{(1)} = y_t^{(1)} + \int_t^T M_\tau I_\tau^{-1} x_\tau d\tau.$$

И далее с учетом введенных выше обозначений (9) и (11):

$$\begin{aligned} \beta_t(y) &= I_t \beta_t^{(0)} = \\ &= I_t E\{-2S_T S_T g_T I_T^{-1} x_T(t, y) + y_T^{(1)}(t, y, 0)\}, \quad (14) \end{aligned}$$

где

$$\begin{aligned} \beta_t^{(0)}(y) &= \beta_t^{(1)}(y, y^{(1)}) - y^{(1)} = \\ &= E\{-2S_T S_T g_T I_T^{-1} x_T(t, y) + y_T^{(1)}(t, y, y^{(1)})\} - \\ &- y^{(1)} = E\{-2S_T S_T g_T I_T^{-1} x_T(t, y) + y_T^{(1)}(t, y, 0)\}. \end{aligned}$$

В последних выражениях учитывается начальное условие, с которым решается уравнение для $y_\tau^{(1)}$ из (13): для переменной $\beta_t^{(1)}$ интегрирование начинается с величины $y_t^{(1)}$ – интеграла, рассчитанного для $0 \leq \tau \leq t$, для переменной $\beta_t^{(0)}$ это слагаемое обнуляется.

Полученное выражение (14) является, во-первых, некоторым объяснением физического смысла коэффициента $\beta_t(y)$, определяющего оптимальное управление в рассматриваемой задаче, а именно: $\beta_t(y)$, аддитивно входящее в оптимальное управление u_t^* , представляет собой линейную комбинацию всех прогнозов $x_\tau(t, y) = E\{y_\tau | \mathcal{F}_t^y\}$ для $t \leq \tau \leq T$, т.е. состояний от текущего момента t вплоть до горизонта управления T . Здесь \mathcal{F}_t^y – σ -алгебра, порожденная значениями y_τ до момента t включительно, с учетом марковского свойства y_t из (1) формально выражает начальное условие $x_t = y_t$ в системе (13).

Отметим, что аналогичное свойство обсуждается в работе [11], где рассмотрена такая же точно

задача, но в дискретном времени. Второе обстоятельство в связи с выражением (14) — это возможность решения уравнения (5) для расчета $\beta_t(y)$ без использования каких-либо (например, сеточных) методов решения уравнений в частных производных, а с использованием компьютерного имитационного моделирования, т. е. с заменой расчетов аппроксимаций частных производных моделированием пучка траекторий процесса x_τ , описываемого уравнением (13), или, что то же самое, процесса y_t , описываемого уравнением (1), и набора дополнительных переменных, задающих фигурирующие в (14) интегралы (с.к.-интеграл от x_τ из (14), задаваемый процессом $y_\tau^{(1)}$, и римановский интеграл I_t). Формальный алгоритм, реализующий эту идею, будет представлен и исследован позже.

Далее, опираясь на полученный результат, можно записать аналогичные выражения и для коэффициента $\gamma_t(y)$, т. е. закрыть уже вопрос полностью, вычислив функцию Беллмана.

Итак, рассмотрим уравнение (6), записав его для удобства в виде:

$$\frac{\partial \gamma_t(y)}{\partial t} + A_t(y) \frac{\partial \gamma_t(y)}{\partial y} + \frac{1}{2} \Sigma_t^2(y) \frac{\partial^2 \gamma_t(y)}{\partial y^2} + L_t(y) = 0, \\ \gamma_T(y) = S_T s_T^2 y^2, \quad (15)$$

где

$$L_t(y) = \beta_t(y) \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 - \\ - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2(y).$$

Здесь задача немного проще, чем в (8), так как нет слагаемого с $\gamma_t(y)$, поэтому преобразования ограничатся одним шагом с новой функцией, как в (11): $\gamma_t^{(1)}(y, y^{(2)}) = \gamma_t(y) + y^{(2)}$ и теми же действиями, которые приведут к уравнению А. Н. Колмогорова (7), порождаемому случайным процессом

$$X_\tau = \begin{pmatrix} x_\tau \\ y_\tau^{(2)} \end{pmatrix}, \quad t \leq \tau \leq T,$$

заданным стохастической дифференциальной системой:

$$\left. \begin{aligned} dx_\tau &= A_\tau(x_\tau) d\tau + \Sigma_\tau(x_\tau) dv_\tau, \\ t &\leq \tau \leq T, \quad x_t = y_t; \\ dy_\tau^{(2)} &= L_\tau(x_\tau) d\tau, \quad y_t^{(2)} = \int_0^t L_\tau(y_\tau) d\tau, \end{aligned} \right\} \quad (16)$$

причем процесс x_τ здесь тот же, что и в (13), так что (13) и (16) можно рассматривать как одну систему для переменной

$$X_\tau = \left(x_\tau, y_\tau^{(1)}, y_\tau^{(2)} \right)^T.$$

Итогом станет следующее представление для $\gamma_t(y)$:

$$\gamma_t(y) = E \left\{ S_T s_T^2 x_T^2(t, y) + y_T^{(2)}(t, y, 0) \right\}. \quad (17)$$

Полученное выражение играет ту же роль, что и (14) для коэффициента $\beta_t(y)$. Отметим, что согласно (17) прогнозированию подлежит еще и второй момент процесса x_τ , описываемого уравнением (13), или процесса y_t . Кроме того, обратим внимание на возможность вычисления $\gamma_t(y)$ вместе с $\beta_t(y)$ по тому же пучку траекторий (с дополнительной переменной $y_\tau^{(2)}$, представляющей с.к.-интеграл от x_τ) состояния y_t . Это замечание завершает исследование оптимального управления, так как наряду с возможностью вычисления самого управления имеется и возможность определения его качества:

$$\min_{U_0^T} J(U_0^T) = J\left((U^*)_0^T\right) = E\{V_0(Y, Z)\} = \\ = E\{\alpha_0 Z^2 + \beta_0(Y)Z + \gamma_0(y)\},$$

которое после решения уравнений (4)–(6) методом имитационного моделирования естественным будет определить усреднением по уже имеющемуся пучку.

5 Представление Фейнмана–Каца коэффициентов функции Беллмана

В этом разделе приводится дополнение к основным полученным результатам — соотношениям (14) и (17), а именно: их несложные преобразования приведут к формуле Фейнмана–Каца [9], которая представляет самый известный результат ассоциации параболического дифференциального уравнения в частных производных (линейного неоднородного второго порядка) со стохастической системой. С этой целью заметим, во-первых, что, доопределив случайный процесс x_τ из (13) или (16) равенством $x_\tau = y_\tau$ на интервале $0 \leq \tau \leq t$, получим стохастически эквивалентные процессы y_t и x_t на интервале $0 \leq t \leq T$. Подчеркивая это, в (13) и (16) в диффузионном слагаемом использовано то же обозначение для винеровского процесса v_t , что и в (1), хотя

они должны быть определены на разных пространствах. Это позволяет использовать в (14) и (17) вместо $x_T(t, y)$ переменную $y_T(t, y)$ и далее в выражениях использовать только y_T , заменяя начальное условие $x_\tau = y_t$ на условное математическое ожидание относительно \mathcal{F}_t^y . Во-вторых, вернем к исходным обозначениям следующие элементы, фигурирующие в (14) и (17):

$$y_T^{(1)}(t, y, 0) = \int_t^T M_\tau I_\tau^{-1} y_\tau d\tau;$$

$$y_T^{(2)}(t, y, 0) = \int_t^T L_\tau(y_\tau) d\tau;$$

$$I_t I_\tau^{-1} = \exp \left\{ - \int_0^t N_s ds \right\} \exp \left\{ \int_0^\tau N_s ds \right\} = \exp \left\{ \int_t^\tau N_s ds \right\};$$

$$I_t I_T^{-1} = \exp \left\{ - \int_0^t N_\tau d\tau \right\} \exp \left\{ \int_0^T N_\tau d\tau \right\} = \exp \left\{ \int_t^T N_\tau d\tau \right\}.$$

Подставляя теперь все вместе в (14) и (17), получаем:

$$\beta_t(y) = E \left\{ -2S_T s_T g_T \exp \left\{ \int_t^T N_\tau d\tau \right\} y_T + \int_t^T \exp \left\{ \int_t^\tau N_s ds \right\} M_\tau y_\tau d\tau \middle| \mathcal{F}_t^y \right\}; \quad (18)$$

$$\gamma_t(y) = E \left\{ S_T s_T^2 y_T^2 + \int_t^T L_\tau(y_\tau) d\tau \middle| \mathcal{F}_t^y \right\}.$$

Каждое из равенств (18) — это формула Фейнмана—Каца [9], записанная с учетом конкретного вида слагаемых и терминальных условий в (8) и (15) соответственно. Практическую же ценность представляет не то, в какой форме записан результат (18) или (14), (17), а дифференциальная система (13), (16), служащая основой для приближенного расчета коэффициентов $\beta_t(y)$ и $\gamma_t(y)$.

6 Заключение

В статье продолжено исследование задачи оптимизации линейного выхода нелинейной дифферен-

циальной системы по квадратичному критерию, начатое в [1]. Свойства оптимального решения формируются коэффициентами α_t , $\beta_t(y)$ и $\gamma_t(y)$, определенными выражением для функции Беллмана $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$. И если коэффициент α_t , описываемый уравнением Риккати, не предоставляет пространства для исследования, то коэффициенты $\beta_t(y)$ и $\gamma_t(y)$, описываемые схожими по структуре линейными неоднородными уравнениями в частных производных второго порядка параболического типа, обнаруживают любопытные свойства. Выражением этих свойств является формула Фейнмана—Каца (18), показывающая связь исследуемых коэффициентов и моментных характеристик некоторой дифференциальной системы. Сама полученная система (13), (16) составляет основной прикладной результат работы, так как позволяет предложить альтернативный (по отношению к традиционному сеточному методу, использованному для расчетов в [2]) метод приближенного решения задачи оптимизации в исходной модели (1)–(3). Формальное изложение этого метода и его практическая апробация составляют дальнейшую задачу исследования, планируемую к последующей публикации.

Литература

1. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. I. Оптимальное решение методом динамического программирования // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 99–106. doi: 10.14357/19922264180314.
2. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. II. Численное решение уравнений динамического программирования // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 9–15. doi: 10.14357/19922264190102.
3. Гухман И. И., Скороход А. В. Теория случайных процессов. Т. III. — М.: Наука, 1975. 496 с.
4. Bismut J.-M. Conjugate convex functions in optimal stochastic control // J. Math. Anal. Appl., 1973. Vol. 44. No. 2. P. 384–404. doi: 10.1016/0022-247X(73)90066-8.
5. Pardoux E., Peng S.G. Adapted solution of a backward stochastic differential equation // Syst. Control Lett., 1990. Vol. 14. No. 1. P. 55–61. doi: 10.1016/0167-6911(90)90082-6.
6. Delong L. Backward stochastic differential equations with jumps and their actuarial and financial applications. — New York, NY, USA: Springer-Verlag, 2013. 288 p. doi: 10.1007/978-1-4471-5331-3.

7. Touzi N. Optimal stochastic control, stochastic target problems, and backward SDEs. — New York, NY, USA: Springer-Verlag, 2013. 214 p. doi: 10.1007/978-1-4614-4286-8.
8. Fahim A., Touzi N., Warin X. A probabilistic numerical method for fully nonlinear parabolic pdes // Ann. Appl. Probab., 2011. Vol. 21. No. 4. P. 1322–1364. doi: 10.1214/10-AAP723.
9. Øksendal B. Stochastic Differential equations. An introduction with applications. — New York, NY, USA: Springer-Verlag, 2003. 379 p. doi: 10.1007/978-3-642-14394-6.
10. Флеминг У., Ришел Р. Оптимальное управление детерминированными и стохастическими системами / Пер. с англ. — М.: Мир, 1978. 316 с. (Fleming W. H., Rishel R. W. Deterministic and stochastic optimal control. — New York, NY, USA: Springer-Verlag, 1975. 222 p.)
11. Босов А. В. Управление линейным выходом дискретной стохастической системы по квадратичному критерию // Изв. РАН. Теория и системы управления, 2016. № 3. С. 19–35. doi: 10.1134/S1064230716030060.

Поступила в редакцию 21.02.19

STOCHASTIC DIFFERENTIAL SYSTEM OUTPUT CONTROL BY THE QUADRATIC CRITERION. III. OPTIMAL CONTROL PROPERTIES ANALYSIS

A. V. Bosov and A. I. Stefanovich

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

Abstract: The investigation of the optimal control problem for the Ito diffusion process and linear controlled output with a quadratic quality criterion is continued. The properties of the optimal solution defined by the Bellman function of the form $V_t(y, z) = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y)$, whose coefficients $\beta_t(y)$ and $\gamma_t(y)$ are described by linear parabolic equations, are studied. For these coefficients, alternative equivalent descriptions are defined in the form of stochastic differential equations and a theoretical-to-probabilistic representation of their solutions, known as the Kolmogorov equation. It is shown that the obtained differential representation is equivalent to the Feynman–Kac integral formula. In the future, the obtained description of the coefficients and, as a result, the solutions of the original control problem can be used to implement an alternative numerical method for calculating them as a result of computer simulation of the solution of a stochastic differential equation.

Keywords: stochastic differential equation; optimal control; Bellman function; linear differential equations of parabolic type; Kolmogorov equation; Feynman–Kac formula

DOI: 10.14357/19922264190307

Acknowledgments

This work was partially supported by the Russian Foundation for Basic Research (grant 19-07-00187-A).

References

1. Bosov, A. V., and A. I. Stefanovich. 2018. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadraticnomu kriteriyu [Stochastic differential system output control by the quadratic criterion. I. Dynamic programming optimal solution]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):99–106. doi: 10.14357/19922264180314.
2. Bosov, A. V., and A. I. Stefanovich. 2019. Upravlenie vykhodom stokhasticheskoy differentsial'noy sistemy po kvadraticnomu kriteriyu [Stochastic differential system output control by the quadratic criterion. II. Dynamic programming equations numerical solution]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):9-15. doi: 10.14357/19922264190102.
3. Gihman, I. I., and A. V. Skorohod. 2012. *The theory of stochastic processes III*. New York, NY: Springer-Verlag. 388 p.
4. Bismut, J.-M. 1973. Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.* 44(2):384–404. doi: 10.1016/0022-247X(73)90066-8.
5. Pardoux, E., and S. G. Peng. 1990. Adapted solution of a backward stochastic differential equation. *Syst. Control Lett.* 14(1):55–61. doi: 10.1016/0167-6911(90)90082-6.
6. Delong, L. 2013. *Backward stochastic differential equations with jumps and their actuarial and financial applications*.

- New York, NY: Springer-Verlag. 288 p. doi: 10.1007/978-1-4471-5331-3.
7. Touzi, N. 2013. *Optimal stochastic control, stochastic target problems, and backward SDEs*. New York, NY: Springer-Verlag. 214 p. doi: 10.1007/978-1-4614-4286-8.
 8. Fahim, A., N. Touzi, and X. Warin. 2011. A probabilistic numerical method for fully nonlinear parabolic pdes. *Ann. Appl. Probab.* 21(4):1322–1364. doi: 10.1214/10-AAP723.
 9. Øksendal, B. 2003. *Stochastic differential equations. An introduction with applications*. New York, NY: Springer-Verlag. 379 p. doi: 10.1007/978-3-642-14394-6.
 10. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p.
 11. Bosov, A. V. 2016. Discrete stochastic system linear output control with respect to a quadratic criterion. *J. Comput. Syst. Sc. Int.* 55(3):349–364.

Received February 21, 2019

Contributors

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

Stefanovich Alexey I. (b. 1983) — junior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AStefanovich@frcsc.ru

О РЕШЕНИИ ПРОБЛЕМЫ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ЗАПАСОМ ДИСКРЕТНОГО ПРОДУКТА В СТОХАСТИЧЕСКОЙ МОДЕЛИ РЕГЕНЕРАЦИИ С НЕПРЕРЫВНО ПРОИСХОДЯЩИМ ПОТРЕБЛЕНИЕМ

П. В. Шнурков¹, Н. А. Вахтанов²

Аннотация: Работа представляет собой вторую и завершающую часть исследования проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации. Основным содержанием работы является вывод аналитических представлений для математического ожидания приращения функционала прибыли, полученной за период регенерации. При этом указанные математические ожидания находятся при всех возможных в данной задаче условиях на решения, принимаемые на периоде регенерации. Полученные аналитические представления позволяют явно выразить стационарный стоимостный показатель эффективности управления, который был определен в первой части исследования. Таким образом, появляется возможность численно решить задачу оптимального управления запасом в рассматриваемой модели.

Ключевые слова: управление запасом дискретного продукта; управляемый регенерирующий процесс; стационарный стоимостный показатель эффективности управления

DOI: 10.14357/19922264190308

1 Введение

В первой части настоящего исследования [1] была построена стохастическая модель управления запасом дискретного продукта в форме регенерирующего процесса. Получено общее представление для стационарного стоимостного показателя эффективности управления. Установлено, что этот показатель по форме представляет собой дробно-линейный интегральный функционал, заданный на множестве дискретных вероятностных распределений, каждое из которых определяет стратегию управления запасом или, что то же самое, стратегию управления введенным регенерирующим процессом. Задача оптимального управления по отношению к указанному функционалу решается на основе теоремы о безусловном экстремуме дробно-линейного интегрального функционала [2, 3]. Доказано, что оптимальное управление является детерминированным и определяется точкой глобального экстремума некоторой функции от дискретного аргумента, который представляет собой параметр управления. Таким образом, проблема оптимального управления сведена к задаче поиска глобального экстремума этой функции.

Вторая часть проведенного исследования посвящена нахождению явного аналитического представления для указанной целевой функции, которая формально есть не что иное, как стационарный сто-

имостный показатель эффективности, т. е. средняя удельная прибыль, при детерминированном управлении.

Чтобы найти аналитическое представление средней удельной прибыли как функции от параметра управления, необходимо провести глубокий анализ процедуры формирования доходов и затрат, связанных с эволюцией исследуемой системы. Такая процедура существенно зависит от значения параметра управления. Проведя соответствующий анализ для всех допустимых значений параметра управления, можно получить явное аналитическое представление для целевой функции средней удельной прибыли. Именно такое представление и получено в данной работе. Окончательное решение задачи оптимального управления, т. е. нахождение значения параметра управления, которое доставляет глобальный максимум целевой функции, может быть выполнено только численными методами для заданного набора исходных параметров математической модели.

2 Вспомогательные теоретические результаты

Прежде всего, приведем обобщение известной формулы для математического ожидания неотрицательной случайной величины.

¹Национальный исследовательский университет «Высшая школа экономики», pshnurkov@hse.ru

²Национальный исследовательский университет «Высшая школа экономики», Vakhtanov1997@mail.ru

Пусть (Ω, \mathcal{A}, P) — исходное вероятностное пространство. Предположим, что на этом пространстве задана неотрицательная случайная величина $\eta = \eta(\omega)$, $\omega \in \Omega$, и произвольное событие $A \in \mathcal{A}$. Обозначим через $F(x) = P(\eta < x)$, $0 < x < \infty$, функцию распределения случайной величины η .

Известен следующий результат о представлении математического ожидания неотрицательной случайной величины η , принадлежащий Б. В. Гнеденко [4]:

$$E\eta = \int_0^\infty x dF(x) = \int_0^\infty P(\eta \geq x) dx = \int_0^\infty (1 - F(x)) dx. \quad (1)$$

Рассмотрим теперь совместное распределение $F(x; A) = P(\eta < x; A)$ и соответствующее математическое ожидание

$$E(\eta; A) = \int_0^\infty x dF(x; A).$$

П. В. Шнурковым доказано следующее утверждение о представлении математического ожидания по совместному распределению неотрицательной случайной величины η и события A :

$$E(\eta; A) = \int_0^\infty P(\eta \geq x; A) dx. \quad (2)$$

Соотношение (2) является обобщением формулы (1), которая непосредственно следует из (2) при $A = \Omega$.

Общая формула (2) будет лежать в основе доказательства утверждения, которое в данной работе будет называться теоремой 1. Это утверждение, в свою очередь, будет использоваться для получения явных представлений для различных вспомогательных вероятностных характеристик, связанных с целевым показателем управления.

Перейдем к формулировке и доказательству указанного утверждения. При этом будем предполагать, что все вводимые стохастические объекты определены на исходном вероятностном пространстве (Ω, \mathcal{A}, P) , которое описывает проводимый случайный эксперимент.

Теорема 1. *Рассмотрим простейший поток однородных событий с параметром λ . Обозначим через ξ_1, \dots, ξ_s интервалы между последовательными событиями этого потока, $\zeta_s = \sum_{i=1}^s \xi_i$ — момент наступления s -го события, $s = 1, 2, \dots$*

Пусть h_r — положительная случайная величина, имеющая заданную функцию распределения $F_r(z)$, которая может зависеть от некоторого вещественного параметра r . Будем интерпретировать h_r как случайный интервал времени. Предположим, что случайная величина h_r при любом фиксированном r не зависит от рассматриваемого простейшего потока и, в частности, от случайных величин ξ_1, \dots, ξ_s . Обозначим через A_s случайное событие, состоящее в том, что на интервале времени h_r произошло ровно s событий простейшего потока, а через $\theta_{r,s} = h_r - \zeta_s$ — случайную величину, которая представляет собой остаточную длительность интервала времени h_r после наступления s -го события потока. Тогда математическое ожидание случайной величины $\theta_{r,s}$, определяемое по совместному распределению с событием A_s , может быть выражено следующей формулой:

$$\tau_{r,s} = E[\theta_{r,s}; A_s] = \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dF_r(z) dx. \quad (3)$$

Доказательство. Заметим, что случайная величина $\theta_{r,s}$ определена на множестве элементарных исходов $\omega \in \Omega$, которые соответствуют событию A_s . Тогда можно воспользоваться формулой (2) для математического ожидания по совместному распределению случайной величины $\theta_{r,s}$ и события A_s :

$$\tau_{r,s} = E[\theta_{r,s}; A_s] = \int_0^\infty P(\theta_{r,s} > x; A_s) dx. \quad (4)$$

В то же время событие A_s можно представить следующим образом:

$$A_s = \left(\zeta_s = \sum_{i=1}^s \xi_i < h_r, \zeta_s + \xi_{s+1} > h_r \right).$$

Выразим совместное распределение случайной величины $\theta_{r,s}$ и события A_s . Для любого $x \geq 0$ имеем:

$$\begin{aligned} P(\theta_{r,s} > x; A_s) &= \\ &= P(\theta_{r,s} > x, \zeta_s < h_r, \zeta_s + \xi_{s+1} > h_r) = \\ &= P(h_r - \zeta_s > x, \zeta_s < h_r, \zeta_s + \xi_{s+1} > h_r) = \\ &= P(h_r - \zeta_s > x, \xi_{s+1} > h_r - \zeta_s). \end{aligned} \quad (5)$$

В силу введенных предположений случайная величина ζ_s имеет распределение Эрланга порядка $s \geq 1$ с параметром λ . Для этой величины известны явные представления функции распределения $G_s(y)$ и ее полного дифференциала:

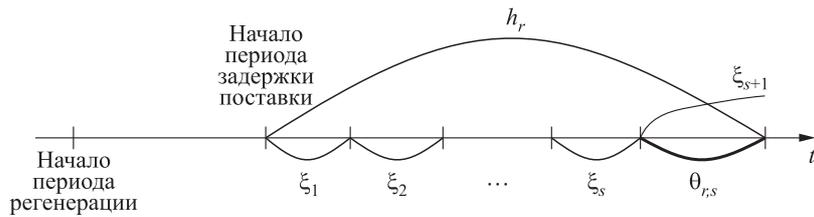


Рис. 1 Графическая иллюстрация к утверждению теоремы 1

$$G_s(y) = P(\zeta_s < y) = 1 - \sum_{i=0}^{s-1} \frac{(\lambda y)^i}{i!} e^{-\lambda y}, \quad s \geq 1;$$

$$dG_s(y) = \lambda \frac{(\lambda y)^{s-1}}{(s-1)!} e^{-\lambda y} dy, \quad s \geq 1.$$

Заметим, что в условиях доказываемой теоремы случайные величины h_r , ζ_s и ξ_{s+1} независимы. Зафиксируем значения случайных величин $h_r = z$ и $\zeta_s = y$. При указанном условии для реализации события $(h_r - \zeta_s > x, \xi_{s+1} > h_r - \zeta_s)$ необходимо, чтобы выполнялось соотношение $z - y > x$, т.е. $0 < y < z - x$. В то же время должно выполняться соотношение $z > x$, так как иначе событие $(h_r - \zeta_s > x)$ невозможно, т.е. имеет вероятность, равную нулю.

Из указанных замечаний следует, что совместная вероятность $P(h_r - \zeta_s > x, \xi_{s+1} > h_r - \zeta_s)$ представима в виде:

$$P(h_r - \zeta_s > x, \xi_{s+1} > h_r - \zeta_s) =$$

$$= \int_x^\infty \int_0^{z-x} e^{-\lambda(z-y)} dG_s(y) dF_r(z) =$$

$$= \int_x^\infty \int_0^{z-x} e^{-\lambda(z-y)} \lambda \frac{(\lambda y)^{s-1}}{(s-1)!} e^{-\lambda y} dy dF_r(z) =$$

$$= \int_x^\infty \int_0^{z-x} \lambda \frac{(\lambda y)^{s-1}}{(s-1)!} e^{-\lambda z} dy dF_r(z) =$$

$$= \int_x^\infty \frac{\lambda^s}{(s-1)!} e^{-\lambda z} \int_0^{z-x} y^{s-1} dy dF_r(z) =$$

$$= \int_x^\infty \frac{\lambda^s}{(s-1)!} e^{-\lambda z} \frac{(z-x)^s}{s} dF_r(z) =$$

$$= \int_x^\infty \frac{\lambda^s}{s!} (z-x)^s e^{-\lambda z} dF_r(z). \quad (6)$$

Из (5) и (6) получаем явное представление для совместного распределения $P(\theta_{r,s} > x; A_s)$, $x > 0$. Остается применить формулу (4):

$$\tau_{r,s} = E[\theta_{r,s}; A_s] = \int_0^\infty P(\theta_{r,s} > x; A_s) dx =$$

$$= \int_0^\infty P(h_r - \zeta_s > x, \xi_{s+1} > h_r - \zeta_s) dx =$$

$$= \int_0^\infty \int_x^\infty \frac{\lambda^s (z-x)^s}{s!} e^{-\lambda z} dF_r(z) dx =$$

$$= \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dF_r(z) dx.$$

Полученное соотношение совпадает с формулой (3). Теорема 1 доказана.

Замечание 1. В ходе дальнейшего анализа рассматриваемой модели управления запасом в роли простейшего потока событий будет выступать поток моментов поступления заявок на приобретение товара. Случайная величина h_r будет играть роль длительности задержки поставки, распределение которой зависит от параметра управления r . Случайная величина $\theta_{r,s}$ будет представлять собой остаточную длительность задержки после поступления последнего на данном периоде регенерации заказа, имеющего фиксированный номер s , считая от момента начала периода задержки. Иллюстрация реализаций соответствующих событий в ходе эволюции рассматриваемой системы приведена на рис. 1.

Замечание 2. Согласно принятым предположениям о рассматриваемой модели [1], случайная величина h_r имеет заданное распределение $H_r(z)$. Таким образом, в дальнейшем при использовании теоремы 1 будем полагать $F_r(z) = H_r(z)$, $z \geq 0$.

3 Аналитические представления для стоимостных характеристик исследуемой модели

Перейдем к доказательству утверждений о конкретных представлениях вероятностных характеристик, связанных со стационарным стоимостным

показателем эффективности управления. Прежде всего введем систему $\{A_s, s = 0, 1, 2, \dots\}$, где $A_s \in \mathcal{A}$ — событие, состоящее в том, что за период задержки в рассматриваемую систему поступило ровно s требований на продукт (клиентов). Из свойств пуассоновского потока и формулы полной вероятности следует:

$$P(A_s) = \int_0^{\infty} \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y), \quad s = 0, 1, 2, \dots$$

Рассмотрим случайную величину $\Delta\gamma_n$, которая представляет собой приращение аддитивного функционала прибыли на очередном периоде регенерации $(t_n, t_{n+1}]$ (см. [1, разд. 4]). Поскольку система $\{A_s, s = 0, 1, 2, \dots\}$ образует полную группу несовместных событий, то по свойству математического ожидания можно записать:

$$E(\Delta\gamma_n) = \sum_{s=0}^{\infty} E(\Delta\gamma_n; A_s).$$

Теперь рассмотрим соответствующие математические ожидания при условии, что параметр управления на данном периоде регенерации принимает фиксированное значение r . Обозначим эти математические ожидания $E_r(\Delta\gamma_n)$ и $E_r(\Delta\gamma_n; A_s)$. Используя упомянутое свойство математического ожидания, получим следующее исходное соотношение:

$$E_r(\Delta\gamma_n) = \sum_{s=0}^{\infty} E_r(\Delta\gamma_n; A_s). \quad (7)$$

Введем дополнительно две случайные величины, связанные с периодом регенерации $(t_n, t_{n+1}]$: $\Delta\gamma_n^{(+)}$ — приращение дохода (кратко — доход) и $\Delta\gamma_n^{(-)}$ — приращение затрат (кратко — затраты). Тогда приращение прибыли на периоде представляется в виде $\Delta\gamma_n = \Delta\gamma_n^{(+)} - \Delta\gamma_n^{(-)}$.

Введем также обозначения $E_r(\Delta\gamma_n^{(+)}; A_s)$ и $E_r(\Delta\gamma_n^{(-)}; A_s)$ для условных математических ожиданий дохода и затрат на периоде регенерации $(t_n, t_{n+1}]$, определяемых по совместному распределению с событием A_s , при условии, что на указанном периоде принято решение r . В силу свойства аддитивности величины $\Delta\gamma_n$ имеет место соотношение:

$$E_r(\Delta\gamma_n; A_s) = E_r(\Delta\gamma_n^{(+)}; A_s) - E_r(\Delta\gamma_n^{(-)}; A_s).$$

Условные математические ожидания $E_r(\Delta\gamma_n)$ необходимы для получения явного аналитического вида показателя эффективности управления I_r .

Следует отметить, что структура формирования случайных величин $\Delta\gamma_n^{(+)}$ и $\Delta\gamma_n^{(-)}$ существенно зависит от значения параметра управления r и числа требований, поступивших за период задержки поставки, т. е. от величины s . В связи с этим разобьем множество возможных значений параметра r на следующие непересекающиеся подмножества и определим на каждом из них явное представление для $E_r(\Delta\gamma_n)$ и I_r :

- $r = N$;
- $1 \leq r < N$;
- $r = 0$;
- $-N_0 < r \leq -1$;
- $r = -N_0$.

В дальнейшем в каждой из указанных областей значений параметра r необходимо найти явные представления для математических ожиданий $E_r(\Delta\gamma_n; A_s)$ при всех значениях $s = 0, 1, 2, \dots$. После этого можно воспользоваться формулой (7).

Ограничения по объему данной статьи не позволяют привести явные формулы для всех областей изменения параметра управления. Полностью эти формулы приведены и доказаны в приложении к настоящей работе [5]. В рамках настоящей статьи ограничимся только некоторыми фрагментами вероятностного анализа, дающими представление о характере данного этапа исследования.

Рассмотрим следующий вариант: $1 \leq r < N$, $1 \leq s < r$ (рис. 2). Заметим, что если параметры r и s принимают значения в указанных областях, то происходят следующие события: заказ на очередную партию поставки осуществляется после прихода некоторого числа требований, но до полного опустошения склада, а число требований, поступивших в период задержки поставки, меньше оставшегося на складе в момент заказа числа единиц продукта.

В первую очередь выпишем математическое ожидание дохода за период регенерации по со-

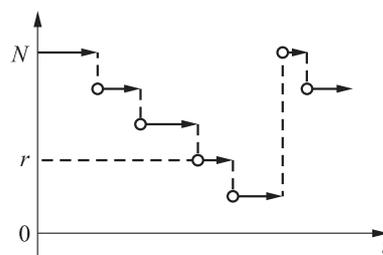


Рис. 2 Графическая иллюстрация эволюции процесса для варианта значений параметров $1 \leq r < N$ и $1 \leq s < r$

вместному распределению с событием A_s . Получим:

$$E_r(\Delta\gamma_n^{(+)}; A_s) = E_r(\Delta\gamma_n^{(+)}|A_s) P(A_s) = c_0(N-r+s) \int_0^\infty \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y). \quad (8)$$

Теперь необходимо проанализировать аналогичное математическое ожидание для расходов на периоде регенерации. Расходы в рассматриваемом случае обусловлены хранением имеющегося товара и приобретением дополнительного товара для пополнения запаса. Приведем общее представление для этой стоимостной характеристики, добавив некоторые пояснения к отдельным компонентам найденной формулы:

$$E_r(\Delta\gamma_n^{(-)}; A_s) = E_r(\Delta\gamma_n^{(-)}|A_s) P(A_s) = \left[\underbrace{\sum_{i=r-s+1}^N \frac{c_1 i}{\lambda}}_{\substack{\text{затраты} \\ \text{на хранение} \\ \text{до поступления} \\ \text{последнего} \\ \text{требования}}} + \underbrace{E_{1*}(\Delta\gamma_n^{(-)}|A_s)}_{\substack{\text{затраты} \\ \text{на хранение} \\ \text{после поступления} \\ \text{последнего} \\ \text{требования}}} + \underbrace{c_2(N-r-s)}_{\substack{\text{затраты} \\ \text{на пополнение} \\ \text{запаса}}} \right] \times P(A_s). \quad (9)$$

Заметим, что

$$\sum_{i=r-s+1}^N \frac{c_1 i}{\lambda} = \frac{c_1(N-r+s)(N+r-s+1)}{2\lambda}. \quad (10)$$

Явное представление для математического ожидания $E_{1*}(\Delta\gamma_n^{(-)}; A_s)$ определяется на основе теоремы 1 (формула (3)):

$$E_{1*}(\Delta\gamma_n^{(-)}; A_s) = c_1(r-s)\tau_{r,s} = c_1(r-s) \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dH_r(z) dx. \quad (11)$$

Подставляя (10) и (11) в соотношение (9), получаем явную формулу для математического ожидания расходов по совместному распределению с событием A_s :

$$E_r(\Delta\gamma_n^{(-)}; A_s) = \left[\frac{c_1(N-r+s)(N+r-s+1)}{2\lambda} + c_2(N-r+s) \right] \int_0^\infty \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y) + c_1(r-s) \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dH_r(z) dx. \quad (12)$$

Теперь воспользуемся формулами (8) и (12) и получим окончательное выражение для математического ожидания приращения функционала прибыли при фиксированных значениях параметров r и s , удовлетворяющих условиям $1 \leq r < N, 1 \leq s < r$:

$$E_r(\Delta\gamma_n; A_s) = E_r(\Delta\gamma_n^{(+)}; A_s) - E_r(\Delta\gamma_n^{(-)}; A_s) = c_0(N-r+s) \int_0^\infty \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y) - \left[\frac{c_1(N-r+s)(N+r-s+1)}{2\lambda} + c_2(N-r+s) \right] \int_0^\infty \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y) - c_1(r-s) \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dH_r(z) dx = (N-r+s) \left[c_0 - \frac{c_1(N+r-s+1)}{2\lambda} - c_2 \right] \int_0^\infty \frac{(\lambda y)^s}{s!} e^{-\lambda y} dH_r(y) - c_1(r-s) \int_0^\infty \frac{\lambda^s}{s!} \int_x^\infty (z-x)^s e^{-\lambda z} dH_r(z) dx. \quad (13)$$

Напомним [1], что показатель эффективности управления I_r представляет собой отношение математического ожидания приращения аддитивного функционала прибыли, полученной за период регенерации, к математическому ожиданию длительности периода регенерации. Оба указанных математических ожидания определяются при условии, что параметр управления принимает фиксированное значение r . В принятых обозначениях

$$I_r = \frac{E_r(\Delta\gamma_n)}{E_r(\Delta t_n)}. \quad (14)$$

При этом в соответствии с [1, соотношение (1)] условное математическое ожидание $E_r(\Delta t_n)$ определяется формулой:

$$E_r(\Delta t_n) = \frac{N-r}{\lambda} + \tau_r = \frac{N-r}{\lambda} + \int_0^\infty (1-H_r(y)) dy.$$

В заключение данного раздела приведем также итоговое представление для показателя эффективности I_r для варианта, когда параметр управления r

принимает значение в области $1 \leq r < N$. Заметим дополнительно, что выражение для $E_r(\Delta\gamma_n)$ в формуле (14) определяется на основе соотношения (7) с учетом, в частности, формулы (13):

$$\begin{aligned}
 I_r = & \\
 = & \frac{1}{(N-r)/\lambda + \int_0^\infty (1-H_r(y)) dy} \left[\left(c_0(N-r) - \right. \right. \\
 & - \frac{c_1(N+r+1)(N-r)}{2\lambda} - c_1 r \int_0^\infty (1-H_r(y)) dy - \\
 & \left. \left. - c_2(N-r) \right) \int_0^\infty e^{-\lambda y} dH_r(y) + \right. \\
 & + \sum_{i=1}^{r-1} \left(\left(c_0(N-r+i) - \right. \right. \\
 & - \frac{c_1(N-r+i)(N+r-i+1)}{2\lambda} - \\
 & \left. \left. - c_2(N-r+i) \right) \int_0^\infty \frac{(\lambda y)^i}{i!} e^{-\lambda y} dH_r(y) - \right. \\
 & \left. - c_1(r-i) \int_0^\infty \frac{\lambda^i}{i!} \int_x^\infty (z-x)^i e^{-\lambda z} dH_r(z) dx \right) + \\
 & + \left(c_0 N - \frac{c_1(N+1)N}{2\lambda} - \right. \\
 & \left. - c_2 N \right) \int_0^\infty \frac{(\lambda y)^r}{r!} e^{-\lambda y} dH_r(y) + \\
 & + \sum_{i=r+1}^{r+N_0-1} \left(\left(c_0(N-r+i) - \right. \right. \\
 & - \frac{c_1(N+1)N}{2\lambda} - c_2(N-r+i) - \\
 & \left. \left. - \frac{c_3(i-r)(i-r-1)}{2\lambda} \right) \int_0^\infty \frac{(\lambda y)^i}{i!} e^{-\lambda y} dH_r(y) - \right. \\
 & \left. - c_3(i-r) \int_0^\infty \frac{\lambda^i}{i!} \int_x^\infty (z-x)^i e^{-\lambda z} dH_r(z) dx \right) + \\
 & + \left(c_0(N+N_0) - \frac{c_1(N+1)N}{2\lambda} - c_2(N+N_0) - \right. \\
 & \left. - \frac{c_3 N_0(N_0-1)}{2\lambda} \right) \int_0^\infty \frac{(\lambda y)^{r+N_0}}{(r+N_0)!} e^{-\lambda y} dH_r(y) - \\
 & - c_3 N_0 \int_0^\infty \frac{\lambda^{r+N_0}}{(r+N_0)!} \int_x^\infty (z-x)^{r+N_0} e^{-\lambda z} dH_r(z) dx +
 \end{aligned}$$

$$\begin{aligned}
 + \sum_{i=r+N_0+1}^\infty & \left(\left(c_0(N+N_0) - \frac{c_1(N+1)N}{2\lambda} - \right. \right. \\
 & - c_2(N+N_0) - \frac{c_3 N_0(2i-2r-N_0-1)}{2\lambda} - \\
 & \left. \left. - c_4^{(i-r-N_0)} \right) \int_0^\infty \frac{(\lambda y)^i}{i!} e^{-\lambda y} dH_r(y) - \right. \\
 & \left. - c_3 N_0 \int_0^\infty \frac{\lambda^i}{i!} \int_x^\infty (z-x)^i e^{-\lambda z} dH_r(z) dx \right) \Big]. \quad (15)
 \end{aligned}$$

4 Решение проблемы оптимального управления

В настоящем исследовании получены явные аналитические представления показателя эффективности управления I_r , аналогичные формуле (15), для всех значений параметра оптимизации $r \in U = \{N, N-1, \dots, 0, -1, \dots, -N_0\}$. Соответствующие представления полностью приведены в приложении [5]. Как было установлено в [1, формулы (12), (13)], показатель I_r совпадает с основной функцией дробно-линейного дискретного интегрального функционала I_α :

$$\begin{aligned}
 I_r = C(r) = & \frac{A(r)}{B(r)}, \\
 r \in U = & \{N, N-1, \dots, 0, -1, \dots, -N_0\}. \quad (16)
 \end{aligned}$$

В соответствии с общим подходом к решению задачи оптимального управления в рассматриваемой модели регенерации (см. [1, разд. 5]) оптимальное управление является детерминированным и определяется точкой глобального экстремума (максимума) функции I_r , которая аналитически задается равенством (16). Поскольку множество допустимых управлений U конечно, глобальный максимум основной функции достигается в некоторой точке $r^* \in U$. Следовательно, оптимальное управление существует и совпадает с точкой глобального максимума r^* . Задача оптимального управления в рассматриваемой стохастической модели сводится к нахождению точки r^* , которое можно осуществить только численным методом.

5 Заключение

Подведем итоги проведенного исследования в целом. В работе [1] была построена стохастическая модель управления дискретным запасом в виде

регенерирующего случайного процесса. Определен общий подход к решению этой задачи на основе теоремы о безусловном экстремуме дробно-линейного интегрального функционала, зависящего от дискретного вероятностного распределения, которое характеризует стратегию управления. В настоящей работе получены явные аналитические представления для основной функции дробно-линейного интегрального функционала, который играет роль показателя эффективности управления. В соответствии с утверждением указанной выше теоремы решение исходной задачи управления существует и определяется точкой глобального максимума заданной основной функции.

Литература

1. Шнурков П. В., Вахтанов Н. А. Исследование проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением и случайной задержкой поставки // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 54–61.
2. Шнурков П. В. О решении задачи безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. Акад. наук. Сер. Математика, 2016. Т. 470. № 4. С. 387–392.
3. Шнурков П. В., Горшенин А. К., Белоусов В. В. Аналитическое решение задачи оптимального управления полумарковским процессом с конечным множеством состояний // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 72–88.
4. Гнеденко Б. В. Курс теории вероятностей. — М.: Либроком, 2011. 488 с.
5. Шнурков П. В., Вахтанов Н. А. Приложение к статьям «Исследование проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением и случайной задержкой поставки» и «Об оптимальном решении проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением», 2019. 69 с. <http://www.ipiran.ru/publications/Приложение.pdf>.

Поступила в редакцию 01.07.19

ON THE SOLUTION OF THE OPTIMAL CONTROL PROBLEM OF INVENTORY OF A DISCRETE PRODUCT IN THE STOCHASTIC MODEL OF REGENERATION WITH CONTINUOUSLY OCCURRING CONSUMPTION

P. V. Shnurkov and N. A. Vakhtanov

National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

Abstract: The article is the second and final part of the research of the optimal control problem of inventory of a discrete product in a stochastic regeneration model. The main content of the work is the derivation of analytical representations for the mathematical expectation of the increment of the functional of profit obtained during the regeneration period. At the same time, these mathematical expectations are determined under different conditions for decisions made during the regeneration period. The obtained analytical representations enable one to explicitly determine the stationary cost indicator of control efficiency, which was introduced in the first part of the research. Thus, it becomes possible to numerically solve the optimal control problem of inventory in the model under consideration.

Keywords: inventory management of a discrete product; controlled regenerative process; stationary cost indicator of control efficiency

DOI: 10.14357/19922264190308

References

1. Shnurkov, P. V., and N. A. Vakhtanov. 2019. Issledovanie problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoy modeli regeneratsii s nepre-ryvno proiskhodyashchim potrebleniem i sluchaynoy za-derzhkoy postavki [Research of the optimal control prob-lem of inventory of a discrete product in stochastic regen-eration model with continuously occurring consumption and random delivery delay]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):54–61.
2. Shnurkov, P. V. 2016. Solution of the unconditional ex-tremum problem for a linear-fractional integral functional

- on a set of probability measures. *Dokl. Math.* 94(2):550–554.
3. Shnurkov, P. V., A. K. Gorshenin, and V. V. Belousov. 2016. Analiticheskoe reshenie zadachi optimal'nogo upravleniya polumarkovskim protsessom s konechnym mnozhestvom sostoyaniy [Analytical solution of the optimal control task of a semi-Markov process with finite set of states]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):72–88.
 4. Gnedenko, B. V. 2011. *Kurs teorii veroyatnostey* [A course in the theory of probability]. Moscow: Librokom. 488 p.
 5. Shnurkov, P. V., and N. A. Vakhtanov. 2019. Prilozhenie k stat'yam "Issledovanie problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoj modeli regeneratsii s nepreryvno proiskhodyashchim potrebleniem i sluchaynoy zaderzhkoy postavki" i "Ob optimal'nom reshenii problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoj modeli regeneratsii s nepreryvno proiskhodyashchim potrebleniem" [Appendix to articles "Research of the optimal control problem of inventory of a discrete product in stochastic regeneration model with continuously occurring consumption and random delivery delay" and "On the optimal solution of the optimal control problem of inventory of a discrete product in stochastic regeneration model with continuously occurring consumption"]. 69 p. Available at: <http://www.ipiran.ru/publications/Приложение.pdf> (accessed May 6, 2019).

Received July 1, 2019

Contributors

Shnurkov Peter V. (b. 1953) — Candidate of Science (PhD) in physics and mathematics, associate professor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; pshnurkov@hse.ru

Vakhtanov Nikita A. (b. 1997) — Master student, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; Vakhtanov1997@mail.ru

ОЦЕНКА УРОВНЯ ЗНАЧИМОСТИ КРИТЕРИЯ ШУИРМАННА ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ БИОЭКВИВАЛЕНТНОСТИ ПРИ НАЛИЧИИ ПРОПУЩЕННЫХ ДАННЫХ*

Т. В. Захарова¹, А. А. Тархов²

Аннотация: Задача проверки гипотезы биоэквивалентности имеет важное значение в фармакокинетике. С ее помощью принимают решение об эквивалентности воспроизведенного лекарственного препарата референтному лекарственному препарату. Одна из проблем исследований биоэквивалентности — наличие пропущенных данных. Так как объем исследуемых данных достаточно мал, то удаление данных субъекта, у которого есть пропущенные данные, нежелательно. Поэтому стоит задача оценить влияние пропущенных данных при принятии решения о биоэквивалентности, а именно: дать оценку уровня значимости. Основным методом проверки гипотезы биоэквивалентности является процедура двух односторонних тестов Шуирманна. В статье дана оценка уровня значимости данной процедуры при наличии пропущенных данных. В явном виде получена компонента оценки уровня значимости, зависящая от уровня полноты данных.

Ключевые слова: биоэквивалентность; уровень значимости; ошибка первого рода; пропущенные данные; процедура двух односторонних тестов Шуирманна

DOI: 10.14357/19922264190309

1 Введение

Предположим, что имеется лекарственный препарат, для которого был проведен набор широкомасштабных клинических исследований, доказавших его безопасность и медицинскую эффективность. Данный препарат будем называть референтным лекарственным препаратом. На основе действующих веществ референтного лекарственного препарата могут быть созданы новые лекарственные препараты с таким же количественным и качественным составом. Далее будем называть их воспроизведенными лекарственными препаратами. Чтобы перенести имеющиеся сведения о безопасности и эффективности референтного лекарственного препарата на воспроизведенный препарат без проведения широкомасштабных исследований, исследуют биоэквивалентность лекарственных препаратов.

Понятие биоэквивалентности тесно связано с понятием биодоступности.

Биодоступность — скорость и степень, с которыми действующее вещество или его активная часть молекулы из дозированной лекарственной формы всасываются и становятся доступными в месте дей-

ствия. Два лекарственных препарата, содержащих одинаковое количество действующего вещества, считаются биоэквивалентными, если они являются фармацевтически эквивалентными или фармацевтически альтернативными и их биодоступность (по скорости и степени) после применения в одинаковой молярной дозе укладывается в заранее установленные допустимые пределы [1].

В процессе сбора данных некоторые полученные значения могут быть утеряны. Так как число испытуемых ограничено, то исключать данные испытуемого, для которого было утеряно одно значение концентрации действующего вещества, нерационально. Вместо этого пропущенные данные заполняют нулем или другим значением, полученным на основе информации о других значениях.

При использовании недостаточно точных методов заполнения данных, таких как заполнение нулем, уменьшается часть значений исследуемых данных. Хотелось бы оценить, как наличие пропущенных данных влияет на проверку гипотезы биоэквивалентности.

В данной работе будет рассмотрена процедура двух односторонних тестов Шуирманна, которая

*Работа выполнена при поддержке РФФИ (проект 18-07-00252).

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, tvzaharova@mail.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, alexeytarkhov@gmail.com

в настоящее время используется при проверке гипотезы биоэквивалентности, при наличии пропусков в исследуемых данных.

2 Задача проверки гипотезы биоэквивалентности

Пусть T — воспроизведенный лекарственный препарат, а R — референтный лекарственный препарат и соответственно μ_T и μ_R — математические ожидания сравнительных характеристик для лекарственных препаратов T и R .

Основными сравнительными характеристиками биоэквивалентности служат максимальная концентрация в крови C_{\max} и площадь под кривой «концентрация — время» AUC (от *англ.* Area Under the Curve).

Будем следовать предположению, что рассматриваемые сравнительные характеристики имеют логнормальное распределение [2–4]. Например, для AUC:

$$AUC_i \sim \text{Log } N(a_i, \sigma_i), \quad i \in \{T, R\}.$$

Пусть θ_1 и θ_2 — соответственно нижний и верхний принятый допустимый предел признания биоэквивалентности. Следовательно, гипотеза о биоэквивалентности может быть записана следующим образом:

$$H_0 : \frac{\mu_T}{\mu_R} \leq q\theta_1 \text{ или } \frac{\mu_T}{\mu_R} \geq \theta_2;$$

$$H_A : \theta_1 < \frac{\mu_T}{\mu_R} < \theta_2.$$

Сделав логарифмическое преобразование, можем перейти к следующей постановке рассматриваемой гипотезы:

$$H'_0 : \mu'_T - \mu'_R \leq q\delta_1 \text{ или } \mu'_T - \mu'_R \geq \delta_2;$$

$$H'_A : \delta_1 < \mu'_T - \mu'_R < \delta_2,$$

где $\delta_1 = \ln \theta_1$ и $\delta_2 = \ln \theta_2$, а μ'_T и μ'_R — математические ожидания логарифмов сравнительных характеристик для лекарственных препаратов T и R . Например, для $\ln AUC_T$ из свойств логнормального распределения следует, что $\mu'_T = a_T$.

Гипотеза H'_0 соответствует небиоэквивалентности исследуемых лекарственных препаратов, в то время как H'_A утверждает, что они биоэквивалентны. Выбор такого порядка основной и альтернативной гипотез обусловлен тем, что в таком случае ошибка первого рода соответствует признанию лекарственных средств биоэквивалентными, хотя на

самом деле они такими и не является. В этом случае пациент несет риск получить препарат, который может не обладать такими же эффективностью и безопасностью, как референтный лекарственный препарат [4].

3 Процедура двух односторонних тестов Шуирманна

Разобьем гипотезы H'_0 и H'_1 на два множества односторонних гипотез:

$$\begin{cases} H_{01} : \mu'_T - \mu'_R \leq q\delta_1; \\ H_{A1} : \mu'_T - \mu'_R > \delta_1; \end{cases}$$

$$\begin{cases} H_{02} : \mu'_T - \mu'_R \geq \delta_2; \\ H_{A2} : \mu'_T - \mu'_R < \delta_2. \end{cases}$$

Процедура двух односторонних тестов заключается в том, что H'_0 отвергаем при уровне значимости α , тем самым устанавливая эквивалентность μ_T и μ_R , только в том случае, если отвергаются обе гипотезы H_{01} и H_{02} при заданном уровне значимости α [2, 3].

Таким образом, два односторонних теста проверяются с использованием односторонних t-критериев, т. е. характеристики биодоступности признаются эквивалентными, если

$$\left. \begin{aligned} t_1 &= \frac{\bar{Y}_T - \bar{Y}_R - \delta_1}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}} > t(\alpha, n_1 + n_2 - 2); \\ t_2 &= \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}} < -t(\alpha, n_1 + n_2 - 2), \end{aligned} \right\} \quad (1)$$

где n_1 и n_2 — число субъектов в последовательностях клинического исследования с перекрестным двухпоследовательным дизайном, $t(\alpha, n_1 + n_2 - 2)$ — $(1 - \alpha)$ -квантиль центрального t-распределения с $n_1 + n_2 - 2$ степенями свободы; $\hat{\sigma}_d$ — обобщенная выборочная дисперсия разностей между периодами (для обоих последовательностей в исследовании), которая является несмещенной оценкой σ_d , причем

$$\sigma_d^2 = \frac{\sigma_w^2}{2},$$

где σ_w — внутрисубъектная вариабельность изучаемых параметров [4].

Процедура двух односторонних тестов эквивалентна подходу с построением доверительного интервала для разности выборочных средних, т. е. получению следующей интервальной оценки:

$$\left(\bar{Y}_T - \bar{Y}_R + t(\alpha, n_1 + n_2 - 2) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; \right. \\ \left. \bar{Y}_T - \bar{Y}_R - t(\alpha, n_1 + n_2 - 2) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right).$$

Признание эквивалентности параметров биодоступности на уровне значимости α может быть сделано, только если полученный доверительный $(1 - 2\alpha)100\%$ -ный интервал для $\mu'_T - \mu'_R$ полностью содержится в интервале (δ_1, δ_2) [2].

Так как рассматриваем сбалансированный дизайн, то $n_1 = n_2 = n$, и, учитывая (1), получаем, что t -критерии принимают вид:

$$t'_1 = \frac{\bar{Y}_T - \bar{Y}_R - \delta_1}{\hat{\sigma}_d \sqrt{2/n}} > t(2n - 2, \alpha); \\ t'_2 = \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d \sqrt{2/n}} < -t(2n - 2, \alpha)$$

и соответствующий доверительный интервал принимает вид:

$$\left(\bar{Y}_T - \bar{Y}_R + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}}; \right. \\ \left. \bar{Y}_T - \bar{Y}_R - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} \right).$$

4 Оценка уровня значимости при наличии пропущенных данных

Рассмотрим выборочное пространство χ — пространство элементарных событий. Статистический критерий разбивает пространство элементарных событий χ на два подмножества:

- (1) область принятия гипотезы χ_0 — множество, состоящее из точек, для которых гипотеза H_0 принимается;
- (2) область отклонения гипотезы χ_A — множество, состоящее из точек, для которых гипотеза H_0 отвергается.

Говорят, что критерий имеет уровень значимости α , если вероятность наступления ошибки первого рода не превышает α , $0 < \alpha < 1$, для $\delta \in \chi_A$:

$$P \{ \text{отклонить } H_0 \text{ при истинной} \\ \text{небиоэквивалентности} \} = \\ = P \{ \text{отклонить } H_0, \delta \in \chi_0 \} \leq \alpha.$$

Рассмотрим задачу при наличии пропусков в данных: пусть q — уровень полноты данных, т. е. доля данных, оставшихся от изначальных дан-

ных, $0 < q \leq 1$ ($1 - q$ — доля пропущенных данных в выборке).

Тогда $\tilde{Y}_T = \bar{Y}_T + \ln(q)$ — выборочное среднее логарифмов сравнительных характеристик для лекарственного препарата T при наличии пропущенных данных.

Критическая область для данной задачи имеет следующий вид:

$$\chi'_A = \left\{ (\tilde{Y}_T - \bar{Y}_R, \hat{\sigma}_d) : \delta_1 + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} < \right. \\ \left. < \tilde{Y}_T - \bar{Y}_R < \delta_2 - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} \right\}.$$

Рассмотрим функцию мощности:

$$\phi'_{\hat{\sigma}_d}(\delta) = P \{ \text{отклонить } H_0 \text{ при истинной} \\ \text{биоэквивалентности} \} = \\ = P \{ (\tilde{Y}_T - \bar{Y}_R, \hat{\sigma}_d) : \chi'_A, \delta \in \chi'_A \}.$$

Фиксируем $\delta = \delta_0$, получаем:

$$\phi'_{\hat{\sigma}_d}(\delta_0) = \\ = P \left((\tilde{Y}_T - \bar{Y}_R, \hat{\sigma}_d) : \delta_1 + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} < \right. \\ \left. < \tilde{Y}_T - \bar{Y}_R < \delta_2 - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} \mid \delta = \delta_0 \right) = \\ = P \left(\delta_1 + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} < \bar{Y}_T + \ln(q) - \right. \\ \left. - \bar{Y}_R < \delta_2 - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{\frac{2}{n}} \right) = \\ = P \left(\frac{\delta_1 + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d \sqrt{2/n}} < \right. \\ \left. < \frac{\bar{Y}_T - \bar{Y}_R - \delta_0}{\sigma_d \sqrt{2/n}} < \right. \\ \left. < \frac{\delta_2 - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d \sqrt{2/n}} \right) = \\ = \{ \text{фиксируем } \hat{\sigma}_d \} = \\ = E \left[P \left(\frac{\delta_1 + t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d \sqrt{2/n}} < \right. \right. \\ \left. \left. < \frac{\bar{Y}_T - \bar{Y}_R - \delta_0}{\sigma_d \sqrt{2/n}} < \right. \right. \\ \left. \left. < \frac{\delta_2 - t(\alpha, 2n - 2) \hat{\sigma}_d \sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d \sqrt{2/n}} \mid \hat{\sigma}_d \right) \right] =$$

$$= E \left[\Phi \left(\frac{\delta_1 + t(\alpha, 2n - 2)\hat{\sigma}_d\sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d\sqrt{2/n}} \right) - \Phi \left(\frac{\delta_2 - t(\alpha, 2n - 2)\hat{\sigma}_d\sqrt{2/n} - \ln(q) - \delta_0}{\sigma_d\sqrt{2/n}} \right) \right],$$

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

Процедура двух односторонних тестов Шуирманна используется в условиях решающего правила 80/125 [2, 4, 5]. Это значит, что $\delta_1 = \ln(0, 80) \approx -0,2231$ и $\delta_2 = \ln(1, 25) \approx 0,2231$; следовательно, $\delta_1 \approx -\delta_2$. Тогда видим, что функция $\phi'_{\sigma_d}(\delta)$ симметрична относительно точки $\delta = -\ln(q)$ и достигает максимума в этой точке.

Тогда

$$\begin{aligned} \max_{\delta \in \chi_0} P\{\text{отклонить } H_0\} &= \phi'_{\sigma_d}(\delta_2) = \\ & \{ \text{подставим } \delta = \delta_2, \delta_1 = -\delta_2, \sigma_d = \hat{\sigma}_d \} = \\ &= P \left(\frac{-2\delta_2 - \ln(q)}{\hat{\sigma}_d\sqrt{2/n}} + t(\alpha, 2n - 2) < \right. \\ & < \left. \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < \frac{-\ln(q)}{\hat{\sigma}_d\sqrt{2/n}} - t(\alpha, 2n - 2) \right) = \\ &= P \left(\frac{-2\delta_2 - \ln(q)}{\hat{\sigma}_d\sqrt{2/n}} + t(\alpha, 2n - 2) < \right. \\ & < \left. \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < -t(\alpha, 2n - 2) \right) + \\ &+ P \left(-t(\alpha, 2n - 2) < \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < \right. \\ & < \left. \frac{-\ln(q)}{\hat{\sigma}_d\sqrt{2/n}} - t(\alpha, 2n - 2) \right) \leq \\ &\leq P \left(\frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < -t(\alpha, 2n - 2) \right) + \\ &+ P \left(-t(\alpha, 2n - 2) < \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < \right. \\ & < \left. \frac{-\ln(q)}{\hat{\sigma}_d\sqrt{2/n}} - t(\alpha, 2n - 2) \right) = \\ &= \alpha + P \left(-t(\alpha, 2n - 2) < \frac{\bar{Y}_T - \bar{Y}_R - \delta_2}{\hat{\sigma}_d\sqrt{2/n}} < \right. \\ & < \left. \frac{-\ln(q)}{\hat{\sigma}_d\sqrt{2/n}} - t(\alpha, 2n - 2) \right) = \alpha + \alpha'. \end{aligned}$$

Полученная оценка показывает, что величина ошибки первого рода не превосходит $\alpha + \alpha'$. При чем в работах [3, 4] показано, что при исполь-

зовании двух односторонних тестов Шуирманна величина вероятности ошибки первого рода не превосходит α . В рассматриваемой постановке уровень значимости критерия повышается на величину α' , что обусловлено наличием пропущенных данных для воспроизводимого лекарственного препарата. Таким образом, риск потенциального выхода на рынок небоэквивалентного лекарственного препарата повышается.

5 Заключение

Процедура двух односторонних тестов Шуирманна — одно из основных средств при проверке гипотезы биоэквивалентности. При исследовании критериев принятия гипотезы биоэквивалентности важную роль играет оценка вероятности наступления ошибки первого рода. Важность ее контроля обусловлена риском пациента получить препарат с несоответствующими эффективностью и безопасностью. В данной статье впервые дана оценка уровня значимости процедуры двух односторонних тестов Шуирманна при наличии пропущенных данных. В частности, в явном виде показана та ее часть, которая зависит от уровня полноты данных. В практическом плане данная оценка может быть использована для корректировки задаваемого уровня значимости при известном уровне полноты данных, чтобы обеспечить гарантированную эффективность и безопасность воспроизведенных лекарств.

Литература

1. Правила проведения исследований биоэквивалентности лекарственных средств Евразийского экономического союза. <http://www.eurasiancommission.org/ru/act/texnreg/deptexreg/konsultComitet/Documents/Правила%20БЭИ%20итор%2020.02.2015%20на%20сайт.pdf>.
2. *Schuirman D. J.* A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability // *J. Pharmacokinet. Biop.*, 1987. Vol. 15. P. 657–680.
3. *Chow Shein-Chung, Liu Jen-pei.* Design and analysis of bioavailability and bioequivalence studies. — Chapman & Hall/CRC, 2009. 735 p.
4. *Драницына М. А., Захарова Т. В., Ниязов Р. Р.* Свойства процедуры двух односторонних тестов для признания биоэквивалентности лекарственных препаратов // *Ремедиум. Журнал о рынке лекарств и медицинской техники*, 2019. № 3. С. 40–47.
5. *Berger R. L., Hsu J. C.* Bioequivalence trials, intersection-union tests and equivalence confidence sets // *Stat. Sci.*, 1996. Vol. 11. No. 4. P. 283–319.

Поступила в редакцию 09.05.19

EVALUATION OF THE SIGNIFICANCE LEVEL IN SCHUIRMANN'S TEST FOR CHECKING THE BIOEQUIVALENCE HYPOTHESIS IN MISSING DATA CONDITIONS

T. V. Zakharova^{1,2} and A. A. Tarkhov¹

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 bioequivalence hypothesis testing is the important task in pharmacokinetics. It helps to make a decision about the equivalence of the reproduced drug to the reference drug. One of the problems of bioequivalence studies is the availability of missing data. A small amount of data entails the inability to delete a data sample with missing data. Therefore, there is a task to estimate the impact of missing data on bioequivalence testing task, in particular, to estimate the significance level. The main method of the bioequivalence hypothesis testing is Schuirmann's two one-sided tests procedure. The article shows the significance level evaluation of this procedure in the case of missing data. The evaluation component, depending on the level of data completeness, is shown in the explicit form.

Keywords: bioequivalence; significance level; type I error; missing data; Schuirmann's two one-sided tests procedure

DOI: 10.14357/19922264190309

Acknowledgments

The paper was supported by the Russian Foundation for Basic Research (project 18-07-00252).

References

1. Pravila provedeniya issledovaniy bioekvivalentnosti lekarstvennykh sredstv Evraziyskogo ekonomicheskogo soyuza. Available at: <http://www.eurasiancommission.org/ru/act/texnreg/deptexreg/konsultComitet/Documents/Pravila%20BEI%20itog%2020.02.2015%20na%20sajt.pdf> (accessed May 8, 2019).
2. Schuirmann, D.J. 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biop.* 15:657–680.
3. Chow, Shein-Chung, and Jen-pei Liu. 2009. *Design and analysis of bioavailability and bioequivalence studies*. Chapman & Hall/CRC. 735 p.
4. Dranitsyna, M. A., T. V. Zakharova, and R. R. Niyazov. 2019. Svoystva protsedury dvukh odnostoronnykh testov dlya priznaniya bioekvivalentnosti lekarstvennykh preparatov [Properties of the two-sided tests procedure for the bioequivalence assessment of medical products]. *Remedium. Zh. o rynke lekarstv i meditsinskoy tekhniki* [Remedium: J. of the Market of Medicines and Medical Equipment] 2019(3):40–47.
5. Berger, R. L., and J. C. Hsu. 1996. Bioequivalence trials, intersection–union tests and equivalence confidence sets. *Stat. Sci.* 11(4):283–319.

Received May 9, 2019

Contributors

Zakharova Tatiana V. (b. 1962) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; 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; tvzaharova@mail.ru

Tarkhov Alexey A. (b. 1995) — master student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; alexeytarkhov@gmail.com

ФОРМАЛИЗАЦИЯ МЕТОДА РАНЖИРОВАНИЯ АЛЬТЕРНАТИВ ДЛЯ ПРОЦЕССА ГРУППОВОГО ПРИНЯТИЯ РЕШЕНИЙ ПРИ АНАЛИЗЕ СОЦИАЛЬНЫХ СЕТЕЙ*

А. А. Гайдамака¹, Н. В. Чухно², О. В. Чухно³, К. Е. Самуйлов⁴, С. Я. Шоргин⁵

Аннотация: Распространение и доступность интернет-технологий позволили по-новому взглянуть на социальные сети. Если раньше этот онлайн-сервис имел в большей степени развлекательный характер, то сегодня, с повышением скоростей передачи данных и появлением возможности общения пользователей в реальном времени, социальные сети, на платформе которых можно легко организовать опрос или провести голосование, становятся мощным механизмом достижения консенсуса в процессе принятия решений. В статье предложен обзор известных моделей группового принятия решений (*англ.* group decision making, GDM) и формальное описание разработанного на его основе алгоритма принятия решений, использующего большой объем данных в социальной сети. Особенностью алгоритма является возможность выбора экспертом из огромного числа альтернатив ограниченного набора, представляющего для него интерес, а также учет масштабирования сети с точки зрения числа экспертов, привлекаемых к процессу группового принятия решений. Представлен метод экстраполяции значений оценок при масштабировании сети. Разработан и исследован кейс для иллюстрации работы алгоритма.

Ключевые слова: групповое принятие решений; анализ социальных сетей; нечеткая логика; набор языковых выражений

DOI: 10.14357/19922264190310

1 Введение

С помощью интернет-технологий люди из разных стран мира могут присоединяться к социальным сетям, создавать новые ресурсы и контент и взаимодействовать с другими пользователями [1]. Социальные сети — одна из самых последних тенденций, обеспечивающих общение, взаимодействие, обмен информацией и связь пользователей в виртуальной среде [2].

В последние годы исследователи в области группового принятия решений проявили интерес к анализу социальных сетей, так как последние могут быть успешно использованы для моделирования процесса взаимодействия между лицами, принимающими решения. Тем не менее социальная сеть имеет некоторые особенности, которые отличают ее от классических сценариев GDM [1].

С одной стороны, внутри социальной сети взаимодействуют тысячи пользователей, однако многие

из них не принимают непосредственного участия в процессе голосования. Кроме того, пользователи общаются в режиме реального времени [3, 4]. С другой стороны, зачастую некоторые участники могут мешать достижению консенсуса, настаивая на своем варианте и не уступая другим экспертам.

Таким образом, групповое принятие решений — это, во-первых, процесс, состоящий в выборе наилучшей альтернативы или множества альтернатив из всех возможных, учитывая мнения, выраженные группой лиц [3, 5]. Во-вторых, это процесс, ориентированный на людей с их субъективностью и неопределенностью в выставлении оценок. Поэтому в исследованиях GDM, как правило, используется теория нечеткой логики (*англ.* fuzzy logic) [6–8].

Традиционные алгоритмы GDM рассматривают небольшую группу экспертов и альтернатив [5]. Однако сегодня чрезвычайно важно разрабатывать такие методы GDM, которые помогут пользователям принимать консенсусные решения [4, 9, 10].

* Публикация подготовлена при поддержке Программы РУДН «5-100» и при финансовой поддержке РФФИ (проекты 18-07-00576 и 18-00-01555 (18-00-01685)).

¹ Российский университет дружбы народов, aagajdamaka@sci.pfu.edu.ru

² Российский университет дружбы народов, nvchukhno@gmail.com

³ Российский университет дружбы народов, olgachukhno95@gmail.com

⁴ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samuylov-ke@rudn.ru

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

Такие методы могут быть использованы в новых услугах сети интернет, например при оценке удобства пользования социальной сетью [5, 9].

В [4] выявлены важные направления дальнейших исследований в рассматриваемой области. Одной из новых задач является разработка структур предпочтений и их применение в социальных сетях. Этой проблеме посвящена настоящая статья, которая организована следующим образом: в разд. 2 с использованием языка UML (Unified Modeling Language) формально описана модель группового принятия решений в социальной сети [11]. Раздел 3 демонстрирует модули алгоритма ранжирования альтернатив. В разд. 4 предлагается метод оценки альтернатив при масштабировании сети. Раздел 5 посвящен анализу кейса при оценке удобства пользования социальной сетью. В заключении подведены итоги и поставлены задачи дальнейших исследований.

2 Модель группового принятия решений в социальных сетях

Любям комфортнее выражать свои предпочтения с помощью слов. Однако компьютеры работают с числами. В связи с этим возникает проблема непонимания между человеком и вычислительной машиной: человеку может быть сложно передать свои идеи компьютеру, а компьютеру понять их и работать с полученной информацией [6, 12, 13]. Работа с неопределенностью всегда была сложной задачей, для решения которой применяются различные аналитические инструменты. В статье исследована лингвистическая модель принятия решений, в которой эксперты выставляют оценки альтернативам, используя набор языковых выражений (*англ.* linguistic term set, LTS) [7, 13, 14]. Подход основан на упорядочении лингвистических терминов в наборе LTS, использует индексирование и предполагает, что все выражения одинаково информативны и симметрично распределены. Каждый эксперт выбирает удобный для собственного оценивания набор LTS. В конечном счете для определения итоговой оценки все предоставленные предпочтения выражаются с помощью одного и того же набора языковых выражений, называемого базовым (*англ.* basic LTS, BLTS) [5, 13].

Обозначим символом e некоторого эксперта, а символом x некоторую альтернативу принимаемого решения. Пронумеруем экспертов и альтернативы. Тогда $E = \{e_1, \dots, e_K\}$ — множество экспертов, такое что $|E| = K < \infty$, где K — число экспертов, и $X = \{x_1, \dots, x_M\}$ — множество альтернатив, такое что $|X| = M < \infty$, где M — число альтернатив. $X_k \subseteq X$ — подмножество альтернатив, выбранных экспертом e_k , $e_k \in E$. Значение предпочтения эксперта e_k для альтернативы x_i по отношению к альтернативе x_j обозначим переменной $p_{ij}(k)$, $i \neq j$, $x_i, x_j \in X$.

Цель исследуемого процесса GDM — упорядочение альтернатив от наилучшей к наихудшей [4]. В статье рассматривается формальная модель процесса GDM, состоящая из двух модулей, как показано на рис. 1 в диаграмме, построенной на языке UML.

3 Модули алгоритма ранжирования альтернатив

Модуль 1 алгоритма «Оценивание альтернатив экспертами» отвечает за вычисление предпочтений экспертов и включает 7 шагов.

1. *Определение множества экспертов $E = \{e_1, \dots, e_K\}$ и множества альтернатив $X = \{x_1, \dots, x_M\}$.*

2. *Выбор экспертом подмножества альтернатив.*

Модель предусматривает, что эксперт e_k выставляет оценки для некоторого подмножества X_k из множества альтернатив X .

3. *Сравнение альтернатив.*

3.1 *Выбор LTS.* Пусть t — уровень иерархии набора языковых выражений LTS и $n(t)$ — число оценок по шкале оценок уровня t , $t = 1, \dots, T$, причем $n(t) < n(t+1)$. Обозначим $S(t) = \{s_1(t), \dots, s_{n(t)}(t)\}$ шкалу уровня t , а $s_l(t)$ символьную оценку из набора оценок этой шкалы, $l \in \{1, \dots, n(t)\}$, $|S(t)| = n(t)$. Считаем оценки упорядоченными по возрастанию значений предпочтений: $s_1(t) < s_2(t) < \dots < s_{n(t)}(t)$.

Заметим, что каждый эксперт самостоятельно делает выбор шкалы, с которой он будет работать. При этом эксперты, которые выбирают шкалу с высокой степенью детализации, вероятнее всего, готовы дать более точную оценку, а в случаях, когда эксперт не чувствует себя уверенным при сравнении альтернатив, он может пользоваться шкалами оценок уровня ниже. Тогда множество экспертов E представимо в виде

$$E = \bigcup_{t=1}^T E(t),$$

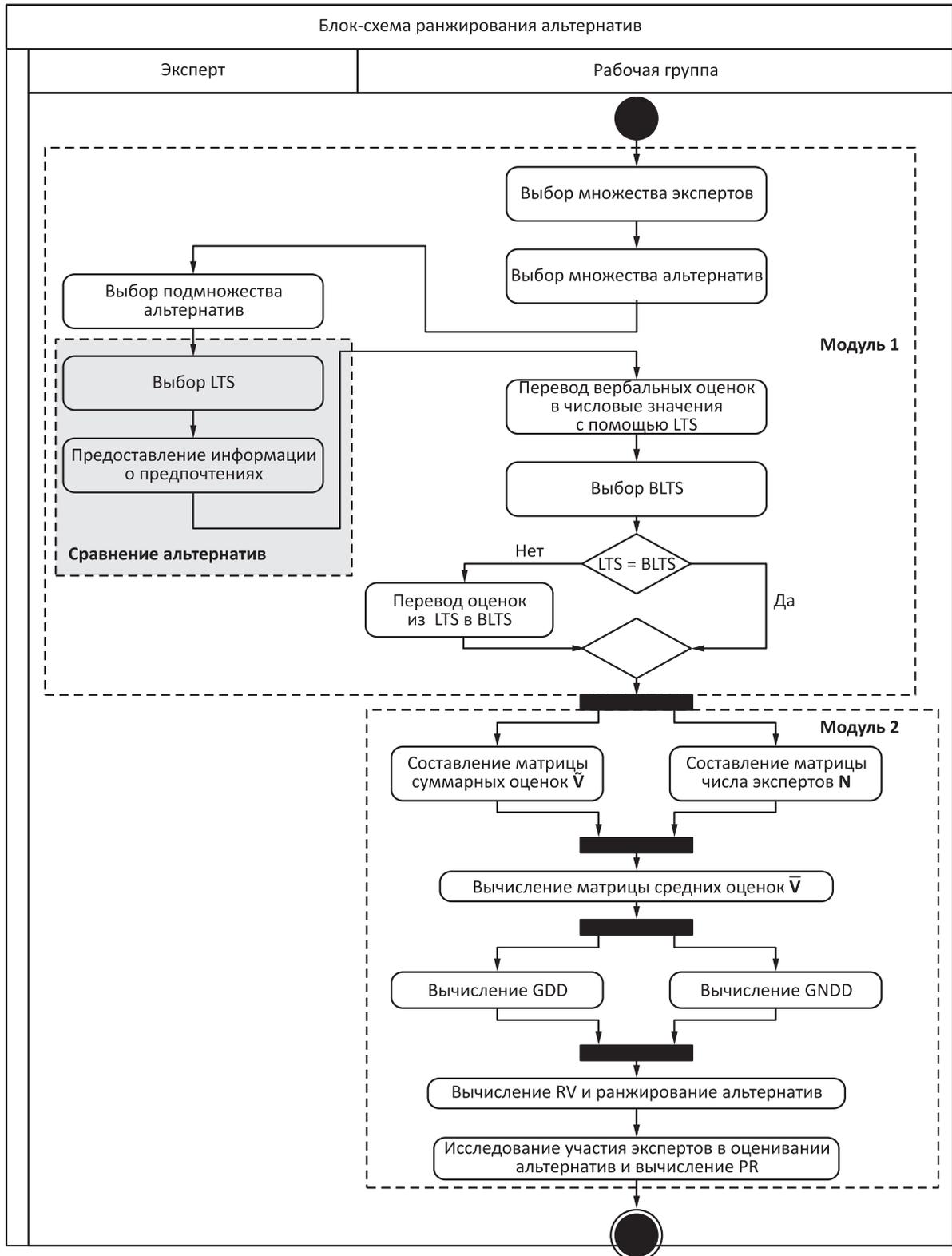


Рис. 1 Модель группового принятия решений — ранжирования альтернатив

где

$$E(t) = \{e_k \in E : \text{эксперт } e_k \text{ выбрал LTS } S(t)\}.$$

3.2 Предоставление информации. Для всех пар (x_i, x_j) , таких что $i \neq j, x_i, x_j \in X$, эксперт e_k указывает свою оценку $s_l(t)$ в качестве значения элемента $p_{ij}(k)$. Полученные значения формируют матрицу предпочтений $\mathbf{P}_k = (p_{ij}(k))_{i,j \in \{1, \dots, M\}}$, $p_{ij}(k) \in \{s_1(t), \dots, s_n(t)(t)\}$.

4. Перевод оценок эксперта e_k в численные значения.

Для перевода оценок может быть использован один из известных методов [8], выбор которого не существен с точки зрения представленных в данной статье результатов. Матрица предпочтений \mathbf{P}_k модифицируется в матрицу предпочтений $\mathbf{V}_k = (v_{ij}(k))_{i,j \in \{1, \dots, M\}}$, например путем замены для каждого элемента символьной оценки $s_l(t)$ на значение l .

5. Выбор базового набора языковых выражений BLTS.

В качестве базового может быть выбран любой из имеющихся наборов языковых выражений LTS, а по умолчанию выбирается шкала оценок с наибольшей детализацией. В предположении, что шкалы упорядочены по возрастанию числа оценок, т. е. $S(1) \prec S(2) \prec \dots \prec S(T)$, по умолчанию базовым набором является набор уровня T , а именно: $S(t') = S(T)$.

6. Перевод оценок из LTS в BLTS.

Пусть $B(t)$ — множество пар вида $(s_l(t), \alpha)$, таких что $s_l(t) \in S(t)$ и $\alpha \in [-0,5, 0,5)$, т. е.

$$B(t) = S(t) \times [-0,5, 0,5). \quad (1)$$

Определим функцию F , такую что

$$F : B(t) \rightarrow B(t').$$

Введем действительную величину $\beta(t') \in (0, n(t'))$. Величина $\beta(t')$ необходима для определения оценки $s_{l'}(t') \in S(t')$, в которую переводится оценка $s_l(t) \in S(t)$, поставленная экспертами по шкале $S(t)$. Заметим, что в (1) для значений α выбран интервал $[-0,5; 0,5)$ потому, что для получения числовой оценки l' величина $\beta(t')$ будет округляться до ближайшего целого по шкале $S(t')$. Чтобы оценке $s_l(t)$ по шкале $S(t)$ поставить в соответствие некоторое численное значение l' по шкале $S(t')$, необходимо определить положение соответствующей величины $\beta(t')$ на отрезке $(0, n(t'))$. При этом величиной α обозначено смещение величины $\beta(t')$ относительно ближайшего целого числа l' , соответствующего символьной оценке $s_{l'}(t') \in S(t')$. Заметим, что для шкалы $S(t)$ величина $\beta(t) = l$ принимает целые значения, $\beta(t) \in \{1, \dots, n(t)\}$, а смещение α равно нулю. На рис. 2 введенные обозначения проиллюстрированы для оценки $l = 2$ по шкале уровня $t = 3$, которой в базовом наборе языковых выражений уровня $t' = 5$ будет соответствовать оценка $l' = 3$.

Будем строить отображение F так, чтобы оно было взаимно однозначным. Поскольку F строится для наборов языковых выражений со шкалами $S(t)$ и $S(t')$, то смещение будем выбирать таким образом, чтобы выполнялось соотношение:

$$\frac{\beta(t')}{\beta(t)} = \frac{n(t')}{n(t)}. \quad (2)$$

Значение $\beta(t')$, соответствующее оценке $s_{l'}(t') \in S(t')$, должно быть получено из (2) по известной после шага 3 модуля 1 оценке $s_l(t) \in S(t)$.

Введем отображение Δ следующим образом:

$$\Delta(\beta(t')) = (s_{l'}(t'), \alpha).$$

Очевидно, что $\alpha = \beta(t') - [\beta(t')]$, $l' = [\beta(t')]$, поэтому, положив

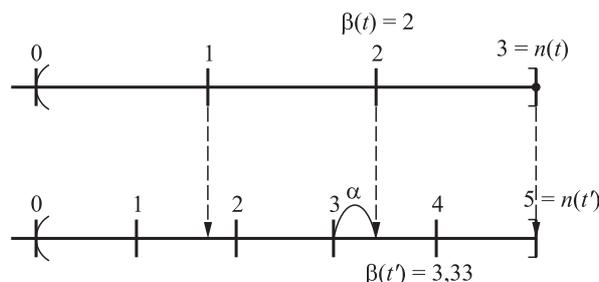


Рис. 2 Иллюстрация величины $\beta(t)$ при переводе оценок из шкалы $S(t)$ в шкалу $S(t')$

$$\Delta(\beta(t')) = (s_{[\beta(t')]}(t'), \beta(t') - [\beta(t')]) ,$$

имеем:

$$\beta(t') = \Delta^{-1}(s_{[\beta(t')]}(t'), \beta(t') - [\beta(t')]) .$$

Поскольку $F = \Delta(\beta(t'))$, $t' = 1, \dots, T$, то получена формула для перевода оценок из шкалы $S(t)$ в $S(t')$ и доказана следующая теорема.

Теорема 1. Если $S(t) \prec S(t')$ и $\beta(t')/\beta(t) = n(t')/n(t)$, то отображение $F : B(t) \rightarrow B(t')$ является взаимно однозначным и определяется формулой:

$$F(s_{\beta(t)}(t), 0) = \Delta(\beta(t')) = (s_{[\beta(t')]}(t'), \beta(t') - [\beta(t')]) ,$$

где

$$\beta(t') = \frac{\beta(t)n(t')}{n(t)} .$$

7. Обновление матрицы предпочтений экспертов.

По матрице предпочтений \mathbf{V}_k должна быть вычислена матрица предпочтений $\tilde{\mathbf{V}}_k = (\tilde{v}_{ij}(k))_{i,j \in 1, \dots, M}$, полученная с помощью алгоритма перевода численных оценок из LTS в BLTS. Результатом работы модуля 1 алгоритма являются матрицы предпочтений экспертов, представленных в численном виде в едином базовом наборе языковых выражений.

Алгоритм перевода численных оценок из LTS в BLTS

Data: Численные оценки экспертов

$$v_{ij}(k) = 1, \dots, n(t), e_k \in E(t), S(t), S(t')$$

procedure $\mathbf{V}_k \rightarrow \tilde{\mathbf{V}}_k$

for all $e_k, k = 1$ to K

if $S(t) \neq S(t')$ **do**

for all $v_{ij}(k)_{i,j \in 1, \dots, M}$ **do**

$$\beta(t) = v_{ij}(k)$$

$$\beta(t') = \frac{\beta(t)n(t')}{n(t)}$$

$$\tilde{v}_{ij}(k) = \beta(t')$$

end for

else $\tilde{v}_{ij}(k) = v_{ij}(k)$

end if

end for

Вывести новые оценки экспертов $\tilde{v}_{ij}(k)$

end procedure $\mathbf{V}_k \rightarrow \tilde{\mathbf{V}}_k$

Модуль 2 алгоритма «Расчет средних оценок, создание рейтинга альтернатив и анализ степени участия экспертов» отвечает за вычисление матрицы усредненных оценок, полученных на основании матриц предпочтений экспертов $\tilde{\mathbf{V}}_k$, а также

за формирование ранжированного списка альтернатив. Матрица $\tilde{\mathbf{V}} = (\tilde{v}_{ij})_{i,j \in 1, \dots, M}$ содержит сумму всех значений предпочтений, предоставленных экспертами:

$$\tilde{\mathbf{V}} = \sum_{k=1}^K \tilde{\mathbf{V}}_k . \quad (3)$$

Матрица \mathbf{N}_k содержит информацию об альтернативах, выбранных для сравнения экспертом e_k :

$$n_{ij}(k) = \begin{cases} 1, & x_i, x_j \in X_k; \\ 0, & x_i, x_j \notin X_k. \end{cases} \quad (4)$$

Тогда матрица $\mathbf{N} = (n_{ij})_{i,j \in 1, \dots, M}$, содержащая информацию о числе экспертов, участвующих в попарном сравнении альтернатив x_i и x_j , может быть вычислена по формуле:

$$\mathbf{N} = \sum_{k=1}^K \mathbf{N}_k . \quad (5)$$

Элементы матрицы усредненных оценок $\bar{\mathbf{V}} = (\bar{v}_{ij})_{i,j \in 1, \dots, M}$ вычисляются следующим образом:

$$\bar{v}_{ij} = \frac{\tilde{v}_{ij}}{n_{ij}} . \quad (6)$$

Для ранжирования альтернатив используются, например, операторы GDD (*англ.* Guided Dominance Degree) и GNDD (*англ.* Guided Non Dominance Degree) [10, 14]. При этом оператор GDD показывает, насколько оцениваемая альтернатива доминирует над всеми остальными, т.е. насколько эта альтернатива лучше всех остальных, а оператор GNDD — насколько оцениваемая альтернатива не доминируется остальными. Усредняя значения GDD и GNDD, получаем значение ранжирования RV (*англ.* Ranking Value):

$$RV_i = \frac{GDD_i + GNDD_i}{2}, \quad i = 1, \dots, M, \quad (7)$$

где

$$GDD_i = \sum_{j=1}^M \bar{v}_{ij}, \quad i = 1, \dots, M;$$

$$GNDD_i = \sum_{j=1}^M \max\{\bar{v}_{ji} - \bar{v}_{ij}, 1\}, \quad i = 1, \dots, M.$$

Из значений ранжирования формируется рейтинг альтернатив по возрастанию степени значимости — предпочтительнее будет альтернатива, которая имеет более высокое значение RV.

Заключительный шаг модуля 2 процесса GDM посвящен оценке достоверности полученных результатов. Коэффициент участия (*англ.* Participation Rate) служит одним из ключевых показателей эффективности процесса группового принятия решений. Высокий коэффициент свидетельствует о том,

что полученные оценки достаточно точны, а результаты процесса GDM надежны. Так как в процессе принятия решения эксперты сравнивают не все альтернативы, а только выбранное ими подмножество альтернатив, то данный коэффициент необходимо вычислить для каждой альтернативы:

$$PR_i = \frac{1}{M} \sum_{j=1}^M n_{ij}, \quad i = 1, \dots, M. \quad (8)$$

Общий коэффициент участия (*англ.* the overall participation rate) определим, как среднее по всем PR_i , т. е.

$$PR = \frac{1}{M} \sum_{i=1}^M PR_i.$$

Результатом работы модуля 2 алгоритма является ранжированный список альтернатив и коэффициенты участия для каждой альтернативы.

4 Метод оценки альтернатив при масштабировании сети

Привлечение дополнительного эксперта к процессу оценивания альтернатив не вызывает необходимости пересчитывать усредненные оценки по

формулам (3)–(6). Пусть \tilde{V}_{K+1} — матрица предпочтений эксперта e_{K+1} . Тогда матрица усредненных оценок \bar{V} для нового набора экспертов $E_{K+1} = E_K \cup \{e_{K+1}\}$ может быть вычислена по формуле:

$$\bar{v}_{ij}(K+1) = \frac{1}{n_{ij}(K+1)} \tilde{v}_{ij}(K+1) + \left(1 - \frac{1}{n_{ij}(K+1)}\right) \bar{v}_{ij}(K).$$

5 Численный эксперимент

При выполнении численного эксперимента был проведен опрос с целью определения, какая из 27 социальных сетей, представленных на рис. 3, наиболее удобна для общения по мнению 20 экспертов.

С использованием изложенной выше методики и алгоритма было разработано программное средство для расчета значений ранжирования по формуле (7), проиллюстрированных на рис. 3.

По формуле (8) для каждой из 27 альтернатив был рассчитан коэффициент участия PR_i , показанный в таблице. Из таблицы видно, что наиболее удобными для общения, по мнению 20 экспертов из России, оказались сети Vkontakte (альтернатива x_1)

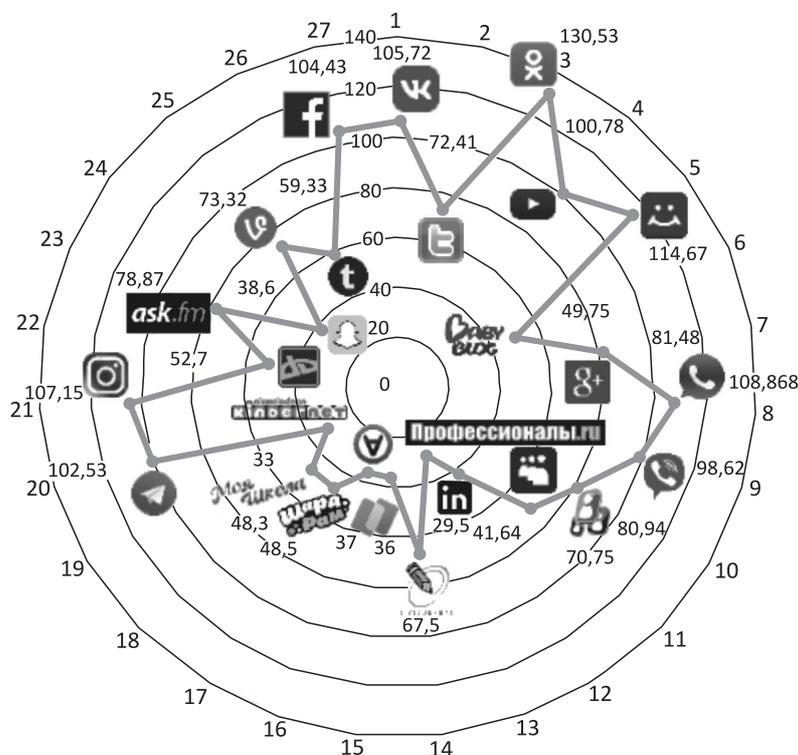


Рис. 3 Лепестковая диаграмма — рейтинг альтернатив

Коэффициент участия
каждой альтернативы

x_i	PR_i
x_1	3
x_2	1,33333
x_3	1,62963
x_4	3
x_5	1,66667
x_6	0,62963
x_7	1,07407
x_8	2,62963
x_9	2,18519
x_{10}	0,92593
x_{11}	0,74074
x_{12}	0,37037
x_{13}	0,18519
x_{14}	0,74074
x_{15}	0,22222
x_{16}	0,22222
x_{17}	0,37037
x_{18}	0,37037
x_{19}	0,18519
x_{20}	1,92593
x_{21}	2,33333
x_{22}	0,62963
x_{23}	1
x_{24}	0,22222
x_{25}	1
x_{26}	0,88889
x_{27}	1,44444

и YouTube (альтернатива x_4), а сеть Kindernet (альтернатива x_{19}) наименее удобна. Общий коэффициент участия $PR = 1,1454$, т.е. в среднем всего 1 эксперт предоставил информацию для каждой пары альтернатив. Поскольку коэффициент участия достаточно низок, полученные результаты не являются надежными. Для получения точных результатов следует привлечь к процессу GDM большее число экспертов.

6 Заключение

С каждым годом пользователи сети интернет становятся более вовлеченными в процесс предоставления информации, что приводит к созданию в рамках социальных сетей среды хранения огромного объема доступных каждому пользователю данных. В статье представлен новый метод группового принятия решений, который предназначен для работы с такими типами сред. Задачей дальнейших исследований является применение метода в масштабных численных экспериментах. При этом

помимо вычисления усредненных величин планируется также оценивать другие показатели эффективности процесса принятия решений, в том числе старшие моменты исследуемых характеристик. Еще одной важной задачей является измерение и оценка достигнутого экспертами консенсуса, поскольку при невысокой степени согласованности экспертов полученный результат нельзя считать надежным.

Литература

1. Wasserman S., Faust K. Social network analysis: Methods and applications. — Cambridge: Cambridge University Press, 1994. Vol. 8. 857 p.
2. Wu J., Xiong R., Chiclana F. Uninorm trust propagation and aggregation methods for group decision making in social network with for tuple information // Knowl.-Based Syst., 2016. Vol. 96. P. 29–39.
3. Urea R., Cabrerizo F. J., Morente-Molinera J. A., Herrera-Viedma E. GDM-R: A new framework in R to support fuzzy group decision making processes // Inform. Sciences, 2016. Vol. 357. P. 161–181.
4. Herrera-Viedma E., Cabrerizo F., Chiclana F., Wu J., Cobo M., Samuylov K. Consensus in group decision making and social networks // Stud. Inform. Control, 2017. Vol. 26. Iss. 3. P. 259–268.
5. Herrera-Viedma E., Cabrerizo F. J., Kacprzyk J., Pedrycz W. A review of soft consensus models in a fuzzy environment // Inform. Fusion, 2014. Vol. 17. P. 4–13.
6. Zadeh L. A. Fuzzy sets // Inform. Control, 1965. Vol. 8. Iss. 3. P. 338–353.
7. Zadeh L. A. Fuzzy logic = computing with words // IEEE T. Fuzzy Syst., 1996. Vol. 4. Iss. 2. P. 103–111.
8. Torra V. Hesitant fuzzy sets // Int. J. Intell. Syst., 2010. Vol. 25. Iss. 6. P. 529–539.
9. Kacprzyk J. Group decision making with a fuzzy linguistic majority // Fuzzy Set. Syst., 1986. Vol. 18. P. 105–118.
10. Chiclana F., Tapia Garcia J., del Moral M. J., Herrera-Viedma E. A statistical comparative study of different similarity measures of consensus in group decision making // Inform. Sciences, 2013. Vol. 221. P. 110–123.
11. Самуйлов К. Е., Серебренникова Н. В., Чукарин А. В., Яркина Н. В. Основы формальных методов описания бизнес-процессов. — М.: РУДН, 2011. 130 с.
12. Tanino T. Fuzzy preference orderings in group decision making // Fuzzy Set. Syst., 1984. Vol. 12. P. 117–131.
13. Herrera F., Martinez L. A. 2-tuple fuzzy linguistic representation model for computing with words // IEEE T. Fuzzy Syst., 2000. Vol. 8. Iss. 6. P. 746–752.
14. Herrera F., Herrera-Viedma E. Choice functions and mechanisms for linguistic preference relations // Eur. J. Oper. Res., 2000. Vol. 120. Iss. 1. P. 144–161.

Поступила в редакцию 01.07.19

FORMALIZATION OF THE ALTERNATIVES RANKING METHOD FOR GROUP DECISION MAKING IN SOCIAL NETWORKS

A. A. Gaidamaka¹, N. V. Chukhno¹, O. V. Chukhno¹, K. E. Samouylov^{1,2}, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia (RUDN University), 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 expansion and accessibility of Internet technology has allowed a new look at social networks. A few decades ago, this online service was more entertaining in nature. However, today, with increasing transmission rates and the possibility of real-time communication, social networks, on which platform a poll or vote can be easily organized, become a powerful mechanism for achieving consensus in decision making process. The paper offers an overview of the known models of group decision making (GDM) and a formal description of the decision-making algorithm developed on the basis of the overview taking into account a large amount of data in a social network. A feature of the algorithm is the ability for an expert to choose a limited subset of interest from a huge number of alternatives, as well as taking into account network scaling in terms of the number of experts involved in the GDM process. The method of extrapolating the values of the estimates for network scaling is proposed. A numerical case for an illustration of the algorithm is developed and presented.

Keywords: group decision making; social network analysis; fuzzy logic; LTS

DOI: 10.14357/19922264190310

Acknowledgments

The publication has been prepared with the support of the "RUDN University Program 5-100" and funded by the Russian Foundation for Basic Research according to the research projects Nos. 18-07-00576 and 18-00-01555 (18-00-01685).

References

1. Wasserman, S., and K. Faust. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press. Vol. 8. 857 p.
2. Wu, J., R. Xiong, and F. Chiclana. 2016. Uninorm trust propagation and aggregation methods for group decision making in social network with for tuple information. *Knowl.-Based Syst.* 96:29–39.
3. Urea, R., F.J. Cabrerizo, J.A. Morente-Molinera, and E. Herrera-Viedma. 2016. GDM-R: A new framework in R to support fuzzy group decision making processes. *Inform. Sciences* 357:161–181.
4. Herrera-Viedma, E., F. Cabrerizo, F. Chiclana, J. Wu, M. Cobo, and K. Samouylov. 2017. Consensus in group decision making and social networks. *Stud. Inform. Control* 26(3):259–268.
5. Herrera-Viedma, E., F.J. Cabrerizo., J. Kacprzyk, and W. Pedrycz. 2014. A review of soft consensus models in a fuzzy environment. *Inform. Fusion* 17:4–13.
6. Zadeh, L. A. 1965. Fuzzy sets. *Inform. Control* 8(3):338–353.
7. Zadeh, L. A. 1996. Fuzzy logic = computing with words. *IEEE T. Fuzzy Syst.* 4(2):103–111.
8. Torra, V. 2010. Hesitant fuzzy sets. *Int. J. Intell. Syst.* 25(6):529–539.
9. Kacprzyk, J. 1986. Group decision making with a fuzzy linguistic majority. *Fuzzy Set. Syst.* 18:105–118.
10. Chiclana, F., J. Tapia Garcia, M.J. del Moral, and E. Herrera-Viedma. 2013. A statistical comparative study of different similarity measures of consensus in group decision making. *Inform. Sciences* 221:110–123.
11. Samouylov, K. E., N. V. Serebrennikova, A. V. Chukarin, and N. V. Jarkina. 2011. *Osnovy formal'nykh metodov opisaniya biznes-protsessov* [Basics of formal methods of describing business processes]. Moscow: RUDN. 130 p.
12. Tanino, T. 1984. Fuzzy preference orderings in group decision making. *Fuzzy Set. Syst.* 12:117–131.
13. Herrera, F., and L. A. Martinez. 2000. 2-tuple fuzzy linguistic representation model for computing with words. *IEEE T. Fuzzy Syst.* 8(6):746–752.
14. Herrera, F., and E. Herrera-Viedma. 2000. Choice functions and mechanisms for linguistic preference relations. *Eur. J. Oper. Res.* 120(1):144–161.

Received July 1, 2019

Contributors

Gaidamaka Anna A. (b. 1997) — Master Student of Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; aagajdamaka@sci.pfu.edu.ru

Chukhno Nadezhda V. (b. 1995) — Master Student of Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; nvchukhno@gmail.com

Chukhno Olga V. (b. 1995) — Master Student of Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; olgachukhno95@gmail.com

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

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

ВЫЧИСЛЕНИЕ ПОКАЗАТЕЛЕЙ ОПТИМАЛЬНЫХ ПО CC-VaR ПОРТФЕЛЕЙ НА РЫНКАХ ОПЦИОНОВ*

Г. А. Агасандян¹

Аннотация: Работа продолжает исследования автора по проблеме применения на финансовых рынках непрерывного критерия VaR (CC-VaR). Речь идет о проецировании идей и методов, разработанных для задачи инвестирования на теоретическом однопериодном рынке с одним базовым активом, на дискретный по страйкам рынок опционов с небольшим числом сценариев. Основное внимание уделяется методам расчета функций распределения дохода и доходности, среднего дохода и средней доходности для оптимального по CC-VaR портфеля опционов и его рандомизированных версий — как полных, так и частичных. Предлагаются эффективные по скорости и точности алгоритмы вычислений для разных вариантов задач оптимизации. Исследование иллюстрируется на примере с бета-распределенными рыночными ценами базового актива и прогнозом инвестора, сопровождаемом графиками.

Ключевые слова: непрерывный критерий VaR (CC-VaR); функция рискованных предпочтений (ф.р.п.); сценарии; опционы; процедура Неймана–Пирсона; индикаторы; баттерфляи; полная и частичная рандомизации; оптимальный портфель; доход; доходность

DOI: 10.14357/19922264190311

1 Введение

Работа продолжает исследования автора по проблеме применения на финансовых рынках непрерывного критерия VaR (CC-VaR) [1, 2]. Речь идет о проецировании идей и методов, разработанных для задачи инвестирования на теоретическом однопериодном рынке с одним базовым активом, на дискретный рынок с небольшим числом сценариев (уже близкий к реальному).

В значительной мере вопросы такого проецирования на сценарные рынки были изучены уже в [1–4]. Здесь основное внимание уделяется распространению аналогичных методов на дискретные по страйкам *опционные* рынки и их рандомизированные версии. Обсуждается инструментарий опционных рынков. Для него предлагается базис из простейших нормированных баттерфляев, в котором можно строить инвестиционные портфели.

Приводятся формулы расчета основных числовых показателей оптимальных портфелей и вычислительные алгоритмы, специально приспособленные к части возникающих при этом задач и демонстрирующие свою эффективность. Исследование иллюстрируется на примере с бета-распределенными рыночными ценами базового актива и прогнозом инвестора. Приводятся графики функций распределения дохода для вариантов инвестиционного портфеля.

2 Теоретический однопериодный δ -рынок и опционы

Напомним основные понятия и обозначения для теоретического однопериодного δ -рынка и распространим его конструкции на опционные инструменты. Известная на начало периода цена базового актива X в его конце образует случайную величину X , принимающую значения x из непрерывного множества $X \subset \mathfrak{R}_+$ (или даже \mathfrak{R}). На рынке обращаются δ -инструменты $D(s)$, $s \in X$, платежная функция которых равна δ -функции относительно s :

$$\pi(x; D(s)) \equiv \delta(x - s).$$

Заданы *прогнозная* $p(x)$ и *стоимостная* $c(x)$, $x \in X$, плотности. На рынке, называемом δ -рынком, можно торговать любым инструментом G с доходом, представимым в виде произвольной неотрицательной измеримой функции $g(x)$, $x \in X$. Ее называем *платежной функцией* инструмента и обозначаем $\pi(x; G)$, $x \in X$, т. е. $g(x) = \pi(x; G)$; также $\pi(x; X) = x$. Цена инструмента G обозначается $|G|$, а средний доход от него — $\|G\|$.

В частности, инструментами δ -рынка являются индикаторы множеств $H\{M\}$, $M \subset X$, играющие роль *базисных* инструментов для *сценарных* рынков. Аналогичную роль на рассматриваемых далее рын-

* Работа выполнена при финансовой поддержке РФФИ (проект 17-01-00816).

¹ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, agasand17@yandex.ru

ках обычных опционов (коллов и путов) играют простейшие нормированные баттерфляи. В терминах теоретического δ -рынка основные опционы, такие как коллы C_s , путы P_s , спрэды быка S_s^+ и медведя S_s^- , баттерфляи B_s , $s \in X$, их платежные функции и теоремы паритета имеют представления:

$$C_s = \int_{\{x \geq s\}} (x - s)D(x) dx, \quad \pi(x; C_s) = \max(0, x - s);$$

$$P_s = \int_{\{x < s\}} (s - x)D(x) dx, \quad \pi(x; P_s) = \max(0, s - x);$$

$$C_s - P_s = X - sU;$$

$$S_{s;h}^+ = C_{s-h} - C_s = hU + P_{s-h} - P_s, \\ \pi(x; S_{s;h}^+) = \min(h, \max(0, x - (s - h))); \quad (1)$$

$$S_{s;h}^- = P_{s+h} - P_s = hU + C_{s+h} - C_s, \\ \pi(x; S_{s;h}^-) = \min(h, \max(0, (s + h) - x)); \quad (2)$$

$$B_{s;h} = S_{s;h}^+ + S_{s;h}^- = C_{s-h} - 2C_s + C_{s+h} = \\ = P_{s-h} - 2P_s + P_{s+h} = \\ = hU + P_{s-h} - P_s - C_s + C_{s+h}, \\ \pi(x; B_{s;h}) = \max(0, h - |x - s|), \quad (3)$$

где s — страйк колла и пута, короткого колла или пута у спрэда и центральный у баттерфляя; h — расстояние между соседними страйками, $h > 0$. (При необходимости формируются и несимметричные баттерфляи.)

Из специально нормированных баттерфляев можно строить сходящиеся к δ -инструментам последовательности, как это делалось в [4] из индикаторов на сценарном рынке. И, наконец, CC-VaR требует, чтобы порождаемый портфелем инвестора доход q удовлетворял неравенствам:

$$P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon \text{ для всех } \varepsilon \in [0, 1].$$

Здесь $P\{M\}$ — вероятностная мера множества M в соответствии с прогнозом инвестора; $\phi(\varepsilon)$ — неотрицательная монотонно возрастающая и непрерывная функция рискованных предпочтений (ф.р.п.) инвестора со значениями в $[0, 1]$.

3 Опционная дискретизация δ -рынка

Напомним [2, 4], что сценарная дискретизация рынка вводится разбиением множества $X = [x_0, x_n)$

на n сценариев $S_i = [x_{i-1}, x_i) \subset X$, $x_{i-1} < x_i$, $i \in I = \{1, \dots, n\}$. Равномерное разбиение выделяется правилом $x_i = x_0 + ih$, $h = (x_n - x_0)/n$, $i \in I$. Рыночные цены базисных инструментов $D_i = H\{S_i\}$, $i \in I$, и прогнозные вероятности сценариев образуют векторы c^S и p^S с компонентами

$$c_i^S = |D_i| = \int_{x_{i-1}}^{x_i} c(x) dx; \\ p_i^S = \|D_i\| = \int_{x_{i-1}}^{x_i} p(x) dx, \quad i \in I. \quad (4)$$

На сценарной основе строится и опционная дискретизация рынка, для которой страйки опционов совпадают с серединами сценариев. С ней связывается также иная система интервалов:

$$K = \{K_i = [\kappa_i, \kappa_{i+1}), \quad i \in I_0\}, \quad (5)$$

где $I_0 = \{0\} \cup I$; κ_i , $i \in I$, — страйки опционов:

$$\kappa_i = \frac{x_{i-1} + x_i}{2} = \frac{i}{n} - \frac{1}{2n},$$

также доопределяем $\kappa_0 = a$, $\kappa_{n+1} = b$. Новые элементы порождают вероятностный вектор p^K с компонентами

$$p_i^K = \int_{\kappa_i}^{\kappa_{i+1}} p(x) dx, \quad i \in I_0, \quad \sum_{i \in I_0} p_i^K = 1. \quad (6)$$

Применяются еще обозначения:

$$I' = I \setminus \{n\} = \{1, \dots, n-1\}; \quad I'' = \{2, \dots, n-2\}.$$

На опционном рынке роль базисных инструментов играют нормированные баттерфляи B_i , $i \in I'$, вместе с двумя крайними нормированными спрэдами. Ограничимся тремя вариантами базисов: (i) только из коллов, (ii) только из путов и (iii) смешанным так называемого естественного происхождения, а именно (индекс i в обозначениях опционов соответствует страйку s_i , $i \in I$):

- (i) $B_1 = U + (C_2 - C_1)/h$, $B_i = (C_{i-1} - 2C_i + C_{i+1})/h$, $i = 2, \dots, n-1$, $B_n = (C_{n-1} - C_n)/h$;
- (ii) $B_1 = (P_2 - P_1)/h$, $B_i = (P_{i-1} - 2P_i + P_{i+1})/h$, $i = 2, \dots, n-1$, $B_n = U + (P_{n-1} - P_n)/h$;
- (iii) выделяется некоторый внутренний страйк, центр рынка, с индексом ν , и для него используется смешанный баттерфляй $B_\nu = U + (P_{\nu-1} - P_\nu - C_\nu + C_{\nu+1})/h$, остальные базисные инструменты с меньшими индексами состоят только из путов, с большими — только из коллов.

Крайние спрэды в базисе также называем (*усеченными*) *баттерфляями*.

Платежные функции базисных инструментов, которым даются специальные обозначения $\beta_i(x)$, $i \in I$, в соответствии с (1)–(3) принимают вид:

$$\begin{aligned} \beta_1(x) &= \max \left(0, \min \left(1, 1 + \frac{s_1 - x}{h} \right) \right); \\ \beta_n(x) &= \max \left(0, \min \left(1, 1 - \frac{s_n - x}{h} \right) \right); \\ \beta_i(x) &= \max \left(0, 1 - \frac{|s_i - x|}{h} \right), \quad i = 2, \dots, n - 1. \end{aligned}$$

Легко видеть, что для всех опционных базисных инструментов, как и для сценарных, интегралы от их платежных функций равны h . Кроме того,

$$\sum_{i \in I} D_i = \sum_{i \in I} B_i = U.$$

Интегрированием платежных функций по мерам с плотностями $c(x)$ и $p(x)$ находятся векторы c^B и p^B с компонентами

$$\begin{aligned} c_i^B &= |B_i| = \int_x \beta_i(x) c(x) dx; \\ p_i^B &= \|B_i\| = \int_x \beta_i(x) p(x) dx, \quad i \in I. \end{aligned} \quad (7)$$

Величина c_i^B означает стоимость i -го базисного баттерфляя, p_i^B — средний доход, который также называем *суррогатной* вероятностью i -го сценария.

4 Алгоритм оптимизации на рынке опционов

Для опционного рынка нахождение весов оптимального портфеля реализует практически тот же дискретный алгоритм, основанный на процедуре Неймана–Пирсона [5], что и для сценарного рынка, только вместо векторов p^S и c^S применяются векторы p^B и c^B , хотя назначение вероятностей производится тем же вектором p^S . Подчеркнем, что сначала речь идет о портфеле с детерминированными весами. Имеем (в сокращенном виде):

$\rho = p^B/c^B$ — вектор (длины n) относительных доходов;

$\xi = O(\rho)$ — вектор, задающий на множестве сценариев позиции компонент вектора ρ в порядке их возрастания;

$\eta = O(\xi)$ — обратный к ξ вектор, задает порядок возрастания компонент ρ ;

$d = p^S(\xi)$ — суперпозиция векторов p^S и ξ , ее компоненты упорядочены по возрастанию компонент вектора ρ (подчеркнем использование вектора p^S);

$T = [t_{ij} = \{1, i \leq j; t_{ij} = 0, i > j\}]$ — треугольная матрица, применяемая для последовательного суммирования компонент векторов, начиная с первой;

$\varepsilon = Td$ — вектор кумулятивных вероятностей для вектора d ; в дополнение к вектору ε принимается также $\varepsilon_0 = 0$.

На основе вектора ε вводятся и последовательные полуинтервалы:

$$\begin{aligned} E_i &= [\underline{\varepsilon}_i, \bar{\varepsilon}_i], \quad \underline{\varepsilon}_i = \varepsilon_{\eta_{i-1}} (= 0, \eta_i = 1), \\ \bar{\varepsilon}_i &= \varepsilon_{\eta_i}, \quad \bar{\varepsilon}_i - \underline{\varepsilon}_i = p_i^S, \quad i \in I. \end{aligned} \quad (8)$$

При назначении вектора b (и $g = b(\eta)$) весов портфеля предлагается вариант модификации ε_{opt} вектора ε , для чего вводится новый вектор d' :

$$\left. \begin{aligned} \varepsilon_{\text{opt}} &= \underline{\varepsilon}(\xi) + d', \quad d'_i = d_i - \frac{i-1}{n-1}, \quad i \in I; \\ b &= \phi(\varepsilon_{\text{opt}}); \quad g = b(\eta). \end{aligned} \right\} \quad (9)$$

Этот вариант связан с опционным характером платежной функции, распространяется на всю область ее значений $[0, 1)$, но применяется только для детерминированной схемы. В большей мере это оправдано, когда минимум портфельного дохода достигается на внутренних сценариях.

Основными *числовыми* показателями портфелей служат его стоимость A (средняя, если применяется рандомизация), средний доход R , средний относительный доход R/A , средняя доходность y и стандартное отклонение δ доходности. Имеем $((u, v)$ — скалярное произведение векторов u и v) запись результатов:

$$\begin{aligned} A^B &= (g, c^B) = (b, c^B(\xi)); \\ R^B &= (g, p^B) = (b, p^B(\xi)); \\ y &= \frac{R}{A} - 1. \end{aligned}$$

В иллюстративных расчетах и при построении графиков на протяжении большей части изложения принимается, что плотности $p(\cdot)$ и $c(\cdot)$ подчиняются бета-распределению с параметрами $(2, 1, 5)$ и $(3, 2, 5)$ соответственно:

$$\begin{aligned} p(x) &= x^2 \frac{(1-x)^{1,5}}{B(3, 2, 5)}; \quad c(x) = x \frac{(1-x)^{0,5}}{B(2, 1, 5)}, \\ x \in X &= [0, 1), \end{aligned} \quad (10)$$

где $B(\cdot, \cdot)$ — бета-функция, $B(3, 2, 5) = 19,6875$, $B(2, 1, 5) = 3,75$; $x_0 = 0$; $x_n = 1$. Эту задачу считаем

прямой, но при проведении некоторого сравнительного анализа приводятся и результаты решения обратной задачи, в которой плотности меняются ролями.

В качестве ф.р.п. инвестора принимается функция $\phi(\varepsilon) = \varepsilon^2$, $\varepsilon \in [0, 1]$, и потому функция распределения доходов оптимального теоретического портфеля [2, 4]

$$F_0(z) = \phi^{\leftarrow}(z) = z^{1/2}, \quad z \in [0, 1].$$

Приближенные значения показателей теоретической континуальной модели, с которыми будем соотносить получаемые далее результаты, получены на дискретной модели с 2000 сценариев [4]:

$$A_0 = 0,265439; \quad R_0 = 0,333729; \\ y_0 = 0,257275; \quad \sigma_0 = 1,1242.$$

Напомним [2], что для показателя R_0 в случае $\phi(\varepsilon) = \varepsilon^\lambda$ точное теоретическое значение равно $(1 + \lambda)^{-1}$, что при $\lambda = 2$ дает $1/3 \approx 0,333333$.

На рис. 1 приведены графики важных для оптимизации функций относительных доходов $\rho(x) = p(x)/c(x)$ в прямой задаче с плотностями (10) (1) и обратной (2), в которой плотности меняются ролями.

Внешний вид этих графиков говорит о том, что инвестор сталкивается с игрой на волатильности. По рыночной терминологии речь идет о ее продаже и покупке соответственно, правда, к ней примешивается и некоторая слабо выраженная игра на понижение в первом случае и на повышение — во втором.

В примерах используется дискретизация, как и в [4], с $n = 5$ сценариями как в сценарной, так и в опционной версиях. (Намеренно выбирается столь малое число сценариев, чтобы высветить

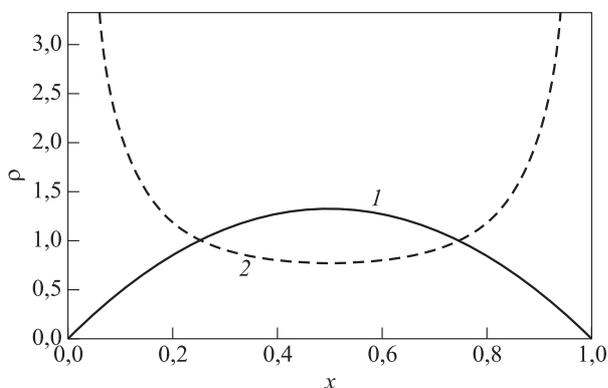


Рис. 1 Графики функции ρ в прямой (1) и обратной (2) задачах

проблемы, связанные именно с дискретизацией, поскольку при возрастании n решение все более подходит на теоретическое для континуальной схемы.)

Интегрированием (4) находятся сценарные векторы:

$$p^S = \{0,0412; 0,206; 0,341; 0,309; 0,104\}; \\ c^S = \{0,070; 0,187; 0,263; 0,284; 0,197\}.$$

Применение алгоритма с p^S и c^S дает векторы сценарного рынка:

$$\xi = \{5, 1, 4, 2, 3\}; \quad \eta = \{2, 4, 5, 3, 1\}; \\ \varepsilon = \{0, 104; 0,145; 0,454; 0,659; 1\}.$$

Рандомизация проводится введением случайных весов портфеля [2, 4]:

$$G = \sum_{i \in I} \omega_i D_i, \quad \omega_i = \phi(\theta_i), \quad (11)$$

где $\theta_i \sim R\{E_i\}$, $i \in I$, — равномерно распределенные случайные величины на последовательных полуинтервалах E_i (8), для которых

$$\underline{e} = \{0,104; 0,454; 0,659; 0,145; 0\}; \\ \bar{e} = \{0,145; 0,659; 1; 0,454; 0,104\}.$$

Функция распределения $F_\zeta(\cdot)$ доходов ζ портфеля (11) определяется правилом:

$$F_{\omega_i; i}(z) = \frac{\phi^{\leftarrow}(z) - \underline{e}_i}{p_i^S}, \quad \phi(\underline{e}_i) \leq z \leq \phi(\bar{e}_i), \quad i \in I; \\ F_\zeta(z) = \phi^{\leftarrow}(z), \quad z \in [0, 1]. \quad (12)$$

При рандомизации вектор ε_{opt} не существует, а из-за случайности величины A для анализа относительного дохода потребуются специальный подход.

При применении алгоритма к опционным рынкам используются векторы, получаемые по формулам (7) и, по необходимости, (6):

$$p^B = \{0,049; 0,204; 0,333; 0,301; 0,113\}; \\ c^B = \{0,075; 0,185; 0,261; 0,280; 0,200\}; \\ p^K = \{0,006; 0,115; 0,288; 0,350; 0,220; 0,021\}.$$

В результате применения алгоритма (с p^B , c^B и частично p^S) получаем:

$$\xi = \{1, 5, 2, 4, 3\}; \quad \eta = \{1, 3, 5, 4, 2\}.$$

В рассматриваемом примере эта пара векторов (ξ, η) отличается от сценарного рынка, что вполне

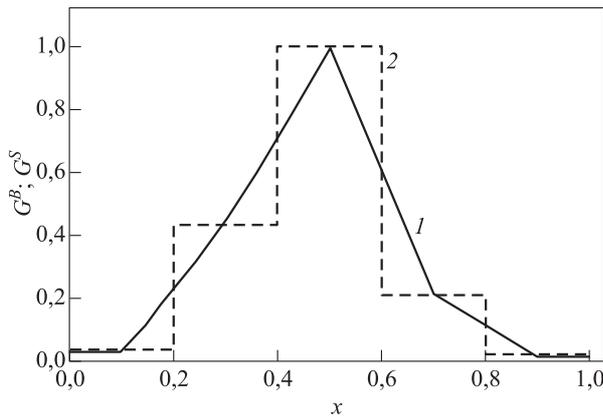


Рис. 2 Графики платежных функций G^B (1) и G^S (2)

возможно при малом числе сценариев. И показатели \mathbf{b} , \mathbf{g} и A для опционного рынка, тем более в связи с (9), должны быть несколько иными по сравнению со сценарным. Для детерминированной схемы имеем:

$$\begin{aligned} \varepsilon_{\text{opt}} &= \{0; 0,114; 0,299; 0,608; 1\}; \\ \mathbf{b} = \phi(\varepsilon_{\text{opt}}) &= \{0; 0,013; 0,090; 0,370; 1\}; \\ \mathbf{g} = \mathbf{b}(\eta) &= \{0,013; 0,370; 1; 0,090; 0\}; \end{aligned} \quad (13)$$

$$\left. \begin{aligned} A &= 0,358244; \quad R = 0,444829; \\ y &= 0,241692; \quad \sigma = 1,162. \end{aligned} \right\} \quad (14)$$

На рис. 2 (для прямой задачи) приводятся графики платежных функций $\pi(x; G^B)$ (1) и $\pi(x; G^S)$ (2), $x \in [0, 1]$. Первая из функций образована $(n + 1)$ -м линейным участком $\psi_i(x)$, $i \in I_0$, по правилу

$$\begin{aligned} \psi_0(x) &= g_1; \\ \psi_i(x) &= \frac{g_i(\kappa_{i+1} - x) + g_{i+1}(x - \kappa_i)}{h}; \\ \psi_n(x) &= g_n; \end{aligned}$$

5 Детерминированный портфель

Оптимальный детерминированный портфель задается вектором весов (13) с нормированными баттерфляями в качестве базисных инструментов:

$$G = \sum_{i \in I} g_i B_i.$$

Функцию распределения дохода $\zeta = \psi(X)$ определяем в системе интервалов \mathbf{K} (5) для $z \in [0, 1]$:

$$F_\zeta(z) = \Phi_0(z) + \Phi_n(z) + \sum_{i=1}^{n-1} \Phi_i(z), \quad (15)$$

где

$$\Phi_i(z) = P \{ \zeta < z, X \in K_i \}, \quad i \in I_0.$$

Для $i = 1$ и n доходы постоянны и равны g_1 и g_n соответственно, и потому (здесь и далее $u(w)$ — характеристическая функция множества $w \geq 0$)

$$\Phi_0(z) = p_0^K u(z - g_1); \quad \Phi_n(z) = p_n^K u(z - g_n).$$

Но для $i \in I'$ уже сказывается опционный характер платежной функции, и потому доходы и частные функции распределения соответственно равны:

$$\psi_i(x) = \frac{g_i(\kappa_{i+1} - x) + g_{i+1}(x - \kappa_i)}{h}, \quad x \in K_i;$$

$$\Phi_i(z) = \int_{\kappa_i}^{\kappa_{i+1}} u(z - \psi_i(x)) p(x) dx.$$

Все слагаемые суммы $F_\zeta(z)$ в (15) находятся простым интегрированием. Из функции распределения можно определять и запись результатов, но в детерминированном случае этого и не требуется, так как она уже получена выше (14).

График F_ζ представлен (под обозначением F_d) на рис. 3 (1), на нем приводится для сравнения и теоретический график F_0 (2).

Функция распределения относительного дохода $\chi = \zeta/A$ находится при этом простым увеличением аргумента в $1/A$ раз:

$$F_\chi(z) = F_\zeta(Az), \quad z \in [0, 1/A].$$

Далее рассматриваются варианты рандомизации. В силу специфики опционного рынка наряду с полной оправдан и вариант частичной рандомизации.

6 Полная рандомизация

При полной рандомизации все веса портфеля базисных баттерфляев ω_i , $i \in I$, являются случайными величинами. В основе определения функции распределения дохода вновь лежит представление (15), но теперь вместо детерминированных весов g_i вводятся случайные величины $\omega_i = \phi(\theta_i)$, $i \in I$, и тогда доход $\zeta = \psi(x; \mathbf{t})$, где $\mathbf{t} = \{t_1, \dots, t_n\}$ — вектор значений случайного вектора θ . Итак,

$$F_\zeta(z) = \sum_{i \in I_0} \Phi_i(z), \quad (16)$$

где

$$\Phi_i(z) = P \{ \zeta \leq z, x \in K_i \}, \quad i \in I_0.$$

Для $i = 0$ и n из (12) следует, что соответственно

$$\Phi_0(z) = \frac{p_0^K (\phi^{\leftarrow}(z) - \underline{e}_1)}{p_1^S}, z \in [\phi(\underline{e}_1), \phi(\bar{e}_1)]; \quad (17)$$

$$\Phi_n(z) = \frac{p_n^K (\phi^{\leftarrow}(z) - \underline{e}_n)}{p_n^S}, z \in [\phi(\underline{e}_n), \phi(\bar{e}_n)]. \quad (18)$$

В каждом из остальных слагаемых в (16), т.е. для $K_i, i \in I'$, следует проводить интегрирование по $x \in X$ и по всем n -мерным векторам t (хотя, по сути, лишь по всем парам $\{t_i, t_{i+1}\}$ значений для $\{\theta_i, \theta_{i+1}\}, t_i \in E_i, t_{i+1} \in E_{i+1}$). Имеем:

$$\begin{aligned} \psi_i(x; t_i, t_{i+1}) &= \\ &= \frac{\phi(t_i)(\kappa_{i+1} - x) + \phi(t_{i+1})(x - \kappa_i)}{h}, \\ & \quad x \in K_i, i \in I'; \quad (19) \end{aligned}$$

$$\begin{aligned} \Phi_i(z) &= \\ &= \iint_{E_i \times E_{i+1}} \left(\int_{\kappa_i}^{\kappa_{i+1}} u(z - \psi_i(x; t_i, t_{i+1})) p(x) dx \right) \frac{dt_i dt_{i+1}}{p_i^S p_{i+1}^S}, \\ & \quad i \in I'. \quad (20) \end{aligned}$$

Здесь численное интегрирование реализуемо для любого значения $z \in [0, 1]$. С его помощью и строится график функции (16) (например, интерполяцией). Он представлен на рис. 3 под обозначением F_f (3).

Использование стандартных программ интегрирования не работает в более сложных случаях, например для *относительного дохода*, когда требуется шестикратное интегрирование. Тем не менее для демонстрации общности подхода протестируем на оценивании функции $F_\zeta(z)$ (16) также контролируемый и учитывающий специфику задачи детерминированный аналог метода Монте-Карло, уже применявшийся в [4] для сценарного рынка.

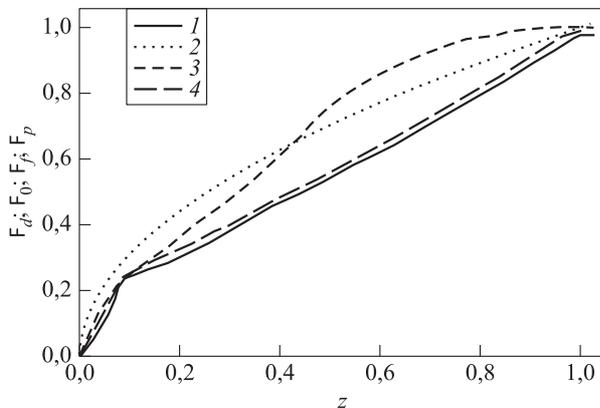


Рис. 3 Графики функций F_d (1), F_0 (2), F_f (3) и F_p (4)

С учетом рандомизации вводятся решетки значений случайных величин $\theta_i, i \in I$, определяемые параметром n_J и множеством $J = \{1, 2, \dots, n_J\}$:

$$t_{ij} = \underline{e}_i + \frac{p_i^S (j - 1/2)}{n_J}, j \in J, i \in I. \quad (21)$$

Для построения решетки реализуемых *доходов* для системы K интервалов (5) (за исключением интервалов K_0 и K_n) образуются дополнительные решетки значений цены базового актива $\pi_{il}, l \in L, i \in I'$, определяемые параметром n_L и множеством $L = \{0, 1, \dots, n_L\}$, и вычисляются соответствующие вероятности:

$$\begin{aligned} r_{0,j_1} &= \phi(t_{1,j_1}); r_{n,j_n} = \phi(t_{n,j_n}), j_1, j_n \in J; \\ r_{i,j_i,j_{i+1},l} &= \phi(t_{i,j_i}) \left(1 - \frac{(l-1/2)}{n_L} \right) + \\ &+ \phi(t_{i+1,j_{i+1}}) \frac{l-1/2}{n_L}, l \in L, j_i \in J, i \in I'; \\ \pi_{i,l} &= \kappa_i + \frac{l-1/2}{nn_L}; \\ p_{i,l}^X &= \int_{\pi_{i,l-1/(2nn_L)}}^{\pi_{i,l+1/(2nn_L)}} p(x) dx; p_0^K + \sum_{\substack{i \in I', \\ l \in L}} p_{i,l}^X + p_n^K = 1. \end{aligned}$$

В результате получается приближение для функции распределения в виде:

$$\begin{aligned} F_\zeta(z) &= \\ &= \frac{p_0^K}{n_J} \sum_{j_1 \in J} u(z - r_{0,j_1}) + \frac{p_n^K}{n_J} \sum_{j_n \in J} u(z - r_{n,j_n}) + \\ &+ \sum_{\substack{i \in I', \\ j_i, j_{i+1} \in J, \\ l \in L}} \frac{p_{i,l}^X}{n_J^2} u(z - r_{i,j_i,j_{i+1},l}). \quad (22) \end{aligned}$$

Построенный по (22) график визуально не отличим от графика F_f на рис. 3, что косвенно подтверждает правильность обоих подходов.

На основе (22) строятся и представления для моментов:

$$\begin{aligned} \mu_m &= \frac{p_0^K}{n_J} \sum_{j_1 \in J} r_{0,j_1}^m + \frac{p_n^K}{n_J} \sum_{j_n \in J} r_{n,j_n}^m + \\ &+ \sum_{\substack{i \in I', \\ j_i, j_{i+1} \in J, \\ l \in L}} \frac{p_{i,l}^X}{n_J^2} r_{i,j_i,j_{i+1},l}^m, m = 1, 2; \end{aligned}$$

$$\sigma^2 = \mu_2 - \mu_1^2.$$

По ним определяются числовые показатели портфеля:

$$E\zeta = 0,326928; E\zeta^2 = 0,161734; \sigma = 0,234204,$$

а средняя сумма EA находится далее при оценивании *относительного* дохода.

Требуемое для расчетов время $\sim nn^2n_L$.

Подобный подход без труда распространяется и на распределение *относительного* дохода χ . Формула (16) переписывается применительно к случайной величине

$$\chi = \psi(x; \theta_1, \dots, \theta_n) = \psi(x; \theta) = \frac{\zeta}{A(\theta)},$$

где $A(\theta) = \sum_{i \in I} c_i^B \phi(\theta_i)$ — случайная инвестиционная сумма со значениями $A(t) = \sum_{i \in I} c_i^B \phi(t_i)$:

$$F_\chi(z) = \sum_{i \in I_0} \Phi_i(z), \quad (23)$$

где

$$\Phi_i(z) = P\{\chi \leq z, x \in K_i\}, \quad i \in I_0,$$

а формулы (17)–(20) приобретают вид:

$$\psi_0(x; t) = \frac{\phi(t_1)}{A(t)}, \quad x \in K_0;$$

$$\psi_n(x; t) = \frac{\phi(t_n)}{A(t)}, \quad x \in K_n,$$

$$\psi_i(x; t) = h \frac{\phi(t_i)(\kappa_i - x) + \phi(t_{i+1})(x - \kappa_{i+1})}{A(t)},$$

$$x \in K_i, \quad i \in I';$$

$$\Phi_i(z) = \int_{\prod_{i=1, \dots, n} E_i} \left(\int_{\kappa_i}^{\kappa_{i+1}} \frac{u(z - \psi_i(x; t)) p(x)}{A(t)} dx \right) \times \frac{\prod_{i=1, \dots, n} dt_i}{\prod_{i=1, \dots, n} p_i^S}, \quad i \in I_0.$$

Решетка для инвестиционной суммы $A(t)$ в соответствии с (21) имеет вид:

$$a_{j_1, \dots, j_n} = \sum_{i \in I} c_i^B \phi(t_{i, j_i}), \quad j_i \in J, \quad i \in I.$$

Функция распределения относительного дохода, а также его моменты и дисперсия определяются приближенными формулами, фактически реализующими шестикратное (при $n = 5$) интегрирование:

$$F_\chi(z) = \frac{p_0^K}{n_J^n} \sum_{j_1, \dots, j_n \in J} u\left(z - \frac{r_{0, j_1}}{a_{j_1, \dots, j_n}}\right) + \frac{p_n^K}{n_J^n} \sum_{j_1, \dots, j_n \in J} u\left(z - \frac{r_{n, j_n}}{a_{j_1, \dots, j_n}}\right) + \sum_{\substack{i \in I', \\ j_1, \dots, j_n \in J, \\ l \in L}} \frac{p_{i, l}^X}{n_J^n} u\left(z - \frac{r_{i, j_i, j_{i+1}, l}}{a_{j_1, \dots, j_n}}\right); \quad (24)$$

$$\mu_m = \frac{p_0^K}{n_J^n} \sum_{j_1, \dots, j_n \in J} \frac{r_{0, j_1}^m}{a_{j_1, \dots, j_n}^m} + \frac{p_n^K}{n_J^n} \sum_{j_1, \dots, j_n \in J} \frac{r_{n, j_n}^m}{a_{j_1, \dots, j_n}^m} + \sum_{\substack{i \in I', \\ j_1, \dots, j_n \in J, \\ l \in L}} \frac{p_{i, l}^X}{n_J^n} \frac{r_{i, j_i, j_{i+1}, l}^m}{a_{j_1, \dots, j_n}^m},$$

$$m = 1, 2; \quad \sigma^2 = \mu_2 - \mu_1^2. \quad (25)$$

Выборочные моменты и дисперсия определяются по (25):

$$E\chi = 1,21373; E\chi^2 = 2,1483; \sigma = 0,821676$$

(при $n_L = 8, n_J = 6$).

Из подобных (25) формул также $EA = 0,268748, E(1/A) = 3,83142$.

Построенный по (24) график $F_\chi(\cdot)$ походит на растянутый по оси абсцисс график $F_\zeta(\cdot)$ с коэффициентом, близким к $E(1/A)$, и в работе не приводится.

Требуемое для расчетов время $\sim nn^2n_L$.

7 Частичная рандомизация

Под частичной рандомизацией понимается вариант рандомизации, в котором случайными становятся лишь веса ω_1 и ω_n базисных баттерфляев в портфеле для крайних страйков. Суть в том, что уже сами платежные функции опционов, исключая эти страйки, реализуют некоторое сглаживание, но лишь для внутренних страйков. И частичной рандомизации может оказаться достаточно.

При частичной рандомизации *доход* $\zeta = \psi(x; t_1, t_n)$. Вновь исходим из (16) и для крайних интервалов $K_i, i = 0, n$, применяем формулы (17) и (18):

$$\Phi_0(z) = p_0^K \frac{\phi^{\leftarrow}(z) - \underline{e}_1}{p_1^S}, \quad z \in [\phi(\underline{e}_1), \phi(\bar{e}_1)];$$

$$\Phi_n(z) = p_n^K \frac{\phi^{\leftarrow}(z) - \underline{e}_n}{p_n^S}, \quad z \in (\phi(\underline{e}_n), \phi(\bar{e}_n)).$$

Для $i \in I''$ применяются формулы детерминированной схемы, а для $i = 1, n-1$ формулы приобретают гибридный вид, сочетающий в себе характерные

черты как рандомизированной, так и детерминированной схем. Имеем:

$$\psi_1(x; t_1, t_n) = h(\psi(t_1)(\kappa_1 - x) + g_2(x - \kappa_2)), \quad x \in K_1;$$

$$\psi_{n-1}(x; t_1, t_n) = h(g_{n-1}(\kappa_{n-1} - x) + \phi(t_n)(x - \kappa_n)), \quad x \in K_{n-1},$$

$$\psi_i(x; t_1, t_n) = h(g_i(\kappa_{i+1} - x) + g_{i+1}(x - \kappa_i)), \quad x \in K_i, \quad i \in I'',$$

$$\begin{aligned} \Phi_i(z) = \\ = \iiint_{E_1 \times E_2} \left(\int_{\kappa_i}^{\kappa_{i+1}} u(z - \psi_i(x; t_1, t_n)) p(x) dx \right) \frac{dt_1 dt_n}{p_1^S p_n^S}, \quad i \in I'. \end{aligned}$$

На этот раз решетка (21) применяется лишь для крайних сценариев:

$$t_{i,j} = \underline{e}_i + p_i^S \frac{(j-1/2)}{n_J}, \quad j \in J, \quad i = 1, n. \quad (26)$$

По ней строится и решетка доходов:

$$r_{0,j_1} = \phi(t_{1,j_1}), \quad r_{n,j_n} = \phi(t_{n,j_n}), \quad j_1, j_n \in J;$$

$$r_{1,j_1,l} = \phi(t_{1,j_1}) \left(1 - \frac{l-1/2}{n_L} \right) + g_2 \frac{l-1/2}{n_L}, \quad j_1 \in J, \quad l \in L;$$

$$r_{n-1,j_n,l} = g_{n-1} \left(1 - \frac{l-1/2}{n_L} \right) + \phi(t_{n,j_n}) \frac{l-1/2}{n_L}, \quad j_n \in J, \quad l \in L;$$

$$r_{i,l} = g_i \left(1 - \frac{l-1/2}{n_L} \right) + g_{i+1} \frac{l-1/2}{n_L}, \quad l \in L, \quad i \in I''.$$

Наконец, находится аппроксимация функции распределения дохода:

$$\begin{aligned} F_\zeta(z) = \sum_{j_1 \in J} \frac{p_0^K}{n_J} u(z - r_{0,j_1}) + \sum_{j_n \in J} \frac{p_n^K}{n_J} u(z - r_{n,j_n}) + \\ + \sum_{\substack{j_1 \in J, \\ l \in L}} \frac{p_{1,l}^X}{n_J} u(z - r_{1,j_1,l}) + \sum_{\substack{j_n \in J, \\ l \in L}} \frac{p_{n-1,l}^X}{n_J} u(z - r_{n-1,j_n,l}) + \\ + \sum_{\substack{i \in I'', \\ l \in L}} \frac{p_{i,l}^X}{n_J^2} u(z - r_{i,l}). \quad (27) \end{aligned}$$

На ее основе строится график функции распределения дохода. Выборочные моменты и дисперсия определяются формулами:

$$\begin{aligned} \mu_m = \sum_{j_1 \in J} \frac{p_0^K}{n_J} r_{0,j_1}^m + \sum_{j_n \in J} \frac{p_n^K}{n_J} r_{n,j_n}^m + \\ + \sum_{\substack{j_1 \in J, \\ l \in L}} \frac{p_{1,l}^X}{n_J} r_{1,j_1,l}^m + \sum_{\substack{j_n \in J, \\ l \in L}} \frac{p_{n-1,l}^X}{n_J} r_{n-1,j_n,l}^m + \\ + \sum_{\substack{i \in I'', \\ l \in L}} \frac{p_{i,l}^X}{n_J^2} r_{i,l}^m, \quad m = 1, 2; \\ \sigma^2 = \mu_2 - \mu_1^2. \end{aligned}$$

Из них находим $E\zeta = 0,436903$, $E\zeta^2 = 0,292945$ и $\sigma = 0,319469$ (при $n_L = 100$ и $n_J = 50$). Требуемое для расчетов время $\sim nn^2 n_L$.

График функции (27) представлен на рис. 3 под обозначением F_p (4).

Для относительного дохода имеем

$$\chi = \psi(x; \theta_1, \theta_n) = \frac{\zeta}{A(\theta_1, \theta_n)},$$

где $A(\theta_1, \theta_n) = c_1^B \phi(\theta_1) + \sum_{l=2, \dots, n-1} c_l^B g_l + c_n^B \phi(\theta_n)$ — случайная инвестиционная сумма со значениями $A(t_1, t_n)$. Вновь исходим из представления (23), при этом для компонент с $i = 0$ и n формулы те же, что и для полной рандомизации, а для $i \in I'$ требуются очевидные изменения:

$$\psi_0(x; \mathbf{t}) = \frac{\phi(t_1)}{A(\mathbf{t})}, \quad x \in K_0;$$

$$\psi_n(x; \mathbf{t}) = \frac{\phi(t_n)}{A(\mathbf{t})}, \quad x \in K_n;$$

$$\psi_1(x; t_1, t_n) = \frac{h(\phi(t_1)(\kappa_1 - x) + g_2(x - \kappa_2))}{A(t_1, t_n)}, \quad x \in K_1;$$

$$\begin{aligned} \psi_{n-1}(x; t_1, t_n) = \\ = \frac{h(g_{n-1}(\kappa_{n-1} - x) + \phi(t_n)(x - \kappa_n))}{A(t_1, t_n)}, \quad x \in K_{n-1}; \end{aligned}$$

$$\psi_i(x; t_1, t_n) = \frac{h(g_i(\kappa_{i+1} - x) + g_{i+1}(x - \kappa_i))}{A(t_1, t_n)}, \quad x \in K_i, \quad i \in I'';$$

$$\begin{aligned} \Phi_i(z) = \\ = \iiint_{E_1 \times E_n} \left(\int_{\kappa_i}^{\kappa_{i+1}} u(z - \phi_i(x; t_1, t_n)) p(x) dx \right) \frac{dt_1 dt_n}{p_1^S p_n^S}, \quad i \in I_0. \end{aligned}$$

На этот раз также образуется двумерная решетка (26), и на ее основе формируется решетка стоимости портфеля:

$$a_{j_1, j_n} = c_1^B \phi(t_{1,j_1}) + \sum_{l=2, \dots, n-1} c_l^B g_l + c_n^B \phi(t_{n,j_n}), \quad j_1, j_n \in J.$$

Аппроксимацию функции распределения для χ дает представление:

$$\begin{aligned}
 F_{\chi}(z) = & \frac{p_0^K}{n_J^2} \sum_{j_1, j_n \in J} u\left(z - \frac{r_{0, j_1}}{a_{j_1, j_n}}\right) + \\
 & + \frac{p_n^K}{n_J^2} \sum_{j_1, j_n \in J} u\left(z - \frac{r_{n, j_n}}{a_{j_1, j_n}}\right) + \\
 & + \sum_{\substack{j_1, j_n \in J, \\ l \in L}} \frac{p_{1, l}^X}{n_J^2} u\left(z - \frac{r_{1, j_1, l}}{a_{j_1, j_n}}\right) + \\
 & + \sum_{\substack{j_1, j_n \in J, \\ l \in L}} \frac{p_{n-1, l}^X}{n_J^2} u\left(z - \frac{r_{n-1, j_n, l}}{a_{j_1, j_n}}\right) + \\
 & + \sum_{\substack{i \in I'', \\ j_1, j_n \in J, \\ l \in L}} \frac{p_{i, l}^X}{n_J^2} u\left(z - \frac{r_{i, l}}{a_{j_1, j_n}}\right).
 \end{aligned}$$

Выборочные моменты и дисперсия определяются формулами:

$$\begin{aligned}
 \mu_m = & \frac{p_0^K}{n_J^2} \sum_{j_1, j_n \in J} \frac{r_{0, j_1}^m}{a_{j_1, j_n}^m} + \frac{p_n^K}{n_J^2} \sum_{j_1, j_n \in J} \frac{r_{n, j_n}^m}{a_{j_1, j_n}^m} + \\
 & + \sum_{\substack{j_1, j_n \in J, \\ l \in L}} \frac{p_{1, l}^X}{n_J^2} \frac{r_{1, j_1, l}^m}{a_{j_1, j_n}^m} + \sum_{\substack{j_1, j_n \in J, \\ l \in L}} \frac{p_{n-1, l}^X}{n_J^2} \frac{r_{n-1, j_n, l}^m}{a_{j_1, j_n}^m} + \\
 & + \sum_{\substack{i \in I'', \\ j_1, j_n \in J, \\ l \in L}} \frac{p_{i, l}^X}{n_J^2} \frac{r_{i, l}^m}{a_{j_1, j_n}^m}, \quad m = 1, 2; \\
 \sigma^2 = & \mu_2 - \mu_1^2.
 \end{aligned}$$

Из них находим $E\chi = 1,22723$, $E\chi^2 = 2,31133$ и $\sigma = 0,897355$ (при $n_L = 50$ и $n_J = 6$), а из аналогичных им формул — также $E A = 0,356003$ и $E(1/A) = 2,80898$. Кстати, $E A \times E(1/A) = 1,0000035$, что косвенно подтверждает близость графиков для детерминированной и частично рандомизированной схем.

Требуемое для расчетов время $\sim n n_J^2 n_L$.

8 Заключение

Завершаем описание алгоритмов нахождения функций распределения доходов и относительного дохода в задаче оптимизации по CC-VaR портфеля опционов. Результаты расчетов на примере с бета-распределениями для прямой задачи (продажи волатильности) проиллюстрированы на графиках рис. 3.

Разумеется, при дискретизации задачи в случае небольшого числа сценариев результаты должны

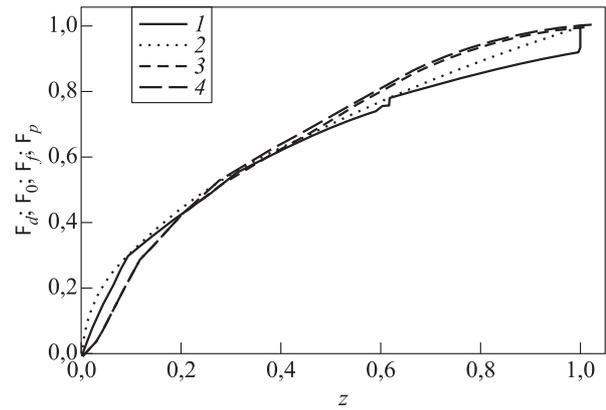


Рис. 4 Функции $F_d(1)$, $F_d(2)$, $F_f(3)$ и $F_p(4)$ для обратной задачи

отличаться от теоретической схемы. И интерес представляет сравнение полученных результатов, в частности с обратной задачей — покупкой волатильности. На рис. 4 представлены для нее аналогичные результаты, полученные тем же алгоритмом. Значительное расхождение результатов в большей степени объясняется различием задач, очевидным по рис. 1.

В прямой задаче частичная рандомизация дает практически тот же результат, что и детерминированная схема (см. рис. 3), что легко объясняется незначительным вкладом в картину доходов крайними сценариями, на которые частичная рандомизация только и распространяется. А потому выбор предстоит делать между детерминированной схемой и полной рандомизацией. При этом полная рандомизация дает более сглаженную и более близкую к теоретической схеме кривую, а детерминированная — более надежную в выполнении CC-VaR.

В обратной задаче (см. рис. 4) все три аппроксимации дают результаты, достаточно близкие к теоретическим, что можно объяснять полноценным вкладом в итоговый результат всех пяти сценариев. И частичная рандомизация, почти совпадающая с детерминированной кривой в области пониженных доходов и с полной рандомизированной — в зоне повышенных, участвует в этом в наибольшей степени и даже представляется наилучшей из трех схем.

В целом, на рынках с малым числом сценариев при рассмотрении вариантов назначения весов портфеля инвестору придется при выборе варианта дополнительно руководствоваться критериями близости функции распределения к теоретической, степени выполнения CC-VaR, средней доходности и пр.

Литература

1. Agasandian G. A. Optimal behavior of an investor in option market // Conference (International) on Neural Networks. The 2002 IEEE World Congress on Computational Intelligence. — Honolulu, Hawaii, 2002. P. 1859–1864.
2. Агасандян Г. А. Применение непрерывного критерия VaR на финансовых рынках. — М.: ВЦ РАН, 2011. 299 с.
3. Агасандян Г. А. Континуальный критерий VaR на многомерных рынках опционов. — М.: ВЦ РАН, 2015. 297 с.
4. Агасандян Г. А. Континуальный критерий VaR на сценарных рынках // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 32–40.
5. Крамер Г. Математические методы статистики / Пер. с англ. — М.: Мир, 1975. 750 с. (Cramer H. Mathematical methods of statistics. — Princeton, NJ, USA: Princeton University Press, 1946. 575 p.)

Поступила в редакцию 18.12.18

PERFORMANCE ESTIMATIONS FOR OPTIMAL-ON-CC-VaR PORTFOLIOS IN OPTION MARKETS

G. A. Agasandyan

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

Abstract: The paper continues investigations of the author about using continuous VaR-criterion (CC-VaR) in financial markets. The problem of projecting ideas and methods elaborated for investments in the ideal theoretical one-period market and its discrete scenario analog onto a discrete-in-strikes option market is considered. The main focus is on the methods of calculating distribution function of income and return relative, and also their mean for option portfolios optimal on CC-VaR and their randomized versions, both full and partial. A discrete optimization algorithm as the result of projecting the theoretical algorithm based on the Newman–Pearson procedure onto scenario market is suggested. The optimal vector of weights derived from this algorithm is applied to the basis of normalized simplest butterflies. If randomizing portfolios are admissible, then special algorithms based on the ideas of the Monte-Carlo method that determine distribution functions of income and return relative are suggested. The exposition is illustrated by examples with beta-distributed underlier’s prices and investor’s probability forecast, which deal with the problems of volatility selling and buying. The respective diagrams are adduced.

Keywords: continuous VaR-criterion (CC-VaR); investor’s risk-preferences function (r.p.f.); Newman–Pearson procedure; scenarios; options; indicators; butterflies; full and partial randomizing; optimal portfolio; income; yield

DOI: 10.14357/19922264190311

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 17-01-00816).

References

1. Agasandian, G. A. 2002. Optimal behavior of an investor in option market. *Conference (International) on Neural Networks. The 2002 IEEE World Congress on Computational Intelligence*. Honolulu, Hawaii. 1859–1864.
2. Agasandyan, G. A. 2011. *Primenenie kontinual'nogo kriteriya VaR na finansovykh ryinkakh* [Application of continuous VaR-criterion in financial markets]. Moscow: CC RAS. 299 p.
3. Agasandyan, G. A. 2011. *Kontinual'nyy kriteriy VaR na mnogomernykh ryinkakh optsiyonov* [Continuous VaR-criterion in multidimensional option markets]. Moscow: CC RAS. 297 p.
4. Agasandyan, G. A. 2018. Kontinual'nyy kriteriy VaR na stsennarnykh ryinkakh [Continuous VaR-Criterion in scenario markets]. *Informatika i ee Primeneniya — Inform. Appl.* 12(1):32–40.
5. Cramer, H. 1946. *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press. 575 p.

Received December 18, 2018

Contributor

Agasandyan Gennady A. (b. 1941) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; agasand17@yandex.ru

ИНТЕРФЕЙСЫ ТРЕТЬЕГО ПОРЯДКА В ИНФОРМАТИКЕ

И. М. Зацман¹

Аннотация: В Европейской стратегии «Информатика для всех», объявленной в Брюсселе в марте 2018 г., различаются две ступени преподавания информатики в системе среднего и высшего профессионального образования. Вторая ступень, ориентированная на изучение информационных трансформаций в искусственных, живых и социальных системах, подразумевает сначала выбор для преподавания некоторой парадигмы информатики, а затем и ее развитие. Необходимость развития обусловлена двумя причинами: во-первых, существенным расширением сферы применения информационных технологий (ИТ), рассматриваемых в учебном процессе, во-вторых, интеграцией методов и средств информатики в учебные программы по другим областям знания, что расширяет спектр изучаемых информационных трансформаций. В отсутствие доминирующей парадигмы информатики и при наличии нескольких ее вариантов вопрос ее выбора как отправной точки развития является дискуссионным. В обзоре “Informatics Education in Europe”, опубликованном в 2017 г. и предваряющем разработку Европейской стратегии «Информатика для всех», перечислены три варианта парадигмы, включая позиционирование информатики как самостоятельной отрасли знания, предложенное Деннингом и Розенблюмом в 2009 г. Подробное описание этого варианта под названием полиадического компьютеринга дано Розенблюмом в его монографии 2013 г. Цель статьи состоит в определении нового понятия «интерфейс 3-го порядка» на основе деления на среды предметной области информатики как полиадического компьютеринга. Актуальность введения этого понятия иллюстрируется на примере процессов управления роботизированной рукой сигналами головного мозга.

Ключевые слова: интерфейсы 3-го порядка; полиадический компьютеринг; среды предметной области информатики; информационные трансформации

DOI: 10.14357/19922264190312

1 Введение

Апрельский номер журнала *Communications of the ACM* за 2019 г. включает специальный раздел «Информатика в Европе», состоящий из статей, посвященных актуальным темам в сфере информатики и вычислительной техники [1]. Одна из статей посвящена вопросам преподавания информатики [2]. В ней анализируется «План действий в области цифрового образования», одобренный Еврокомиссией в 2018 г. [3], а также отчет о состоянии преподавания дисциплин информатики в Европе, включая РФ [4].

Авторы отчета предложили ряд рекомендаций, которые послужили бы улучшению ситуации с изучением дисциплин информатики в Европейском регионе, на основе которых был подготовлен доклад “Informatics for All: The strategy”. Основное внимание в докладе уделяется преподаванию информатики в системе среднего и высшего профессионального образования, которая рассматривается как фундаментальная наука со своими собственными научными методами [5].

Во многих отношениях основные идеи этого доклада совпадают с инициативой президента Оба-

мы [6]. Важнейшим элементом Европейского доклада “Informatics for all: The strategy”, который отличает его от этой инициативы, является двухступенчатая стратегия преподавания информатики: (1) информатика как самостоятельный предмет; (2) интеграция методов и средств информатики в предметные области других наук [2].

Для данной статьи наибольший интерес представляет вторая ступень. Реализация планов преподавания информатики на второй ступени, скорее всего, потребует существенного расширения ее предметной области, так как появляется необходимость рассматривать информационные трансформации не только в искусственных, но и в живых и социальных системах. Такой спектр информационных трансформаций будет охватывать объекты разных сред и, что важно отметить, разной природы. Сегодня с точки зрения теоретического описания уже существующих ИТ необходимо различать как минимум пять сред разной природы: нейросреду (сигналы активности мозга в нейрокоммуникационных технологиях); ментальную (концепты в технологиях представления знаний человека); социоинформационную (тексты на естественных языках, таблицы, рисунки, растровые и векторные

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

изображения и т. д. в технологиях их кодирования); ДНК (представление информации и данных с помощью синтезированных цепочек ДНК); цифровую (компьютерные коды). При этом теоретическое описание технологий и интерфейсов будущих поколений может потребовать включения в рассмотрение и сред другой природы, отличной от пяти перечисленных.

Основная цель статьи состоит в описании границ между средами предметной области информатики как полиадического компьютеринга, концепция которого предполагает расширение ее предметной области на информационные трансформации в живых и социальных системах. Предлагаемое описание сред и границ дает возможность определить понятие «интерфейс 3-го порядка» и ответить на вопрос: «Зачем оно нужно в информатике?» Актуальность введения этого понятия иллюстрируется на примере процессов управления роботизированной рукой сигналами головного мозга.

2 О предметной области информатики

На первых этапах становления информатики объектами ее исследований были в основном компьютеры и окружающие их явления. Например, в 1967 г. А. Ньюэлл, А. Дж. Перлис и Х. А. Саймон писали: «Объектам и явлениям соответствуют те науки, которые их изучают. Появились компьютеры. Следовательно, назначение информатики — это изучение компьютеров» [7].

В 1963 г. Сол Горн предлагал более широкий взгляд на предметную область новой развивающейся науки [8]. Однако на первых этапах ее становления основными объектами предметной области были компьютерные коды, их генерация и преобразование в цифровой среде, а также явления, окружающие компьютеры.

В конце прошлого столетия изучение информационных трансформаций в живых системах стали также относить к предметной области информатики вместе с генерацией и преобразованиями компьютерных кодов в цифровой среде. Использование слова *computing* вместо двух слов *computer science* стало подчеркивать более широкую предметную область информатики. «Первоначальное определение информатики — изучение явлений, окружающих компьютеры, — сейчас устарело. Она изучает и естественные, и искусственные информационные процессы. Информатика включает компьютерную науку, разработку программного обеспечения, ИТ, информационную науку и информационные системы» [9].

Такое расширение предметной области потребовало пересмотра понятия *computation*. «Нам придется пересмотреть понятие *computation*, которое должно быть в состоянии учитывать также и обработку информации, происходящую в природе. . . На самом деле исследования информационных трансформаций в живых системах уже привели к пересмотру ряда парадигм, лежащих в основе традиционного понятия *computation*» [10].

Другое направление расширения предметной области информатики было связано с исследованиями информационных процессов в социальных системах. Проиллюстрируем это направление цитатой из статьи Марка Снира: «Быстрое развитие информационных технологий стимулирует периодический пересмотр теоретического фундамента их разработки и сейчас настало время провести очередной пересмотр. . . Раньше, когда с компьютерами имел дело ограниченный круг лиц, можно было рассчитывать на то, что те, кто профессионально общаются с компьютерами, могут к ним адаптироваться. В наши дни, когда значительно больше людей взаимодействуют с цифровыми устройствами и информационно-компьютерными системами, информационные технологии тесно вплетены во многие когнитивные и социальные процессы. В этих условиях нельзя игнорировать эти процессы при их создании. Кроме того, подобное вплетение является интересным объектом для исследования» [11].

Один из подходов к пересмотру теоретического фундамента разработки ИТ в 2013 г. был предложен Полом Розенблюмом [12]. В рамках этого подхода он описал четыре этапа расширения предметной области информатики, взяв за основной объект исследований информационные трансформации в искусственных, живых и социальных системах. При этом он использовал деление всех научных знаний на самом верхнем уровне на четыре отрасли науки [13], обозначив их литерами: информатика (C), физические науки (P), науки о жизни (L) и социальные науки (S). Далее он рассматривает информационные трансформации, охватывающие две, три или четыре отрасли, включая информатику, обозначая каждый этап расширения предметной области и вербально, и с использованием этих четырех литер, следующим образом: диадический компьютеринг (C + P, C + L, C + S); триадический компьютеринг (C + P + S, C + P + L, C + L + S); тетрадический компьютеринг (P + S + L + C); полиадический компьютеринг (объединение всех перечисленных выше расширений).

В завершение этого раздела отметим в истории науки аналогию между эволюционным расшире-

нием предметных областей геологии и информатики. Развитие геологии как науки о Земле началось в XVII в. с работы Николая Стенона [14]. Отдельные модели развития Земли, которые были впервые разработаны в геологии, затем использовались и для описания эволюции планет и других тел Солнечной системы. Появились новые научные дисциплины, такие как «геология Луны», «геология Марса» и «геология Венеры», которые описывают модели эволюции этих тел Солнечной системы.

Аналогичным образом могут создаваться теоретические модели для описания, например, взаимосвязей «мозг–компьютер» (brain–computer interface — BCI) или других видов интерфейсов, которые появятся в ИТ будущих поколений вследствие расширения сферы их применения [15]. Их описание может быть во многом выполнено по аналогии с моделями человеко–машинных интерфейсов (human–computer interface — HCI), разработанными и уже широко используемыми в системах и средствах информатики.

В следующем разделе статьи будет дано описание взаимосвязей «мозг–компьютер» с использованием сочетания традиционных взаимосвязей на границах двух сред, которые будем называть интерфейсами второго порядка, с интерфейсами 3-го порядка на границах трех сред.

3 Интерфейсы третьего порядка

Сначала приведем краткое описание взаимосвязей «мозг–компьютер» из работы [12], которое будет использоваться далее при обосновании необходимости введения понятия «интерфейс 3-го порядка»: «Если мы хотим представить все богатство взаимосвязей между мозгом и компьютером, нам необходимо рассмотреть информационные трансформации, относящиеся ко всей предметной области тетрадического компьютеринга ($P + S + L + C$). Эти интерфейсы включают не

только трансформации сигналов мозга (L) и генерацию компьютерных кодов (C), но также физические устройства (P), которые обеспечивают преобразование сигналов мозга в коды компьютера, плюс концепты знаний человека (S), которые могут быть представлены сигналами, отражающими активность мозга, и эти сигналы в конечном итоге и будут использоваться для управления компьютером или другими устройствами».

До описания понятия «интерфейс 3-го порядка» на примере взаимосвязей между мозгом и компьютером сначала рассмотрим пять ранее перечисленных сред. Выделим их из предметной области тетрадического компьютеринга, которая позволяет охватить максимальный спектр информационных трансформаций в искусственных, живых и социальных системах (рис. 1).

Основное отличие выделения сред разной природы из предметной области тетрадического компьютеринга от деления всей науки Деннингом и Розенблумом на четыре отрасли знания [13] состоит в том, что объекты одной и той же среды могут изучаться одновременно в разных научных дисциплинах. Например, концепты встречаются в задачах представления знаний в информатике [15–17], они являются объектами исследований в науках о жизни [18] и социогуманитарных науках [19], но относятся они только к одной ментальной среде знаний человека. Отметим, что пять сред, перечисленных в начале статьи, присутствуют в предметных областях и тетрадического компьютеринга (слева на рис. 1), и полиадического компьютеринга, но они не описаны в работе [12]. В работе [15] было показано, что кроме пяти перечисленных могут быть выделены и включены в рассмотрение среды и другой природы. Однако из пяти сред, условно обозначенных в виде прямоугольников справа на рис. 1, четырех (кроме ДНК-среды) будет достаточно для описания интерфейсов 3-го порядка, рассматриваемых в статье.

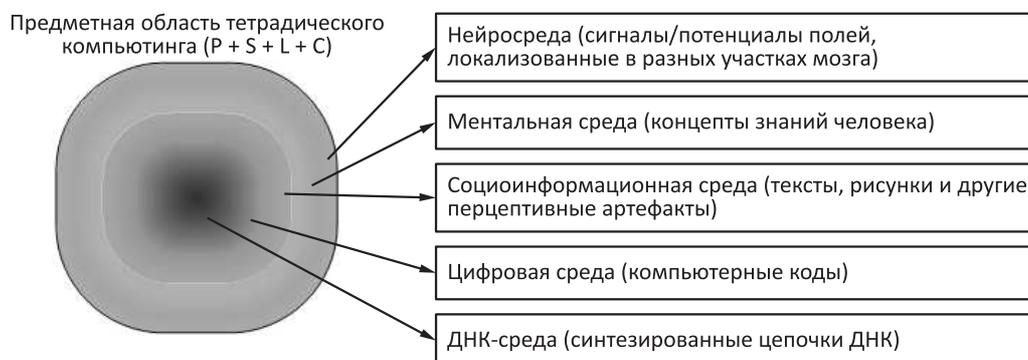


Рис. 1 Пять сред разной природы в предметной области тетрадического компьютеринга

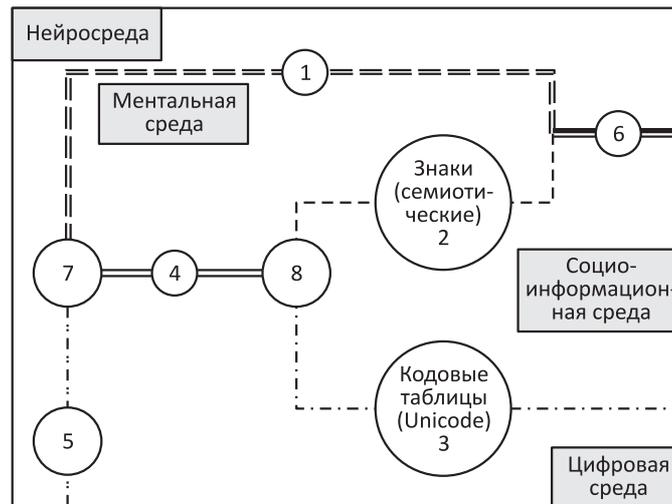


Рис. 2 Четыре среды разной природы и границы между ними

Эти интерфейсы рассмотрим на примере процессов управления роботизированной рукой. В этом примере расположим четыре среды по-другому, чтобы были видны границы между ними (рис. 2). На этом рисунке на границах между двумя средами кругами условно обозначены интерфейсы 2-го порядка шести разных видов (они пронумерованы от 1 до 6), а на границах между тремя средами обозначены интерфейсы 3-го порядка двух видов, которые по определению обеспечивают связи между объектами трех сред разной природы (7 и 8). В этой статье другие виды интерфейсов 3-го и более высокого порядка не рассматриваются. Интерфейсы вида № 7 связывают объекты нейросреды, ментальной и цифровой сред, а вида № 8 — ментальной, социоинформационной и цифровой сред.

Если направление информационных трансформаций на границах сред не учитывать, то для четырех сред число видов интерфейсов 2-го порядка равно шести. Все они были описаны в [15]. Число видов интерфейсов 3-го порядка равно четырем. Из шести видов интерфейсов 2-го порядка наиболее известными и широко используемыми являются: № 2 на границе ментальной и социоинформационной сред, описывающий взаимосвязи между концептами и формами их представления [19], и № 3, включающий человеко-машинные интерфейсы кодирования растровых и векторных изображений, а также символов, составляющих вербальные формы представления знаний человека, например с помощью таблиц Unicode.

Интерфейсы вида № 1 описывают связи потенциалов фокальных полей или других сигналов нейросреды как индикаторов активности головного мозга с концептами знаний человека в тех слу-

чаях, когда эти связи могут быть установлены. Интерфейсы вида № 4 теоретически предназначены для описания связей между концептами ментальной среды и компьютерными кодами цифровой среды. Однако интерфейсы этого вида, как будет показано ниже, практически невозможно вычленишь из интерфейсов третьего порядка вида № 7 между концептами, сигналами нейросреды и компьютерными кодами. Интерфейсы вида № 5 описывают связи сигналов нейросреды с компьютерными кодами. Интерфейсы вида № 6 описывают связи сигналов нейросреды с формами представления знаний человека в виде текстов, изображений и других перцептивных форм.

Для описания процессов управления роботизированной рукой понадобятся интерфейсы 2-го порядка трех видов (1, 4 и 5) и один вид интерфейсов 3-го порядка (7).

Рисунок 3 содержит упрощенную схему сред и границ с интерфейсами «мозг–компьютер», которую будем использовать, чтобы получить ответ на основной вопрос статьи: «Зачем нужны интерфейсы 3-го порядка?»

На схеме условно обозначены связи между концептами знаний человека (ментальная среда) и соответствующими сигналами нейросреды (№ 1). С нейрофизиологической точки зрения эти сигналы отражают активность разных участков головного мозга, но не концепты. Согласно Б. Баарсу и Н. Гейдж, «чрезвычайно трудно обнаружить конкретные места коры мозга для определенных концептов» [18]. Однако в некоторых случаях (определенные категории животных и инструментов, а также числа) связи между концептами и локализованными сигналами уже установлены [20, 21].

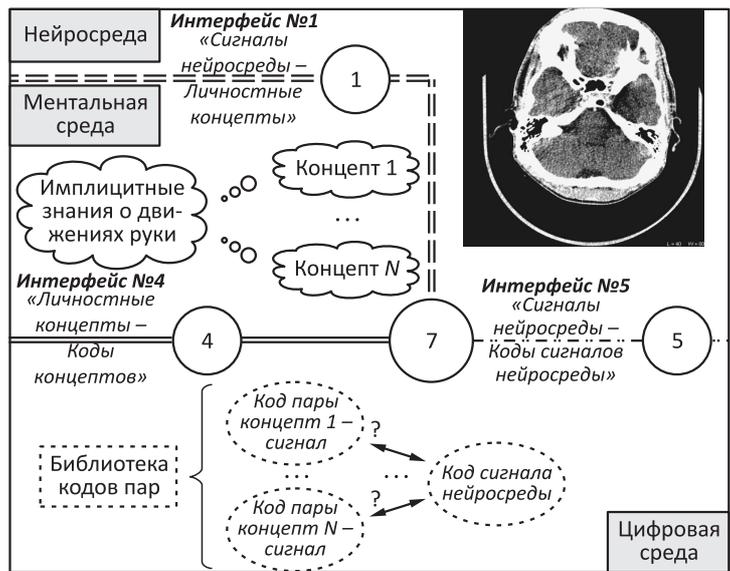


Рис. 3 Три среды и границы с интерфейсами для управления роботизированной рукой

При этом одному концепту может соответствовать сочетание сигналов.

Процессы управления роботизированной рукой служат примером формирования и последующего использования локализованных сигналов нейросреды на личностном уровне. Для конкретного человека связь между его имплицитными личностными концептами, релевантными его намерениям, и сигналами нейросреды, регистрируемыми с помощью интерфейсов «мозг–компьютер», устанавливается в процессе обучения системы управления роботизированной рукой с использованием методов измерения мозговой активности, например электроэнцефалографии (ЭЭГ) или электрокортикографии¹. Проведенные эксперименты показали, что после обучения системы этот человек может управлять ею с помощью сигналов нейросреды, регистрируемых с помощью метода ЭЭГ [23].

Интерфейсы второго порядка вида № 4 теоретически предназначены для описания связей между концептами ментальной среды и компьютерными кодами цифровой среды. Эти связи практически невозможно вычленишь из интерфейса третьего порядка вида № 7 между концептами, сигналами нейросреды и компьютерными кодами. Невозможность вычленения является следствием того, что чрезвычайно трудно обнаружить конкретные участки мозга, зарегистрированные сигналы в которых

соответствуют определенному концепту или концептам, о чем писали Б. Баарс и Н. Гейдж [18]. Однако приведенный эксперимент показал, что использование компьютерной системы управления роботизированной рукой в сочетании с аппаратом для ЭЭГ дает возможность сформировать библиотеку кодов пар «имплицитный концепт – сигнал нейросреды» (на рис. 3 и далее слова «имплицитный» и «нейросреда» в парах опущены). Эта библиотека кодов описывает интерфейсы именно 3-го порядка вида № 7 на границе между нейросредой, ментальной и цифровой средами.

Отсюда и следует ответ на основной вопрос статьи: «Зачем нужны интерфейсы 3-го порядка в информатике?» — «Для описания взаимодействий объектов на границах трех сред, когда невозможна их декомпозиция на интерфейсы 2-го порядка». При этом отметим, что интерфейсы второго порядка вида № 5 (в отличие от № 4) достаточно просто вычлениваются из интерфейсов вида № 7 и они могут использоваться независимо от интерфейсов 2-го порядка других видов или в сочетании с ними. Они применяются для установления связей между сигналами нейросреды и компьютерными кодами. Однако интерфейсы 2-го порядка не предназначены для установления связей с имплицитными личностными концептами. Поэтому в процессе обучения системы для конкретного человека

¹ Исследования, проведенные за последние три десятилетия, показали, что можно использовать сигналы мозга для передачи информации о своих намерениях компьютеру, если применять интерфейсы «мозг–компьютер». Системы, на них основанные, измеряют характеристики мозговой активности. Сначала чаще всего использовались методы ЭЭГ. В первое десятилетие нашего века значительное число исследователей стало использовать метод электрокортикографии для регистрации активности непосредственно с поверхности мозга [22]. В этой работе представлено описание проблемы сбора сигналов, протоколов и результатов научных исследований, ориентированных на решение задач нейропротезирования.

используются интерфейсы 3-го порядка вида № 7 и формируется библиотека компьютерных кодов пар «концепт—сигнал» (на рис. 3 они обозначены как «концепт₁—сигнал», ..., «концепт_N—сигнал»), соответствующих личностным концептам именно этого человека и составляющих систему его имплицитных знаний о движениях руки (см. рис. 3). Эта библиотека формируется в процессе обучения до начала управления рукой. Когда библиотека сформирована, человек может начать управление роботизированной рукой с помощью сигналов, регистрируемых методом ЭЭГ, что получило экспериментальное подтверждение. Полученный код сигнала сравнивается с кодами пар библиотеки. По их совпадению и определяется именно то движение, которое намеревался выполнить человек, и оно выполняется роботизированной рукой.

4 Заключение

В статье дана краткая характеристика шести видов интерфейсов второго порядка на границах двух из четырех сред разной природы, входящих в предметную область информатики: нейросреды, ментальной, социоинформационной и цифровой сред. Предлагаемое выделение сред планируется использовать в будущем для концептуально единого описания системы интерфейсов 2-го, 3-го и более высоких порядков.

Новое понятие «интерфейс 3-го порядка» вводится на примере взаимосвязей между имплицитными личностными концептами, сигналами нейросреды и компьютерными кодами. При этом были рассмотрены четыре среды из пяти упомянутых (кроме ДНК-среды, информационные трансформации с участием объектов которой планируются рассмотреть в другой статье). Наблюдаемое и возможное будущее увеличение числа сред предметной области информатики является во многом следствием расширения ее предметной области на информационные трансформации в живых и социальных системах.

В статье актуальность введения понятия «интерфейсы 3-го порядка» иллюстрировалась на примере процессов управления роботизированной рукой сигналами нейросреды. При этом был упомянут и еще один вид интерфейсов 3-го порядка — № 8 между объектами ментальной, социоинформационной и цифровой сред. На его основе планируется определить систему одновременного кодирования концептов (ментальная среда) и форм их представления (социоинформационная среда) в цифровой среде систем искусственного интеллекта, включая личностные и коллективные, а также темпоральные концепты знаний человека [16, 17]. Будет рассмот-

рена еще одна причина необходимости использования интерфейсов 3-го порядка: в общем случае информационные трансформации на границах № 2 и № 3 (см. рис. 2), которые широко используются в системах и средствах информатики, являются асимметричными. Многозначные слова и синонимы могут служить примерами проявления асимметрии на границе № 2 [24]. Определение интерфейсов вида № 8 служит основой разработки системы кодирования каждой из пар «значение знака — форма знака» [19] одним кодом, что должно обеспечивать представление разных значений и каждого из синонимов уникальными кодами в цифровой среде.

Литература

1. *Fatourou P., Hankin C.* Welcome to the Europe region special section // *Commun. ACM*, 2019. Vol. 62. No. 4. P. 30.
2. *Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E.* Informatics as a fundamental discipline for the 21st century // *Commun. ACM*, 2019. Vol. 62. No. 4. P. 58–63.
3. The European Commission. Digital Education Action Plan, 2018. 31 p. https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en.
4. The Committee on European Computing Education. Informatics education in Europe: Are we all in the same boat? — New York, NY, USA: ACM, 2017. Technical Report. 251 p. <https://doi.org/10.1145/3106077>.
5. *Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E.* Informatics for all: The strategy. — New York, NY, USA: ACM, 2018. 16 p.
6. The White House. Office of the Press Secretary. FACT SHEET: President Obama Announces Computer Science for All Initiative, 2016. 12 p. <https://obamawhitehouse.archives.gov/the-press-office/2016/01/30/fact-sheet-president-obama-announces-computer-science-all-initiative-0>.
7. *Newell A., Perlis A., Simon H.* Computer science // *Science*, 1967. Vol. 157. No. 3795. P. 1373–1374.
8. *Gorn S.* The computer and information sciences: A new basic discipline // *SIAM Rev.*, 1963. Vol. 5. No. 2. P. 150–155.
9. *Denning P.* Computing is a natural science // *Commun. ACM*, 2007. Vol. 50. No. 7. P. 13–18.
10. *Rozenberg G.* Computer science, informatics, and natural computing — personal reflections // *New computational paradigms: Changing conceptions of what is computable* / Eds. S. B. Cooper, B. Löwe, A. Sorbi. — New York, NY, USA: Springer, 2008. P. 373–379.
11. *Snir M.* Computer and information science and engineering: One discipline, many specialties // *Commun. ACM*, 2011. Vol. 54. No. 3. P. 38–43.
12. *Rosenbloom P.* On computing: The fourth great scientific domain. — Cambridge: MIT Press, 2013. 308 p.

13. Denning P., Rosenbloom P. Computing: The fourth great domain of science // *Commun. ACM*, 2009. Vol. 52. No. 9. P. 27–29.
14. Стенон Н. О твердом, естественно содержащемся в твердом / Пер. с лат. — М.: Изд-во АН СССР, 1957. 152 с. (*Stenonis N. De solido intra solidum naturaliter contento.* — Florentiae: Ex typographia sub signo Stellae, 1669. <http://www.biusante.parisdescartes.fr/histmed/medica/cote?06916>).
15. Зацман И. М. Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2014. № 11. С. 1–15.
16. Зацман И. М., Косарик В. В., Курчавова О. А. Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применение, 2008. Т. 2. Вып. 3. С. 54–69.
17. Zatsman I. Tracing emerging meanings by Computer: Semiotic framework // 13th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2012. Vol. 2. P. 1298–1307.
18. Baars B., Gage N. Cognition, brain, and consciousness: Introduction to cognitive neuroscience. — Amsterdam: Academic Press/Elsevier, 2010. 677 p.
19. Eco U. A theory of semiotics. — Bloomington: Indiana University Press, 1976. 356 p.
20. Caramazza A., Mahon B. Z. The organization of conceptual knowledge: The evidence from category-specific semantic deficits // *Trends Cogn. Sci.*, 2003. Vol. 7. No. 8. P. 354–361.
21. Dehaene S., Molko N., Cohen L., Wilson A. J. Arithmetic and the brain // *Current Opin. Neurobiol.*, 2004. Vol. 14. No. 2. P. 218–224.
22. Schalk G., Leuthardt E. C. Brain–computer interfaces using electrocorticographic signals // *IEEE Reviews Biomedical Eng.*, 2011. Vol. 4. No. 1. P. 140–154.
23. Sunny T. D., Aparna T., Neethu P., Venkateswaran J., Vishnupriya V., Vyas P. S. Robotic arm with brain–computer interfacing // *Proc. Tech.*, 2016. Vol. 24. P. 1089–1096.
24. Зацман И. М. Концептуальный поиск и качество информации. — М.: Наука, 2003. 272 с.

Поступила в редакцию 01.07.19

THIRD-ORDER INTERFACES IN INFORMATICS

I. M. Zatsman

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

Abstract: The European Strategy “Informatics for All,” formally launched in Brussels in March 2018, distinguishes two tiers of teaching informatics in the system of secondary and higher education. The second tier, focused on the study of informational transformations in artificial, living, and social systems, involves choosing an informatics paradigm for teaching, and then its development. The need for development is due to two reasons: firstly, a significant expansion of the scope of applications of information technologies considered in educational processes, and secondly, the integration of methods and tools of informatics into curricula in other areas of knowledge, which expands the range of information transformations. In the absence of a dominant informatics paradigm and the presence of several of its variants, the question of its choosing as the starting point of development is disputable. The “Informatics education in Europe” review, published in 2017 and preceding the development of the European strategy “Informatics for All,” lists three paradigm variants, including positioning informatics as the fourth great domain of science, proposed by Denning and Rosenbloom in 2009. A detailed description of this variant under the name of polyadic computing was given by Rosenbloom in his book in 2013. The goal of the paper is to define a new concept of “third-order interface” based on the one-natured division of the domain of informatics as polyadic computing. The relevance of the concept is illustrated by the example of robotic arm control using brain–computer interfaces.

Keywords: third-order interface; polyadic computing; one-natured media of informatics domain; information transformations

DOI: 10.14357/19922264190312

References

1. Fatourou, P., and C. Hankin. 2019. Welcome to the Europe region special section. *Commun. ACM* 62(4):30.
2. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2019. Informatics as a fundamental discipline for the 21st century. *Commun. ACM* 62(4): 58–63.
3. The European Commission. 2018. Digital Education Action Plan. Available at: https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en (accessed April 2, 2019).

4. The Committee on European Computing Education. 2017. Informatics education in Europe: Are we all in the same boat? New York, NY: ACM. Technical Report. 251 p. Available at: <https://dl.acm.org/citation.cfm?id=3106077> (accessed April 2, 2019).
5. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2018. *Informatics for all: The strategy*. New York, NY: ACM. 16 p.
6. The White House. Office of the Press Secretary. 2016. FACT SHEET: President Obama Announces Computer Science for All Initiative. 12 p. Available at: <https://obamawhitehouse.archives.gov/the-press-office/2016/01/30/fact-sheet-president-obama-announces-computer-science-all-initiative-0> (accessed April 2, 2019).
7. Newell, A., A. Perlis, and H. Simon. 1967. Computer science. *Science* 157(3795):1373–1374.
8. Gorn, S. 1963. The computer and information sciences: A new basic discipline. *SIAM Rev.* 5(2):150–155.
9. Denning, P. 2007. Computing is a natural science. *Commun. ACM* 50(7):13–18.
10. Rozenberg, G. 2008. Computer science, informatics, and natural computing — personal reflections. *New computational paradigms: Changing conceptions of what is computable*. Eds. S. B. Cooper, B. Löwe, and A. Sorbi. New York, NY: Springer. 373–379.
11. Snir, M. 2011. Computer and information science and engineering: one discipline, many specialties. *Commun. ACM* 54(3):38–43.
12. Rosenbloom, P. 2013. *On computing: The fourth great scientific domain*. Cambridge: MIT Press. 308 p.
13. Denning, P., and P. Rosenbloom. 2009. Computing: The fourth great domain of science. *Commun. ACM* 52(9):27–29.
14. Stenonis, N. 1669. *De solido intra solidum naturaliter contento*. Florentiae: Ex typographia sub signo Stellae. Available at: <http://www.biusante.parisdescartes.fr/histmed/medica/cote?06916> (accessed April 2, 2019).
15. Zatsman, I. 2014. Table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):233–246.
16. Zatsman, I., V. Kosarik, and O. Kurchavova. 2008. Zadachi predstavleniya lichnostnykh i kollektivnykh kontseptov v tsifrovoy srede [Problems of representation of personal and collective concepts in the digital medium]. *Informatika i ee Primeneniya — Inform. Appl.* 2(3):54–69.
17. Zatsman, I. 2012. Tracing emerging meanings by computer: Semiotic framework. *13th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 2:1298–1307.
18. Baars, B., and N. Gage. 2010. *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Amsterdam: Academic Press/Elsevier. 677 p.
19. Eco, U. 1976. *A theory of semiotics*. Bloomington: Indiana University Press. 356 p.
20. Caramazza, A., and B. Z. Mahon. 2003. The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends Cogn. Sci.* 7(8): 354–361.
21. Dehaene, S., N. Molko, L. Cohen, and A. J. Wilson. 2004. Arithmetic and the brain. *Current Opin. Neurobiol.* 14(2):218–224.
22. Schalk, G., and E. C. Leuthardt. 2011. Brain–computer interfaces using electrocorticographic signals. *IEEE Reviews Biomedical Eng.* 4(1):140–154.
23. Sunny, T. D., T. Aparna, P. Neethu, J. Venkateswaran, V. Vishnupriya, and P. S. Vyas. 2016. Robotic arm with brain–computer interfacing. *Proc. Tech.* 24:1089–1096.
24. Zatsman, I. 2003. *Kontseptual'nyy poisk i kachestvo informatsii* [Conceptual retrieval and quality of information]. Moscow: Nauka. 272 p.

Received July 1, 2019

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

АРХИТЕКТУРА СИСТЕМЫ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА

В. А. Нуриев¹

Аннотация: Описывается архитектура системы нейронного машинного перевода. Актуальность объекта исследования обусловлена тем, что сейчас доминирующей парадигмой в области автоматизированного перевода стал машинный перевод с использованием искусственных нейронных сетей. Качество нейронного перевода в значительной мере превосходит результаты, обеспечиваемые предыдущим поколением машинных переводчиков, однако и оно все еще далеко отстоит от качества переводов, выполненных человеком-экспертом. Для его улучшения необходимо более четкое понимание устройства системы нейронного машинного перевода. В общем случае архитектура системы нейронного машинного перевода включает в себя две рекуррентные нейронные сети (РНС), одна из которых ответственна за обработку входной текстовой последовательности, а другая — за формирование выходного текста-перевода. Часто в систему также встраивается механизм внимания, позволяющий оптимизировать работу с протяженными входными последовательностями. Архитектура системы описывается на примере переводного сервиса Google (translate.google.com, Google’s Neural Machine Translation system, GNMT), так как он стал одним из самых востребованных в мире на данный момент: ежедневно сервис обрабатывает около 143 млрд слов более чем на 100 языках. В заключение формулируются некоторые перспективы исследования.

Ключевые слова: нейронный машинный перевод; искусственные нейронные сети; рекуррентные нейронные сети; механизм внимания; архитектура системы машинного перевода; Google’s Neural Machine Translation system

DOI: 10.14357/19922264190313

1 Введение

В настоящее время доминирующей парадигмой в области автоматизированного перевода стал машинный перевод с использованием искусственных нейронных сетей [1–4], пришедший на смену статистическим методам². Нейронные сети инкорпорированы в переводные системы компаний Google, Microsoft, Яндекс, DeepL, PROMT и др. В общем случае архитектура такой системы включает в себя две РНС (recurrent neural networks), одна из которых ответственна за обработку входной текстовой последовательности, а другая — за формирование выходного текста-перевода. Часто в систему также встраивается механизм внимания, позволяющий оптимизировать работу с протяженными входными последовательностями [1, 7].

Качество нейронного перевода в значительной мере превосходит результаты, обеспечиваемые предыдущим поколением машинных переводчиков, однако и оно все еще далеко отстоит от качества переводов, выполненных человеком-экспертом. Для его улучшения необходимо более

четкое понимание устройства системы нейронного машинного перевода.

Цель статьи, следовательно, — описать архитектуру такой системы. Она описывается на примере переводного сервиса Google — GNMT³, так как он стал одним из самых востребованных в мире на данный момент: ежедневно сервис обрабатывает около 143 млрд слов более чем на 100 языках [8]. В качестве основных источников материала используются статьи, авторство которых принадлежит разработчикам GNMT [9, 10].

2 GNMT и недостатки нейронного машинного перевода

Первые попытки спроектировать нейронный машинный переводчик выявили несколько его существенных недостатков. По сравнению с системами статистического машинного перевода скорость его обучения оказалась значительно ниже. Из-за

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

²О системах статистического машинного перевода см., например, [5, 6].

³Интернет-версия переводчика.

большого числа вовлеченных параметров больше времени требовалось на формирование выходной информации. Кроме того, он оказался неэффективен при работе с редкими словами, а в отдельных случаях ему вообще не удавалось перевести все слова входного предложения.

Для решения проблемы перевода редких слов предлагались разные подходы, например применение специального механизма внимания, чтобы копировать редкие слова и в исходной форме воспроизводить их в выходном тексте-переводе [11]. Практическое применение этого подхода не принесло ожидаемых результатов: простое копирование редких слов не всегда оказывается лучшим способом перевода — оно не подходит, когда согласно конвенциям переводящего языка требуется транскрибирование.

Разработка системы нейронного машинного перевода GNMT была нацелена на учет указанных недостатков. В этой системе задействованы РНС с долгой краткосрочной памятью (ДКП, long short-term memory) [12, 13]. Они имеют восьмислойную структуру, где связь между слоями осуществляется, в том числе, за счет остаточных соединений (residual connections), обеспечивающих градиентный поток (gradient flow) [14].

Для достижения параллелизма в обработке данных внимание нижнего уровня декодирующей сети подключается к верхнему уровню кодирующей сети. Чтобы уменьшить время вывода, используются вычисления с низкой точностью (low-precision arithmetic), а в целях дополнительного ускорения — специальное оборудование (Google's Tensor Processing Unit — тензорный процессор Google). Перевод редких слов осуществляется путем их представления на входе и выходе в форме набора составных элементов (частей слова — wordpieces) (подробнее см. в [15]).

Разбивка на составные элементы позволяет достичь баланса в выборе минимальной переводной единицы, не спускаясь на уровень отдельных букв, что чрезвычайно увеличило бы входные и выходные последовательности, и между тем не оставаясь на уровне слова, если слово оказывается редким и незнакомым. Кроме того, это помогает отказаться от необходимости вводить специальную обработку незнакомых слов.

Для решения проблемы неполного перевода используется лучевой поиск (beam search), с помощью которого во время декодирования соизмеряются возможные варианты (hypotheses) разной протяженности, и штраф за пропуск слова, что должно обеспечивать перевод всей полученной на входе информации.

3 Архитектура системы нейронного машинного перевода GNMT

Спроектированная модель использует распространенную схему обучения «последовательность к последовательности» (sequence-to-sequence learning framework) с механизмом внимания (см. подробнее [1, 16]). Она включает в себя три компонента: кодирующую и декодирующую сети (кодер/декодер), а также сеть внимания. Кодирующая сеть трансформирует исходное предложение в список векторов, причем на каждый входной символ приходится один вектор. Обработывая этот список векторов, декодер производит поочередно по одному символу до тех пор, пока не достигнет специального символа, ассоциированного с окончанием предложения (special end-of-sentence symbol). Кодирующий и декодирующий сегменты соединяются модулем внимания (attention module), ответственным за способность декодера фокусироваться на различных отрезках исходного предложения в процессе декодирования.

В нотации при описании архитектуры GNMT для обозначения векторов (\mathbf{v}, \mathbf{o}_i) используются выделенные полужирным строчные буквы, для обозначения матриц (\mathbf{U}, \mathbf{W}) — выделенные полужирным прописные буквы для представления последовательностей (X, Y) и отдельных символов в них (x_1, x_2) — прописные буквы и строчные буквы соответственно.

Пусть (X, Y) — это пара, состоящая из исходного и переводного предложений; $X = x_1, x_2, x_3, \dots, x_M$ — последовательность из M символов в исходном предложении, а $Y = y_1, y_2, y_3, \dots, y_N$ — последовательность из N символов в переводном предложении. Тогда кодер — это функция следующей формы:

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M = \text{КодерРНС}(x_1, x_2, x_3, \dots, x_M).$$

В этом уравнении $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M$ — список векторов фиксированного размера, причем число составляющих этот список членов то же, что и число символов в исходном предложении (M в данном примере). Применяя цепное правило, условную вероятность последовательности $P(Y|X)$ можно разложить как

$$P(Y|X) = P(Y|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M) = \prod_{i=1}^N P(y_i|y_0, y_1, y_2, \dots, y_{i-1}; \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M),$$

где y_0 — специальный ассоциированный с началом предложения символ (“beginning of sentence” symbol), который ставится в начале каждого переводного предложения.

При выводе вычисляется вероятность следующего символа с учетом кодировки исходного предложения и членов выходной последовательности, уже прошедших декодирование: $P(y_i | y_0, y_1, y_2, y_3, \dots, y_{i-1}; \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M)$.

Декодер представляет собой сочетание РНС и слоя softmax. Рекуррентная нейронная сеть создает скрытое состояние \mathbf{y}_i для следующего прогнозируемого символа, которое затем проходит через слой softmax, чтобы сгенерировать распределение вероятности по выходным символам-кандидатам. Согласно экспериментальным данным, для улучшения качества перевода в системах нейронного машинного перевода необходимо обеспечить достаточную глубину кодирующей и декодирующей РНС; только так РНС могут учитывать тонкие нюансы исходного и переводящего языков. Похожие наблюдения были сделаны в [16]: глубокие РНС с ДКП значительно превосходят по производительности неглубокие (о ДКП см. подробнее [12]). Поэтому здесь, как и, например, в [17], для кодирующей и декодирующей РНС используется глубокая сеть с ДКП.

Модуль внимания сформирован, как в [1]. Пусть \mathbf{y}_{i-1} будет выходным сигналом декодирующей РНС для предыдущего временного шага декодирования (здесь используется выходной сигнал нижнего слоя декодера). Контекст внимания \mathbf{a}_i для текущего временного шага вычисляется по следующей формуле:

$$\mathbf{a}_i = \sum_{t=1}^M p_t \mathbf{x}_t.$$

Здесь

$$p_t = \frac{\exp(s_t)}{\sum_{t=1}^M \exp(s_t)} \quad \forall t, 1 \leq t \leq M,$$

где

$$s_t = \text{Функция внимания}(\mathbf{y}_{i-1}, \mathbf{x}_t) \quad \forall t, 1 \leq t \leq M,$$

а Функция внимания — это сеть прямой связи с одним скрытым слоем.

Как уже говорилось выше, глубокие РНС с ДКП часто показывают лучший результат, чем неглубокие. Однако простое наложение слоев в ДКП работает только для определенного числа слоев, за его пределами сеть становится слишком медленной и труднообучаемой — вероятно, из-за проблем взрывного и исчезающего градиента (о них см.

в [18, 19]). Эксперименты показывают, что для решения широкомасштабных задач перевода простое наложение слоев в ДКП хорошие результаты дает при четырех слоях, удовлетворительные — при шести и очень плохие, когда их число превышает восемь.

Руководствуясь идеей о том, что для увеличения скорости обучения нейронной сети ее слои можно представить как остаточные функции со ссылкой на входные данные слоя, а не как функции без ссылок (см. подробнее в [14, 20, 21]), разработчики GNMT ввели остаточные соединения между организованными в последовательность слоями ДКП. Пусть ДКП_{*i*} и ДКП_{*i+1*} будут *i*-м и (*i* + 1)-м слоями ДКП в последовательности, параметры которых — \mathbf{W}^i и \mathbf{W}^{i+1} соответственно. Тогда на временном шаге *t* для организованных в последовательность слоев ДКП без остаточных соединений имеем:

$$\begin{aligned} \mathbf{c}_t^i, \mathbf{m}_t^i &= \text{ДКП}_i(\mathbf{c}_{t-1}^i, \mathbf{m}_{t-1}^i, \mathbf{x}_t^{i-1}; \mathbf{W}^i); \\ \mathbf{c}_t^{i+1}, \mathbf{m}_t^{i+1} &= \text{ДКП}_{i+1}(\mathbf{c}_{t-1}^{i+1}, \mathbf{m}_{t-1}^{i+1}, \mathbf{x}_t^i; \mathbf{W}^{i+1}), \end{aligned}$$

где \mathbf{x}_t^i — данные, поступающие на вход ДКП_{*i*} на временном шаге *t*:

$$\mathbf{x}_t^i = \mathbf{m}_t^i,$$

а \mathbf{m}_t^i и \mathbf{c}_t^i — скрытые состояния и состояния памяти ДКП_{*i*} на временном шаге *t* соответственно.

При введении остаточных соединений между ДКП_{*i*} и ДКП_{*i+1*} эти уравнения принимают следующий вид:

$$\begin{aligned} \mathbf{c}_t^i, \mathbf{m}_t^i &= \text{ДКП}_i(\mathbf{c}_{t-1}^i, \mathbf{m}_{t-1}^i, \mathbf{x}_t^{i-1}; \mathbf{W}^i); \\ \mathbf{c}_t^{i+1}, \mathbf{m}_t^{i+1} &= \text{ДКП}_{i+1}(\mathbf{c}_{t-1}^{i+1}, \mathbf{m}_{t-1}^{i+1}, \mathbf{x}_t^i; \mathbf{W}^{i+1}), \end{aligned}$$

где

$$\mathbf{x}_t^i = \mathbf{m}_t^i + \mathbf{x}_t^{i-1}.$$

Остаточные соединения значительно улучшают поток градиента при обратном проходе, что позволяет обучать очень глубокие сети кодера и декодера. В большинстве экспериментов авторами использовалось 8 слоев ДКП для кодера и декодера, между тем применение остаточных соединений дает возможность обучать сети гораздо большей глубины.

Информация, необходимая для перевода некоторых слов, может содержаться в любом месте входного текстового фрагмента. Обычно информационная развертка происходит слева направо как на входе, так и на выходе, однако в зависимости от языковой пары информация (контекст) для перевода конкретного выходного слова может распределяться или даже разделяться между определенными областями входного фрагмента. Учитывая это, чтобы получить наилучший возможный контекст

в каждом отдельном случае, для кодера целесообразно использовать двунаправленную РНС (см. подробнее в [1, 22]). Вместе с тем, чтобы обеспечить максимально возможное распараллеливание во время вычислений, двунаправленные соединения в GNMT используются только для нижнего уровня кодера — все его остальные уровни являются однонаправленными. Двунаправленный нижний слой кодера, в свою очередь, состоит из двух слоев, один из которых обрабатывает исходное предложение слева направо, другой — справа налево. Затем их выходные данные объединяются и подаются на следующий (однаправленный) слой.

Для увеличения скорости обучения в системе GNMT применяется параллелизм модели и параллелизм данных. Параллелизм данных достигается обучением n копий модели одновременно, применяется алгоритм распределенного обучения Doudour SGD (stochastic gradient descent) [23]. Все n копий оперируют одним и тем же набором параметров модели, при этом каждая копия асинхронно обновляет параметры, используя комбинацию адаптивного алгоритма Adam (adaptive moment estimation) [24] и алгоритма SGD. При экспериментальной проверке n , как правило, составляло около 10. Каждая модель обрабатывала минипакет (mini-batch) из m пар предложений за единицу времени, m при этом в большинстве случаев равнялось 128.

В дополнение к параллелизму данных для повышения скорости вычисления градиента на каждой копии используется параллелизм модели. Сети кодера и декодера разделены послойно и размещаются на разных графических процессорах, запуская каждый слой на отдельном графическом процессоре. Поскольку все слои, кроме первого слоя кодера, однонаправленные, слой $i + 1$ может начать вычисления до того, как слой i полностью их завершит, что ускоряет обучение. Слой softmax также разделен, причем каждый раздел отвечает за подмножество символов в выходном словаре.

Параллелизм модели накладывает определенные ограничения на ее архитектуру. Например, нельзя для каждого слоя кодера иметь двунаправленный слой ДКП: это уменьшит параллелизм между последующими слоями, поскольку каждый очередной слой не сможет начать вычисления, пока предыдущий их не закончит в обоих направлениях. В результате параллельно можно будет использовать только два графических процессора — один

для прямого направления и один для обратного. В механизме внимания, чтобы достичь максимального параллелизма при работе декодера, выходное представление его нижнего слоя выровнено с выходным представлением верхнего слоя кодера. Выравнивание верхнего слоя декодера с верхним слоем кодера препятствовало бы распараллеливанию в сети декодера, и применение более одного графического процессора для декодирования не имело бы смысла.

4 Заключительные замечания

В настоящей статье на примере GNMT рассмотрена только архитектура системы нейронного машинного перевода. Для лучшего понимания механизмов ее работы и, в частности, того, как она производит спецификацию переводных вариантов, также важно иметь представление, например, о принципах ее обучения (об этом см. [9, 10]). Не вызывает сомнений, что дальнейшее изучение систем нейронного перевода необходимо, так как, несмотря на их явные преимущества по сравнению со статистическими машинными переводчиками, опытные данные показывают недостаточный уровень качества переводов, которое они обеспечивают¹.

В заключение следует отметить, что одним из перспективных направлений в оптимизации систем нейронного машинного перевода представляется разработка способов интеграции языковых моделей в их архитектуру [25].

Литература

1. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate // Conference (International) on Learning Representations. — San Diego, CA, USA, 2015. P. 1–15. <https://arxiv.org/pdf/1409.0473.pdf>.
2. Castilho Sh., Moorkens J., Gaspari F., Calixto I., Tinsley J., Way A. Is neural machine translation the new state of the art? // Prague Bull. Math. Linguistics, 2017. Vol. 108. No. 1. P. 109–120.
3. Zhao Y., Wang Y., Zhang J., Zong C. Cost-aware learning rate for neural machine translation // Chinese computational linguistics and natural language processing based on naturally annotated big data / Eds. M. Sun, X. Wang, B. Chang, D. Xiong. — Cham: Springer International Publishing, 2017. P. 85–93.

¹На это указывают и сами создатели GNMT. Например, с этим связаны эксперименты с обучением трансферного типа (transfer learning experiments), а также попытки отказаться от интерлингвального подхода к переводу (когда сначала производится перевод на язык-посредник типа английского и затем с него осуществляется перевод на желаемый выходной язык) и использовать возможности «zero-shot translation» (переводить с одного языка на другой, не используя примеры для этой языковой пары в процессе обучения) [10].

4. Zamora-Martinez F., Castro-Bleda M.J. Efficient embedded decoding of neural network language models in a machine translation system // *Int. J. Neural Syst.*, 2018. Vol. 28. Iss. 9. Art. No. 1850007. P. 1–15.
5. Koehn Ph. *Statistical machine translation*. — New York, NY, USA: Cambridge University Press, 2010. 447 p.
6. Hearne M., Way A. *Statistical machine translation: A guide for linguists and translators* // *Lang. Linguist. Compass*, 2011. Vol. 5. No. 5. P. 205–226.
7. Choi H., Cho K., Bengio Y. Fine-grained attention mechanism for neural machine translation // *Neurocomputing*, 2018. Vol. 284. P. 171–176.
8. Davenport C. Google Translate processes 143 billion words every day. androidpolice.com, 2018. <https://www.androidpolice.com/2018/10/09/google-translate-processes-143-billion-words-every-day>.
9. Wu Y., Schuster M., Chen Z., et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv.org, 2016. <https://arxiv.org/pdf/1609.08144.pdf>.
10. Johnson M., Schuster M., Le Q. V., et al. Google's multilingual neural machine translation system: Enabling zero-shot translation // *T. Assoc. Computational Linguistics*, 2017. Vol. 5. P. 339–351.
11. Sébastien J., Kyunghyun C., Memisevic R., Bengio Y. On using very large target vocabulary for neural machine translation // 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing of the Asian Federation of Natural Language Processing Proceedings. — Beijing, China: The Association for Computer Linguistics, 2015. Vol. 1. P. 1–10.
12. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Comput.*, 1997. Vol. 9. No. 8. P. 1735–1780.
13. Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM // *Neural Comput.*, 2000. Vol. 12. No. 10. P. 2451–2471.
14. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // *IEEE Conference on Computer Vision and Pattern Recognition*. — Las Vegas, NV, USA: IEEE, 2016. P. 770–778.
15. Schuster M., Nakajima K. Japanese and Korean voice search // *IEEE Conference (International) on Acoustics, Speech and Signal Processing*. — Piscataway, NJ, USA: IEEE, 2012. P. 5149–5152.
16. Sutskever I., Vinyals O., Le Q. V. Sequence to sequence learning with neural networks // 27th Conference (International) on Neural Information Processing Systems Proceedings. — Cambridge, MA, USA: MIT Press, 2014. Vol. 2. P. 3104–3112.
17. Luong M.-T., Sutskever I., Le Q. V., Vinyals O., Zaremba W. Addressing the rare word problem in neural machine translation // 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings. — Beijing, China: Association for Computational Linguistics, 2015. Vol. 1. P. 11–19.
18. Hochreiter S., Bengio Y., Frasconi P., Schmidhuber J. Gradient flow in recurrent nets: The difficulty of learning long-term dependencies // *A field guide to dynamical recurrent networks* / Eds. J. F. Kolen, S. Kremer. — Los Alamitos, CA, USA: IEEE Press, 2001. P. 237–243.
19. Pascanu R., Mikolov T., Bengio Y. On the difficulty of training recurrent neural networks. arXiv.org, 2013. <https://arxiv.org/pdf/1211.5063v2.pdf>.
20. Fahlman S. E., Lebiere C. The cascade-correlation learning architecture // *Advances in neural information processing systems* / Ed. D. S. Touretzky. — San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990. Vol. 2. P. 524–532.
21. Srivastava R. K., Greff K., Schmidhuber J. Highway networks. arXiv.org, 2015. <https://arxiv.org/pdf/1507.06228.pdf>.
22. Schuster M., Paliwal K. Bidirectional recurrent neural networks // *IEEE T. Signal Proces.*, 1997. Vol. 45. No. 11. P. 2673–2681.
23. Dean J., Corrado G. S., Monga R., et al. Large scale distributed deep networks // *Advances in neural information processing systems*, 2012. Vol. 25. <https://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf>.
24. Kingma D. P., Ba J. Adam: A method for stochastic optimization. arXiv.org, 2014. <https://arxiv.org/pdf/1412.6980.pdf>.
25. Gulcehre C., Firat O., Xu K., Cho K., Bengio Y. On integrating a language model into neural machine translation // *Comput. Speech Lang.*, 2017. Vol. 45. P. 137–148.

Поступила в редакцию 30.06.19

ARCHITECTURE OF A MACHINE TRANSLATION SYSTEM

V. A. Nuriev

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 describes architecture of a Neural Machine Translation (NMT) system. The subject is brought up since NMT, i. e., translation using artificial neural networks, is now a leading Machine Translation paradigm. The NMT systems manage to deliver much better quality of output than the machine translators of the previous

generation (statistical translation systems) do. Yet, the translation they produce still may contain various errors and it is relatively inaccurate compared with human translations. Therefore, to improve its quality, it is important to see more clearly how an NMT system is built and works. Commonly, its architecture consists of two recurrent neural networks, one to get the input text sequence and the other to generate translated output (text sequence). The NMT system often has an attention mechanism helping it cope with long input sequences. As an example, Google's NMT system is taken as the Google Translate service is one of the most highly demanded today, it processes around 143 billion words in more than 100 languages per day. The paper concludes with some perspectives for future research.

Keywords: neural machine translation; artificial neural networks; recurrent neural networks; attention mechanism; architecture of a machine translation system; Google's Neural Machine Translation system

DOI: 10.14357/19922264190313

References

- Bahdanau, D., K. Cho, and Y. Bengio. 2015. Neural machine translation by jointly learning to align and translate. *Conference (International) on Learning Representations*. San Diego, CA. 1–15. Available at: <https://arxiv.org/pdf/1409.0473.pdf> (accessed June 20, 2019).
- Castilho, Sh., J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley, and A. Way. 2017. Is neural machine translation the new state of the art? *Prague Bull. Math. Linguistics* 108(1):109–120.
- Zhao, Y., Y. Wang, J. Zhang, and C. Zong. 2017. Cost-aware learning rate for neural machine translation. *Chinese computational linguistics and natural language processing based on naturally annotated big data*. Eds. M. Sun, X. Wang, B. Chang, and D. Xiong. Cham: Springer International Publishing. 85–93.
- Zamora-Martinez, F., and M. J. Castro-Bleda. 2018. Efficient embedded decoding of neural network language models in a machine translation system. *Int. J. Neural Syst.* 28(9):1850007(1–15).
- Hearne, M., and A. Way. 2011. Statistical machine translation: A guide for linguists and translators. *Lang. Linguist. Compass* 5(5):205–226.
- Koehn, Ph. 2010. *Statistical machine translation*. New York, NY: Cambridge University Press. 447 p.
- Choi, H., K. Cho, and Y. Bengio. 2018. Fine-grained attention mechanism for neural machine translation. *Neurocomputing* 284:171–176.
- Davenport, C. 2018. Google Translate processes 143 billion words every day. Available at: <https://www.androidpolice.com/2018/10/09/google-translate-processes-143-billion-words-every-day> (accessed June 20, 2019).
- Wu, Y., M. Schuster, Z. Chen, *et al.* 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. Available at: <https://arxiv.org/pdf/1609.08144.pdf> (accessed June 20, 2019).
- Johnson, M., M. Schuster, Q. V. Le, *et al.* 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *T. Assoc. Computational Linguistics* 5:339–351.
- Sébastien, J., C. Kyunghyun, R. Memisevic, and Y. Bengio. 2015. On using very large target vocabulary for neural machine translation. *53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing of the Asian Federation of Natural Language Processing Proceedings*. Beijing, China: The Association for Computer Linguistics. 1:1–10.
- Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9(8):1735–1780.
- Gers, F. A., J. Schmidhuber, and F. Cummins. 2000. Learning to forget: Continual prediction with LSTM. *Neural Comput.* 12(10):2451–2471.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV: IEEE. 770–778.
- Schuster, M., and K. Nakajima. 2012. Japanese and Korean voice search. *IEEE Conference (International) on Acoustics, Speech and Signal Processing*. Piscataway, NJ: IEEE. 5149–5152.
- Sutskever, I., O. Vinyals, and Q. V. Le. 2014. Sequence to sequence learning with neural networks. *27th Conference (International) on Neural Information Processing Systems Proceedings*. Cambridge, MA: MIT Press. 2:3104–3112.
- Luong, M.-T., I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba. 2015. Addressing the rare word problem in neural machine translation. *53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings*. Beijing, China: Association for Computational Linguistics. 1:11–19.
- Hochreiter, S., Y. Bengio, P. Frasconi, and J. Schmidhuber. 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *A field guide to dynamical recurrent networks*. Eds. J. F. Kolen, and S. Kremer. Los Alamitos, CA: IEEE Press. 237–243.
- Pascanu, R., T. Mikolov, and Y. Bengio. 2013. On the difficulty of training recurrent neural networks. [arxiv.org. Available at: https://arxiv.org/pdf/1211.5063v2.pdf](https://arxiv.org/pdf/1211.5063v2.pdf) (accessed June 20, 2019).
- Fahlman, S. E., and C. Lebiere. 1990. The cascade-correlation learning architecture. *Advances in neural information processing systems*. Ed. D. S. Touretzky. San Francisco, CA: Morgan Kaufmann Publishers Inc. 2:524–532.
- Srivastava, R. K., K. Greff, and J. Schmidhuber. 2015. Highway networks. [arxiv.org. Available at: https://arxiv.org/pdf/1507.06228.pdf](https://arxiv.org/pdf/1507.06228.pdf) (accessed June 20, 2019).

22. Schuster, M., and K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE T. Signal Proces.* 45(11):2673–2681.
23. Dean, J., G.S. Corrado, R. Monga, *et al.* 2012. Large scale distributed deep networks. *Advances in neural information processing systems*. Vol. 25. Available at: <https://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf> (accessed June 20, 2019).
24. Kingma, D.P., and J. Ba. 2014. Adam: A method for stochastic optimization. arxiv.org. Available at: <https://arxiv.org/pdf/1412.6980.pdf> (accessed June 20, 2019).
25. Gulcehre, C., O. Firat, K. Xu, K. Cho, and Y. Bengio. 2017. On integrating a language model into neural machine translation. *Comput. Speech Lang.* 45:137–148.

Received June 30, 2019

Contributor

Nuriev Vitaly A. (b. 1980) — Candidate of Science (PhD) in philology, 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; nurieff.v@gmail.com

МЕТОДИКА ПОИСКА ИМПЛИЦИТНЫХ ЛОГИКО-СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ В ТЕКСТЕ*

А. А. Гончаров¹, О. Ю. Инькова²

Аннотация: Рассматривается методика поиска имплицитных логико-семантических отношений (ЛСО) в текстах надкорпусной базы данных (НБД) коннекторов. На примере направления перевода «русский → французский» описаны стадии процесса поиска: (1) выбор анализируемого ЛСО и создание массива аннотаций тех коннекторов русского языка, которые могут считаться прототипическими средствами выражения этого ЛСО; (2) анализ полученного массива аннотаций и выявление основных переводных эквивалентов для данных коннекторов во французском языке; (3) использование двуязычного поиска в НБД, причем для текстов на русском языке исключаются коннекторы, аннотированные на первой стадии, а для текстов на французском проводится поиск по языковым единицам, представляющим собой наиболее частотные переводные эквиваленты этих коннекторов, определенные на второй стадии; (4) аннотирование найденных таким образом пар фрагментов параллельных текстов; (5) анализ сформированных аннотаций, дающий возможность произвести категоризацию случаев, когда ЛСО в русском языке является имплицитным. Предлагаемая методика поиска в НБД позволяет получить новое знание об имплицитных ЛСО.

Ключевые слова: выявление имплицитной информации; коннекторы; контрастная лингвистика; корпусная лингвистика; надкорпусные базы данных; логико-семантические отношения

DOI: 10.14357/19922264190314

1 Введение

Кросслингвистическая НБД коннекторов (подробнее о структуре НБД см. [1, 2], об используемой системе терминов см. [3, 4]) разрабатывается в ИПИ ФИЦ ИУ РАН на основе текстов параллельных подкорпусов Национального корпуса русского языка (НКРЯ) с учетом особенностей этих языковых единиц, основная функция которых состоит в выражении ЛСО в тексте [5].

В процессе работы исследователи столкнулись со случаями, когда в параллельных текстах на одном из двух языков ЛСО выражается коннектором, а на другом языке либо используются отличные от коннекторов средства, либо ЛСО не выражено никакой языковой единицей, т. е. является имплицитным.

Такие случаи представляют интерес как для лингвистики, так и для разработки методов автоматической обработки текстов и повышения качества машинного перевода, поскольку позволяют получить новое знание о языковых средствах, которые используются для выражения того или иного ЛСО.

В то же время выявление случаев, когда то или иное ЛСО не выражено в тексте эксплицитно (или выражено не своим прототипическим³ показателем), является довольно сложной и трудоемкой задачей. Самый распространенный на сегодняшний день способ выявления таких случаев — сплошная ручная разметка, требующая значительных временных и человеческих ресурсов [7–10].

Подобные случаи не являются, однако, редкими и исключительными. Таблица 1 содержит данные по числу аннотаций переводных соответствий (далее — ПС), сформированных в НБД, где в одной из частей были проставлены признаки, указывающие на отсутствие эксплицитного показателя ЛСО. Это явление может быть вызвано рядом факторов, описание которых выходит за рамки данной статьи (о некоторых из них см. [11]). Признаки, присваиваемые аннотации при разметке во всех таких случаях, объединены в кластер *Zero*.

Основной причиной столь значительного численного перевеса аннотаций, в которых признаки из кластера *Zero* проставлены именно во франко-

*Работа выполнена в Институте проблем информатики ФИЦ ИУ РАН при поддержке РФФИ (проект по гранту № 18-07-00192).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

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

³Прототипическим считается коннектор, который специализируется на выражении данного ЛСО, наиболее часто упоминается как его показатель в грамматиках и словарях, а значит, может считаться репрезентативным образцом показателя данного ЛСО (см. понятие «лучшего представителя» в когнитивной лингвистике и психологии, например, в работе [6]).

Таблица 1 Распределение аннотаций ПС с признаками из кластера *Zero* в НБД коннекторов до начала эксперимента по описываемой методике (указано число аннотаций, сформированных в НБД до 11.03.2019 включительно (учитываются только кортежи первого типа, подробнее о категориях кортежей см. [1]). Аннотации в рамках эксперимента формировались начиная с 12.03.2019)

Направление перевода	Общее число аннотаций ПС до начала эксперимента	Язык	Число аннотаций ПС с признаками из кластера <i>Zero</i>	Примерная доля аннотаций ПС с признаками из кластера <i>Zero</i>
Русский → французский	12 667	Русский	94	0,7%
		Французский	1438	11,4%
Французский → русский	3049	Французский	435	14,2%
		Русский	64	2,1%

язычной части, является прежде всего тот факт, что поиск пар фрагментов параллельных текстов для аннотирования осуществлялся, как правило, в текстах на русском языке по эксплицитной языковой единице для обоих направлений перевода. В результате было сформировано 1438 аннотаций в направлении «русский → французский» и 435 аннотации в направлении «французский → русский», где ЛСО имплицировано в тексте на французском языке. Цель статьи состоит в постановке и решении задачи поиска в НБД подобных случаев в русскоязычном тексте, их последующего аннотирования и анализа. Для эффективного решения этой задачи и была разработана рассматриваемая методика.

2 Исходные данные для постановки задачи поиска

Первые две стадии методики обеспечивают получение исходных данных для решения задачи: во-первых, выбор анализируемого ЛСО и создание массива аннотаций тех коннекторов русского языка, которые могут считаться прототипическими средствами выражения этого ЛСО; во-вторых, анализ полученного массива аннотаций и выявление в переводах на французский язык основных функционально эквивалентных фрагментов (ФЭФ) данных коннекторов русского языка. В настоящей статье методика поиска случаев имплицитных ЛСО иллюстрируется на примере ЛСО причины. К основным коннекторам, выражающим данное отношение в русском языке, относятся *потому что*, *ибо*, *поскольку*, которые наиболее часто упоминаются в словарях и грамматиках (например, [12, с. 577–583]). При поиске контекстов не учитывалось наличие запятой между частями коннектора *потому*

что и возможное сочетание *поскольку* с компонентом *то*, однако при аннотировании эти коннекторы фиксировались в той форме, в которой они употреблены в тексте. Именно поэтому далее приводятся данные не по трем, а по пяти коннекторам русского языка, для которых в сумме сформировано 350 аннотаций в НБД (для направления перевода «русский → французский»): *ибо* (50 аннотаций), *поскольку* (46 аннотаций), *поскольку||то*¹ (4 аннотации), *потому что* (196 аннотаций) и *потому, что* (54 аннотации). Поскольку в рамках данной статьи перечисленные коннекторы анализируются в текстах оригинала, они предстают в качестве стимулов перевода (СП). Соответствие СП и ФЭФ будем называть моделью перевода (МП).

Этот объем аннотаций оказался достаточным для выявления основных ФЭФ перечисленных коннекторов во французском языке. Таблица 2 содержит информацию о тех ПС, где коннектору русского языка соответствует какой-либо коннектор французского языка (т. е. случаи, когда перевод коннектора является конгруэнтным; о понятиях «конгруэнтный» и «дивергентный» перевод применительно к коннекторам см. [13]). В первом столбце перечислены коннекторы, зафиксированные хотя бы один раз как ФЭФ выбранных 5 коннекторов русского языка. Следующие 5 столбцов содержат число ПС, в которых ФЭФ из первого столбца соответствует каждому из анализируемых СП. Последний столбец отражает общее число ПС для того или иного ФЭФ. Именно эта информация является определяющей при выборе единицы французского языка, которая будет использована для поиска ПС, где ЛСО причины в русском языке выражается имплицитно.

Из табл. 2 видно, что среди 28 ФЭФ (не все из которых, впрочем, выражают именно ЛСО причины), упорядоченных по частотности, со значитель-

¹Используемая при перечислении коннекторов двойная вертикальная черта «||» указывает на то, что коннектор состоит из двух или более расположенных дистантно компонентов, каждый из которых вводит отдельный фрагмент текста; одиночная вертикальная черта «|» говорит о том, что языковые единицы (элементы), составляющие коннектор или его компонент, разделены текстом, но находятся при этом в рамках одного и того же фрагмента.

Таблица 2 Фрагмент таблицы по МП, где и в оригинале, и в переводе использован коннектор

ФЭФ французского языка, представляющий собой коннектор	СП в русском языке					Общее число ПС
	ибо	поскольку	поскольку то	потому что	потому, что	
1. <i>car</i>	35	5	0	65	0	105
2. <i>parce que</i>	1	1	0	63	35	100
3. <i>puisque</i>	3	3	0	6	0	12
4. <i>comme</i>	0	5	2	2	0	9
5. <i>dans la mesure où</i>	0	5	0	0	0	5
...
28. <i>vu</i>	0	0	0	1	0	1
ВСЕГО	42 (из 50)	29 (из 46)	2 (из 4)	156 (из 196)	47 (из 54)	276 (из 350)

ным перевесом выделяются два коннектора французского языка — *car* и *parce que*, употребленные в 105 и 100 случаях соответственно. После того как наиболее частотный ФЭФ найден (коннектор *car*), можно переходить к основной стадии предлагаемой методики — непосредственно поиску.

3 Двухязычный поиск в надкорпусной базе данных

Надкорпусная база данных коннекторов — кросслингвистическая, и при работе с ней имеется возможность осуществлять двухязычный поиск одновременно с использованием как текстов оригинала, так и текстов перевода (виды поиска, используемые в шести существующих НБД, описаны в [14]). Для обнаружения примеров использования конкретной МП применяется поиск по текстам на обоих языках: так, средствами НБД можно найти все контексты, где фрагмент на русском языке содержит коннектор *потому что*, а фрагмент на французском — *car* (по состоянию на 15.06.2019 таких контекстов 221).

Можно с уверенностью утверждать лишь то, что эти контексты *содержат* данные языковые единицы, но функциональное соответствие между заданными в поисковом запросе единицами в некоторых случаях может и отсутствовать. К примеру, если осуществить поиск пар фрагментов параллельных

текстов, которые одновременно содержат *поскольку* в русской части и *car* во французской, будет найдено 6 таких пар, однако только в 5 из них между этими коннекторами есть функциональное соответствие. В оставшейся паре используются оба коннектора, но *поскольку* переведен при помощи *pour autant que*, а СП для французского *car* является *потому что* (см. табл. 3, где эти единицы выделены полужирным).

Несмотря на наличие такого «шума», этот вид поиска позволяет проверить возможность использования той или иной МП и, если случаи ее использования были найдены, сформировать аннотации в рамках НБД.

Однако сама по себе функция двухязычного поиска еще не позволяет целенаправленно находить имплицитные и/или выраженные косвенно (ср. понятие «альтернативные лексикализации» в [15], а также группы случаев, выделяемые в разд. 2) ЛСО. Для этих целей в НБД реализован поиск, при котором ищутся такие пары контекстов, где во фрагменте на одном из языков употреблена какая-либо конкретная языковая единица, а во фрагменте на другом языке отсутствуют частотные СП или ФЭФ этой единицы.

Так, на основе приведенных в разд. 2 данных по 350 аннотациям, где для выражения ЛСО причины используются коннекторы *ибо*, *поскольку* и *потому что*, было установлено, что французский кон-

Таблица 3 Пример результата двухязычного поиска, в котором между единицами, включенными в поисковый запрос, отсутствует соответствие

[...] поскольку мне ни читать, ни тем более писать таких писем никогда не доводилось, я с большим интересом пробежал их глазами, пока они мне не приелись, потому что накал Арнольдовой страсти от письма к письму угасал, сменившись вскоре житейской прозой [...]. [А. Вайнер, Г. Вайнер. Эра милосердия (1975)]	[...] pour autant que je n'avais jamais eu à lire, encore moins à écrire, de telles lettres, je les parcourus avec beaucoup d'intérêt, jusqu'au moment où elles commencèrent à m'ennuyer, car l'ardeur de la passion d'Arnold diminuait de lettre en lettre pour faire place à une banale prose [...]. [Trad. par J.-P. Dussaussois et E. Avrorine, en coll. avec J.-G. Synakiewicz. Revu et corrigé par B. Durupt (2005)]
---	---

Таблица 4 Пара фрагментов, где для фрагмента на русском языке проставлен признак *Absent FragmCNT*

<p>Прокуратор начал с того, что пригласил первосвященника на балкон, с тем чтобы укрыться от безжалостного зноя, но Каифа вежливо извинился и объяснил, что сделать этого не может. [М. А. Булгаков. Мастер и Маргарита (1929–1940)]</p>	<p>Le procureur commença par inviter le grand prêtre à venir jusqu'à la terrasse couverte, afin de s'y abriter de la chaleur impitoyable, mais Caïphe s'excusa poliment, en expliquant qu'il ne le pouvait pas, <i>car on était à la veille des fêtes</i>. [Trad. par C. Ligny (1968)]</p>
--	--

нектор *car* чаще всего используется для выражения этого ЛСО. Исходя из предположения, что, будучи столь частотным средством выражения ЛСО причины во французском языке, *car* может быть ФЭФ не только для перечисленных коннекторов русского языка, в НБД был осуществлен поиск таких пар фрагментов, где во французском языке используется *car*, а в русском коннекторы *ибо*, *поскольку*, *потому что*, напротив, отсутствуют¹. По этому запросу было найдено 444 пары фрагментов параллельных текстов, из которых аннотации ПС были сформированы для 381. Несмотря на то что некоторые из найденных контекстов являются нерелевантными для целей исследования (в первую очередь речь идет о случаях, когда в русскоязычной части содержится какой-либо другой коннектор причины, не попавший в список исключаемых из выдачи единиц), данный вид поиска дает возможность решить поставленную задачу. После того как работа по формированию аннотаций для найденных контекстов завершена, можно переходить к их анализу.

4 Полученные результаты

В 47,2% сформированных аннотаций (180 случаев) для русскоязычной части был проставлен один из признаков кластера *Zero* (перечисленных ниже), тогда как для всех направлений перевода, представленных в НБД, процент таких аннотаций в несколько раз ниже (0,7% — более чем в 67 раз меньше — для рассматриваемого направления перевода «русский → французский», см. табл. 1). Столь представительный объем аннотаций, причем для одного и того же ЛСО, позволил уточнить систему признаков лингвистического аннотирования, используемую в НБД для категоризации контекстов, когда ЛСО является имплицитным или выражено косвенно. Вместо одного признака *Zero*, который использовался в предыдущей методике аннотирования во всех случаях, когда языко-

вая единица, передающая семантику коннектора, отсутствовала, в новой методике предлагается использовать систему из трех признаков, отражающих характеристики описываемых контекстов (в скобках приводятся количественные данные по описываемому материалу).

1. *Absent CNT* — коннектор, выражающий ЛСО, отсутствует, но может быть восстановлен без изменения синтаксической структуры фрагмента текста (40,7%, 155 случаев); см. пример в табл. 6 ниже.
2. *Absent FragmCNT* — фрагмент, соответствующий фрагменту с коннектором, отсутствует (2,3%, 9 случаев); см. пример в табл. 4, где при переводе добавлен выделенный полужирным курсивом фрагмент ‘так как был канун праздников’.
3. *DifferStr FragmCNT* — фрагмент, соответствующий фрагменту с коннектором, имеет иную, чем в оригинале, структуру, не допускающую использования коннектора (4,2%, 16 случаев); см. пример в табл. 5, где в переводе фрагмент переведен как ‘воры легко его [замок] выворотили, так как дерево было гнилым’.

Так как лишь первый признак (*Absent CNT*) отмечает случаи имплицитного ЛСО, то аннотации с признаками *Absent FragmCNT* и *DifferStr FragmCNT* при последующем анализе не рассматриваются, поскольку наличие этих признаков дает информацию не об аннотируемом контексте как таковом, а о переводческих решениях (что не входит в цели исследования).

Случаи косвенного выражения ЛСО причины представляется возможным разделить на три группы (подробнее см. [16]):

- (1) использование семантически нагруженных знаков препинания, а именно двоеточия и тире (18,4%, 70 случаев);

¹ Ввиду того, что в текстах, содержащихся в НБД, омонимия не снята, из результатов выдачи были исключены фрагменты, содержащие в русскоязычной части существительное *автобус* во всех формах, поскольку оно может переводиться существительным *car* ‘туристический автобус’, омонимичным коннектору *car*, по которому производился поиск в текстах на французском языке. Данная мера нацелена на сокращение числа нерелевантных контекстов. Также из выдачи исключались фрагменты, содержащие частицу *ведь*, которая не аннотировалась на первой стадии описанной методики, но которая способна выполнять функцию коннектора и выражать ЛСО причины, особенно в разговорной речи.

Таблица 5 Пара фрагментов, где для фрагмента на русском языке проставлен признак *DifferStr FragmCNT*

Воры легко выворотили замок из подгнившего дерева вместе с петлями. [А. Вайнер, Г. Вайнер. Эра милосердия (1975)]	Les voleurs l'avaient facilement arraché car le bois était pourri. [Trad. par J.-P. Dussaussois et E. Avrorine, en coll. avec J.-G. Synakiewicz. Revu et corrigé par B. Durupt (2005)]
--	--

Таблица 6 Пример использования эксплицитного показателя ЛСО обоими переводчиками для передачи ЛСО, имплицированного в оригинале

Не просите меня петь, я не спою уже больше так... [И. А. Гончаров. Обломов (1848–1859)]	Ne me demandez plus de chanter, car je ne serais plus capable de chanter ainsi. . . [Trad. par A. Adamov (1959)]
	Ne me demandez pas de chanter, car je ne saurais plus chanter comme ça. . . [Trad. par L. Jurgenson (1988)]

- (2) использование грамматических конструкций и лексических средств, не являющихся коннекторами, но выражающих семантику причины (16,5%, 63 случая);
- (3) использование коннекторов, для которых выражение ЛСО причины является нехарактерным (17,9%, 68 случаев)¹.

Отметим, что в рамках аналогичного, но моноязычного проекта PDTB (Pen Discourse Tree-Bank) многие из этих случаев рассматриваются как имплицитное ЛСО, а не его косвенное выражение. Авторы руководства по аннотированию в PDTB 3.0 [10] отмечают, что при восстановлении «имплицитного коннектора» разрешается модифицировать исходный контекст в процессе аннотирования для того, чтобы предложение оставалось грамматически правильным:

*The Nicaraguan president, citing (Implicit = as a result of) attacks by the U.S.-backed rebels, suspended a 19-month-old cease-fire. . .*²

Так, в приведенном примере [10, с. 35] курсивом и полужирным выделены связываемые фрагменты текста, причем форма *citing* («ссылаясь на») исключена из их состава. После этого изменения и добавления коннектора *as a result of* («в результате чего-либо») авторы относят данный случай к группе отношения причины. В рамках НБД единицей, выражающей ЛСО, здесь считалась бы как раз форма *citing* (грамматическое средство выражения ЛСО), что позволило бы не менять контекст в зависимости от целей исследования, а анализировать фрагменты реальных текстов.

¹ В рамках данной статьи к этой же группе относятся и те случаи, когда в оригинале используется коннектор, выражающий ЛСО, отличное от ЛСО причины, а также случаи, когда коннектор, способный выражать ЛСО причины (например, коннектор *a to*), выражает какое-либо другое отношение.

² Президент Никарагуа, ссылаясь на нападения, совершенные поддерживаемыми США повстанцами, приостановил режим прекращения огня, продлившийся 19 месяцев. . . (перевод наш — А. Г., О. И.).

Отдельного рассмотрения заслуживают случаи (выходящие за рамки статьи), когда французскому коннектору причины *car* в русском языке соответствуют коннекторы, выражающие какое-либо другое ЛСО. Анализ таких соответствий может позволить лингвистам, с одной стороны, провести более четкие границы между различными ЛСО, а с другой — определить, каковы возможные объективные причины разногласий среди исследователей относительно оснований выделения и количества логических и риторических отношений в рамках различных подходов.

При наличии в НБД для одного и того же текста более чем одного перевода имплицитное ЛСО причины может иметь эксплицитные показатели в переводах. В табл. 6 приводится контекст, где ЛСО причины в русском языке устанавливается имплицитно — «не просите, потому что я все равно не спою», — тогда как оба переводчика употребляют французский коннектор *car*.

Результаты анализа аннотаций говорят о том, что случаи, когда ЛСО причины является имплицитным или выражено косвенными средствами, не являются исключительными, а встречаются в текстах систематически. Несмотря на неявный характер, ЛСО верно интерпретируется реципиентом, что подтверждается данными переводов (в особенности примерами, подобными приведенному в табл. 6).

5 Заключение

Апробация новой методики поиска на примере лишь одного французского коннектора демонстри-

рует целенаправленный характер выявления нового лингвистического знания, которое позволит сформировать типологию случаев, когда ЛСО выражено косвенно или является имплицитным, с использованием методов контрастивного анализа в процессе аннотирования. Для этого представляется целесообразным распространить описанную методику на другие языковые единицы (например, на *parce que*, второй по частотности коннектор причины в НБД) и, шире, на другие ЛСО. Стимулы перевода, относящиеся к группе косвенных средств выражения ЛСО, также могут стать объектом поиска на следующих этапах исследования. Более детальный семантический анализ результатов позволит лингвистам определить и возможные причины выражения ЛСО в тексте без использования их прототипических показателей. Таким образом, предлагаемая методика, ориентированная на использование НБД, предназначена для поиска и исследования случаев, когда ЛСО в тексте выражено косвенными средствами или является имплицитным, а также для последующего формирования их типологии, и была разработана в рамках проекта РФФИ, цель которого состоит в создании методов и средств информатики для целенаправленного формирования новых и развития существующих лингвистических типологий с использованием параллельных текстов [17–20].

Литература

1. Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А. Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
2. Дурново А. А., Зацман И. М., Ложилова Е. Ю. Кросс-лингвистическая база данных для аннотирования логико-семантических отношений в тексте // Системы и средства информатики, 2016. Т. 26. № 4. С. 124–137.
3. Зализняк Анна А., Зацман И. М., Инькова О. Ю. Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–108.
4. Зацман И. М., Кружков М. Г. Надкорпусная база данных коннекторов: развитие системы терминов проектирования // Системы и средства информатики, 2018. Т. 28. № 4. С. 156–167.
5. Инькова-Манзотти О. Ю. Коннекторы противопоставления во французском и русском языках (сопоставительное исследование). — М.: Информэлектро, 2001. 434 с.
6. Rosch E., Mervis C. B., Gray W. D., Johnson D. M., Boyes-Braem P. Basic objects in natural categories // Cognitive Psychol., 1976. Vol. 7. P. 382–439.
7. Carlson L., Marcu D. Discourse tagging reference manual. 2001. 87 p. <ftp://128.9.176.20/isi-pubs/tr-545.pdf>.
8. PDTB Research Group. The Penn Discourse Treebank 2.0 Annotation Manual. — Philadelphia, PA, USA: Institute for Research in Cognitive Science, University of Pennsylvania, 2008. Technical Report IRCS-08-01. <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
9. Ho-Dac L.-M., Péry-Woodley M.-P. Annotation des structures discursives: l'expérience ANNODIS // 4e Congrès Mondial de Linguistique Française. SHS Web of Conferences / Eds. F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, S. Prévost. — Berlin, 2014. Vol. 8. P. 2647–2661.
10. Webber B., Prasad R., Lee A., Joshi A. The Penn Discourse Treebank 3.0 Annotation Manual, 2019. <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
11. Семантика коннекторов: контрастивное исследование / Под науч. ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 368 с.
12. Русская грамматика: в 2 т. Том II. Синтаксис / Под ред. Н. Ю. Шведовой. — М.: Наука, 1980. 709 с.
13. Инькова О. Ю. Аннотирование параллельных текстов: понятие «дивергентный перевод» // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2019. Вып. 18(25). С. 227–238.
14. Гончаров А. А., Инькова О. Ю., Кружков М. Г. Методология аннотирования в надкорпусных базах данных // Системы и средства информатики, 2019. Т. 29. № 2. С. 148–160.
15. Prasad R., Joshi A., Webber B. Realization of discourse relation by other means: Alternative lexicalizations // 23rd Conference (International) on Computational Linguistics Proceedings. — Beijing, China, 2010. P. 1023–1031.
16. Гончаров А. А., Инькова О. Ю. Способы выражения причинных отношений в русском языке: опыт анализа с использованием кросслингвистической надкорпусной базы данных // Русская грамматика: активные процессы в языке и речи: Междунар. научный симпозиум: Сб. статей, 2019 (в печати).
17. Зацман И. М. Имплицированные знания: основания и технологии извлечения // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 74–82.
18. Зацман И. М. Стадии целенаправленного извлечения знаний, имплицированных в параллельных текстах // Системы и средства информатики, 2018. Т. 28. № 3. С. 175–188.
19. Зацман И. М. Целенаправленное развитие систем лингвистических знаний: выявление и заполнение лакун // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 91–98.
20. Гончаров А. А., Зацман И. М. Информационные трансформации параллельных текстов в задачах извлечения знаний // Системы и средства информатики, 2019. Т. 29. № 1. С. 180–193.

Поступила в редакцию 30.06.19

METHODS FOR IDENTIFICATION OF IMPLICIT LOGICAL-SEMANTIC RELATIONS IN TEXTS

A. A. Goncharov and O. Yu. Inkova

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 methods for identification of implicit logical-semantic relations (LSR) in parallel texts of the Supracorpora Database (SCDB) of Connectives. The stages of the search process are described based on the Russian–French translations: (i) selection of an LSR to be analyzed and creation of an array of annotations of Russian connectives considered as prototypical means for expressing this LSR; (ii) analysis of the produced array of annotations and identification of common equivalents for translating Russian connectives into French; (iii) utilizing the bilingual search functions of the SCDB with exclusion of Russian connectives annotated during the first stage and with specification of the most frequent French language units identified during the second stage; (iv) annotation of the pairs of fragments of parallel texts found as the result of the third stage; and (v) analysis of the array annotations produced at the fourth stage in order to identify and categorize instances of implicit LSR. The proposed SCDB-based search methods make it possible to gather new data on implicit LSRs.

Keywords: identifying implicit information; connectives; contrastive linguistics; corpus linguistics; supracorpora databases; logical-semantic relations

DOI: 10.14357/19922264190314

Acknowledgments

The study has been conducted at the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences with financial aid from the Russian Foundation for Basic Research (Grant No. 18-07-00192).

References

- Zatsman, I. M., O. Yu. Inkova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie kross-yazykovykh znaniy o konnektorakh v nadkorporusnykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
- Durnovo, A. A., I. M. Zatsman, and E. Yu. Loshchilova. 2016. Krosslingvisticheskaya baza dannykh dlya annotirovaniya logiko-semanticheskikh otnosheniy v tekste [Cross-lingual database for annotating logical-semantic relations in the text]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(4):124–137.
- Zaliznyak, Anna A., I. M. Zatsman, and O. Yu. Inkova. 2017. Nadkorporusnaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database on connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–106.
- Zatsman, I. M., and M. G. Kruzhkov. 2018. Nadkorporusnaya baza dannykh konnektorov: razvitie sistemy terminov proektirovaniya [Supracorpora database of connectives: Design-oriented evolution of the term system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):156–167.
- Inkova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh. So-*
- postavitel'noe issledovanie* [Connectives of opposition in French and Russian. Comparative research]. Moscow: Informelektro. 434 p.
- Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychol.* 7:382–439.
- Carlson, L., and D. Marcu. 2001. Discourse tagging reference manual. 87 p. Available at: <ftp://128.9.176.20/isi-pubs/tr-545.pdf> (accessed July 29, 2019).
- PDTB Research Group. 2008. The Penn Discourse Treebank 2.0 Annotation Manual. Philadelphia, PA: Institute for Research in Cognitive Science, University of Pennsylvania. Technical Report IRCS-08-01. Available at: <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
- Ho-Dac, L.-M., and M.-P. Péry-Woodley. 2014. Annotation des structures discursives: l'expérience ANNODIS. *4e Congrès Mondial de Linguistique Française. SHS Web of Conferences*. Eds. F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefter, and S. Prévost. Berlin. 8:2647–2661. Available at: https://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01286.pdf.
- Webber, B., R. Prasad, A. Lee, and A. Joshi. 2019. The Penn Discourse Treebank 3.0 Annotation Manual. Available at: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.

11. Inkova, O. Yu., ed. 2018. *Semantika konnektorov: kontrastivnoe issledovanie* [Semantics of connectives: A contrastive study]. Moscow: TORUS PRESS. 368 p.
12. Shvedova, N. Yu., ed. 1980. *Russkaya grammatika: v 2 t. T. II. Sintaksis* [Russian Grammar. In 2 vols. Vol. II. Syntax]. Moscow: Nauka. 709 p.
13. Inkova, O. Yu. 2019. Annotirovanie parallel'nykh tekstov: ponyatie "divergentnyy perevod" [Annotation of parallel texts: The concept of divergent translation]. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"* 18(25):227–238.
14. Goncharov, A. A., O. Yu. Inkova, and M. G. Kruzhkov. 2019. Metodologiya annotirovaniya v nadkorpusnykh bazakh dannykh [Annotation methodology of supracorpus databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):148–160.
15. Prasad R., A. Joshi, and B. Webber. 2010. Realization of discourse relation by other means: Alternative lexicalizations. *23rd Conference (International) on Computational Linguistics Proceedings*. Beijing, China. 1023–1031.
16. Goncharov, A. A., and O. Yu. Inkova. 2019 (in print.) Sposoby vyrazheniya prichinnykh otnosheniy v russkom yazyke: opyt analiza s ispol'zovaniem krosslingvisticheskoy nadkorpusnoy bazy dannykh [Means of expressing causal relations in Russian: Analysis using a cross-linguistic supracorpus database]. *Russkaya grammatika: aktivnye protsessy v yazyke i rechi. Mezhdun. nauchnyy simpozium. Sbornik statey* [Russian Grammar: Active Processes in Language and Discourse: Scientific Symposium (International) Proceedings].
17. Zatsman, I. M. 2018. Implitsirovannyye znaniya: osnovaniya i tekhnologii izvlecheniya [Implied knowledge: Foundations and technologies of explication]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):74–82.
18. Zatsman, I. M. 2018. Stadii tselenapravlennoy izvlecheniya znaniy, implitsirovannykh v parallel'nykh tekstakh [Stages of goal-oriented discovery of knowledge implied in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):175–188.
19. Zatsman, I. M. 2019. Tselenapravlennoe razvitie sistem lingvisticheskikh znaniy: vyyavlenie i zapolnenie lakun [Goal-oriented development of linguistic knowledge systems: Identifying and filling lacunae]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):91–98.
20. Goncharov, A. A., and I. M. Zatsman. 2019. Informatsionnye transformatsii parallel'nykh tekstov v zadachakh izvlecheniya znaniy [Information transformations of parallel texts in knowledge extraction]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):180–193.

Received June 30, 2019

Contributors

Goncharov Alexander A. (b. 1994) — junior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; a.gonch48@gmail.com

Inkova Olga Yu. (b. 1965) — Doctor of Science (PhD) in philology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; olyainkova@yandex.ru

ПЕРСОНАЛЬНЫЙ КОГНИТИВНЫЙ АССИСТЕНТ: КОНЦЕПЦИЯ И ПРИНЦИПЫ РАБОТЫ*

И. В. Смирнов¹, А. И. Панов², А. А. Скрынник³, Е. В. Чистова⁴

Аннотация: Предложена концепция когнитивного персонального ассистента. Когнитивный ассистент выступает виртуальным интеллектуальным агентом, обладающим своей собственной картиной мира (КМ) и строящим КМ пользователя, которому он помогает решать различные задачи. Описана архитектура когнитивного ассистента, рассмотрены основные функции, которые он должен реализовывать, и представлены основные методы и технологии, которые используются при построении такого рода ассистентов. Рассмотрены две предметные области, в которых использование когнитивных ассистентов наиболее перспективно.

Ключевые слова: когнитивный ассистент; образовательный ассистент; медицинский ассистент; знаковая картина мира; обработка естественного языка; диалоговая система; сценарий; планирование

DOI: 10.14357/19922264190315

1 Введение

Разработки в области интеллектуальных ассистентов — помощников человека при совершении им каждодневных задач (выбор товаров и услуг, поиск в интернете, прокладка маршрута и навигация во время вождения, голосовое управление бытовыми устройствами и т. п.) — получили в последнее время новый импульс своего развития в связи с появлением новых методов анализа естественного языка и глубокого обучения. Методы распознавания и синтеза речи, лингвистического анализа и синтеза текстов достигли сегодня такого уровня, который обеспечил создание промышленных голосовых помощников, таких как Siri, Cortana, Alexa, Алиса. Однако до сих пор не решена задача создания полноценного ассистента, который бы не только действовал реактивно в ответ на запросы пользователя, но и вел бы себя проактивно, предсказывая поведение пользователя, а также обладал целеполаганием и внутренней мотивацией по достижению поставленных перед ним целей. Во многом это связано с тем, что основной упор при создании современных ассистентов делается преимущественно на поддержание вопросно-ответной коммуникации, т. е. создание диалоговых систем.

Для решения же поставленной цели генерации проактивных целенаправленных действий ас-

систент должен обладать элементами искусственного сознания, собственной моделью поведения и строить аналогичную по сложности модель поведения пользователя на основе понимания разговорного языка. Только в таком случае интеллектуальный агент сможет стать полноценным помощником человека, а не только голосовым интерфейсом для разных приложений, например поисковых.

В настоящей работе предлагается концепция персонального когнитивного ассистента, т. е. такого виртуального ассистента, который обладал бы проактивным целенаправленным поведением и моделировал поведение собеседника. Описана архитектура, принципы работы, основные функции когнитивного ассистента, предложены методы реализации этих функций. Возможные варианты использования когнитивного ассистента представлены для двух предметных областей: ассистирование в процессе обучения пользователя (образовательный ассистент) и ассистирование в процессе поддержания здоровья пользователя (виртуальный тренер здоровья).

2 Архитектура когнитивного ассистента

На рис. 1 представлена принципиальная схема устройства и работы когнитивного ассистента.

* Работа выполнена при частичной финансовой поддержке РФФИ (проект № 18-29-22027).

¹ Институт проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, ivs@isa.ru

² Институт проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский физико-технический институт (государственный университет), pan@isa.ru

³ Институт проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук, skrynnik@isa.ru

⁴ Институт проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, chistova@isa.ru

]

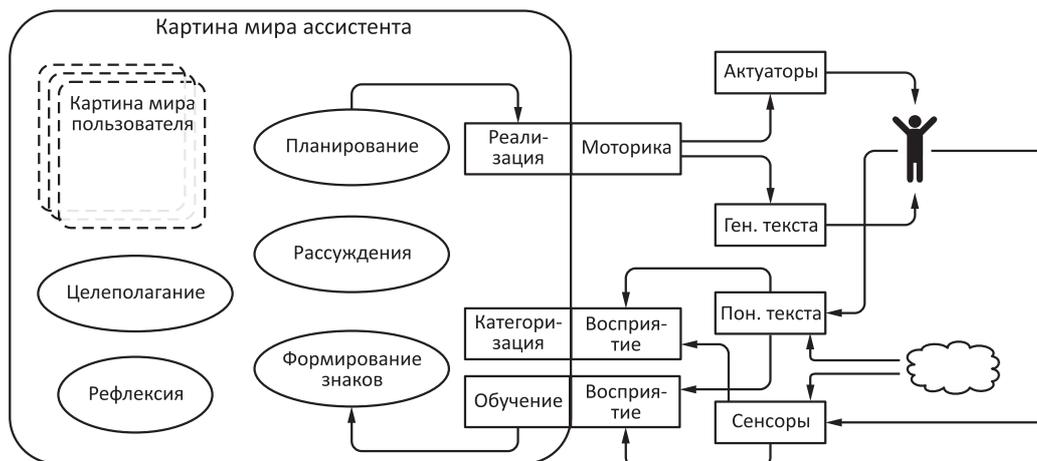


Рис. 1 Архитектура когнитивного ассистента

Основным компонентом агента является его КМ, которая моделируется на основе знакового подхода [1, 2]. Основным компонентом знаковой КМ служит *знак*, представляющий собой на синтаксическом уровне описание модели (по [1]) четверку

$$s = \langle n, p, m, a \rangle,$$

где $n \in N$; $p \in P$; $m \in M$; $a \in A$. Здесь N — множество имен, представляющее собой множество слов конечной длины в некотором алфавите; P — множество замкнутых атомарных формул языка исчисления предикатов первого порядка, которое называется *множеством образов*; M — множество значений; A — множество личностных смыслов.

В случае так называемой житейской КМ, компонента образа знака участвует в процессе распознавания и категоризации. Значения представляют фиксированные, сценарные знания интеллектуального агента о предметной области и окружающей среде, а множество личностных смыслов характеризует его предпочтения и текущий деятельностный контекст. Компонента имени осуществляет связывание остальных компонент знака в единое целое (именование).

На структурном уровне описания знаковой КМ каждая компонента знака представляет собой множество каузальных матриц, которые представляют собой структурированное множество ссылок на другие знаки, либо элементарные компоненты (в случае образа — это первичный признак или данные с сенсоров, в случае личностного смысла — это операционный состав действия). Каузальная матрица позволяет кодировать информацию для представления как декларативных, так и процедурных знаний. Множество компонент знака образуют четыре типа каузальных сетей — специального

типа семантических сетей. Моделирование функций планирования и рассуждения осуществляется за счет введения понятия активности (множества активных знаков или каузальных матриц) и правил распространения активности по различным типам сетей [3]. В процессе работы той или иной когнитивной функции формируются новые каузальные матрицы, которые могут затем сохраняться в составе компонент нового знака аналогично сохранению опыта в системах, основанных на прецедентах.

Отдельное множество знаков в КМ ассистента отвечает за моделирование КМ пользователя. Это представление ассистента о свойствах, возможностях и целях другого субъекта хранится в тех же структурах, что и основная информация о предметной области. В КМ ассистента присутствуют специальные знаки «Я» и «Другой» («Пользователь»), которые позволяют ассистенту различать информацию, относящуюся непосредственно к нему или другому субъекту.

На основе знаковой КМ ассистент способен строить свои собственные планы действий, в которые он может включать действия (реакцию) других субъектов, в том числе пользователя (см. п. 2.2.1). Картина мира ассистента описывает его назначение, цели, возможные действия и сценарии, личностные смыслы, оценки достижения целей. Модель КМ пользователя строится за счет выявления его сценариев и личностных смыслов, ценностей, предпочтений, привычек и т.п. Когнитивный ассистент общается с собеседником с учетом этих двух КМ.

Основными структурами на множестве знаков, которые позволяют генерировать проактивные действия, являются сценарии — переиспользуемые аб-

страктные последовательности действий и ситуаций [1]. Когнитивный ассистент имеет сценарии собственных действий и сценарии действий пользователя. Таким образом, ассистент имеет представление о последовательности действий, необходимых для решения той или иной задачи, что позволяет ему ассистировать пользователю при решении задач, подсказывая дальнейшие действия с учетом персональных сценариев, характерных для конкретного пользователя.

Когнитивный ассистент требует предварительного обучения и настройки за счет внешних источников, по которым он формирует базовые сценарии деятельности, характерные для данной предметной области. Автоматическое или автоматизированное пополнение КМ сценариями происходит за счет двух основных источников текстовой и нетекстовой (видео, изображения, данные с различных сенсоров) информации: внешние источники (интернет, коллекции документов из данной предметной области и т.п.) и пользователь, который передает ассистенту текст в виде запросов и описания задач либо через программный интерфейс, который по желанию пользователя может регистрировать его различные характеристики (например, биометрические данные, психологические черты личности [4], действия с интерфейсом или действия с другими устройствами).

Автоматическое извлечение и формирование сценариев по тексту происходит с помощью подходов, описанных в п. 2.2.2. Допустимо построение КМ ассистента и вручную экспертом предметной области.

Одна из важнейших особенностей ассистента — его способность генерировать текст на естественном языке и вести связный диалог (см. п. 2.2.3).

2.1 Режимы работы

Когнитивный ассистент работает в нескольких режимах, автоматически выбирая режим в зависимости от текущего диалога или высказывания:

- диалог на свободную тему (chit-chat);
- вопросно-ответный режим;
- целеориентированный режим.

В диалоговом режиме ассистент поддерживает беседу на свободную тему, отвечает на приветствия, спрашивает о настроении пользователя. В этом режиме ассистент также выполняет простейшие просьбы, например рассказать анекдот или выдать прогноз погоды.

В вопросно-ответном режиме ассистент находит точный ответ на поставленный пользователем

вопрос. Вопрос может задаваться на естественном языке в свободной форме. При поиске ответа предполагается учитывать интересы и текущее настроение пользователя.

Целеориентированный режим предполагает помощь ассистента в решении конкретных задач пользователя. Этот режим в значительной степени использует КМ ассистента и пользователя и зависит от назначения ассистента.

2.2 Основные методы

2.2.1 Планирование поведения

На этапе синтеза плана деятельности когнитивный ассистент рекурсивно создает все возможные планы по достижению конечной ситуации, которая описывает целевое состояние ассистента и пользователя. Для этого ассистентом рассматриваются все знаки, которые входят в описание текущей ситуации $z_{\text{sit-cur}}$, и с помощью процесса распространения активности по сети значений [3, 5] активируются процедурные матрицы действий. Затем актуализируются матрицы действий, заменяются ссылки на знаки ролей и типов объектов на ссылки конкретных объектов задачи. Далее следует шаг выбора действий, которые эвристически были оценены как наиболее подходящие в ситуации $z_{\text{sit-cur}}$ для достижения ситуации $z_{\text{sit-goal}}$. После этого из эффектов каждого действия и ссылок на знаки, которые входят в текущую ситуацию, строится $z_{\text{sit-cur}+1}$, которая описывает состояние агента после применения действия. В план добавляется рассматриваемое действие и $z_{\text{sit-cur}}$, затем проверяется вхождение $z_{\text{sit-goal}}$ в $z_{\text{sit-cur}+1}$. Если матрицы текущего состояния включают матрицы целевого состояния, то алгоритм сохраняет найденный план как один из возможных; если матрицы целевого состояния не входят, то функция поиска плана рекурсивно повторяется.

2.2.2 Автоматическое формирование сценариев

Основной задачей для когнитивного ассистента является задача автоматизированного формирования КМ, прежде всего сценариев решения задач и сети значений знаков. Предполагается, что основным источником для формирования КМ служат наборы текстов. Для конструирования сценариев по текстам предлагается использовать подходы открытого извлечения информации из текстов [6, 7].

Сеть на значениях в КМ состоит из концептов и связей между ними. Концепт представляет собой предмет или явление в рамках предметной области. Разные лексические единицы могут ссылаться на

один и тот же концепт. Например, «центральное обрабатывающее устройство» и «процессор».

Между концептами существуют два типа связей:

- (1) таксономические связи образуют иерархии концептов. Например, «попугай — это птица». «Птица» — гипероним по отношению к «попугай»;
- (2) нетаксономические связи являются предикатами, описывающими взаимодействие концептов. Например, «Эксперт размечает корпус».

Основные шаги пополнения сети значений.

1. Выполняется графематический, морфологический и синтаксический анализ текстов. На этапе графематического анализа происходит выделение предложений из текста и выделение слов из предложений (токенизация). Морфологический анализ позволяет получить леммы (нормальные формы) и морфологические признаки для каждого слова из текста. В результате синтаксического анализа текстов генерируются синтаксические деревья зависимостей для каждого предложения.
2. Извлекаются всевозможные именные группы из синтаксических деревьев зависимостей с помощью следующего алгоритма:
 - выполняется поиск вершины синтаксического дерева (как правило, это глагол);
 - происходит спуск по дереву до первого существительного;
 - найденное существительное сохраняется вместе с потомками в качестве именной группы, вплоть до первого слова, часть речи которого не входит в следующий список: существительное, прилагательное, местоимение, числительное, имя собственное, союз, наречие, причастие.
3. Извлекаются термины предметной области путем кластеризации выделенных именных групп с учетом метрики C-value:

$$C\text{-Value}(a) = \begin{cases} \log_2 |a| \cdot \text{freq}(a), & \text{если именная группа} \\ & \text{не вложена в другие;} \\ \log_2 |a| \cdot \text{freq}(a) - \frac{1}{p(T_a)} \sum_{b \in T_a} \text{freq}(b) & \text{иначе,} \end{cases}$$

где a — именная группа; $|a|$ — число слов в именной группе; $\text{freq}(a)$ — частота встречаемости a ;

T_a — именные группы, в которые входит a ;

$p(T_a)$ — число именных групп, содержащих a .

Данная метрика учитывает пересечения лексики между словосочетаниями и позволяет выделять многословные термины.

4. Для извлечения концептов выполняется кластеризация полученных терминов. В качестве признаков используются векторные представления слов [8]. Для словосочетаний используются усредненные векторы.
5. С помощью набора эвристик из полученных концептов извлекаются таксономические связи. Эвристики в первую очередь опираются на предлоги, союзы и морфологические признаки.
6. Кандидаты в таксономические связи извлекаются путем поиска глагольных групп в синтаксическом дереве зависимостей. Поиск выполняется между лексическими единицами, связанными с концептами.
7. На конечном этапе таксономические связи извлекаются путем кластеризации нескольких признаков кандидатов: векторное представление глагольной группы, идентификаторы концептов, между которыми найдена глагольная группа.
8. На основе ряда эвристик по тексту формируются последовательности найденных троек (концепт — глагольная группа — концепт).

В результате выявленные в тексте последовательности троек и представляют собой сценарии решения задач с необходимыми участниками и орудиями действий.

2.2.3 Вопросно-ответный режим

Вопросно-ответный режим предполагается реализовать с помощью технологий Exactus [9], основанных на реляционно-ситуационном и семантико-синтаксическом анализе текста [10, 11]. Реляционно-ситуационный анализ текста представляет семантику текста в виде семантической сети, а семантика предложения или высказывания при этом представляется в виде совокупности предикатных слов, их аргументов и семантических ролей. Семантические сети строятся для вопроса и каждого текста, в котором может находиться формулировка ответа, затем происходит сопоставление семантических сетей вопроса и текстов, вычисляется релевантность текстов вопросу. Таким образом, для работы необходим набор текстов, в которых потенциально могут содержаться ответы на вопросы, т. е. ответы ищутся, а не генерируются. Вопросно-ответный поиск реализуется в два этапа: на первом выполняется семантический поиск предложений,

содержащих формулировку ответа на поставленный вопрос; на втором этапе из предложения выделяется фрагмент, являющийся точным ответом на поставленный вопрос.

В работе [8] было показано, что учет семантической структуры предложения в вопросно-ответном поиске значительно повышает качество поиска ответов по сравнению с лексическим критерием ранжирования, а также позволяет извлекать сам ответ на вопрос. В 2010 г. технология Eхastus была представлена на российском семинаре по оценке методов информационного поиска РОМИП в дорожке вопросно-ответного поиска и показала высокие результаты по всем метрикам [12].

2.2.4 Диалоговый режим

Для реализации режима диалога на свободную тему на русском языке предлагается использовать порождающие подходы на основе различных нейросетевых моделей. В работе [13] показано, что добавление к модели seq2seq механизма внимания повышает качество и грамматическую согласованность генерируемых реплик в диалоге. В будущем возможна комбинация данного подхода к генерации ответов и использование баз знаний для того, чтобы модель оперировала более конкретными представлениями в каждой области.

3 Применение в образовательном процессе

Один из вариантов использования ассистента, строящего модель КМ собеседника, а также обладающего одним из вариантов КМ, — его применение в процессах онлайн-обучения.

Образовательные онлайн-системы (к примеру, Coursera, Stepik, Logiclik, Examer и др.) сейчас используют методы искусственного интеллекта для улучшения качества обучения пользователей при решении следующих задач:

1. Выбор образовательной траектории, подстройка блоков курса, иными словами, адаптация программы обучения по выбранной пользователем теме в зависимости от когнитивных особенностей этого пользователя.
2. Организация frequently asked questions (FAQ) (раздела вопросов-ответов) по данному курсу, в котором ассистент заменяет преподавателя при ответе на стандартные вопросы по курсу.
3. Подсказки учителю по время подготовки или проведения урока, которые формируются в зависимости от реакции аудитории на прошлые

занятия, либо в зависимости от когнитивных особенностей преподавателя.

4. Прокторинг — отслеживание поведения пользователя во время просмотра курса или выполнения задания с учетом его поведенческих характеристик, подача ему предупреждений или советов.
5. Автоматическая проверка выполненного задания, выдача рекомендаций по исправлению ошибок или по выполнению дополнительного задания на одну из тем, по которой пользователь допустил ошибку, с учетом его эмоциональных и когнитивных особенностей.
6. Формирование советов по дальнейшему усвоению курса как для пользователя, так и для преподавателя.
7. Мотивация пользователя путем генерации соответствующих его состоянию и психологическим особенностям реплик или вывод его на диалог.

Все упомянутые выше поведенческие, психологические и когнитивные особенности должны определяться когнитивным ассистентом в рамках КМ пользователя, модель которой он строит. Гипотеза авторов состоит в том, что весь комплекс упомянутых функций и корректный учет всех особенностей пользователя системой онлайн-образования невозможно обеспечить без использования специальных методов моделирования его знаковой КМ. При этом для реализации некоторых функций (например, 4 или 7) необходимо, чтобы агент сам обладал КМ (возможно, статичной, возможно, эволюционирующей), которая позволяла бы ему вести более полноценные диалоги и демонстрировать собственные психологические особенности для достижения большего эффекта от общения с пользователем.

На рис. 2 приведен примерный вид интерфейса программной реализации такого ассистента, который выполняет роль помощника при обучении игры в шахматы.

Ассистент предварительно обучен на текстах шахматных книг (в том числе детских и художественных, в которых встречается большее разнообразие простых языковых конструкций) — по ним он формирует базовую КМ потенциального пользователя и свою собственную КМ учителя, используя методы п. 2.2.1. Также агент обладает функционалом классических шахматных программ (система



Рис. 2 Пример интерфейса образовательного когнитивного ассистента по игре в шахматы

просчета вариантов и т.п.). Виды формируемых сценариев: шахматные розыгрыши в стандартных позициях, правила оценки той или иной позиции, общие сценарии поведения учителя (того, кто советует, направляет, подсказывает) и ученика (того, кто что-то решает, приобретает знания). Ассистент учится во время взаимодействия с учеником, обновляя как свою КМ, так и информацию о КМ пользователя (какие ошибки совершает, что интересно).

Информация, которую получает ассистент от пользователя: выбор какой-либо темы обучения (дебют, окончания, тактика, стратегия, разбор партий и т.п.), ходы при решении предложенных задач и время от времени вопросы и ответы в текстовом поле. Также возможно отслеживание по видеоканалу поведения пользователя в процессе решения задач.

Основные задачи ассистента: по сформированным (и пополняемым в процессе работы) в КМ сценариям для пользователя ассистент (с учетом особенностей своей собственной КМ) выдает мотивирующие реплики пользователю, задает ему вопросы, отвечает на его вопросы, предлагает подсказки в различных ситуациях, дает ему советы после партии, предлагает ему новые задачи в рамках выбранного курса для ликвидации определенных недостатков в знаниях пользователя, предлагает ему пройти необходимый ему новый курс (например, король и пешка против короля).

4 Применение в здоровьесбережении

Другим вариантом использования когнитивного ассистента является его применение в процессах

здоровьесбережения. Сегодня наблюдается активный рост объема научных исследований и разработок в области здоровьесбережения — профилактики и поддержания здоровья людей. Основная цель этого направления — анализ особенностей здоровья конкретного человека и подбор и проведение персонализированных профилактических мероприятий до появления первых симптомов возможных заболеваний. Как можно более ранний прогноз возникновения заболеваний служит снижению риска патологии. В настоящее время разработаны интеллектуальные системы, поддерживающие процесс здоровьесбережения на всех стадиях [14]. Основные задачи технологий здоровьесбережения включают:

- сбор данных о состоянии здоровья и образе жизни человека из различных источников;
- интеллектуальный анализ данных о состоянии здоровья и образе жизни человека для выявления проблем со здоровьем, оценку персонализированных рисков ухудшения здоровья человека;
- подбор персональных рекомендаций по изменению образа жизни конкретного человека в зависимости от состояния его здоровья, образа жизни, проблемных зон и индивидуальных особенностей;
- формирование персонального плана профилактических мероприятий;
- мотивацию человека к выполнению рекомендаций, отслеживанию изменений в образе жизни и рисках заболеваний.

Персональный когнитивный ассистент в виде виртуального персонального тренера здоровья может использоваться при решении всех указанных задач. Предполагается, что виртуальный тренер здоровья постоянно мониторит психическое и физическое состояние пользователя, сообщая пользователю о критических изменениях. Он может быть интегрирован с различными гаджетами (фитнестрекерами) и мобильными приложениями, может анализировать активность в социальных сетях, учитывать пищевые предпочтения пользователя с помощью фуд-трекера.

Одним из основных способов получения информации об образе жизни и здоровье человека будет проактивный диалог ассистента с пользователем. Например, предполагается, что ассистент сам инициирует диалог — спрашивает, как настроение сегодня, просит ответить на вопросы психологического теста, напоминает о необходимости запланированных действий, например приема пищи, отдыха, физической активности. Когнитивный ассистент может ответить на любой вопрос по теме

здоровьесбережения или отправить пользователя к информации, содержащей ответ на вопрос, обладает эмоциями и выражает их с помощью смайлов или эмодзи, учитывает КМ и текущее настроение пользователя при выражении эмоций. При формировании плана профилактических мероприятий ассистент также должен учитывать КМ пользователя, его предпочтения.

5 Заключение

Сегодня существуют много виртуальных ассистентов и чат-ботов, нуждающихся в интеллектуальной начинке, которая была бы основана на моделировании когнитивных функций человека и повышала качество работы ассистентов. Когнитивный ассистент отличается от аналогов наличием формализованных знаний о части окружающего мира и способах решения определенных задач, а также способностью учитывать КМ пользователя в процессе ассистирования ему.

В результате дальнейших работ планируется создание семейства интеллектуальных ассистентов и соответствующих технологий, легко настраиваемых на решение новых задач. На основе предложенной концепции персонального когнитивного ассистента предполагается разработка масштабируемой программной платформы для создания ассистентов различного назначения и их настройки под конкретные задачи и предметные области. Такие ассистенты будут способны к общению с человеком (или другим виртуальным ассистентом) на естественном языке и смогут встраиваться в различные программы (мессенджеры, социальные сети) или искусственные технические устройства (например, робототехнические ассистенты). Когнитивные ассистенты наиболее востребованы в автономных интеллектуальных устройствах или их коалициях, действующих в опасных средах с отложенной коммуникацией с человеком.

Литература

1. *Осипов Г. С., Панов А. И.* Отношения и операции в знаковой картине мира субъекта поведения // Искусственный интеллект и принятие решений, 2017. № 4. С. 5–22.
2. *Осипов Г. С., Чудова Н. В., Панов А. И., Кузнецова Ю. М.* Знаковая картина мира субъекта поведения. — М.: Физматлит, 2018. 264 с.
3. *Панов А. И.* Целеполагание и синтез плана поведения когнитивным агентом // Искусственный интеллект и принятие решений, 2018. № 2. С. 21–35.
4. *Stankevich M., Smirnov I., Ignatiev N., Grigoriev O., Kiselnikova N.* Analysis of big five personality traits by processing of social media users activity features // CEUR Workshop Procee., 2018. Vol. 2277. P. 162–166.
5. *Киселев Г. А., Панов А. И.* Знаковый подход к задаче распределения ролей в коалиции когнитивных агентов // Труды СПИИРАН, 2018. № 2. С. 161–187.
6. *Шелманов А. О., Исаков В. А., Станкевич М. А., Смирнов И. В.* Открытое извлечение информации из текстов. Часть I. Постановка задачи и обзор методов // Искусственный интеллект и принятие решений, 2018. № 2. С. 47–67.
7. *Шелманов А. О., Девяткин Д. А., Исаков В. А., Смирнов И. В.* Открытое извлечение информации из текстов. Часть II. Извлечение семантических отношений с помощью машинного обучения без учителя // Искусственный интеллект и принятие решений, 2019. № 2. С. 39–49.
8. *Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J.* Distributed representations of words and phrases and their compositionality // Adv. Neur. Inf., 2013. Vol. 26. P. 3111–3119.
9. *Шелманов А. О., Каменская М. И., Ананьева И. В., Смирнов И. В.* Семантико-синтаксический анализ текстов в задачах вопросно-ответного поиска и извлечения определений // Искусственный интеллект и принятие решений, 2016. № 4. С. 47–61.
10. *Осипов Г. С., Смирнов И. В., Тихомиров И. А.* Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений, 2008. № 2. С. 3–10.
11. *Смирнов И. В., Шелманов А. О., Кузнецова Е. С., Храмоин И. В.* Семантико-синтаксический анализ естественных языков. Часть II. Метод семантико-синтаксического анализа текстов // Искусственный интеллект и принятие решений, 2014. № 1. С. 11–24.
12. *Завьялова О. С., Киселёв А. А., Осипов Г. С., Смирнов И. В., Тихомиров И. А., Соченков И. В.* Система интеллектуального поиска и анализа информации «ЕХАСТУС» на РОМИП-2010 // Тр. Российского семинара по оценке методов информационного поиска. — Казань: Казанский ун-т, 2010. С. 49–69.
13. *Чистова Е. В., Шелманов А. О., Смирнов И. В.* Применение глубокого обучения к моделированию диалога на естественном языке // Тр. Института системного анализа РАН, 2019. Т. 69. № 1. С. 105–115.
14. *Grigoriev O. G., Kobrinskii B. A., Osipov G. S., Molodchenkov A. I., Smirnov I. V.* Health management system knowledge base for formation and support of a preventive measures plan // Procedia Comput. Sci., 2018. Vol. 145. P. 238–241.

Поступила в редакцию 28.01.19

PERSONAL COGNITIVE ASSISTANT: CONCEPT AND KEY PRINCIPALS

I. V. Smirnov^{1,2}, A. I. Panov^{1,3}, A. A. Skrynnik¹, and E. V. Chistova^{1,2}

¹Institute of Artificial Intelligence Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 9 60-letiya Oktyabrya Prosp., Moscow 117312, Russian Federation

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

³Moscow Institute of Physics and Technology (State University), 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation

Abstract: The paper proposes the concept of cognitive personal assistant. The cognitive assistant is a virtual intelligent agent that has its own sign-based world model and builds a world model of the user, which it helps to solve various problems. The architecture of the cognitive assistant is described, the main functions that it should implement are considered, and the main methods and technologies that are used in the construction of such assistants are presented. Two subject areas in which the use of cognitive assistants is the most promising are considered.

Keywords: cognitive assistant; educational assistant; medical assistant; sign-based worldview; natural language processing; script; dialog system; planning

DOI: 10.14357/19922264190315

Acknowledgments

The study was partially funded by the Russian Foundation for Basic Research (project No. 18-29-22027).

References

1. Osipov, G. S., and A. I. Panov. 2018. Relationships and operations in agent’s sign-based model of the world. *Scientific Technical Information Processing* 45(5):1–14.
2. Osipov, G. S., A. I. Panov, and N. V. Chudova. 2014. Behavior control as a function of consciousness. I. World model and goal setting. *J. Comput. Sys. Sc. Int.* 53(4):517–529.
3. Panov, A. I. 2017. Behavior planning of intelligent agent with sign world model. *Biol. Inspir. Cogn. Arc.* 19:21–31.
4. Stankevich, M., I. Smirnov, N. Ignatiev, O. Grigoriev, and N. Kiselnikova. 2018. Analysis of big five personality traits by processing of social media users activity features. *CEUR Workshop Procee.* 2277:162–166.
5. Kiselev G. A., and A. I. Panov. 2018. Znakovyy podkhod k zadache raspredeleniya roley v koalitsii kognitivnykh agentov [Sign-based approach to the task of role distribution in the coalition of cognitive agents]. *Trudy SPIIRAN [SPIIRAS Proceedings]* 57:161–187.
6. Shelmanov, A. O., V. A. Isakov, M. A. Stankevich, and I. V. Smirnov. 2018. Otkrytoe izvlechenie informatsii iz tekstov. Chast’ I. Postanovka zadachi i obzor metodov [Open information extraction. Part I. The task and the review of the state of the art]. *Iskusstvennyy intellekt i prinyatie resheniy [Artificial Intelligence and Decision Making]* 2:47–67.
7. Shelmanov, A. O., D. A. Devyatkin, V. A. Isakov, and I. V. Smirnov. 2019. Otkrytoe izvlechenie informatsii iz tekstov. Chast’ II. Izvlechenie semanticheskikh ot-nosheniy s pomoshch’yu mashinnogo obucheniya bez uchitelya [Open information extraction from texts. Part 2. Extraction of semantic relations using unsupervised machine learning]. *Iskusstvennyy intellekt i prinyatie resheniy [Artificial Intelligence and Decision Making]* 2:39–49.
8. Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. *Adv. Neur. Inf.* 26:3111–3119.
9. Shelmanov, A. O., M. I. Kamenskaya, I. V. Ananyeva, and I. V. Smirnov. 2017. Semantic-syntactic analysis for question answering and definition extraction. *Scientific Technical Information Processing* 44(6):412–423.
10. Osipov, G. S., I. V. Smirnov, and I. A. Tikhomirov. 2010. Relational-situational method for text search and analysis and its applications. *Scientific Technical Information Processing* 6:432–437.
11. Smirnov, I. V., A. O. Shelmanov, E. S. Kuznetsova, and I. V. Khrainin. 2014. Semantiko-sintaksicheskii analiz estestvennykh yazykov. Chast’ II. Metod semantiko-sintaksicheskogo analiza tekstov [Semantic-syntactic analysis of natural languages. Part II. Method for semantic-syntactic analysis of texts]. *Iskusstvennyy intellekt i prinyatie resheniy [Artificial Intelligence and Decision Making]* 1:11–24.
12. Zavjalova, O. S., A. A. Kiselyov, G. S. Osipov, I. V. Smirnov, I. A. Tikhomirov, and I. V. Sochenkov. 2010. Sistema intellektual’nogo poiska i analiza informatsii

- “EXACTUS” na ROMIP-2010 [The system of information search and analysis EXACTUS on ROMIP-2010]. *Tr. Rossiyskogo seminara po otsenke metodov informatsionnogo poiska* [ROMIP-2010 Proceedings]. Kazan: Kazan State University. 49–69.
13. Chistova, E. V., A. O. Shelmanov, and I. V. Smirnov. 2019. Primenenie glubokogo obucheniya k modelirovaniyu dialoga na estestvennom yazyke [Natural language dialogue modelling with deep learning]. *Proceedings of the Institute for Systems Analysis of the Russian Academy of Sciences* 69(1):105–115.
14. Grigoriev, O. G., B. A. Kobrinskii, G. S. Osipov, A. I. Molodchenkov, and I. V. Smirnov. 2018. Health management system knowledge base for formation and support of a preventive measures plan. *Procedia Comput. Sci.* 145:238–241.

Received January 28, 2019

Contributors

Smirnov Ivan V. (b. 1978) — Candidate of Science (PhD) in physics and mathematics; Head of Department, Institute of Artificial Intelligence Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 9, 60-letiya Oktyabrya Prosp., Moscow 117312, Russian Federation; associate professor, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ivs@isa.ru

Panov Aleksandr I. (b. 1987) — Candidate of Science (PhD) in physics and mathematics; senior scientist, Institute of Artificial Intelligence Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 9, 60-letiya Oktyabrya Prosp., Moscow 117312, Russian Federation; deputy head of laboratory, Moscow Institute of Physics and Technology (State University), 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; pan@isa.ru

Skrynnik Alexey A. (b. 1993) — junior scientist, Institute of Artificial Intelligence Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 9, 60-letiya Oktyabrya Prosp., Moscow 117312, Russian Federation; skrynnik@isa.ru

Chistova Elena V. (b. 1996) — programmer, Institute of Artificial Intelligence Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 9, 60-letiya Oktyabrya Prosp., Moscow 117312, Russian Federation; student, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; chistova@isa.ru

ПРИМЕНЕНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ МОМЕНТОВ КОНЕЧНЫХ НОРМАЛЬНЫХ СМЕСЕЙ*

А. К. Горшенин¹, В. Ю. Кузьмин²

Аннотация: Проведено сравнение нейронных сетей прямого распространения и рекуррентных модификаций для решения задачи построения прогнозов непрерывных значений для математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса конечных нормальных смесей. Рассмотрены 14 различных архитектур нейронных сетей, включая и LSTM (Long-Short Term Memory). Для повышения скорости обучения использованы высокопроизводительные вычислительные средства. Продемонстрировано, что на рассматриваемых данных наилучшие результаты для всех моментных характеристик в смысле качества прогнозирования в стандартных метриках (среднеквадратичная ошибка, функция потерь, средняя абсолютная ошибка) достигаются с использованием двух рекуррентных архитектур — с одним скрытым слоем из 100 нейронов и тремя слоями по 50 нейронов.

Ключевые слова: рекуррентные нейронные сети; прогнозирование; глубокое обучение; высокопроизводительные вычисления; CUDA

DOI: 10.14357/19922264190316

1 Введение

Одним из наиболее популярных и востребованных методов анализа данных и по настоящий день остается EM (expectation–maximization) алгоритм. Обычно рассматриваются различные его модификации, однако общий принцип наличия E- и M-шагов остается без изменений. В частности, он применяется для анализа параметров смешанных вероятностных моделей в различных прикладных задачах. Также можно упомянуть кластерный анализ [1–3] и работу с цензурированными данными на основе масштабных смесей нормальных законов [4]. Остается популярным направление исследований, ориентированное на повышение скорости работы EM-алгоритма с помощью различных модификаций — за счет внедрения искусственного шума [5], блочных реализаций для организации параллельных вычислений [6], приближений на основе методов Монте-Карло по схеме марковских цепей [7]. Актуальными остаются вопросы о числе компонент в смешанных моделях [8–10], снижении размерности параметрического пространства [11].

В статье [12] для описания эволюции турбулентных процессов в магнитоактивной высокотемпера-

турной плазме использован метод скользящего разделения конечных нормальных смесей [13]. Исследование моментных характеристик аппроксимирующих распределений позволило проанализировать нелинейную стадию развития турбулентности, ее насыщения, образования вихрей и их хаотизации.

В статье [14] исследованы вопросы построения прогнозов моментов конечных нормальных смесей с помощью нейронных сетей прямого распространения в смысле классической задачи кластеризации. Непрерывные данные разбивались на некоторые промежутки, а нейронная сеть определяла попадание в тот или иной класс. К достоинствам данного метода стоит отнести полученную высокую точность (вплоть до 99,7%) прогнозирования. Однако точные значения моментных характеристик (т. е. решение задачи регрессии), которые бы представляли особый интерес для проведения исследований в физике турбулентной плазмы, в упомянутой работе не определялись.

В данной статье будут рассмотрены 14 различных топологий нейронных сетей [15], включая рекуррентную модификацию LSTM [16]. Они отличаются между собой количеством скрытых слоев и числом нейронов в них. Все они предназначены

* Работа выполнена при поддержке РФФИ (проекты 18-29-03100 и 19-07-00352) и Стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-538.2018.5). Для ускорения обучения был использован гибридный высокопроизводительный вычислительный комплекс ФИЦ ИУ РАН: <http://hhpcc.frccsc.ru>.

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, agorshenin@frccsc.ru

² ООО «Вай2Гео», shadesilent@yandex.ru

для решения задачи построения прогнозов непрерывных значений для математического ожидания (Exp), дисперсии (Var), коэффициентов асимметрии (Skew) и эксцесса (Kurt) конечных нормальных смесей. Кроме того, в работе обсуждается повышение скорости обучения нейросетей с использованием средств гибридного высокопроизводительного вычислительного комплекса (ГВБК).

2 Базовые архитектуры нейронных сетей для решения задачи прогнозирования

В этом разделе опишем конфигурации нейронных сетей, выбранных для прогнозирования различных моментов конечных смесей нормальных законов с функцией распределения следующего вида ($x \in \mathbb{R}$, $a_i(n) \in \mathbb{R}$, $\sigma_i(n) > 0$, $p_i(n) \geq 0$, $\sum p_i(n) = 1$, $i = \overline{1, k(n)}$, $n = 1, 2, \dots$):

$$F(x) = \sum_{i=1}^{k(n)} \frac{p_i(n)}{\sigma_i(n)\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(y - a_i(n))^2}{2\sigma_i^2(n)}\right\} dy.$$

Зависимость от параметра n (номера шага) в этой формуле соответствует принятому в методе скользящего разделения смесей (СРС-методе) изучению эволюции распределения данных во времени в режиме сдвигающегося окна. В качестве тестовых выборок будут использоваться ряды из статьи [12]. Явные выражения для моментных характеристик, включая матричные представления, приведены в работах [14, 17].

Для исследования моментных характеристик были рассмотрены два типа архитектур — сети прямого распространения (многослойный перцептрон) и рекуррентные сети с долгой краткосрочной памятью LSTM. Строится прогноз на 1 шаг, при этом в качестве входных данных используются 50 предшествующих наблюдений. Были рассмотрены следующие архитектуры нейронной сети:

- I: один скрытый слой, 60 нейронов;
- II: один скрытый слой, 100 нейронов;
- III: два скрытых слоя по 20 нейронов в каждом;
- IV: два скрытых слоя по 50 нейронов в каждом;
- V: два скрытых слоя по 100 нейронов в каждом;
- VI: три скрытых слоя по 20 нейронов в каждом;
- VII: три скрытых слоя по 50 нейронов в каждом.

Обучение проводилось на протяжении 750 эпох либо останавливалось при отсутствии убывания

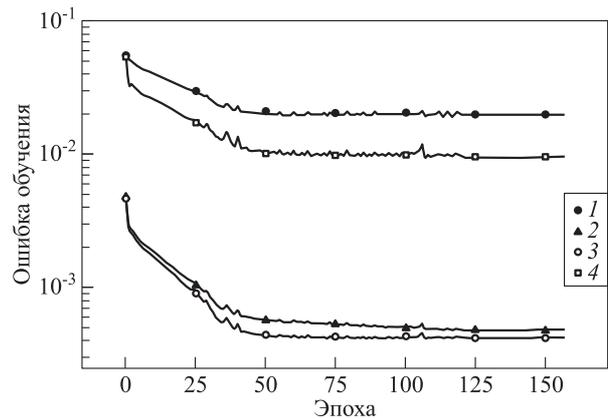


Рис. 1 Пример изменения величины ошибок в процессе обучения (логарифмическая шкала): 1 — RMSE; 2 — Loss; 3 — MSE; 4 — MAE

функции потерь, основанной на среднеквадратичной ошибке, в течение 35 эпох подряд. На рис. 1 показано изменение величины различных использованных метрик в зависимости от эпохи обучения для коэффициента эксцесса, анализируемого с помощью рекуррентной модификации архитектуры VII. Очевидно, что в данном случае нет необходимости в значительном числе шагов, однако подобная ситуация наблюдалась не для всех рядов. Кроме того, через некоторое число эпох после выхода на локальное «плато» обучение может быть продолжено.

Результаты представлены для метода оптимизации Adam [18]; использование SGD [15] и AdaDelta [19] не давало ощутимых преимуществ. В качестве функции активации сетей прямого распространения выбрана ReLU (Rectified Linear Unit) [20]. Для повышения точности прогнозов использовано изменение скорости обучения нейронной сети при достижении «плато» точности (метод ReduceLROnPlateau библиотеки Keras). Для рекуррентных сетей в качестве функции активации применены гиперболический тангенс и так называемая рациональная сигмоида $x/(1 + |x|)$, для которой получены лучшие результаты на тестовых данных. Эффект переобучения не наблюдался. Применение дропаут-слоев [21] не привело к повышению качества результатов, поэтому в окончательных вариантах архитектур они не использовались.

3 Применение нейронных сетей для прогнозирования моментов

В данном разделе рассмотрим результаты применения описанных выше архитектур к задаче

прогнозирования математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса конечных нормальных смесей. Для сравнения результатов прогнозирования использованы среднеквадратичная ошибка RMSE, функция потерь на основе MSE и L^2 -регуляризации, а также средняя абсолютная ошибка MAE. Данные нормализованы (сдвинуты и нормированы) так, чтобы все наблюдения принадлежали сегменту $[0, 1]$.

На рис. 2 представлены величины ошибок RMSE и MAE моделей, полученных в результате обучения 14 архитектур на основе базовых типов I–VII. Символ «а» рядом с римскими цифрами используется для обозначения LSTM.

Из рис. 2 видно, что использование рекуррентных архитектур во всех случаях уменьшает значение ошибки (точные величины для RMSE приведены в табл. 1, наименьшие значения в каждом столбце выделены полужирным шрифтом), при этом в среднем для всех рядов — в 1,33 и 1,31 раза (RMSE и MAE соответственно). Для математического ожидания и коэффициента эксцесса на лучших LSTM-архитектурах ошибка в среднем меньше на 10%–20%, а для дисперсии и коэффициента асимметрии — на 30%–60%. Этот эффект более наглядно проявляется для функции потерь (рис. 3).

В среднем для всех рядов в данной метрике разница составляет 20,3 раза, а в отдельных случаях (см.

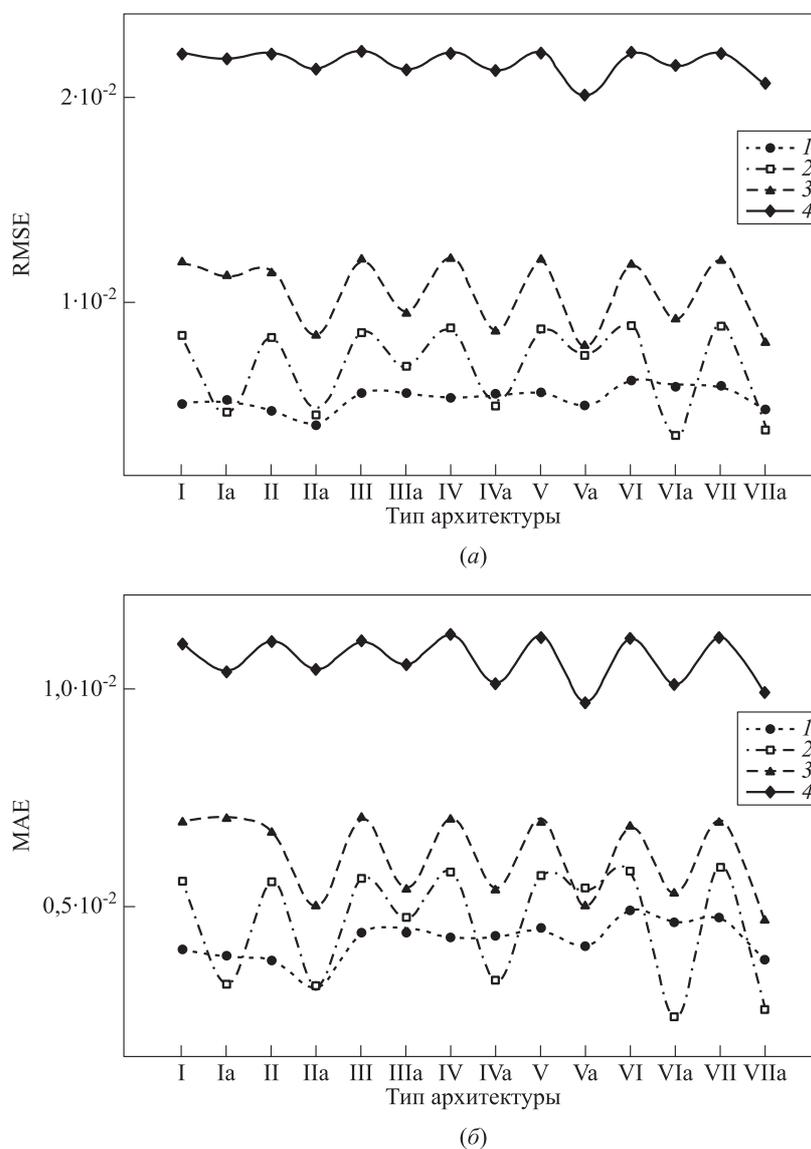
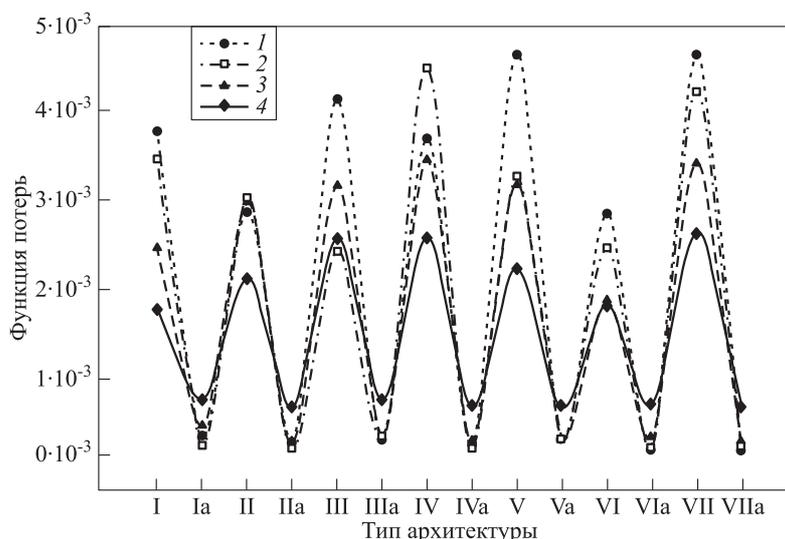


Рис. 2 Сравнение среднеквадратичной (а) и средней абсолютной (б) ошибок для различных архитектур: 1 — математическое ожидание; 2 — дисперсия; 3 — коэффициент асимметрии; 4 — коэффициент эксцесса

Таблица 1 Значения метрики RMSE и функции потерь Loss (мантиссы, порядок — 10^{-3}) для различных архитектур

Архитектура		Момент							
		Exp		Var		Skew		Kurt	
Тип	LSTM	RMSE	Loss	RMSE	Loss	RMSE	Loss	RMSE	Loss
I	Нет	5,03	3,78	8,43	3,45	11,92	2,42	22,18	1,71
	Да	5,08	0,23	4,5	0,11	11,34	0,34	21,94	0,65
II	Нет	4,67	2,85	8,25	3	11,5	2,97	22,18	2,07
	Да	4,04	0,16	4,54	0,05	8,38	0,14	21,44	0,55
III	Нет	5,56	4,14	8,51	2,39	12,02	3,15	22,26	2,53
	Да	5,53	0,19	6,9	0,22	9,56	0,23	21,42	0,64
IV	Нет	5,31	3,7	8,7	4,53	12,14	3,46	22,22	2,54
	Да	5,45	0,17	4,86	0,09	8,59	0,18	21,34	0,57
V	Нет	5,53	4,68	8,61	3,23	11,93	3,16	22,18	2,19
	Да	4,94	0,19	7,37	0,2	7,72	0,18	20,06	0,55
VI	Нет	6,14	2,82	8,78	2,41	11,79	1,81	22,19	1,74
	Да	5,85	0,08	3,38	0,09	9,23	0,22	21,57	0,59
VII	Нет	5,85	4,67	8,89	4,24	12,03	3,41	22,15	2,6
	Да	4,71	0,07	3,71	0,11	8,01	0,17	20,69	0,55


Рис. 3 Сравнение величины функции потерь для различных архитектур: 1 — математическое ожидание; 2 — дисперсия; 3 — коэффициент асимметрии; 4 — коэффициент эксцесса

математическое ожидание для конфигураций VII и VIIa) получается более чем 65-кратное уменьшение величины ошибки. Отдельно можно выделить конфигурации IIa и VIIa (т. е. рекуррентные сети), в которых для всех моментных характеристик сразу в обеих метриках получены либо наименьшие среди всех, либо близкие к этому значения. Таким образом, применение рекуррентных архитектур ведет к значительному повышению качества обучения в любой из рассматриваемых метрик.

Для сравнения архитектур определим индикатор, который представим как сумму отношений ошибки в некоторой метрике на данной архитектуре к минимальной ошибке прогноза для данного

ряда (случай нулевой величины на практике является почти недостижимым). Очевидно, что его наименьшее значение — 4.

На тестовых данных минимальные значения данного индикатора достигаются на архитектуре IIa для математического ожидания, VIa для дисперсии и Va для коэффициентов асимметрии и эксцесса. При этом оптимальными в общем случае можно признать конфигурации II и IIa, поскольку для них получается минимальная средневзвешенная сумма ошибок для сетей прямого распространения (4,001) и второе по величине значение после архитектуры VIIa, для которой введенный индикатор равен 4,49, среди всех рекуррентных сетей.

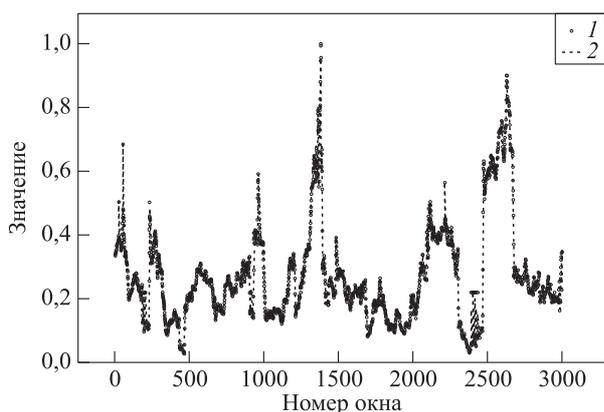


Рис. 4 Коэффициент эксцесса и прогнозы архитектуры VIIa: 1 — данные; 2 — модель

Для иллюстрации качества приближения данных обученными моделями на рис. 4 продемонстрированы значения для коэффициента эксцесса (нормализованные данные) и аппроксимация ряда с помощью предсказаний, сделанных с применением рекуррентной архитектуры VIIa.

4 Повышение эффективности обучения нейронных сетей за счет использования высокопроизводительных вычислительных средств

В данном разделе будут обсуждаться вопросы, связанные с временными затратами на обучение

базовых архитектур I–VII. Результаты разд. 3 означают, что использование рекуррентных сетей позволило существенным образом повысить качество аппроксимации исходных данных. Однако нельзя не принимать во внимание тот факт, что усложнение конфигурации влечет за собой и дополнительную вычислительную нагрузку при обучении. В табл. 2 приведены временные затраты для «обычной» и LSTM-архитектур, при этом не учитываются накладные расходы на предварительную обработку данных и инициализацию графических видеокарт.

В среднем сети прямого распространения обучались за 664 эпохи, в то время как рекуррентные модификации — за 687. Для рассматриваемых рядов было установлено, что для LSTM-конфигураций необходимое время обучения в среднем превышает результаты для классических в 47 раз (минимальное значение — 6 раз, максимальное — 90) в зависимости от архитектуры и анализируемого ряда. Из отмеченных в разд. 3 конфигураций IIa и VIIa лучшее время показывает именно IIa — около 30 мин в среднем на моментную характеристику, в то время как для VIIa получается более 55 мин. Таким образом, для более быстрой обработки рядов можно порекомендовать именно архитектуру IIa.

Эффект от использования средств ГВБК проявляется прежде всего при обучении нейронных сетей прямого распространения. Так, среднее время на одну эпоху для сетей прямого распространения составляет 2 с для настольного решения (процессор Core i7, видеокарта NVIDIA GTX970M), а для ГВБК — 0,1 с. Однако и для рекуррентных конфигураций получено пятикратное ускорение, а именно: с 25 до 5 с, — которое позволяет рассматривать их в качестве инструмента решения реальных прикладных задач.

Таблица 2 Время, затраченное на обучение на ГВБК (в секундах)

Архитектура		Момент			
Тип	LSTM	Exp	Var	Skew	Kurt
I	Нет	38,08	32,99	303,79	67,72
	Да	2081,98	1932,59	1872,37	2034,13
II	Нет	51,21	31,99	37,93	71,48
	Да	2170,9	1999,57	1985,41	1079,17
III	Нет	45,61	49,82	56,26	80,56
	Да	3114,08	2813,48	2815,11	1362,48
IV	Нет	63,83	36,16	57,23	80,04
	Да	3108,84	2839,28	2870,9	1680,07
V	Нет	58,79	40,7	61,33	79,21
	Да	3194,87	2983,34	2988,7	3159,44
VI	Нет	73,9	42,52	71,38	61,83
	Да	3897,8	3826,78	3756,39	2435,18
VII	Нет	71,72	275,35	68,82	61,96
	Да	3534,68	4018,36	3890,21	1909,27

5 Заключение

В работе проведено сравнение результатов для 14 различных архитектур нейронных сетей в задаче прогнозирования первых четырех моментных характеристик конечных смесей нормальных распределений. Выбраны две лучшие конфигурации (с одним скрытым слоем из 100 нейронов и с тремя слоями по 50 нейронов), для которых получены наиболее точные значения во всех рассматриваемых метриках. При этом на обучение конфигурации с одним слоем затрачивается меньшее время, поэтому она может быть использована в задачах, для которых более критично быстрое действие, а не точность аппроксимации. Используемые подходы в достаточной степени универсальны, поэтому полученные архитектуры могут быть успешно применены для анализа временных рядов различной природы.

Литература

1. Yang M.-Sh., Lai Ch.-Yo, Lin C.-Y. A robust EM clustering algorithm for Gaussian mixture models // Pattern Recogn., 2012. Vol. 45. Iss. 11. P. 3950–3961.
2. Cai T. T., Ma J., Zhang L. CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality // Ann. Stat., 2019. Vol. 47. Iss. 3 P. 1234–1267.
3. Ben Hassen H., Masmoudi K., Masmoudi A. Model selection in biological networks using a graphical EM algorithm // Neurocomputing, 2019. Vol. 349. P. 271–280.
4. Zeller C. B., Cabral C. R. B., Lachos V. H., Benites L. Finite mixture of regression models for censored data based on scale mixtures of normal distributions // Adv. Data Anal. Classif., 2019. Vol. 13. Iss. 1. P. 89–116.
5. Osoba O., Mitaim S., Kosko B. The noisy Expectation–Maximization algorithm // Fluct. Noise Lett., 2013. Vol. 12. Iss. 3. Art. No. 1350012.
6. Lee S. X., Leemaqz K. L., McLachlan G. J. A block EM algorithm for multivariate skew normal and skew t-mixture models // IEEE T. Neur. Net. Lear., 2018. Vol. 29. Iss. 11. P. 5581–5591.
7. Wu D., Ma J. An effective EM algorithm for mixtures of Gaussian processes via the MCMC sampling and approximation // Neurocomputing, 2019. Vol. 331. P. 366–374.
8. Verbeek J. J., Vlassis N., Krose B. Efficient greedy learning of Gaussian mixture models // Neural Comput., 2003. Vol. 15. Iss. 2. P. 469–485.
9. Горшенин А. К. Проверка статистических гипотез в модели расщепления компоненты // Вестник Московского университета. Сер. 15: Вычислительная математика и кибернетика, 2011. Вып. 4. С. 26–32.
10. Liu C., Li H.-C., Fu K., Zhang F., Dacu M., Emery W. J. A robust EM clustering algorithm for Gaussian mixture models // Pattern Recogn., 2019. Vol. 87. P. 269–284.
11. Yu L., Yang T., Chan A. B. Density-preserving hierarchical EM algorithm: Simplifying Gaussian mixture models for approximate inference // IEEE T. Pattern Anal., 2019. Vol. 41. Iss. 6. P. 1323–1337.
12. Batanov G. M., Borzosekov V. D., Gorshenin A. K., Kharchev N. K., Korolev V. Yu., Sarskyan K. A. Evolution of statistical properties of microturbulence during transient process under electron cyclotron resonance heating of the L-2M stellarator plasma // Plasma Phys. Contr. F., 2019. Vol. 61. Iss. 7. Art. No. 075006.
13. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011. 512 с.
14. Gorshenin A. K., Kuzmin V. Yu. Improved architecture of feedforward neural networks to increase accuracy of predictions for moments of finite normal mixtures // Pattern Recogn. Image Anal., 2019. Vol. 29. No. 1. P. 79–88.
15. Buduma N. Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. — Sebastopol, CA, USA: O’Reilly Media, 2017. 298 p.
16. Greff K., Srivastava R. K., Koutnik J., Steunebrink B. R., Schmidhuber J. LSTM: A search space Odyssey // IEEE T. Neur. Net. Lear., 2017. Vol. 28. Iss. 10. P. 2222–2232.
17. Горшенин А. К. Концепция онлайн-комплекса для стохастического моделирования реальных процессов // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 72–81.
18. Kingma D., Ba J. Adam: A method for stochastic optimization // 3rd Conference (International) for Learning Representations // arXiv:1412.6980, 2015. 13 p.
19. Zeiler M. D. ADADELTA: An adaptive learning rate method // arXiv:1212.5701, 2012. 6 p.
20. Glorot X., Bordes A., Bengio Y. Deep sparse rectifier neural networks // J. Mach. Learn. Res., 2011. Vol. 15. P. 315–323.
21. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting // J. Mach. Learn. Res., 2014. Vol. 15. P. 1929–1958.

Поступила в редакцию 04.09.19

APPLICATION OF RECURRENT NEURAL NETWORKS TO FORECASTING THE MOMENTS OF FINITE NORMAL MIXTURES

A. K. Gorshenin^{1,2} and V. Yu. Kuzmin³

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

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

³“Wi2Geo LLC”, 3-1 Mira Ave., Moscow 129090, Russian Federation

Abstract: The article compares the application of feedforward and recurrent neural networks to forecasting continuous values of expectation, variance, skewness, and kurtosis of finite normal mixtures. Fourteen various architectures of neural networks are considered. To increase training speed, the high-performance computing cluster is used. It is demonstrated that the best forecasting results based on standard metrics (root-mean-square error, mean absolute errors, and loss function) are achieved on the two LSTM (Long-Short Term Memory) networks: with 100 neurons in one hidden layer and 50 neurons in each three hidden layers.

Keywords: recurrent neural networks; forecasting; deep learning; high-performance computing; CUDA

DOI: 10.14357/19922264190316

Acknowledgments

The research is partially supported by the Russian Foundation for Basic Research (projects 18-29-03100 and 19-07-00352) and the RF Presidential scholarship program (project No. 538.2018.5). The calculations were performed using Hybrid high-performance computing cluster of FRC CSC RAS (<http://hhpcc.frccsc.ru>).

References

1. Yang, M.-Sh., Ch.-Yo Lai, and C.-Y. Lin. 2012. A robust EM clustering algorithm for Gaussian mixture models. *Pattern Recogn.* 45(11):3950–3961.
2. Cai, T. T., J. Ma, and L. Zhang. 2019. CHIME: Clustering of high-dimensional Gaussian mixtures with em algorithm and its optimality. *Ann. Stat.* 47(3):1234–1267.
3. Ben Hassen, H., K. Masmoudi, and A. Masmoudi. 2019. Model selection in biological networks using a graphical EM algorithm. *Neurocomputing* 349:271–280.
4. Zeller, C. B., C. R. B. Cabral, V. H. Lachos, and L. Benites. 2019. Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Adv. Data Anal. Classif.* 13(1):89–116.
5. Osoba, O., S. Mitaim, and B. Kosko. 2013. The noisy Expectation-Maximization algorithm. *Fluct. Noise Lett.* 12(3):1350012.
6. Lee, S. X., K. L. Leemaqz, and G. J. McLachlan. 2018. A block EM algorithm for multivariate skew normal and skew t-mixture models. *IEEE T. Neur. Net. Lear.* 29(11):5581–5591.
7. Wu, D., J. Ma. 2019. An effective EM algorithm for mixtures of Gaussian processes via the MCMC sampling and approximation. *Neurocomputing* 331:366–374.
8. Verbeek, J. J., N. Vlassis, and B. Krose. 2003. Efficient greedy learning of Gaussian mixture models. *Neural Comput.* 15(2):469–485.
9. Gorshenin, A. K. 2011. Testing of statistical hypotheses in the splitting component model. *Mosc. Univ. Comput. Math. Cybernetics* 35(4):176–183.
10. Liu, C., H.-C. Li, K. Fu, F. Zhang, M. Datcu, and W. J. Emery. 2019. A robust EM clustering algorithm for Gaussian mixture models. *Pattern Recogn.* 87:269–284.
11. Yu, L., T. Yang, and A. B. Chan. 2019. Density-preserving hierarchical EM algorithm: Simplifying Gaussian mixture models for approximate inference. *IEEE T. Pattern Anal.* 41(6):1323–1337.
12. Batanov, G. M., V. D. Borzosekov, A. K. Gorshenin, N. K. Kharchev, V. Yu. Korolev, and K. A. Sarskyan. 2019. Evolution of statistical properties of microturbulence during transient process under electron cyclotron resonance heating of the L-2M stellarator plasma. *Plasma Phys. Contr. F.* 61(7):075006.
13. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Moscow University Publishing House. 512 p.
14. Gorshenin, A. K., and V. Yu. Kuzmin. 2019. Improved architecture of feedforward neural networks to increase accuracy of predictions for moments of finite normal mixtures. *Pattern Recogn. Image Anal.* 29(1):68–77.
15. Buduma, N. 2017. *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. Sebastopol, CA: O'Reilly Media. 298 p.

16. Greff, K., R. K. Srivastava, J. Koutnik, B. R Steunebrink, and J. Schmidhuber. 2017. LSTM: A search space Odyssey. *IEEE T. Neur. Net. Lear.* 28(10):2222–2232.
17. Gorshenin, A. K. 2016. Kontseptsiya onlayn-kompleksa dlya stokhasticheskogo modelirovaniya real'nykh protsessov [Concept of online service for stochastic modeling of real processes]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):72–81.
18. Kingma, D., and J. Ba. 2015. Adam: A method for stochastic optimization. *3rd Conference (International) for Learning Representations*. arXiv:1412.6980. 13 p.
19. Zeiler, M. D. 2012. ADADELTA: An adaptive learning rate method. arXiv:1212.5701. 6 p.
20. Glorot, X., A. Bordes, and Y. Bengio. 2011. Deep sparse rectifier neural networks. *J. Mach. Learn. Res.* 15:315–323.
21. Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15:1929–1958.

Received September 4, 2019

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; agorshenin@frccsc.ru

Kuzmin Victor Yu. (b. 1986) — Head of Development, “Wi2Geo LLC,” 3-1 Mira Ave., Moscow 129090, Russian Federation; shadesilent@yandex.ru

МЕТОДЫ МОДЕЛИРОВАНИЯ И ВИЗУАЛЬНОГО ПРЕДСТАВЛЕНИЯ КОНФЛИКТА В МАЛОМ КОЛЛЕКТИВЕ ЭКСПЕРТОВ, РЕШАЮЩИХ ПРОБЛЕМЫ (ОБЗОР)

С. Б. Румовская¹, И. А. Кириков²

Аннотация: Малые коллективы экспертов как естественный коллективный интеллект поддержки принятия решений (гетерогенный коллектив) эффективно решают сложные проблемы. При этом такая форма взаимодействия между экспертами, как конфликт, порождает позитивные изменения в коллективе: развитие группы, диагностику отношений, снятие напряжения, сплачивание группы, а также способствует сохранению коллектива. В человеческом мышлении огромную роль играют заготовки-схемы стандартных ситуаций, использование которых существенно ускоряет рассуждения. Визуализация конфликтной ситуации делает возникшие противоречия контрастными, видимыми, предоставляя новую информацию для разрешения конфликтов, делая их легкоуправляемыми и позволяя контролировать влияние на них субъективных предпочтений. Рассмотрено понятие конфликта в малых коллективах, его особенности, структура и динамика, а также подходы к моделированию и визуальному представлению конфликтологического аспекта групповой динамики экспертов, решающих проблемы.

Ключевые слова: малый коллектив экспертов; конфликт; модели конфликта; визуализация конфликта

DOI: 10.14357/19922264190317

1 Введение

Естественный коллективный интеллект (гетерогенный коллектив) поддержки принятия решений [1] — малая группа экспертов, которой присущи неоднородность, разнообразие, сотрудничество, дополнительность и относительность знаний. Подобные коллективы эффективно решают сложные проблемы. Ввиду этого малые группы, проблемы взаимодействия людей внутри них, а также моделирование их взаимодействия занимают особое место в широком спектре направлений современной науки. Часто встречающаяся форма организации малых коллективов — совещания, построенные по принципу «круглого стола» [2] с целью выявления и решения проблем. Разного рода конфликты порождают дискуссии, глубина которых позволяет получить более продуманные и согласованные решения. Координация работы экспертов в группе лицом, принимающим решения (ЛПР), позволяет повысить качество решений, а самоорганизация в группе определяет способность чутко реагировать на изменения во внешней среде, корректируя свое функционирование и принимаемые решения.

В [3] представлена система для оперативно-производственного планирования — моделирова-

лась координация ЛПР (начальником производственного отдела) коллективной работы главного конструктора, главного технолога, начальника отдела материального снабжения, начальника электромеханического цеха, начальника отдела продаж с целью повышения качества оперативных планграфиков мелкосерийного производства. Относительная погрешность результатов решения задачи с координацией — менее 1%, а без — до 36%.

В [3] также описана модель самоорганизации в группе для транспортной логистики гибридными интеллектуальными многоагентными системами (ГиМАС). Алгоритм функционирования системы динамически перестраивается, вырабатывая релевантный сложной задаче метод решения и сокращая среднюю суммарную себестоимость и время доставки грузов в день на 7,2% и 12,13% соответственно; среднее время построения маршрутов уменьшилось на 23,14%.

Долгое время социологи и психологи считали, что конфликты — негативное явление и их надо устранять. Однако работа Г. Зиммеля [2] способствовала развитию идеи наличия позитивных изменений, порождаемых конфликтами: сохранение социальной системы, развитие группы, диагностика отношений, снятие напряжения, сплачивание группы и др.

¹Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

²Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

Таким образом, моделирование развития и разрешения конфликта в рамках методологии ГиМАС позволит спроектировать функционирование системы релевантно групповой динамике коллектива экспертов, решающих проблему, и тем самым существенно повысить качество принимаемых решений. При этом визуализация конфликтной ситуации сделает возникшие противоречия контрастными, видимыми, предоставляя новую информацию для разрешения конфликтов в реальном коллективе экспертов. В работе рассмотрено понятие конфликта в малых коллективах, его особенности, структура и динамика, а также подходы к моделированию и визуальному представлению конфликтологического аспекта групповой динамики экспертов, решающих проблемы.

2 Малый коллектив экспертов: понятие и классификация

Попытки определить малую группу экспертов сводились к субъективному пониманию и фокусировке на тех или иных сторонах группового процесса (определенных априори либо эмпирическим): М. Шоу [4] (психологическая составляющая), Р. Браун [5] и Г. М. Андреева [6] (социальная составляющая).

Приведем определение, включающее в себя психологическую и социальную составляющие: «малая группа» — элементарное звено структуры социальных отношений, обретающее через непосредственные межличностные контакты структурные, динамические, феноменологические характеристики, отражающие признаки группы как целостной системы социальных и психологических отношений. Понятия «малая группа» и «малый коллектив» идентичны.

Верхняя граница размерности коллектива (нижняя — 2 участника) определяется с двух точек зрения:

- (1) в соответствии с требованиями реализации ее основной функции [6] — верхняя граница не может быть обозначена априори;
- (2) относительно успешного руководства группой — верхняя граница соответствует «магическому числу» Дж. Миллера (7 ± 2), так как при численности свыше 10 чел. возрастает число подгрупп и вероятность противостояния ЛПР, осложняется координация.

Есть различные классификации малых коллективов [6], в частности английский психолог М. Аргайл выделяет [7]: *семью; подростково-юношеские коллективы; рабочие коллективы* — модель малых

групп с четкой трудовой направленностью и доминирующими отношениями делового характера; *комитеты и группы по решению проблем* — модель коммуникативных малых коллективов, задача которых принимать эффективные решения, а их участники должны владеть навыками организации информационного обмена, достижения внутригруппового согласия и т. п.; *тренинговые и терапевтические группы*.

3 Конфликт: понятие, структура и типология

Конфликт [8, 9] — столкновение противоположно направленных интересов, целей, взглядов и т. п. (обострение противоречия) при взаимодействии и взаимоотношении сторон, воспринимаемое субъектом как значимая для него психологическая проблема, требующая своего разрешения и вызывающая активность, направленную на его преодоление. Структура конфликта представлена на рис. 1 [10, 11].

Стороны конфликта — это субъекты социального взаимодействия: в состоянии конфликта (основные участники, оппоненты); или же явно или неявно поддерживающие конфликтующих (группы поддержки); или оказывающие эпизодическое влияние на конфликт (другие участники — подстрекатели, организаторы и т. д.). Объект находится на пересечении личных или групповых интересов субъектов. Предмет конфликта — это противоречие, из-за которого и для разрешения которого возникает конфликт. Образ конфликтной ситуации — это отображение предмета конфликта в сознании субъектов. Мотивы конфликта — это потребности, интересы, цели и т. д., подталкивающие субъектов к конфликту.

Конфликты классифицируются по различным признакам (рис. 2) [8, 10].

Конструктивные возникают при объективных противоречиях, когда цели и потребности сторон едины (развитие коллектива). *Деструктивные* — причины их субъективны, вызывают напряженность и разрушают коллектив.

Конфликт отношений (социальный конфликт) — разногласие между членами группы по личным вопросам и проблемам, не относящимся к выполняемой работе (связан с несовместимостью и враждебностью между субъектами).

Межличностные конфликты бывают пяти типов (см. рис. 2). При этом конфликты внутри устойчивой подгруппы более скоротечны и чаще

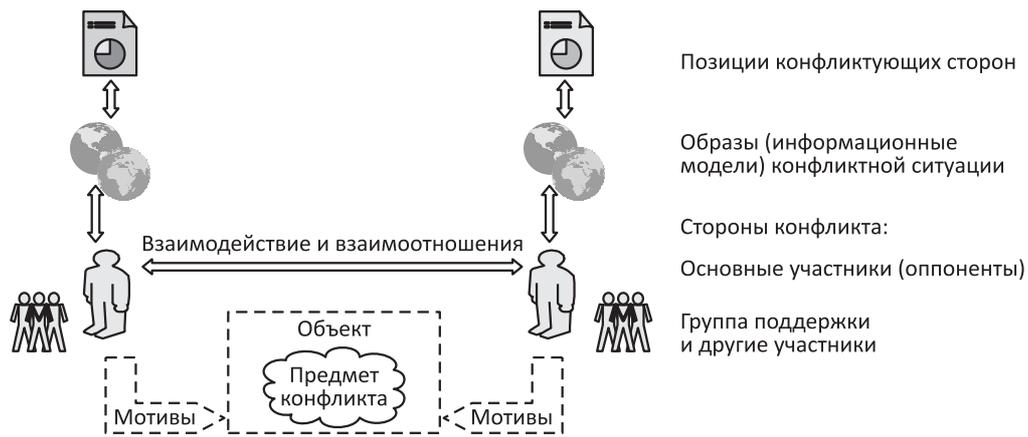


Рис. 1 Структура конфликта

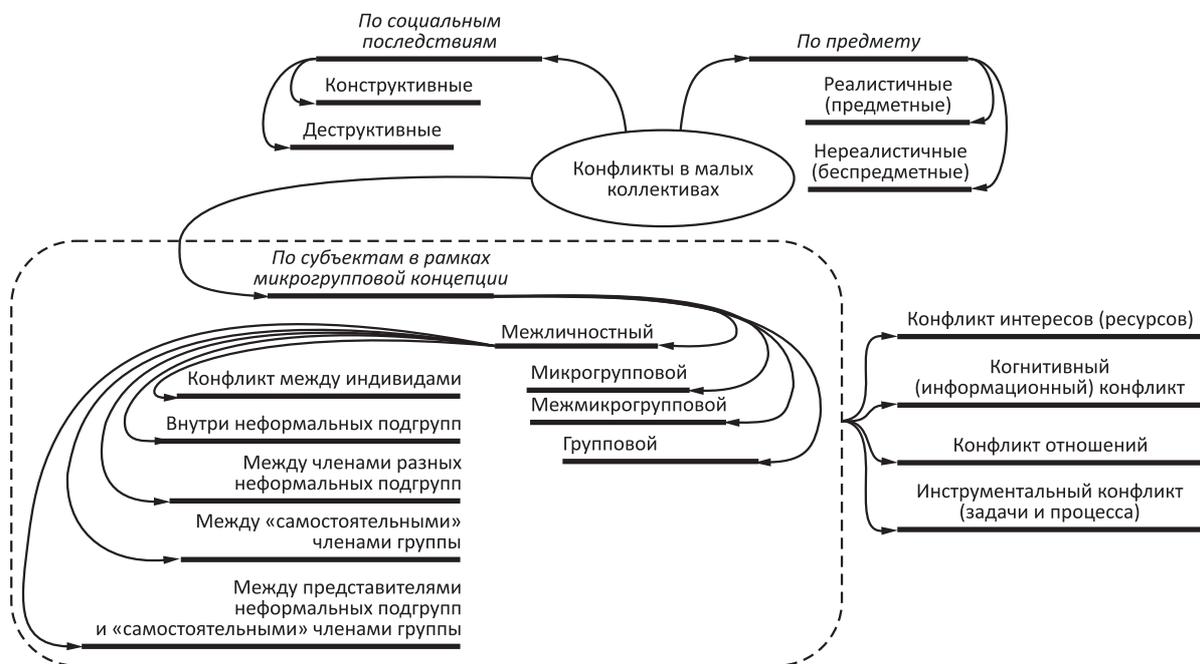


Рис. 2 Типология конфликта в малом коллективе

имеют иное значение и последствие для их участников — расцениваются как конфликт «между своими», семейный конфликт. *Микрогрупповой* — конфликт между индивидом и неформальной подгруппой. *Межмикрогрупповой* — конфликт между неформальными подгруппами в группе. *Групповой* — конфликт между индивидом и группой.

Конфликт интересов обусловлен мотивационными факторами и ситуацией, в которой цели каждого члена не совпадают с целями других. *Конфликт ресурсов* предполагает отсутствие соглашения между субъектами по поводу ресурсов. *Когнитивный конфликт* — цели всех членов группы совпадают, но

их позиции различны (фокус на интеллектуальных или оценочных проблемах, связи между внутригрупповым конфликтом и деятельностью группы).

Инструментальный (деловой) конфликт:

- (1) *конфликт задачи* — связан с различием точек зрения на групповые цели и задачи (помогают открытое обсуждение и споры);
- (2) *конфликт процесса* — возникает по поводу проблем, возникающих относительно технологии и способов решения поставленной задачи, распределения между субъектами ролей и ответственности.

4 Причины, позитивные функции и динамика конфликта

Есть разные точки зрения [9, 12] на классификацию причин конфликтов. Приведем причины по К. Левину [9]:

- *степень удовлетворенности потребностей* — неудовлетворенные потребности часто становятся доминирующими, увеличивая вероятность конфликтов;
- *«пространство свободного движения»* (достаточное пространство — условие удовлетворения индивидуальных потребностей и адаптации к группе, а ограниченность ведет к росту напряжения);
- *внешний барьер* — наличие или отсутствие возможностей выйти из неприятной ситуации (отсутствие провоцирует конфликт);
- *совпадение или расхождение целей членов группы* — конфликты зависят от степени противоречия целей участников и от их готовности к компромиссу.

Позитивные функции конфликта [8]:

- обеспечивает уникальность и автономность каждого из взаимодействующих субъектов, а также развитие отношений между ними;
- предоставляет информацию о возможностях противодействующих субъектов;
- высвобождает накапливающееся внутреннее напряжение, сохраняя связи;
- актуализирует разные позиции и мнения по поводу возникающих проблем и тем самым способствует поиску оптимальных способов их решения;
- усиливает групповую/микрогрупповую идентичность и сплоченность.

Динамика конфликта [10, 11] отражается в двух понятиях: этапы конфликта и фазы конфликта. Основные этапы конфликта:

- (1) *возникновение объективной проблемной ситуации* (появление противоречия);
- (2) *осознание проблемы* хотя бы одним из участников (если разрешить проблему неконфликтными методами не получается, то возникает предконфликтная ситуация);
- (3) *начало открытого конфликтного взаимодействия (инцидент)*;
- (4) *развитие открытого конфликта (эскалация)* — открыто заявляются позиции и выдвигаются требования. Завершается этот этап *сбалансированным противодействием* — когда силовые методы не дают результата, интенсивность борьбы снижается, но действия по достижению согласия еще не предпринимаются;
- (5) *разрешение конфликта*.

Основные фазы конфликта:

- (1) начальная фаза;
- (2) фаза подъема;
- (3) пик конфликта;
- (4) фаза спада.

Фазы могут повторяться циклически (рис. 3) [10]. При этом возможности разрешения конфликта в каждом последующем цикле сужаются.

5 Методы моделирования конфликта

В открытой печати встречаются подходы к моделированию межгрупповых и межгосударственных конфликтов [11, 13–15], которые позволяют заменить непосредственный анализ конфликтов анализом свойств и характеристик их моделей, а также прогнозировать и оценивать события в реальном времени. Их опыт можно перенести на моделирование конфликтов в малых коллективах при их

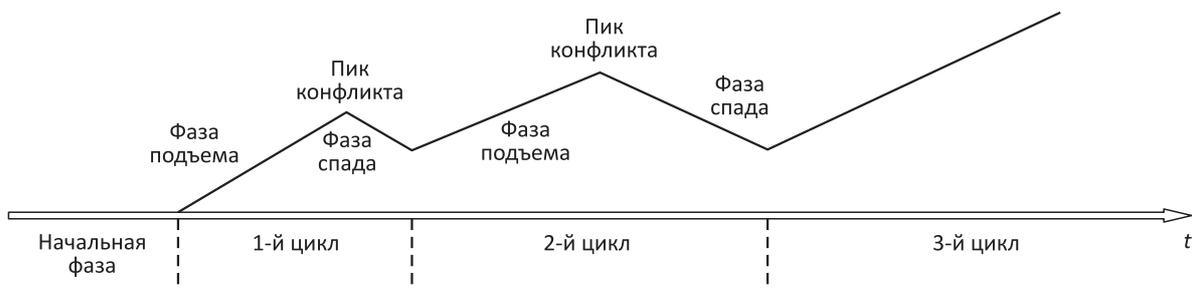


Рис. 3 Фазы конфликта

классификации в рамках микрогрупповой концепции [8]. В [5] представлен развернутый обзор моделирования военных конфликтов и выделяются:

- (1) *описательные модели военных действий* — основываются на методах теории вероятностей и статистической теории решений (принятие решений в условиях «природной» неопределенности), теории надежности и теории массового обслуживания, теории экспертных оценок, а также качественный анализ соответствующих динамических систем и исследование их структурной устойчивости;
- (2) *имитационные модели* — основываются на аппарате марковских цепей, дифференциальных уравнений (ланчестеровские модели), конечных автоматов или методах распределенного искусственного интеллекта;
- (3) *оптимизационные модели военных действий* используют аппарат линейного и динамического программирования, теории оптимального управления, дискретной оптимизации;
- (4) *модели принятия решений* — индивидуального (основной акцент обычно делается на многокритериальном принятии решений) и коллективного (акцент на использовании теории игр).

Многоагентное моделирование — методология, применяемая для поддержки принятия решений, анализа и изучения сложных систем, состоящих из отдельных, функционирующих независимо друг от друга индивидов [16]. Изучается влияние взаимодействий индивидов на системные характеристики в целом.

Вероятностные распределения — способ описания переменных через указание доли элементов совокупности с данным значением переменной [17].

Модели целенаправленного поведения — использование целевых функций для анализа, прогнозирования и планирования социальных процессов [18]. Модели имеют вид задачи математического программирования.

Статистические исследования зависимостей — прежде всего это регрессионные модели, представляющие связь зависимых и независимых переменных в виде функциональных отношений [19].

Теоретические модели предназначены для логического анализа тех или иных содержательных концепций, когда затруднена возможность измерения основных параметров и переменных (возможные межгосударственные конфликты и др.) [17].

6 Методы визуализации конфликтов

Эффективность разрешения конфликтов во многом определяется правильностью и полнотой их анализа. Рассмотрим основные методы, предлагаемые зарубежными и отечественными психологами.

Картографический анализ конфликтов Х. Корнелиус и Ш. Фейр [20]. Проблема обычно записывается в виде противоречия в центре карты (рис. 4, а), которая делится по числу сторон конфликта. У каждого участника конфликта выясняются потребности, интересы и опасения. Все основные элементы конфликта (проблема, позиции, условия, образы, интересы, потребности, опасения, исходы и др.) упорядочиваются и систематизируются.

Метод декартовых координат [21]. Р. Декарт считал, что наше сознание имеет определенную структуру и его можно представить как пространство (рис. 4, б), обозначенное математическими

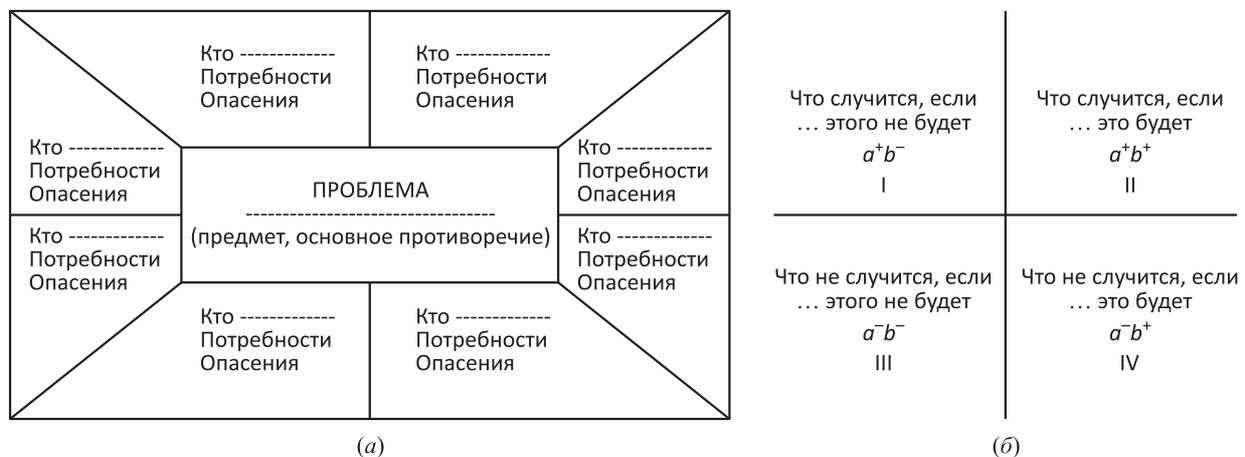


Рис. 4 Карта анализа конфликта (а) и метод декартовых координат (б)

Конфликтограмма

Проблемные вопросы	Мои	Оппонента
Проблема	Балансный лист ресурсов участников конфликта	
Цели, интересы	Мои интересы	Интересы партнера
Препятствия	Ресурсы	Ресурсы
Опасения	Времени	Свободы
Сильные стороны	Уважения	Моральный
Возможность поддержки	Материальный	Здоровья
Недостающая информация	Статусный	Информационный
Какую личную потребность удовлетворяют	Любви	Материальный
Эмоции		
Личностные ресурсы		
Ресурсы, которые можно обменять		

Рис. 5 Ресурсный анализ конфликта: конфликтограмма и балансный лист ресурсов

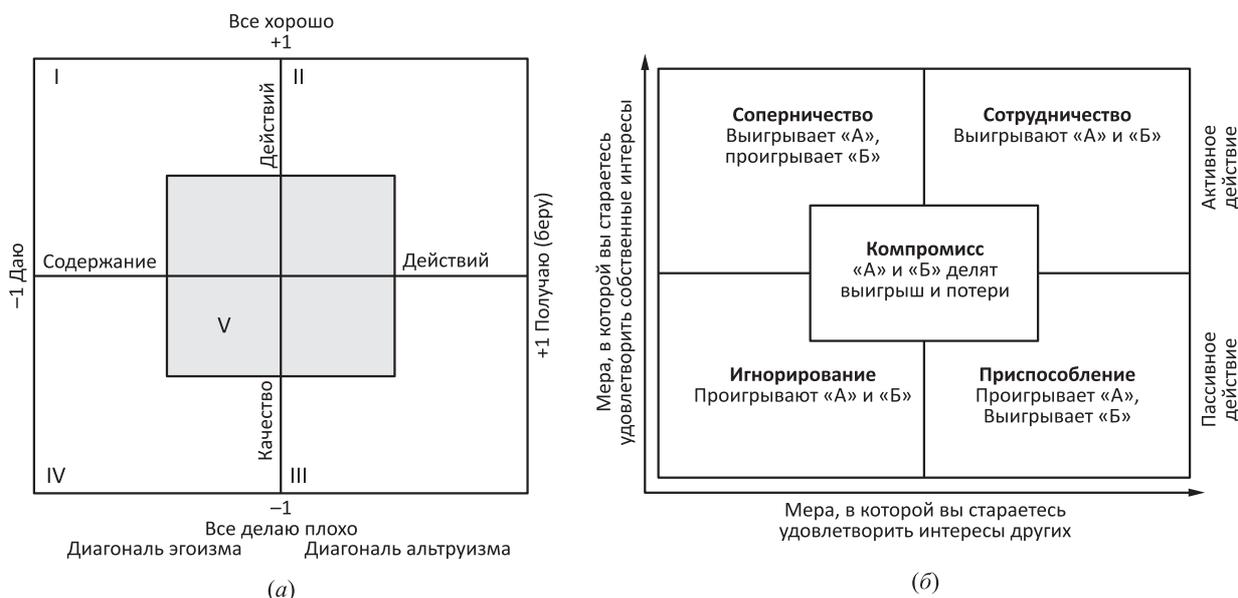


Рис. 6 Модели социального поведения (а) и сетка К. Томаса – Л. Килмена (б)

символами (а, б). Все это пространство поделено на четыре квадрата с помощью двух пересекающихся прямых. В каждый квадрат надо записать не менее десяти ответов, которые сопоставляются и анализируются для поиска решения выхода из конфликта.

Ресурсный анализ конфликта В. Н. Ковалева (РАКК) [22] базируется на заполнении всеми субъектами конфликтограммы и балансного листа ресурсов сторон (рис. 5), которые позволяют посмотреть на конфликт глазами оппонента, установить возможность обмена ресурсами и компромисса.

Межличностные отношения в конфликте в рамках РАКК представляются моделями социального поведения оппонента (рис. 6): четверть I — альтруизм (оказание помощи другим); II — эгоизм (приобретение всего лучшего без обмена ресурсами); III — альтруизм (оказание помощи через избавление от трудностей других); IV — эгоизм (избавле-

ние от всего плохого). В квадрате V в соизмеримом отношении встречаются все формы поведения. Завершающий этап — оценка сторонами того, что они отдали и что приобрели, и взаимное согласие с полученными результатами. Для анализа стилей поведения личности в конфликте психологи пользуются сеткой К. Томаса и Р. Килмена (рис. 6, б) — она дает возможность спрогнозировать исходы конфликта, остроту его протекания, возможные ресурсные потери.

Затем анализируются возможные типы поведения и реакции на них (рис. 7), и каждой стороной заполняется таблица проектов предложений. Стороны приступают к совместному анализу предложений и заключают соглашение.

Также для визуализации конфликта можно использовать дерево целей и дерево проблем [23]. Дерево целей — это структурированная, построенная по иерархическому принципу совокупность целей

Комплексная матрица анализа возможных типов поведения						
Возможный подход	Плюсы подхода	Минусы подхода	Возможная реакция другой стороны	Вероятность реакции	Желательность реакции	Итог, оценка
Уклонение						
Приспособление	Проекты предложений по выбору стиля поведения в конфликте					
Конкуренция						
Компромисс				Ваши предложения к оппоненту	Аргументация предложений	
Сотрудничество						

Рис. 7 Комплексная матрица анализа возможных типов поведения и проекты предложений

конфликтующей системы. Если представить предмет конфликта как проблему, то можно построить дерево проблем, а затем и дерево решений. В [24] рассматривается диаграмма разрешения конфликтов, в которой совместно используются деревья текущей реальности, будущей реальности, перехода, разрешения конфликтов «Грозовая туча».

7 Заключение

Аналитический обзор по материалам открытой печати показал разнообразие подходов к определению понятия конфликта, его классификации, а также причин и функций конфликта. При этом для анализа конфликтной ситуации с целью ее разрешения нет единого подхода к визуализации субъектов конфликтной ситуации, их характеристик, взаимодействия и взаимоотношений. Модели конфликтов строятся в рамках различных методов и подходов: описательные, имитационные, оптимизационные и модели принятия решений — однако найдены были только модели военных конфликтов. Используя опыт моделирования последних [14, 16–21] с позиции микрогрупповой концепции А. В. Сидоренкова [10], будет разработана модель конфликта в ГиМАС.

Литература

1. Колесников А. В. Гетерогенные естественные и искусственные системы // Интегрированные модели и мягкие вычисления в искусственном интеллекте / Под ред. В. Б. Тарасова. — М.: Физматлит, 2013. Т. 1. С. 86–103.
2. Зиммель Г. Конфликт современной культуры / Пер. с нем. — Птг.: Начатки знаний, 1923. 40 с. (*Simmel G. Der Konflikt der modernen Kultur.* — Munchen: Duncker & Humblot, 1918. 60 p.)
3. Колесников А. В., Кириков И. А., Листопад С. В. Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор. — М.: ИПИ РАН, 2014. 189 с.
4. Shaw M. Group dynamics: The psychology of small group behavior. — New York, NY, USA: McGraw-Hill, 1981. 531 p.
5. Brown R. Group processes. — 2nd ed. — Oxford: Wiley-Blackwell, 1996. 442 p.
6. Андреева Г. М. Социальная психология. — М.: Аспект-пресс, 2009. 393 с.
7. Small groups and social interaction / Eds. H. Blumberg, A. Hare, V. Kent, M. Davies. — Chichester: Wiley, 1983. Vol. 1. 478 p.
8. Сидоренков А. В. Конфликт в малой группе: понятие, функции, виды и модель // Северо-Кавказский психологический вестник, 2008. Т. 6. № 4. С. 22–28.
9. Гришина Н. В. Психология конфликта. — 2-е изд. — СПб.: Питер, 2008. 544 с.
10. Емельянов С. М. Практикум по конфликтологии. — 3-е изд. — СПб.: Питер, 2009. 384 с.
11. Анцупов А. Я., Баклановский С. В. Конфликтология в схемах и комментариях. — 2-е изд. — СПб.: Питер, 2009. 304 с.
12. Робер М.-А., Тильман Ф. Психология индивида и группы / Пер. с фр. — М.: Прогресс, 1988. 256 с. (*Robert M. A., Tilman F. Psycho. Connaitre l'individu et le groupe aujourd'hui.* — Bruxelles: Vie ouvriere, 1980. 307 p.)
13. Epstein J. M., Steinbruner J. D., Parker M. T. Modeling civil violence: An agent-based computational approach // P. Natl. Acad. Sci. USA, 2002. Vol. 99. Suppl. 3. P. 7243–7250. doi: 10.1073/pnas.092080199.
14. Новиков Д. А. Иерархические модели военных действий // Управление большими системами, 2012. № 37. С. 25–62.
15. Клаус Н. Г., Свечкарев В. П. Агентные модели локальных этнических конфликтов (на примере осетино-ингушского конфликта в селе Тарское) // Инженерный вестник Дона, 2013. № 4. С. 138. <http://www.ivdon.ru/magazine/archive/n4y2013/1983>.
16. Клаус Н. Г., Свечкарев В. П. Многоагентное моделирование конфликтных ситуаций. — Ростов на Дону: СКНЦВШ ЮФУ, 2013. 148 с.
17. Саати Т. Л. Математические модели конфликтных ситуаций / Пер. с англ. под ред. И. А. Ушакова. — М.: Сов. радио, 1977. 302 с. (*Saaty T. Mathematical models of arms control and disarmament: Application of mathematical structures in politics.* — John Wiley&Sons, 1968. 190 p.)

18. Покорная О. Ю., Покорная И. Ю., Прядкин Д. В. Математическое моделирование оптимальных стратегий в условиях конфликта // Молодой ученый, 2011. № 4-1. С. 16–19. <https://moluch.ru/archive/27/2896>.
19. Tibshirani R. Regression shrinkage and selection via the Lasso // J. Roy. Stat. Soc. B Met., 1996. Vol. 58. No. 1. P. 267–288.
20. Корнелиус Х., Фейр Ш. Выиграть может каждый. Как разрешать конфликты / Пер. с англ. — Киев: Наукова думка, 2006. 344 с. <http://www.conflict-resolve.org/conflictresolve.pdf>. (Cornelius H., Faire S. Everyone can win: How to resolve conflict. — West Roseville, New South Wales, Australia: Simon & Schuster, 1989. 192 p.)
21. Мириманова М. С. Конфликтология. — 2-е изд. — М.: Академия, 2004. 320 с.
22. Ковалев В. Н., Лопатина Н. Н., Лебедеко Д. Н. Сборник ситуационных задач по конфликтологии. — Севастополь: СФ МГУ, 2013. 91 с.
23. Готин С. В., Калоша Л. П. Логико-структурный подход и его применение для анализа и планирования деятельности. — М.: Вариант, 2007. 118 с.
24. Деттмер У. Теория ограничений Голдратта: Системный подход к непрерывному совершенствованию / Пер. с англ. — 4-е изд. — М.: Альпина Паблшер, 2014. 443 с. <https://econ.wikireading.ru/64894>. (Dettmer W. Goldratt's theory of constraints: A systems approach to continuous improvement. — ASQ Quality Press, 2006. 377 p.)

Поступила в редакцию 11.06.19

METHODS OF MODELING AND VISUAL REPRESENTATION OF A CONFLICT IN A SMALL COLLECTIVE OF EXPERTS SOLVING PROBLEMS (REVIEW)

S. B. Rumovskaya and I. A. Kirikov

Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236022, Russian Federation

Abstract: Small collectives of experts as natural collective decision support intellect (heterogeneous collective) solve problems effectively. In addition, the form of interaction between experts as conflict generates positive changes in collective such as development of the group, diagnostics of relations, tension reduction, and consolidation of the group and inspire saving the collective. Preset sketches of standard situations play a huge role in human reasoning. The use of them highly promote reasoning. Visualization of conflict situation makes appeared contradictions contrast and observable, giving a new information of resolving the conflict. This makes them easy-to-handle and gives the opportunity to control the impact on the conflict of subjective preference. The notion and particularities of the conflict in small collectives, its structure, dynamics, and approaches to modelling and visual presentation of the conflict aspect in group dynamics of experts solving problems are reviewed.

Keywords: small collective of experts; conflict; model of a conflict; visualization of a conflict

DOI: 10.14357/19922264190317

References

1. Kolesnikov, A. V. 2013. Geterogennyye estestvennyye i iskusstvennyye sistemy [Natural and artificial heterogeneous systems]. *Integrirovannyye modeli i myagkie vychisleniya v iskusstvennom intellekte* [Integrated models and soft computing in artificial intelligence]. Moscow: Fizmatlit. 1:86–103.
2. Simmel, G. 1918. *Der Konflikt der modernen Kultur*. Munchen: Duncker & Humblot. 60 p.
3. Kolesnikov, A. V., I. A. Kirikov, and S. V. Listopad. 2014. *Gibridnyye intellektual'nyye sistemy s samoorganizatsiyey: koordinatsiya, soglasovannost', spor* [Hybrid artificial systems with self-organization: Coordination, conformance, row]. Moscow: IPI RAN. 189 p.
4. Shaw, M. 1981. *Group dynamics: The psychology of small group behavior*. New York, NY: McGraw-Hill. 531 p.
5. Brown, R. 1996. *Group processes*. 2nd ed. Oxford: Wiley-Blackwell. 442 p.
6. Andreeva, G. M. 2009. *Sotsial'naya psikhologiya* [Social psychology]. Moscow: Aspect-press. 393 p.
7. Blumberg, H., A. Hare, V. Kent, and M. Davies, eds. 1983. *Small groups and social interaction*. Chichester: Wiley. Vol. 1. 478 p.
8. Sidorenkov, A. V. 2008. Konflikt v maloy gruppe: ponyatie, funktsii, vidy i model' [Conflict in a small group: Concept, functions, forms and model]. *Severo-Kavkazskiy psikhologicheskiy vestnik* [North Caucasian Psychological Annals] 6(4):22–28.

9. Grishina, N. V. 2008. *Psikhologiya konflikta* [Psychology of conflict]. SPb.: Piter. 544 p.
10. Emel'yanov, S. M. 2009. *Praktikum po konfliktologii* [Tutorial at conflictology]. SPb.: Piter. 384 p.
11. Antsupov, A. Ya., and S. V. Baklanovskiy. 2009. *Konfliktologiya v skhemakh i kommentariyakh* [Conflictology in schemes and comments]. SPb.: Piter. 304 p.
12. Robert, M. A., and F. Tilman. 1980. *Psycho. Connaître l'individu et le groupe aujourd'hui*. Bruxelles: Vie ouvriere. 307 p.
13. Epstein, J. M., J. D. Steinbruner, and M. T. Parker. 2002. Modeling civil violence: An agent-based computational approach. *P. Natl. Acad. Sci. USA* 99(Suppl. 3):7243–7250. doi: 10.1073/pnas.092080199.
14. Novikov, D. A. 2012. Ierarkhicheskie modeli voennykh deystviy [Hierarchical models of combat]. *Upravlenie bol'shimi sistemami* [Control of Large Systems] 37:25–62.
15. Klaus, N. G., and V. P. Svechkarev. 2013. Agentnye modeli lokal'nykh etnicheskikh konfliktov (na primere ose-tino-ingushskogo konflikta v sele Tarskoe) [Agent based modeling of the social conflict: Ossetino-Ingushskiy conflict in the area of village Tarskoe]. *Inzhenernyy vestnik Dona* [Engineering Annals of Don] 4:138. Available at: <http://www.ivdon.ru/magazine/archive/n4y2013/1983/> (accessed May 27, 2019).
16. Klaus, N. G., and V. P. Svechkarev. 2013. *Mnogoagentnoe modelirovanie konfliktnykh situatsiy* [Multiagent modeling of conflict situations]. Rostov-on-Don: NCSC HS SFEDU. 148 p.
17. Saaty, T. 1968. *Mathematical models of arms control and disarmament: Application of mathematical structures in politics*. John Wiley&Sons. 190 p.
18. Pokornaya, O. Yu., I. Yu. Pokornaya, and D. V. Pryadkin. 2011. Matematicheskoe modelirovanie optimal'nykh strategiy v usloviyakh konflikta [Mathematical modeling of optimal strategies in conflict]. *Molodoy uchenyy* [Young Scientist] 4-1:16–19. Available at: <https://moluch.ru/archive/27/2896/> (accessed May 27, 2019).
19. Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *J. Roy. Stat. Soc. B Met.* 58(1):267–288.
20. Cornelius, H., and S. Faire. 1989. *Everyone can win: How to resolve conflict*. West Roseville, New South Wales, Australia: Simon & Schuster. 192 p.
21. Mirimanova, M. S. 2004. *Konfliktologiya* [Conflictology]. Moscow: Academy. 320 p.
22. Kovalev, V. N., N. N. Lopatina, and D. N. Lebedenko. 2013. *Sbornik situatsionnykh zadach po konfliktologii* [The collection of situational tasks on conflictology]. Sevastopol: SB MSU. 91 p.
23. Gotin, S. V., and L. P. Kalosha. 2007. *Logiko-strukturnyy podkhod i ego primenenie dlya analiza i planirovaniya deyatelnosti* [Logical-structural approach and its application for the analysis and planning of activities]. Moscow: Variant. 118 p.
24. Dettmer, W. *Goldratt's theory of constraints: A systems approach to continuous improvement*. ASQ Quality Press 2006. 377 p. Available at: <https://econ.wikireading.ru/64894> (accessed May 27, 2019).

Received June 11, 2019

Contributors

Rumovskaya Sophiya B. (b. 1985) — Candidate of Sciences (PhD) in technology, scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236022, Russian Federation; sophiyabr@gmail.com

Kirikov Igor A. (b. 1955) — Candidate of Sciences (PhD) in technology; director, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236022, Russian Federation; baltbipiran@mail.ru

РАЗРАБОТКА МЕТОДА ФОРМИРОВАНИЯ ПРИЗНАКОВОГО ПРОСТРАНСТВА И МОДЕЛИ ДЛЯ ОЦЕНКИ И ПРОГНОЗИРОВАНИЯ АНТРОПОГЕННОГО ВЛИЯНИЯ НА ОКРУЖАЮЩУЮ СРЕДУ (НА ПРИМЕРЕ ЛЕСНОГО ФОНДА НЕФТЕДОБЫВАЮЩЕГО РЕГИОНА)*

В. В. Бурлуцкий¹, А. В. Якимчук², А. В. Мельников³, А. Л. Царегородцев⁴,
С. В. Волошин⁵

Аннотация: Работа посвящена разработке системного метода оценки и прогнозирования влияния природных и антропогенных воздействий на окружающую среду, включающего процедуры преобразования исходных информационных массивов, формирования нейросетевой модели, ее обучения и тестирования. Метод применен для анализа последствий антропогенных воздействий на окружающую среду в Ханты-Мансийском автономном округе — Югре.

Ключевые слова: анализ данных; машинное обучение; нейронные сети; пространственный анализ; географические информационные системы; риск-ориентированный подход

DOI: 10.14357/19922264190318

1 Введение

Проблема формирования признакового пространства, математической модели и подходящих алгоритмов машинного обучения для оценки и прогнозирования влияния природных и антропогенных факторов на окружающую среду весьма актуальна как с теоретической, так и с прикладной точки зрения. В особенности решение этой проблемы важно для нефтедобывающих регионов, так как производственные технологии в них сопряжены с загрязнением окружающей среды.

Первая часть исходной проблемы — формирование признакового пространства — связана с доступными количественными показателями техногенных аварий [1–3]. Здесь важной является трансформация исходных информационных массивов в цифровые характеристики риска влияния антропогенных факторов.

Вторая и третья части исходной проблемы — модель и алгоритм обучения — связаны друг с другом и сводятся к задачам классификации, в которых используется какая-либо нейросетевая модель в сочетании с подходящим алгоритмом обучения. В настоящее время разработано большое число различных алгоритмов классификации: метод k ближайших соседей [4], случайный лес [5], стохастиче-

ский градиентный спуск [6], метод опорных векторов [7].

В данной работе для решения поставленной проблемы предлагается метод трансформации исходных разнородных динамически организованных информационных массивов, полученных путем дистанционного зондирования земной поверхности с использованием средств географических информационных систем (ГИС).

На основе преобразованных информационных массивов формируется трехслойная нейронная сеть и проводится ее обучение и тестирование. Обученная сеть используется для прогнозирования последствий природных и антропогенных воздействий на окружающую среду.

Разработанные методы были применены для оценивания и прогнозирования антропогенных рисков на территории Ханты-Мансийского автономного округа.

2 Структуризация и предварительная обработка разнородных данных

Все единицы информационных массивов привязаны к ячейкам поверхностной сетки. Для каждо-

* Работа выполнена при поддержке Научного фонда ЮГУ (проект 13-01-20/25).

¹ Югорский НИИ информационных технологий, г. Ханты-Мансийск, BurlutskyVV@uriit.ru

² Югорский государственный университет, г. Ханты-Мансийск, YakimchukAV@uriit.ru

³ Югорский НИИ информационных технологий, г. Ханты-Мансийск, andmelnikov1956@yandex.ru

⁴ Югорский НИИ информационных технологий, г. Ханты-Мансийск, TsaregorodtsevAL@uriit.ru

⁵ Югорский государственный университет, г. Ханты-Мансийск, Voloshinsv@uriit.ru

го признака набора данных была определена шкала измерения с учетом метода обработки: текстовые признаки были заменены на числовые, непрерывные значения были нормализованы по максимальному значению.

Ключевым атрибутом набора данных служат поверхностные координаты соответствующего информационного объекта.

Для проверки корректности заполнения базы использовались методы визуального анализа: гистограммы и картографический анализ.

Для визуального анализа использовался boxplot, компактно изображающий распределение. Такой вид диаграммы в удобной форме показывает медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

3 Модель оценки рисков

В результате проведенного анализа была построена 4-слойная нейросетевая модель оценки рисков [8]. В качестве данных для входного слоя использовалась векторизация признаков, что позволило существенно сократить число входных нейронов.

Число нейронов во втором слое составило 166, в третьем — 83, в четвертом — 40, и выходной слой был с двумя нейронами, которые фиксировали положительную и отрицательную реакцию модели.

Информационные массивы были разделены на три части, которые использовались для обучения модели, тестирования и прогнозирования.

4 Применение разработанного метода для оценки и прогнозирования рисков антропогенных факторов в нефтедобывающем регионе

В качестве подходящего полигона был избран Ханты-Мансийский автономный округ — Югра, который лидирует среди нефтедобывающих регионов по добыче нефти и газа. При этом более 95% территории автономного округа представляют собой земли лесного фонда, которые находятся под постоянным негативным воздействием предприятий нефтегазового комплекса.

Для построения модели прогнозирования рисков загрязнения использовались данные об авариях и инцидентах, транспортной инфраструктуре, населенных пунктах [9]. Вся информация привязывалась к 5-километровой поверхностной сетке, и процесс интеграции данных заключался в унификации географических координат различных объектов, построении общих классификаторов, удалении дублирующей информации и приведении к единому виду текстовых признаков.

В результате был получен первичный массив данных о загрязнениях окружающей среды за период 2012–2017 гг. Объем данной информативной выборки составил 26 015 записей. Общее число элементарных участков составило 22 054.

Рисунок 1 изображает распределение аварий по районам, а также показывает медиану (или, ес-

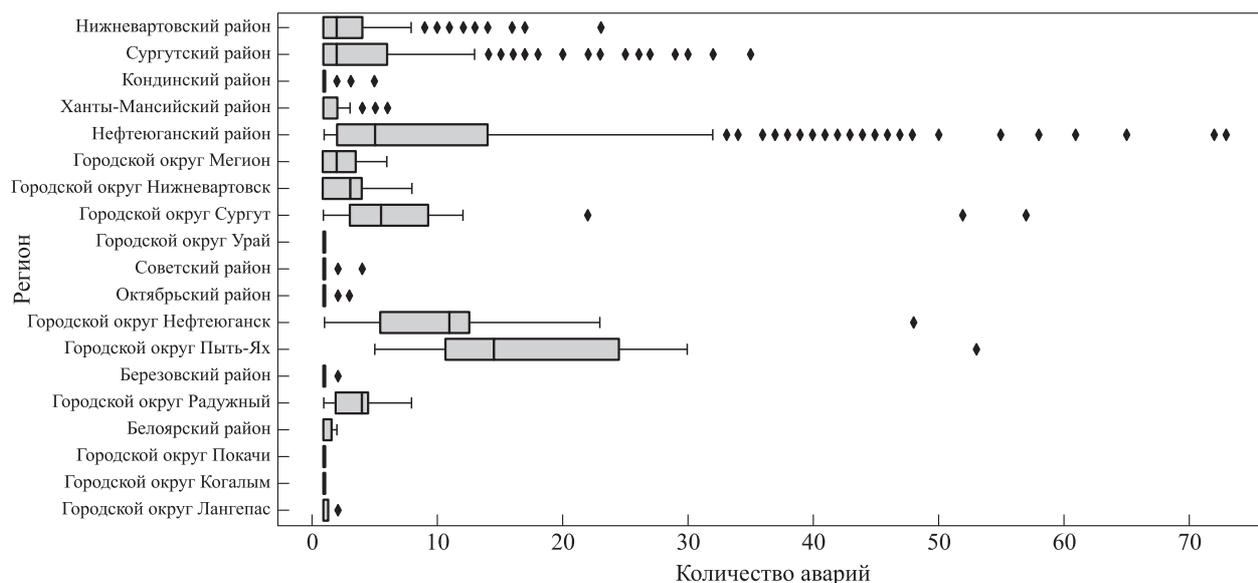


Рис. 1 Распределение аварий по районам

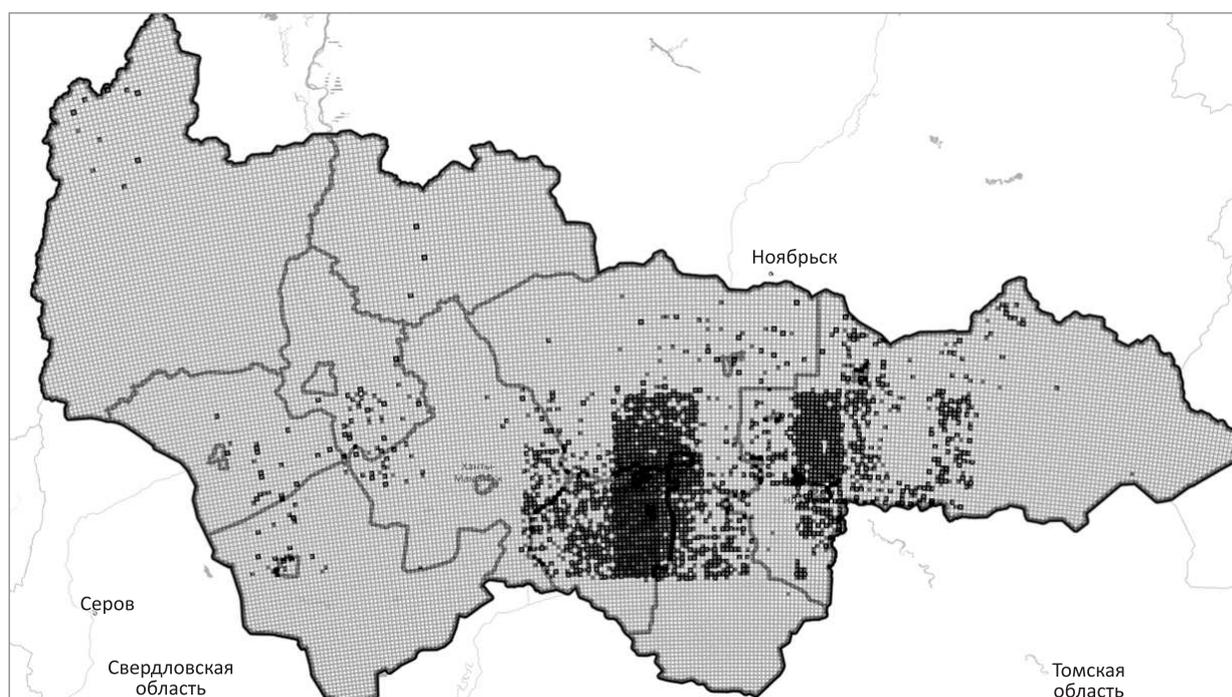


Рис. 2 Визуализация исходного набора данных

ли нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

Объекты, привязанные к элементарным участкам, представлены на рис. 2.

На следующем этапе набор данных был проанализирован на наличие неинформативных признаков, для этого использовался метод главных компонент.

В результате проведенного анализа и предобработки данных был получен базовый набор данных, объем которого составил 22 054.

Для тестирования модели использовался метод кросс-валидации по k блокам [10]. Исходный набор данных разбивался на 10 одинаковых по размеру блоков. Из 10 блоков один оставлялся для тестирования модели, а оставшиеся 9 блоков использовались как тренировочный набор. Процесс повторялся 10 раз, и каждый из блоков один раз использовался как тестовый набор. В итоге получилось 10 результатов, по одному на каждый блок, они усреднялись или комбинировались каким-либо другим способом и дали одну оценку. Результаты представлены в таблице.

Средняя оценка составила 90,68.

Преимущество такого способа перед случайным сэмплением в том, что все наблюдения используются и для тренировки, и для тестирования

Оценка методом кросс-валидации

Номер модели	Полученная оценка
1	91,5
2	90,3
3	90,8
4	91,2
5	89,9
6	90,6
7	91,4
8	89,5
9	91,2
10	90,4

модели, при этом каждое наблюдение используется для тестирования только один раз.

Получившаяся карта с прогнозированными авариями представлена на рис. 3.

5 Заключение

Разработан системный метод оценки и прогнозирования природных и антропогенных воздействий на окружающую среду, включающий процедуры преобразования исходных информационных массивов, формирования нейросетевой прогнозирующей модели, ее обучения и тестирования. Метод применен для прогнозирования техногенных аварий в нефтедобывающем регионе — Югре.

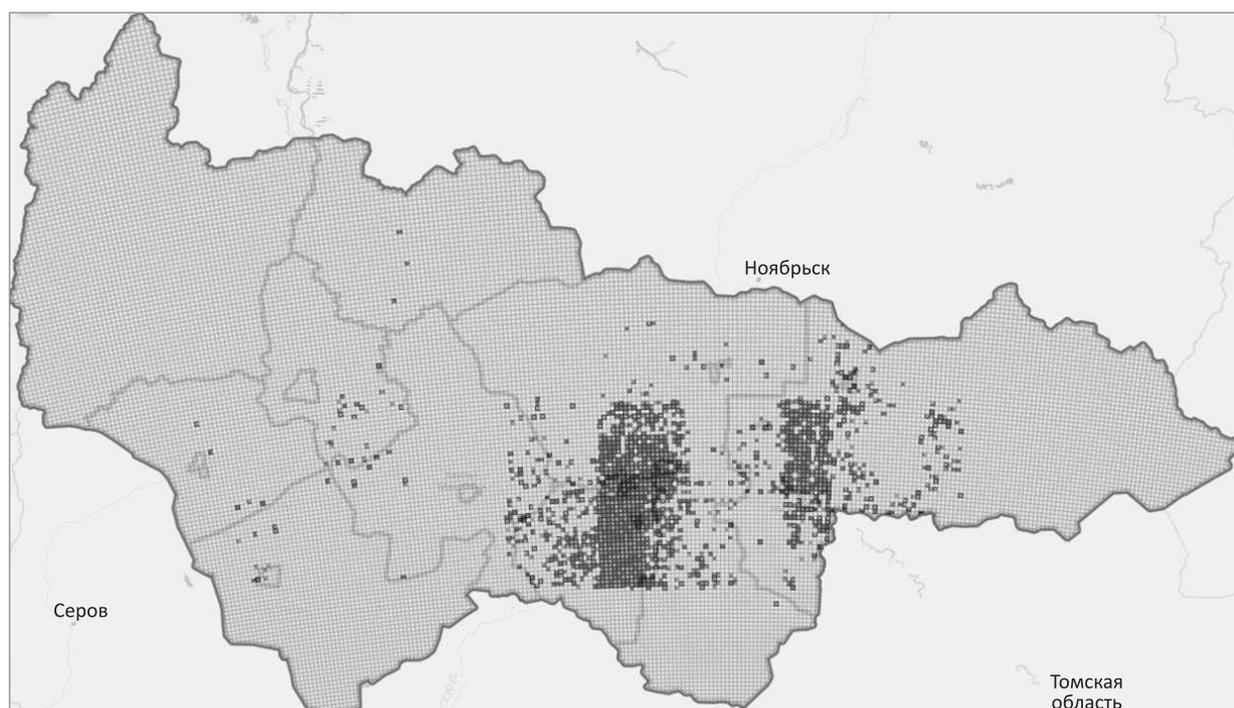


Рис. 3 Визуализация прогноза аварий

Литература

1. Guikema S. D. Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory // *Reliab. Eng. Syst. Safe.*, 2009. Vol. 94. Iss. 4. P. 855–860. doi: 10.1016/j.ress.2008.09.003.
2. Шокин Ю. И., Москвичев В. В., Ничепорчук В. В. Методика оценки антропогенных рисков территорий и построения картограмм рисков с использованием геоинформационных систем // *Вычислительные технологии*, 2010. Т. 15. № 1. С. 120–131.
3. Гуменюк В. И., Кармишин А. М., Киреев В. А. О количественных показателях опасности техногенных аварий // *Научно-технические ведомости СПбГПУ*, 2013. № 2(171). С. 281–288.
4. Колесенков А. Н., Костров Б. В., Ручкин В. Н. Нейронные сети мониторинга чрезвычайных ситуаций по данным ДЗЗ // *Известия Тульского государственного университета. Технические науки*, 2014. № 5. С. 220–225.
5. Stone C. J. Consistent nonparametric regression // *Ann. Stat.*, 1977. Vol. 5. Iss. 4. P. 595–620.
6. Breiman L. Random forests // *Mach. Learn.*, 2001. Vol. 45. Iss. 1. P. 5–32.
7. Robbins H., Siegmund D. A convergence theorem for non negative almost supermartingales and some applications // *Optimizing methods in statistics* / Ed. J. S. Rustagi. — New York, NY, USA: Academic Press, 1971. P. 233–257.
8. Cortes C., Vapnik V. Support-vector networks // *Mach. Learn.*, 1995. Vol. 20. Iss. 3. P. 273–297.
9. Hong H., Pradhan B., Jebur M. N. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines // *Environ. Earth Sci.*, 2016. Vol. 75. Iss. 1. P. 40.
10. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection // *14th Joint Conference (International) on Artificial Intelligence Proceedings*. — Montréal, Québec, Canada, 1995. Vol. 2. P. 1137–1143.

Поступила в редакцию 02.10.18

DEVELOPMENT OF A METHOD FOR THE FORMATION OF ATTRIBUTE SPACE AND A MODEL FOR THE ASSESSMENT AND PREDICTION OF ANTHROPOGENIC INFLUENCE ON THE ENVIRONMENT (ON THE EXAMPLE OF THE FOREST FUND OF THE OIL-PRODUCING REGION)

V. V. Burlutskiy¹, A. V. Yakimchuk², A. V. Melnikov¹, A. L. Tsaregorodtsev¹, and S. V. Voloshin²

¹Yugra Research Institute of Information Technologies, 151 Mira Str., Khanty-Mansiysk 628011, Russian Federation

²Yugra State University, 16 Chekhova Str., Khanty-Mansiysk 628012, Russian Federation

Abstract: The work is devoted to the development of a systematic method for assessing and predicting the influence of natural and anthropogenic impacts on the environment, including the procedures for the transformation of initial data store, the formation of the neural network model, its training, and testing. The method is used to analyze the consequences of anthropogenic impacts on the environment in the Khanty-Mansiysk Autonomous Okrug — Yugra.

Keywords: data analysis; machine learning; neural networks; spatial analysis; geographic information systems; risk-based approach; control and supervision

DOI: 10.14357/19922264190318

Acknowledgments

This work was supported by the Science Foundation of Yugra State University under grant No. 13-01-20/25.

References

1. Guikema, S. D. 2009. Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. *Reliab. Eng. Syst. Safe.* 94(4):855–860. doi: 10.1016/j.res.2008.09.003.
2. Shokin, Y. I., V. V. Moskvichev, and V. V. Nicheporchuk. 2010. Metodika otsenki antropogennykh riskov territoriy i postroeniya kartogramm riskov s ispol'zovaniem geoinformatsionnykh sistem [Technique for estimation of anthropogenous risks for territories and construction of risks cartograms using geoinformation systems]. *Computational Technologies* 15(1):120–131.
3. Gumenyuk, V. I., A. M. Karmishin, and V. A. Kireev. 2013. O kolichestvennykh pokazatelyakh opasnosti tekhnogenykh avariyy [About quantitative indicators of danger man-made accidents]. *Nauchno-tekhnicheskie vedomosti SPbPU* [St. Petersburg State Polytechnic University J. Engineering Science Technology] 2(171):281–288.
4. Kolesenkov, A. N., B. V. Kostrov, and V. N. Ruchkin. 2014. Neyronnye seti monitoringa chrezvychaynykh situatsiy po dannym DZZ [Neural network monitoring for emergencies according ERS]. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki* [Transactions of Tula State University. Technical Sciences] 5:220–225.
5. Stone, C. J. 1977. Consistent nonparametric regression. *Ann. Stat.* 5(4):595–620.
6. Breiman, L. 2001. Random forests. *Mach. Learn.* 45(1):5–32.
7. Robbins, H., and D. Siegmund. 1971. A convergence theorem for non negative almost supermartingales and some applications. *Optimizing methods statistics*. Ed. J. S. Rustagi. New York, NY: Academic Press. 233–257.
8. Cortes, C., and V. Vapnik. 1995. Support-vector networks. *Mach. Learn.* 20(3):273–297.
9. Hong, H., B. Pradhan, and M. N. Jebur. 2016. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* 75(1):40.
10. Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *14th Joint Conference (International) on Artificial Intelligence Proceedings*. Montréal, Québec, Canada. 2:1137–1143.

Received October 2, 2018

Contributors

Burlutskiy Vladimir V. (b. 1975) — Candidate of Science (PhD) in physics and mathematics, Head of the Center of Information and Analytical Systems, Yugra Research Institute of Information Technologies, 151 Mira Str., Khanty-Mansiysk 628011, Russian Federation; BurlutskyVV@uriit.ru

Yakimchuk Aleksandr V. (b. 1994) — PhD student, Yugra State University, 16 Chekhova Str., Khanty-Mansiysk 628012, Russian Federation; YakimchukAV@uriit.ru

Melnikov Andrey V. (b. 1956) — Doctor of Science in technology, professor, Director, Yugra Research Institute of Information Technologies, 151 Mira Str., Khanty-Mansiysk 628011, Russian Federation; andmelnikov1956@yandex.ru

Tsaregorodtsev Aleksandr L. (b. 1979) — Candidate of Science (PhD) in technology, First Deputy Director, Yugra Research Institute of Information Technologies, 151 Mira Str., Khanty-Mansiysk 628011, Russian Federation; TsaregorodtsevAL@uriit.ru

Voloshin Semen V. (b. 1991) — PhD student, Yugra State University, 16 Chekhova Str., Khanty-Mansiysk 628012, Russian Federation; voloshinsv@uriit.ru

НАУЧНЫЙ РЕЗУЛЬТАТ КАК ИНФОРМАЦИОННЫЙ ОБЪЕКТ В КОНТЕКСТЕ СИСТЕМЫ УПРАВЛЕНИЯ НАУЧНЫМИ СЕРВИСАМИ*

А. П. Сучков¹

Аннотация: Обсуждается проблема формализации одного из важнейших понятий для системы научных сервисов — научный результат, который является квинтэссенцией всех научных исследований и основным информационным объектом для процессов постановки научной проблемы, формулировки научных гипотез, мониторинга научно-технической информации. Формализованное понятие «научный результат», интегрирующее всю информационную структуру научного исследования, используется при организации значительного числа научных сервисов: сервисов извлечения фактов и знаний (извлечение фактов, понятий, связей и формализация фактографических данных на основе лингвистического анализа слабоструктурированной информации); интеллектуального поиска информации; тематического индексирования; наукометрического анализа; анализа фронта исследований; сервисов коммуникаций (сервисов работы с научно-технической информацией (НТИ) и коммуникаций в научном сообществе; интеллектуального анализа специализированных социальных сетей и других средств научных коммуникаций); проверки заимствований. Особенно это актуально для междисциплинарных исследований, где обеспечивается поиск и подбор релевантных научных положений и инструментария в смежных областях научного знания.

Ключевые слова: научный результат; информационная модель; научные сервисы; алгоритмы идентификации; интеллектуальный поиск; междисциплинарные исследования

DOI: 10.14357/19922264190319

1 Введение

В настоящий момент развитие народного хозяйства все больше связывается с внедрением подходов и технологий цифрового развития. Российская наука обладает всеми необходимыми компонентами для рассмотрения ее в качестве отрасли цифровой экономики: наличие развитой вычислительной и телекоммуникационной инфраструктуры, системы центров коллективного пользования, других организационных структур, нормативной базы, высокого уровня компетенций и высококвалифицированных научных коллективов. При этом все перечисленные компоненты объективно продвигаются по пути «цифровизации», и наука как отрасль экономики также становится «цифровой».

Другим не менее важным обстоятельством является объективная смена подходов к проведению научных исследований, а именно: переход к новой парадигме в научных исследованиях, основанной на анализе накопленных данных в конкретных предметных областях, естественно, в формализованном цифровом виде. Проведение таких исследований

становится неотъемлемой частью различных областей науки, экономики, бизнеса на основе инструментария интенсивного использования данных.

Научные и образовательные организации России обладают широкой, распределенной по территории страны сетью центров коллективного пользования (ЦКП) и уникальных научных установок (УНУ), обладающих колоссальным спектром научных услуг в различных областях науки [1]. Необходима систематизация этих услуг и повышение эффективности их использования на основе создания современной исследовательской инфраструктуры, которая предоставляла бы широкий спектр возможностей по научным сервисам не только для научных организаций, но и для внешних пользователей.

В ФИЦ ИУ РАН на основе платформенного подхода создается прототип подобной инфраструктуры, который представляет собой совокупность трех компонентов. Первый — центр компетенций, в котором концентрируются знания в конкретной области. Второй — материально-техническая среда в виде высокопроизводительного центра обработки

* Работа выполнена при частичной поддержке РФФИ (проект 18-29-03091).

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

данных с гибридной архитектурой. Наконец, третий — это совокупность научных сервисов, которые создаются и накапливаются на этой платформе научными подразделениями ФИЦ ИУ РАН [2].

Успешное продвижение цифровизации научных исследований во многом обусловлено возможностью создания эффективной системы широкодоступных научных сервисов.

Научный сервис — совокупность действий (процессов) и средств обеспечения процессов (ресурсов) по обслуживанию выполнения конкретных работ и реализации проектов научно-исследовательского и прикладного характера путем сервисной деятельности и предоставления потребителю (исследователям, специалистам или организациям) оборудования, расходных материалов, информационно-коммуникационных ресурсов, обеспечивающих ресурсы, продуктов интеллектуальной научной деятельности и обслуживающих человеческих ресурсов (субъекты сервисной деятельности). Результатом сервисной деятельности является услуга [3]. Очевидно, что научный сервис опирается как на цифровые технологии (автоматическое и автоматизированное предоставление услуг), так и на использование интеллектуальных и обслуживающих человеческих ресурсов. Система научных сервисов должна обеспечивать поддержку процессов автоматизированного подбора релевантных сервисов и осуществления различных формальных и неформальных коммуникаций исследователя и государства, научного сообщества и бизнеса.

Научным коллективом ФИЦ ИУ РАН в рамках проводимых исследований обоснованы [3] концептуальные и системотехнические подходы к созданию системы управления научными сервисами (СУС), которые обеспечиваются ЦКП и УНУ академических институтов.

В статье обсуждается проблема формализации одного из важнейших понятий для системы научных сервисов — научный результат, который является квинтэссенцией всех научных исследований и основным информационным объектом для процессов постановки научной проблемы, формулировки научных гипотез, мониторинга НТИ. Формализованное понятие «научный результат», интегрирующее всю информационную структуру научного исследования, используется при организации значительного числа научных сервисов: сервисов извлечения фактов и знаний (извлечение фактов, понятий, связей и формализация фактографических данных на основе лингвистического анализа слабоструктурированной информации); интеллектуального поиска информации; тематического индексирования; наукометрического анализа; анализа фронта исследований; сервисов комму-

никаций (сервисов работы с НТИ и коммуникаций в научном сообществе; интеллектуального анализа специализированных социальных сетей и других средств научных коммуникаций); проверки заимствований, антиплагиата. Особенно это актуально для междисциплинарных исследований, где обеспечивается поиск и подбор релевантных научных положений и инструментария в смежных областях научного знания.

Другим важнейшим вопросом, связанным с формализацией данного понятия, представляется постановка проблемы создания реестра научных результатов, который мог бы стать ключевым ресурсом системы научных сервисов. В связи с этим также рассматриваются задачи и алгоритмы идентификации научных результатов на построенной информационной модели данного понятия.

2 Информационная структура процессов научного исследования

В современном представлении научное исследование — взаимоувязанная совокупность процессов, включающая ряд типовых компонентов [4, 5]:

- формулировка научной проблемы;
- предварительный анализ доступной НТИ (факты, теории, гипотезы);
- формулировка и сравнительный анализ исходных гипотез;
- планирование научных исследований;
- организация и проведение эксперимента;
- анализ и обобщение полученных результатов;
- проверка исходных гипотез, принятие решений;
- формулирование фактов и положений, их обоснование и описание (получение продукта знаний).

Систематизация процессов научного исследования позволяет сформировать ориентировочный набор научных сервисов, который может лечь в основу системы цифровизации науки. В [5] обосновывается перечень и состав основных групп научных сервисов, которые могут быть использованы в процессах научных исследований:

- сервисы интеллектуального поиска информации и мониторинга НТИ;
- сервисы извлечения фактов и знаний;
- аналитические сервисы;

- сервисы коммуникаций;
- сервисы планирования научного исследования;
- сервисы доступа к услугам ЦКП и УНУ;
- сервисы подготовки публикаций.

Все эти группы научных сервисов с необходимостью должны опираться на формализованную информационную модель процессов научного исследования. Анализ информационной структуры научного исследования позволяет выделить ее основные информационные объекты. В их числе:

- объект исследования — материальная или мыслимая сущность, по поводу которой изучаются ее неизвестные свойства или сам факт существования;
- субъект исследования — ученый-исследователь;
- средства исследования — инструменты исследования (метод, алгоритм, модель, техническое устройство и т. п.);
- научный результат — обоснование неизвестных свойств или факта существования объекта исследования или средства исследования.

Центральным объектом такой модели служит понятие «результат научного исследования», который непосредственно взаимосвязан со всеми другими информационными объектами.

3 Формализация понятия «научный результат»

Законодательно закреплено следующее определение: научный и (или) научно-технический ре-

зультат — продукт научной и (или) научно-технической деятельности, содержащий новые знания или решения и зафиксированный на любом информационном носителе.

Понятие научных результатов подразделяется на два основных вида: во-первых, научных результатов — идей, сформулированных в виде научных положений, и, во-вторых, результатов методического (методологического) и предметного уровня — научного инструментария (методов, моделей, методик, экспериментальных установок), научных эффектов, результатов экспериментов, устройств, технических и организационных систем и др. [6].

Для формирования XML-модели понятия «Научный результат» возможно применение следующих структур данных (табл. 1).

ГРНТИ (<http://grnti.ru>) — Государственный рубрикатор НТИ (прежнее наименование — Рубрикатор ГАСНТИ) — представляет собой универсальную иерархическую классификацию областей знания, принятую для систематизации всего потока НТИ. На основе Рубрикатора построена система локальных (отраслевых, тематических, проблемных) рубрикаторов в органах НТИ.

Объект исследования определяется субъектом исследования и обычно дается в перечне ключевых слов публикации. Новое свойство объекта исследования формулируется на естественном языке и не может быть лингвистически формализовано в виде классификаторов, словарей терминов в силу своей новизны.

Так как перечень видов научных результатов формулируется впервые, проведено исследование

Таблица 1 Структура объекта «Научный результат»

Структуры/подструктуры	Типы	Описание
нр:НаучныйРезультат	Корневой	Продукт научной и (или) научно-технической деятельности, содержащий новые знания или решения и зафиксированный на любом информационном носителе
нр:ОбластьЗнаний	Массив, текст, по классификатору ГРНТИ	Область научной и (или) научно-технической деятельности, для междисциплинарных исследований — множественная
нр:ОбъектИсследования	Текстовый	Материальная или мыслимая сущность, по поводу которой изучаются ее неизвестные свойства или сам факт существования
нр:СредствоИсследования	По классификатору, текстовый	Инструменты исследования (метод, алгоритм, модель, техническое устройство и т. п.)
нр:СвойствоОбъектаИсследования	Текст	Открытое новое свойство объекта исследования
нр:ВидРезультата	Структура, текстовый	Иерархический классификатор научных результатов
нр:БиблиографическаяСсылка	Структура, текстовый	Библиографическое описание публикации по ГОСТ Р 7.0.100-2018

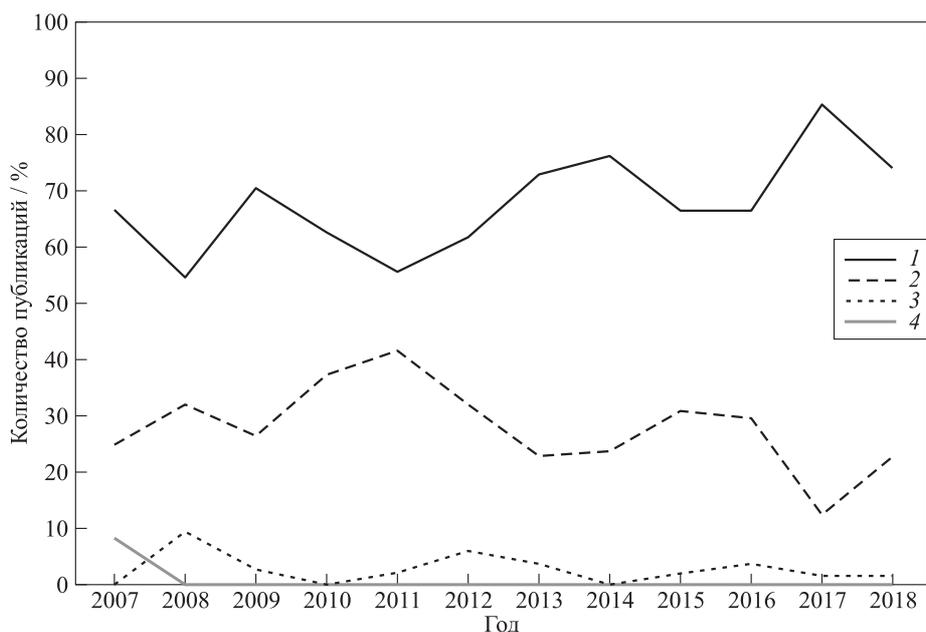


Рис. 1 Распределение публикаций по четырем видам научных результатов: 1 — инструмент; 2 — проблема; 3 — объект; 4 — обзор

корпуса публикаций научного журнала РАН [7] за период времени 2007–2018 гг. с целью обоснования их классификации. Общее число публикаций составило 534. Первоначально в перечень типов научных результатов было включено четыре типа (рис. 1) и анализ дал следующее распределение долей:

- научная проблема (0,2% публикаций);
- новые свойства объекта исследования (30%);
- инструмент исследования (69%);
- систематизация (2,8%).

Очевидно, что при такой типизации возникает существенный перекос в сторону изучения инстру-

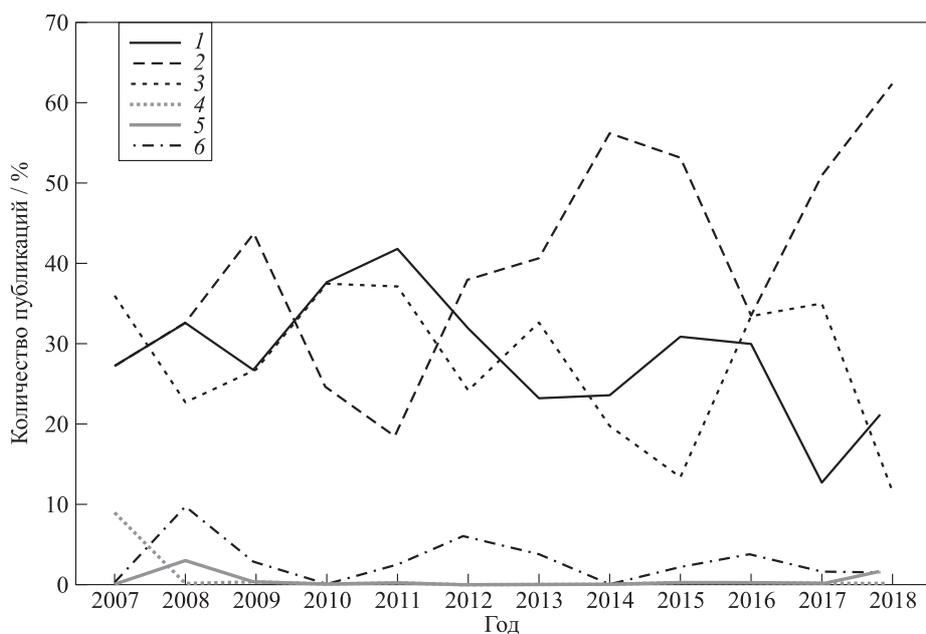


Рис. 2 Распределение публикаций по шести видам научных результатов: 1 — проблема; 2 — инструмент; 3 — объект&инструмент; 4 — новый объект; 5 — новые свойства; 6 — систематизация

Таблица 2 Структура объекта «Вид Научного Результата»

Структуры/подструктуры	Типы	Описание
вр: Вид Результата:	Структура, текстовый	Классификатор видов
Научное Положение: <i>Научная Проблема</i> <i>Новый Объект Исследования</i> <i>Новые Свойства Объекта Исследования</i> <i>Объект & Инструмент исследования</i>	Структура, текстовый	Научные результаты — идеи, сформулированные в виде научных положений
Научный Инструментарий: <i>Инструмент Исследования</i> <i>Систематизация</i>	Структура, текстовый	Методы, модели, методики, экспериментальные установки, а также научные эффекты, результаты экспериментов, технические и организационные системы. Систематизация включает в себя аналитические обзоры, классификации, результаты системного анализа

ментов исследования. Это объясняется тем, что в случае, когда объектом исследования служит научный инструментарий (методы, модели, методики, экспериментальные установки), зачастую получают и результаты о новых свойствах изучаемых материальных или мыслимых сущностей, т. е. научный результат содержит в себе оба вышеупомянутых вида. Поэтому тип «научный инструмент» был разделен на два типа: инструмент исследования как таковой и инструмент исследования с полученными результатами в отношении объекта исследования. Также тип «новые свойства объекта исследования» разделен на два: новый объект исследования и новые свойства объекта исследования. Повторный анализ корпуса публикаций дал следующее распределение их по предложенному перечню типов научных результатов (рис. 2):

- *научная проблема* (0,4%);
- *новый объект исследования* (0,2%);
- *новые свойства объекта исследования* (28%);
- *инструмент исследования* (42%);
- *объект & инструмент исследования* (26,6%);
- *систематизация* (2,8%).

Предложенный состав классификатора можно признать удовлетворительным, так как он, во-первых, позволил классифицировать все статьи из довольно представительного корпуса публикаций за 12 лет, во-вторых, позволил выявить даже такие редкие виды научных результатов, как постановка научной проблемы и систематизация и, в-третьих, позволил осуществить разумное разбиение самого обширного раздела публикации — изучение научного инструментария (до 70% публикаций). Формализованное описание классификатора представлено в табл. 2.

4 Алгоритмы идентификации научных результатов

Формализация обсуждаемого понятия позволяет очертить схему алгоритма процесса идентификации научного результата на основе мониторинга доступного массива НТИ. Научные сервисы, основанные на этом алгоритме, могут позволить решать следующие задачи: определение новизны полученного результата, подбор схожих результатов по различным критериям, поиск результатов исследований в смежных отраслях знаний, поиск научного инструментария. Можно выделить следующие примеры необходимых для этого источников данных, представляющих собой отечественные и зарубежные базы данных (БД) НТИ:

- публикации на английском языке ресурса arXiv.org (1 млн 250 тыс. документов) по естественным наукам;
- англоязычная Википедия (5 млн 164 тыс. статей);
- БД по американским патентам (USPTO с 2002 по 2016 гг.): 2 млн 965 тыс. патентов;
- БД по международным патентам (WIPO): 2 млн 384 тыс. патентов.

Другим значимым источником в этой области служат отечественные и зарубежные электронные реферативные БД:

- БД ВИНТИ РАН;
- электронный каталог ГПНТБ России;
- Scopus (SciVerse Scopus) — библиографический индекс, Elsevier;
- Web of Science — библиографический индекс, Thomson Reuters;

- NTIS (National Technical Information Service) — политехническая БД, U.S. Department of Commerce;
- Life Sciences Collection — БД по естественным наукам, Cambridge Scientific Abstracts;
- Biological Abstracts — БД по биологии, Biological Abstracts Inc.;
- Biotechnology & Bioengineering — БД по биотехнологии и биоинженерии, Cambridge Scientific Abstracts and Engineering Information Inc.;
- Chemical Engineering and Biotechnology Abstracts — БД по химической инженерии и биотехнологии, The Royal Society of Chemistry;
- Compendex Plus — БД по техническим наукам, Engineering Information Inc.;
- Corporate & Industry Research Reports (CIRR) — БД аналитических отчетов фирм, JA Micropublishing;
- Current Contents — еженедельное обозрение содержания научных периодических изданий, Institute for Scientific Information;
- Derwent Biotechnology Abstracts — БД по биотехнологии, Derwent Publications Ltd.;
- MEDLINE — БД по медицине и биологии, National Library of Medicine
- TOXLINE — БД по токсикологии, National Library of Medicine, Swedish Nat. Chemicals Inspectorate.

Не претендуя на полноту и окончательность этого перечня ресурсов, отметим их тематическое разнообразие, значительные объемы и необходимость наличия глубоких научных компетенций для эффективного использования таких данных. Так как понятие научного результата не полностью поддается формализации, а значительные объемы связанной с ним информации выражены на естественном языке и содержатся в плохо структурированных массивах НТИ, первостепенную роль в алгоритмах идентификации играют методы искусственного интеллекта и интеллектуального поиска. Существуют известные методы интеллектуального поиска, например [8], которые включают следующие:

- единый каталог документов — с помощью технологий каталогизации и категоризации система позволяет создать единый каталог документов из всех источников с понятной структурой и удобной навигацией;
- полнотекстовый поиск по содержанию и атрибутам — при формировании результатов обработки запроса поиск ключевых слов ведется в содержании документа, а также в значениях атрибутов (полей) учетной карточки документа;
- поиск с учетом морфологии — морфологический поиск позволяет найти ключевое слово в документах не только в строго заданном виде, но и во всех его морфологических формах (с учетом рода, числа и склонения по падежам);
- поиск с фасетными фильтрами — управление размером выборки документов с помощью группы из нескольких фильтров (фасетов), которые представляют различные характеристики (тип документа, автор, дата создания и др.);
- поиск с учетом словарей синонимов — с использованием словарей синонимов, а также данных о семантической близости слов, полученных с помощью методов дистрибутивной семантики;
- контекстный поиск — позволяет найти документы, содержащие ключевые слова, если они расположены не далее указанного расстояния друг от друга.

Алгоритм состоит из ряда последовательных шагов (рис. 3).

Во-первых, алгоритм опирается на единый реестр научных результатов, который используется и пополняется в ходе выполнения алгоритма.

Во-вторых, в ходе интерактивного диалога формируются запросы различной степени формализации о сути искомого научного результата, которые подаются на вход процедур интеллектуального поиска информации.

В-третьих, процедуры поиска формируют свои поисковые предписания в соответствии с особенностями заложенных алгоритмов искусственного интеллекта (морфологический и синтаксический разбор, синонимия и контекстный поиск).

В-четвертых, по этим предписаниям осуществляется поиск в подключенных источниках НТИ с использованием единого каталога документов. На выходе этих поисковых процедур выдается набор схожих научных результатов, ранжированных по степени релевантности. В случае пустого множества искомого результата признается новым и заносится в единый реестр.

Итак, обоснование структуры и состава информационно-модели понятия «научный результат», а также алгоритмов идентификации позволяет обеспечить новыми возможностями всю систему научных сервисов: интеллектуальный поиск информации; тематическое индексирование; наукометрический анализ; анализ фронта исследований; сервисы коммуникаций (сервисы работы с НТИ и коммуникаций в научном сообществе; интеллектуального анализа специализированных социальных сетей и других средств научных коммуникаций); проверку заимствований. Особенно это

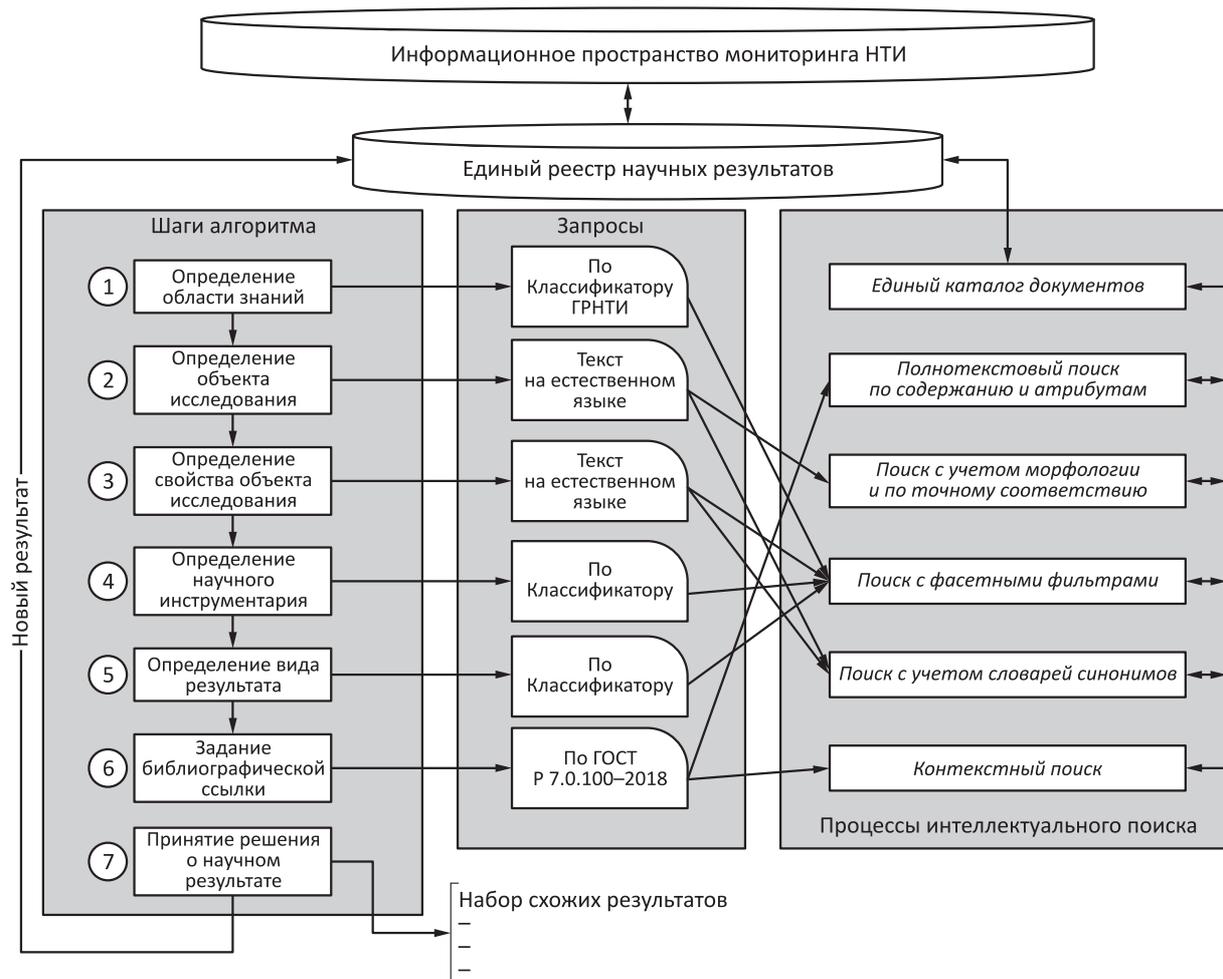


Рис. 3 Схема алгоритма идентификации научного результата

актуально для междисциплинарных исследований, где обеспечивается поиск и подбор релевантных научных положений и инструментария в смежных областях научного знания. Сформированная информационная модель может лечь в основу создания реестра научных результатов, который мог бы стать ключевым ресурсом системы научных сервисов.

Литература

1. Зацаринный А. А., Киселев Э. В., Козлов С. В., Колин К. К. Информационное пространство цифровой экономики России. Концептуальные основы и проблемы формирования / Под общ. ред. А. А. Зацаринного. — М.: ФИЦ ИУ РАН, 2018. 236 с.
2. Зацаринный А. А., Волович К. И., Кондрашев В. А. Методологические вопросы управления научными сервисами научных и образовательных организаций Российской Федерации // Радиолокация, навигация, связь: Сб. трудов XXIII Междунар. научн.-технич. конф. — Воронеж: Вэлборн, 2017. Т. 1. С. 7–14.
3. Исследование вопросов управления результатами научно-исследовательской деятельности организаций, подведомственных ФАНО России, и научными сервисами сети ЦКП ФАНО. — М.: ФИЦ ИУ РАН, 2016. Отчет о НИР «Сервис-У». 437 с.
4. Ушаков Е. В. Введение в философию и методологию науки. — М.: Экзамен, 2005. 528 с.
5. Зацаринный А. А., Кондрашев В. А., Сучков А. П. Система научных сервисов как актуальный компонент научных исследований // Системы и средства информатики, 2019. Т. 29. № 1. С. 23–38.
6. Понятие, сущность и классификация исследований. https://studopedia.ru/6_66034_ponyatie-sushchnost-i-klassifikatsiya-issledovaniy-rol-issledovaniy-v-razvitiimenedzhmenta.html.
7. Системы и средства информатики, 2007–2018.
8. Интеллектуальный поиск. <https://www.naumen.ru/products>.

Поступила в редакцию 30.05.19

THE SCIENTIFIC RESULT AS THE INFORMATION OBJECT IN THE CONTEXT OF THE SCIENTIFIC SERVICES SYSTEM MANAGEMENT

A. P. Suchkov

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 discusses the problem of formalization of one of the most important concepts for the system of scientific services — the scientific result, which is the quintessence of all scientific research and the main information object for the processes of formulation of scientific problems, formulation of scientific hypotheses, and monitoring of scientific and technical information (STI). The formalized concept of “scientific result” which integrates the entire information structure of scientific research is used in the organization of a significant number of scientific services: services of extraction of facts and knowledge (extraction of facts, concepts, connections, and formalization of factual data based on linguistic analysis of semistructured information); intelligent information search; thematic indexing; analysis of the front of research; communication services (services of work with STI and communication in the scientific community; intellectual analysis of specialized social networks and other means of scientific communication; and verification of borrowings). This is especially true for interdisciplinary research providing search and selection of relevant scientific provisions and tools in related fields of scientific knowledge.

Keywords: scientific result; information model; scientific services; identification algorithms; intellectual search; interdisciplinary research

DOI: 10.14357/19922264190319

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (project 18-29-03091).

References

1. Zatsarinnyy, A. A., E. V. Kiselev, S. V. Kozlov, and K. K. Kolin. 2018. *Informatsionnoe prostranstvo tsifrovoy ekonomiki Rossii. Kontseptual'nye osnovy i problemy formirovaniya* [Information space of the digital economy of Russia. Conceptual framework and problems of formation]. Moscow: FRC CSC RAS. 236 p.
2. Zatsarinny, A. A., K. I. Volovich, and V. A. Kondrashev. 2017. Metodologicheskie voprosy upravleniya nauchnymi servisami nauchnykh i obrazovatel'nykh organizatsiy Rossiyskoy Federatsii [Methodological issues of scientific services management of scientific and educational organizations of the Russian Federation] // 23rd Scientific and Technical Conference (International) “Radar, Navigation, Communication” Proceedings. Voronezh. 1:7–14.
3. FRC CSC RAS. 2016. Issledovanie voprosov upravleniya rezul'tatami nauchno-issledovatel'skoy deyatel'nosti organizatsiy, podvedomstvennykh FANO Rossii, i nauchnymi servisami seti TsKP FANO [Study of issues related to the management of the results of research activities of organizations under the jurisdiction of the FASO of Russia, and scientific services of the network of FASO collective centers]. Moscow. Otchet o NIR «Servis-U» [Research Report “Service-U”]. 437 p.
4. Ushakov, E. V. 2005. *Vvedenie v filosofiyu i metodologiyu nauki* [Introduction to philosophy and methodology of science]. Moscow: Ekzamen. 528 p.
5. Zatsarinny, A. A., V. A. Kondrashev, and A. P. Suchkov. 2017. Sistema nauchnykh servisov kak aktual'nyy komponent nauchnykh issledovaniy [The system of scientific services as an actual component of scientific research]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):23–38.
6. Concept, essence and classification of research. Available at: https://studopedia.ru/6_66034_ponyatie-sushchnosti-i-klassifikatsiya-issledovaniy-rol-issledovaniy-v-razvitiimenedzhmenta.html (accessed May 30, 2019).
7. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. 2007–2019.
8. Smart search. Available at: <https://www.naumen.ru/products/> (accessed May 30, 2019).

Received May 30, 2019

Contributor

Suchkov Alexander P. (b. 1954) — Doctor of Science in technology, 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; ASuchkov@ipiran.ru

Агасандян Геннадий Аршавирович (р. 1941) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бурлуцкий Владимир Владимирович (р. 1975) — кандидат физико-математических наук, руководитель Центра информационно-аналитических систем Югорского НИИ информационных технологий, г. Ханты-Мансийск

Вахтанов Никита Андреевич (р. 1997) — студент магистратуры Московского института электроники и математики Национального исследовательского университета «Высшая школа экономики»

Волошин Семён Викторович (р. 1991) — аспирант Югорского государственного университета, г. Ханты-Мансийск

Гайдамака Анна Александровна (р. 1997) — студентка магистратуры кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Гончаров Александр Анатольевич (р. 1994) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Захарова Татьяна Валерьевна (р. 1962) — кандидат 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

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой Вологодского государственного университета; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Вологодского научного центра Российской академии наук

Инькова Ольга Юрьевна (р. 1965) — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра

«Информатика и управление» Российской академии наук

Киселева Ксения Михайловна (р. 1992) — кандидат физико-математических наук, научный сотрудник Вологодского государственного университета

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Кузьмин Виктор Юрьевич (р. 1986) — руководитель Департамента разработки ООО «Вай2Гео»

Лебедев Алексей Викторович (р. 1971) — доктор физико-математических наук, доцент кафедры теории вероятностей механико-математического факультета Московского государственного университета им. М. В. Ломоносова

Мартынов Олег Павлович (р. 1996) — студент факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Мельников Андрей Витальевич (р. 1956) — доктор технических наук, профессор, директор Югорского НИИ информационных технологий, г. Ханты-Мансийск

Нуриев Виталий Александрович (р. 1980) — кандидат филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Павлов Юрий Леонидович (р. 1949) — доктор физико-математических наук, профессор, главный научный сотрудник Института прикладных математических исследований Федерального исследовательского центра «Карельский научный центр Российской академии наук»

Панов Александр Игоревич (р. 1987) — кандидат физико-математических наук, старший научный сотрудник Института проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук; заместитель заведующего лабораторией Московского физико-технического института (государственного университета)

Румовская София Борисовна (р. 1985) — кандидат технических наук, научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой, директор Института Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сатин Яков Александрович (р. 1978) — кандидат физико-математических наук, доцент Вологодского государственного университета

Скрынник Алексей Александрович (р. 1993) — младший научный сотрудник Института проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук

Смирнов Иван Валентинович (р. 1978) — кандидат физико-математических наук, доцент, заведующий отделом Института проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Российского университета дружбы народов

Стефанович Алексей Игоревич (р. 1983) — главный специалист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сучков Александр Павлович (р. 1954) — доктор технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тархов Алексей Андреевич (р. 1995) — студент магистратуры кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Царегородцев Александр Леонидович (р. 1979) — кандидат технических наук, первый заместитель директора Югорского НИИ информационных технологий, г. Ханты-Мансийск

Чистова Елена Викторовна (р. 1996) — программист Института проблем искусственного интеллекта Федерального исследовательского центра «Информатика и управление» Российской академии наук, студентка Российского университета дружбы народов

Чухно Надежда Викторовна (р. 1995) — студентка магистратуры кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Чухно Ольга Викторовна (р. 1995) — студентка магистратуры кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Шнурков Петр Викторович (р. 1953) — кандидат физико-математических наук, доцент Национального

исследовательского университета «Высшая школа экономики»

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Якимчук Александр Васильевич (р. 1994) — аспирант Югорского государственного университета, г. Ханты-Мансийск

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 15 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal”–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (Timoshenko S. P., Young D. H., Weaver W. *Vibration problems in engineering*. — 4th ed. — N.Y.: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . . ”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 15 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazoidachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets.
All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
 - an article and additional materials have to be attached using the “attach” function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)
<http://www.ipiran.ru/english/journal.asp>