

В. А. Каштанов, О. Б. Зайцева. О МИНИМАКСНЫХ ПОДХОДАХ В ЗАДАЧАХ БЕЗОПАСНОСТИ	55
А. Н. Кириллов. СТАБИЛИЗАЦИЯ УПРАВЛЯЕМЫХ ДИНАМИЧЕСКИХ СИСТЕМ ЗА КОНЕЧНОЕ ВРЕМЯ	68
А. О. Олейников. ЧИСЛЕННАЯ ОПТИМИЗАЦИЯ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ В СТАЦИОНАРНОЙ СЛУЧАЙНОЙ СРЕДЕ	73
Е. С. Петрова. О ПРЕДЕЛЬНОМ ПОВЕДЕНИИ ЧИСЛА ДЕРЕВЬЕВ ЗАДАННОГО ОБЪЕМА В СЛУЧАЙНОМ НЕПОМЕЧЕННОМ НЕКОРНЕВОМ ЛЕСЕ	79
А. А. Печников. ПРИМЕНЕНИЕ ВЕБОМЕТРИЧЕСКИХ МЕТОДОВ ДЛЯ ИССЛЕДОВАНИЯ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА НАУЧНОЙ ОРГАНИЗАЦИИ (НА ПРИМЕРЕ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН)	86
А. Л. Рабинович, Д. В. Журкин. СУЩЕСТВЕННАЯ ВЫБОРКА ПРИ МОДЕЛИРОВАНИИ НЕПРЕРЫВНОГО СПЕКТРА КОНФОРМАЦИЙ МАКРОМОЛЕКУЛ МЕТОДОМ МОНТЕ-КАРЛО	96
Е. В. Хворостянская. О СЛУЧАЙНЫХ ПУАССОНОВСКИХ ЗАПОЛНЕНИЯХ ЯЧЕЕК	112
С. П. Чистяков. СЛУЧАЙНЫЕ ЛЕСА: ОБЗОР	117
Правила для авторов	137

ТРУДЫ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН. № 1, 2013



Труды

КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА
РОССИЙСКОЙ АКАДЕМИИ НАУК

№ 1, 2013

ISSN 1997-3217

transactions.krc.karelia.ru

Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Выпуск 4

СОДЕРЖАНИЕ	
Предисловие	3
Б. З. Белашев. АНАЛИЗ РАСПРЕДЕЛЕНИЙ СЕЙСМИЧЕСКИХ СОБЫТИЙ КАТАЛОГА «NEIC»	4
Г. А. Борисов, Т. П. Тихомирова. МЕТОДЫ МИНИМИЗАЦИИ ПОТЕРЬ ЭНЕРГИИ И МОЩНОСТИ В ТОПЛИВНО-ЭНЕРГЕТИЧЕСКОМ ХОЗЯЙСТВЕ РЕГИОНА	12
В. Н. Гридин, А. Ю. Голубин, М. Н. Петрова. ОПТИМАЛЬНЫЕ СТРАТЕГИИ СТРАХОВАНИЯ ИНДИВИДУАЛЬНЫХ РИСКОВ И СУММАРНОГО УЩЕРБА ДЛЯ СТАТИЧЕСКОЙ МОДЕЛИ	17
А. Л. Забровский. КРИТЕРИЙ ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАТЕЛЬНЫХ МУЛЬТИМЕДИЙНЫХ ПОТОКОВ, ТРАНСЛИРУЕМЫХ В РЕАЛЬНОМ РЕЖИМЕ ВРЕМЕНИ	26
Ю. В. Заика. МОДЕЛЬ ТЕРМОДЕСОРБЦИИ ВОДОРОДА ИЗ МЕТАЛЛООРГАНИЧЕСКОЙ СТРУКТУРЫ	33
К. А. Замураев, А. В. Прасолов. МОДЕЛЬ РЕГУЛИРОВАНИЯ ЭКОНОМИКИ САНКТ-ПЕТЕРБУРГА ЗА СЧЕТ ИНВЕСТИЦИОННОГО ПОТОКА	38
Н. В. Каблукова, А. В. Соколов. МАТЕМАТИЧЕСКИЙ АНАЛИЗ ОДНОГО СПОСОБА ПРЕДСТАВЛЕНИЯ ДВУХ FIFO-ОЧЕРЕДЕЙ В ОБЩЕЙ ПАМЯТИ	46

Карельский научный центр
Российской академии наук

ТРУДЫ

КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РОССИЙСКОЙ АКАДЕМИИ НАУК

Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Выпуск 4

Петрозаводск
2013

Научный журнал
**Труды Карельского научного центра
Российской академии наук**
№ 1, 2013
Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
вып. 4

Scientific Journal
**Proceedings of the Karelian Research Centre
of the Russian Academy of Sciences**
No 1, 2013
MATHEMATICAL MODELING AND INFORMATION
TECHNOLOGIES Series
Vol. 4

Главный редактор
А. Ф. Титов

Редакционный совет

А. М. Асхабов, В. Т. Вдовицын, Т. Вихавайнен, А. В. Воронин, С. П. Гриппа, Э. В. Ивантер, А. С. Исаев,
В. Т. Калинин, В. И. Крутов, А. М. Крышень (зам. главного редактора), Е. В. Кудряшова,
В. В. Мазалов, Ф. П. Митрофанов, И. И. Муллонен, Н. Н. Немова, В. В. Окрепилов, О. Н. Пугачев,
Ю. В. Савельев, Н. Н. Филатов, А. И. Шишкин, В. В. Щипцов

Editor-in-Chief
A. F. Titov

Editorial Council

A. M. Askhabov, V. T. Vdovitsyn, T. Vihavainen, A. V. Voronin, S. P. Grippa, E. V. Ivanter, A. S. Isaev,
V. T. Kalinnikov, V. I. Krutov, A. M. Kryshen' (Deputy Editor), E. V. Kudryashova, V. V. Mazalov,
F. P. Mitrofanov, I. I. Mullonen, N. N. Nemova, V. V. Okrepilov, O. N. Pugachyov, Yu. V. Saveliev,
N. N. Filatov, A. I. Shishkin, V. V. Schiptsov

Редакционная коллегия серии «Математическое моделирование и информационные технологии»
В. Т. Вдовицын, Ю. В. Заика, А. Н. Кириллов, В. Ф. Колчин, В. В. Мазалов (отв. редактор), Ю. Л. Павлов
(зам. отв. редактора), Л. А. Петросян, А. В. Соколов, Т. П. Тихомирова (отв. секретарь)

Editorial Board of the «Mathematical Modeling and Information Technologies» Series

V. T. Vdovitsyn, Yu. V. Zaika, A. N. Kirillov, V. F. Kolchin, V. V. Mazalov (Editor-in-Charge), Yu. L. Pavlov
(Deputy Editor-in-Charge), L. A. Petrosian, A. V. Sokolov, T. P. Tikhomirova (Executive Secretary)

ISSN 1997-3217

Зав. редакцией Н. В. Михайлова
Адрес редакции: 185910 Петрозаводск, ул. Пушкинская, 11
тел. (8142)780109; факс (8142)769600
E-mail: trudy@krc.karelia.ru
Электронная полнотекстовая версия: <http://transactions.krc.karelia.ru>

ПРЕДИСЛОВИЕ

В этом выпуске серии «Математическое моделирование и информационные технологии» продолжается публикация статей участников Восьмой Петрозаводской международной конференции «Вероятностные методы в дискретной математике» и Тринадцатого Всероссийского симпозиума по прикладной и промышленной математике, которые проходили летом 2012 г. в Петрозаводске. Проведение Восьмой Петрозаводской международной конференции «Вероятностные методы в дискретной математике» поддержано РФФИ. Информация об этих научных мероприятиях была опубликована в разделе «Хроника» предыдущего выпуска нашей серии. Статьи посвящены главным образом вопросам математического моделирования в разных областях науки и практики. В них рассматриваются вопросы применения математических методов для решения конкретных задач геологии, энергетики, физики, биологии, страховании рисков и ущербов, а также при анализе вычислительных устройств и сетей телекоммуникаций. В некоторых статьях обсуждаются вероятностные методы комбинаторного анализа. Особое место занимает работа В. А. Каштанова и О. Б. Зайцевой, в которой предлагаются новые подходы к решению весьма актуальной в наше время проблеме безопасности. Кроме того, впервые в нашей стране публикуется обзорная статья о методе анализа данных «Случайные леса». Этот мощный современный метод прикладной статистики появился недавно, но он уже получил широкое распространение во многих странах при решении разнообразных задач и позволил получить ряд интересных и важных результатов, имеющих большое практическое значение. К сожалению, в России метод случайных лесов почти не используется, и мы надеемся, что публикация обзора будет способствовать его распространению и развитию.

УДК 550.34,551.24

АНАЛИЗ РАСПРЕДЕЛЕНИЙ СЕЙСМИЧЕСКИХ СОБЫТИЙ КАТАЛОГА «NEIC»

Б. З. Белашев

*Институт геологии Карельского научного центра РАН
Петрозаводский государственный университет*

По данным каталога землетрясений «NEIC» построены распределения сейсмических событий и оценены их пространственные, временные, энергетические и спектральные параметры. Показано, что современная сейсмичность имеет характерные особенности, присущие земной коре, мантии, доменам литосферы. Выявлены долгоживущие центры сейсмической активности.

Ключевые слова: землетрясение, каталог, сейсмичность, кора, мантия, домен, гипоцентр, магнитуда, распределение, спектр, бифуркация.

B. Z. Belashev. ANALYSIS OF THE DISTRIBUTION OF SEISMIC EVENTS FROM THE NEIC CATALOG

Using data from the «NEIC» catalog of earthquakes the distributions of the seismic events were built, and their space, time, energy and spectral parameters were assessed. We show that the current seismicity has specific characteristics pertinent to the Earth's crust, mantle and lithosphere domains. The long-lived centers of seismic activity were identified.

Key words: earthquake, catalog, seismicity, Earth's crust, mantle, domain, hypocenter, magnitude, distribution, spectra, bifurcation.

ВВЕДЕНИЕ

Изучение современной тектоники опирается на каталоги землетрясений. Известный каталог Национальной сейсмической службы США «NEIC» [9] содержит столбцы данных, отвечающих году, месяцу, дате, времени, широте, долготе, глубине и магнитуде землетрясения. Структура каталога позволяет строить разнообразные распределения событий и по ним изучать сейсмическую активность Земли. Цель работы — на основе данных каталога «NEIC» построить распределения событий и извлечь из них информацию о современном сейсмическом процессе.

МАТЕРИАЛЫ И МЕТОДЫ

Для изучения сейсмического процесса и оценки его параметров были использованы данные каталога «NEIC» за период 1973–2006 гг. Распределения N сейсмических событий получали, подсчитывая число событий $f_k(x)$ переменной x в диапазоне ее изменения для k событий по формуле:

$$f_k(x) = f_{k-1}(x) + \delta_k(x), \quad (1)$$

$$k = 1, \dots, N; f_0(x) = 0$$

$$\delta_k(x) = 1, \text{ если } h < x \leq h + \Delta h$$

$$\delta_k(x) = 0, \text{ если } x \leq h \text{ или } x > h + \Delta h.$$

Здесь k — номер события, h — отсчет гистограммы, Δh — шаг гистограммы. Помимо

распределений событий на основе генеральной совокупности каталог удобен для построения выборочных распределений событий для земной коры, мантии и тектонических доменов, относящихся к внутренним частям океанических плит, зонам сжатия, континентального и океанического рифтов [5]. Построение и обработку распределений событий вели в системе компьютерной математики «Matlab» [1]. Структуру распределений выявляли при помощи алгоритмов [6]. При прогнозировании зависимостей применяли метод «Гусеница», как в прямом, так и в обратном направлении [2]. Параметры метода, соответствующие наименьшей ошибке прогноза в обратном направлении, использовали в прямом направлении. Для получения спектров распределений в них устраняли аппроксимируемые полиномами тренды, а затем применяли метод максимума энтропии Бурга [7]. В этом методе оценки текущего значения $y(l)$ распределения и ошибки предсказания $\varepsilon(l) = y(l) - x(l)$ представляют линейной комбинацией p предшествующих значений $x(l - k)$.

$$y(l) = \sum_{k=1}^p a(k)x(l - k), \quad (2)$$

$$\varepsilon(l) = \sum_{k=0}^p a(k)x(l - k) \text{ при } a(0) = -1. \quad (3)$$

Коэффициенты комбинации $a(k)$ ищут из условия минимума дисперсию ошибки предсказания. Для аддитивного белого шума фильтр с такими коэффициентами является «отбеливающим»: превращает входные данные в белый шум на выходе. Спектр мощности процесса выражают через спектральную мощность шума и частотную характеристику фильтра:

$$P(m) = \left[\frac{\sigma^2}{1 + \sum_{k=1}^p a(k) \exp(-jm\Delta\omega\Delta t)} \right]^2 \quad (4)$$

Здесь $\Delta\omega$ и Δt представляют шаги дискретизации данных по угловой частоте и времени. Такой спектр отвечает минимальной информации на выходе фильтра и имеет минимальное число пиков, не являющихся артефактами и совместимых с исходными данными. Усредненный спектр получали, вычисляя периодограмму с окном в 20 отсчетов. Особенности нестационарного сейсмического процесса изучали с помощью вейвлет-алгоритма [6] с ортогональными вейвлетами Добеши и Симлета, по скачкам параметров сигналов, выявля-

ющих моменты смены режима процесса, называемые бифуркациями [3]. Интерпретацию результатов вели, сравнивая распределения, полученные при разных условиях, а также с данными литературных источников.

РЕЗУЛЬТАТЫ

Распределения событий каталога «NEIC» по магнитуде представлены на рис. 1.

При построении распределений сейсмических событий по глубине (рис. 2 а, б) для фиксации нуля логарифмической шкалы к содержанию отсчетов добавляли единицу. В полученных распределениях проявляется упорядоченность гипоцентров землетрясений по глубине. Гипоцентры приурочены к глубинам, чередующимся с шагом примерно 50 км.

Эти распределения отвечают большинству событий в зонах сжатия и отличаются от распределений событий в других доменах, имеющих меньшую мощность сейсмоактивной литосферы. Для областей континентального рифта в распределениях сейсмических событий по глубине присутствуют пики 7, 18, 30 и 60 км, а в для областей океанического рифта и океанических плит – 7 и 30 км. При изучении распределений эпицентров сейсмических событий по широте и долготе отбирали группы 10000 событий из начала и конца каталога, соответствующих началу и концу предполагаемого цикла сейсмической активности. Положение эпицентра задавали географическими координатами, а цвет отображающей его точки – в зависимости от магнитуды события или глубины гипоцентра. В начале цикла преобладали события больших магнитуд и глубин, в конце цикла – события малых глубин и небольших магнитуд. Нестационарность распределения событий оказалась заметно выраженной для распределения событий с глубинами гипоцентров, не превышающих 100 км (рис. 3 а). Условно такие события будем считать коровыми, а события с глубинами гипоцентров больше 100 км – мантийными. Распределение мантийных событий по времени дает рис. 3 б, спектры распределений событий по времени для земной коры, мантии и тектонических доменов – рис. 3 в, г. Периоды сейсмичности тектонических доменов представлены в табл. 1.

По форме спектры распределений событий для зон континентального и океанического рифтов близки. Заметны отличия спектров распределения событий земной коры и мантии. В распределении событий по годам с гипоцентрами глубиной не более 100 км

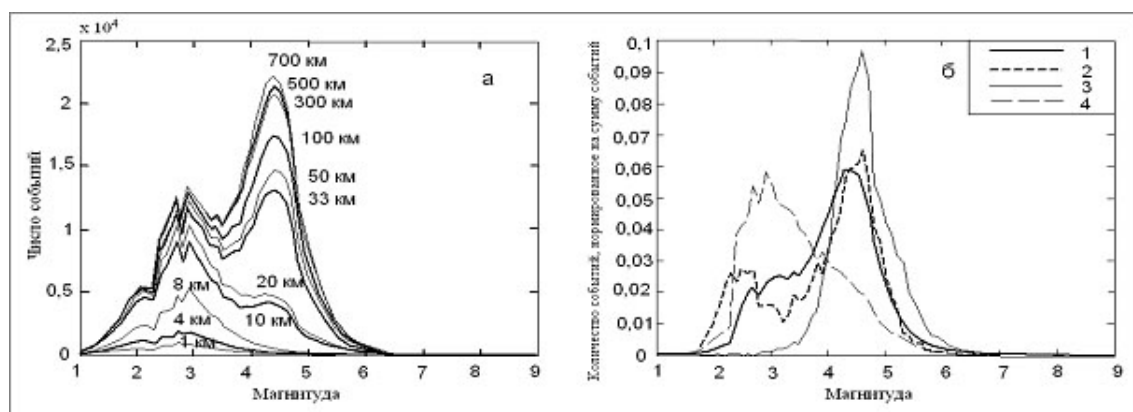


Рис. 1. Распределения сейсмических событий по магнитуде для разных диапазонов глубин (а) и разных тектонических доменов (1 – зоны сжатия, 2 и 3 – зоны континентального и океанического рифтов, 4 – события внутри океанических плит)

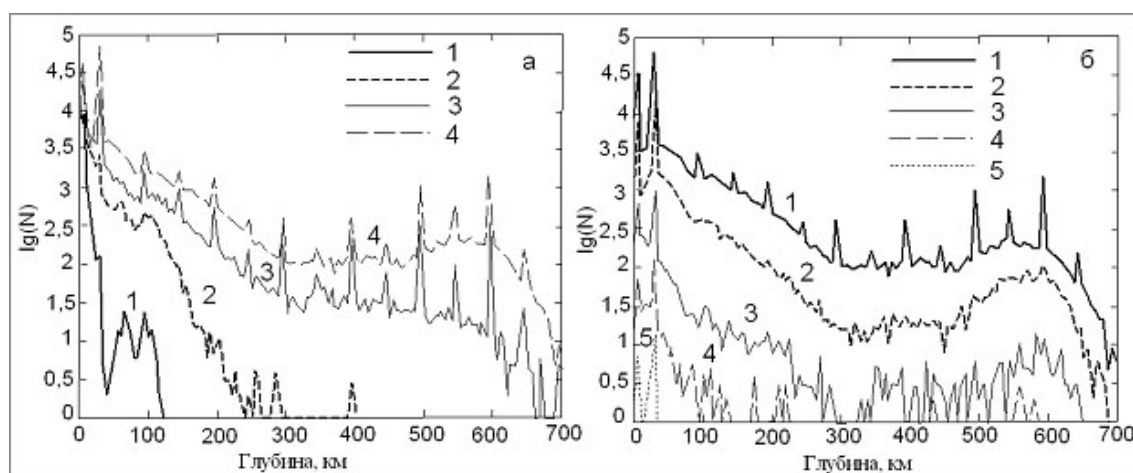


Рис. 2. Распределения сейсмических событий по глубине для фиксированных интервалов магнитуды M : а) 1 – $0,9 < M < 2,1$; 2 – $1,9 < M < 3,1$; 3 – $2,9 < M < 4,1$; 4 – $3,9 < M < 5,1$; б) 1 – $3,9 < M < 5,1$; 2 – $4,9 < M < 6,1$; 3 – $5,9 < M < 7,1$; 4 – $6,9 < M < 8,1$; 5 – $7,9 < M < 9,1$

Таблица 1. Периоды сейсмичности для тектонических доменов

Зоны континентального рифта	15,7	10,4	7,0	5,2	4,0	3,5	3,2	2,6	2,3	2,1
Зоны океанического рифта	15,7	10,5	7,4	5,7	4,5	3,9	3,1	2,6	2,3	–
Зоны сжатия	16,7	9,7	8,4	5,0	4,7	3,3	3,0	2,3	2,2	–
Зоны внутри океанических плит	15,8	9,7	7,9	5,2	4,5	3,5	3,1	2,4	–	2,1

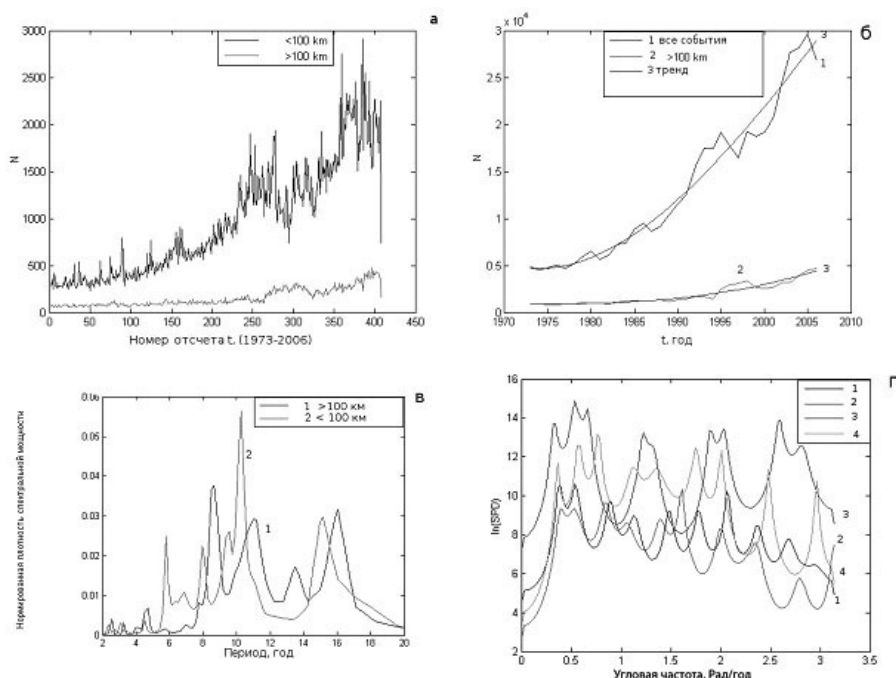


Рис. 3. Временные распределения сейсмических событий: а) шаг отсчета – месяц; б) распределения событий по годам с трендами, аппроксимируемыми полиномами 3 степени; в) периодограммы Бурга для коровых и мантийных событий; г) периодограммы Бурга для тектонических доменов 1 – зоны континентального рифта, 2 – зоны океанического рифта, 3 – зоны сжатия, 4 – события внутри океанических плит

выявлены локальные особенности, соответствующие 1998 и 2004 гг., а в распределения по годам событий с гипоцентрами с глубинами выше 100 км – 2004 г. (рис. 4). На рис. 5 показаны распределения событий по месяцам, отвечающие на вопрос: "В каком месяце землетрясения происходят чаще?". Для событий с гипоцентрами, не более 100 км, такими месяцами являются апрель-май, октябрь-ноябрь, а для событий с гипоцентрами свыше 100 км – август. Участки с наибольшим количеством землетрясений в году рассматривали как центры сейсмической активности. В этих участках удобно изучать процессы накопления напряжений и подготовки землетрясений. Центры сейсмической активности 1973 – 1980 гг. показаны на рис. 6 а, а на рис. 6 б приведено распределение эпицентров землетрясений вблизи одного из них – центра сейсмической активности 1973 г. с координатами 33° с. ш. и 141° в. д.

ОБСУЖДЕНИЕ

Формы распределения событий по магнитуде носят дискретный характер в зависимости от глубины очага. Заметна тенденция увеличения роста магнитуды события с ростом

глубины его гипоцентра (рис. 1 а). Разложение сложного контура распределения сейсмических событий по магнитуде на компоненты в виде функций Гаусса не реалистично. Детали полученных распределений можно описать, используя в качестве базиса разложение компонентов, соответствующие распределениям сейсмическим событиям по магнитуде для тектонических доменов (рис. 1 б). Обнаруженную в распределениях событий по глубине приуроченность гипоцентров землетрясений к глубинам, чередующихся с шагом 50 км, (см. рис. 2 а, б), можно объяснить с позиций радиальных стоячих волн напряжений, узлы и пучности которых соответствуют малым и повышенным вероятностям событий. Такие волны напряжений возможны в зонах сжатия, имеющих большую радиальную протяженность. В других тектонических доменах распределения значительно более короткие и редуцированные, как например, у океанических плит с малой толщиной коры и простым строением [2]. Изменения в распределении эпицентров событий в начале и конце 33-летнего цикла свидетельствует о том, что события, происходящие в разных местах в разное время, не являются независимыми. Переход от сильных глубинных землетрясений к многочисленным слабым поверхностным землетрясениям

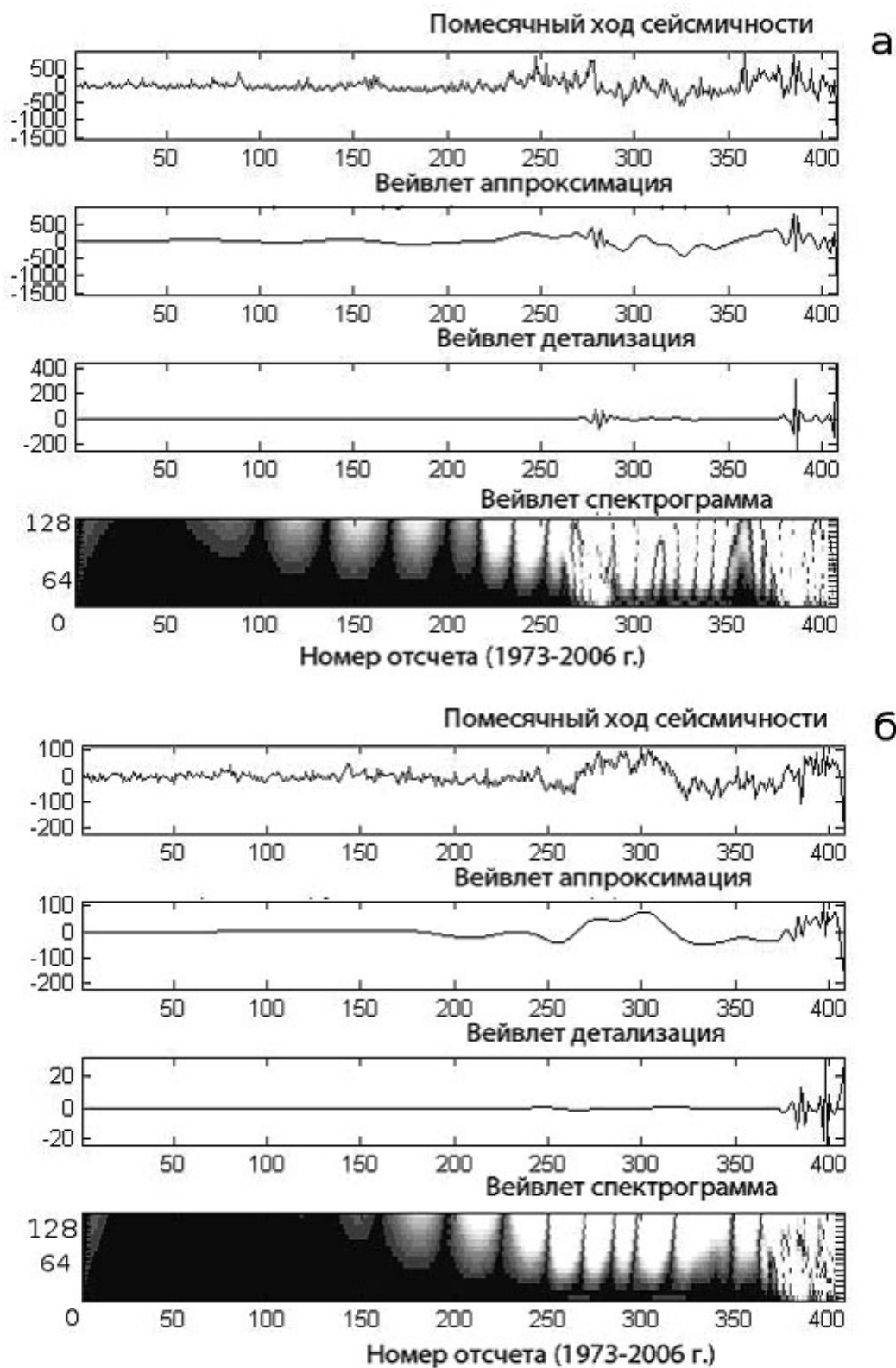


Рис. 4. Бифуркации в распределениях сейсмических событий по времени без трендов для событий с глубинами гипоцентров, не превышающих 100 км (а) и более 100 км (б)

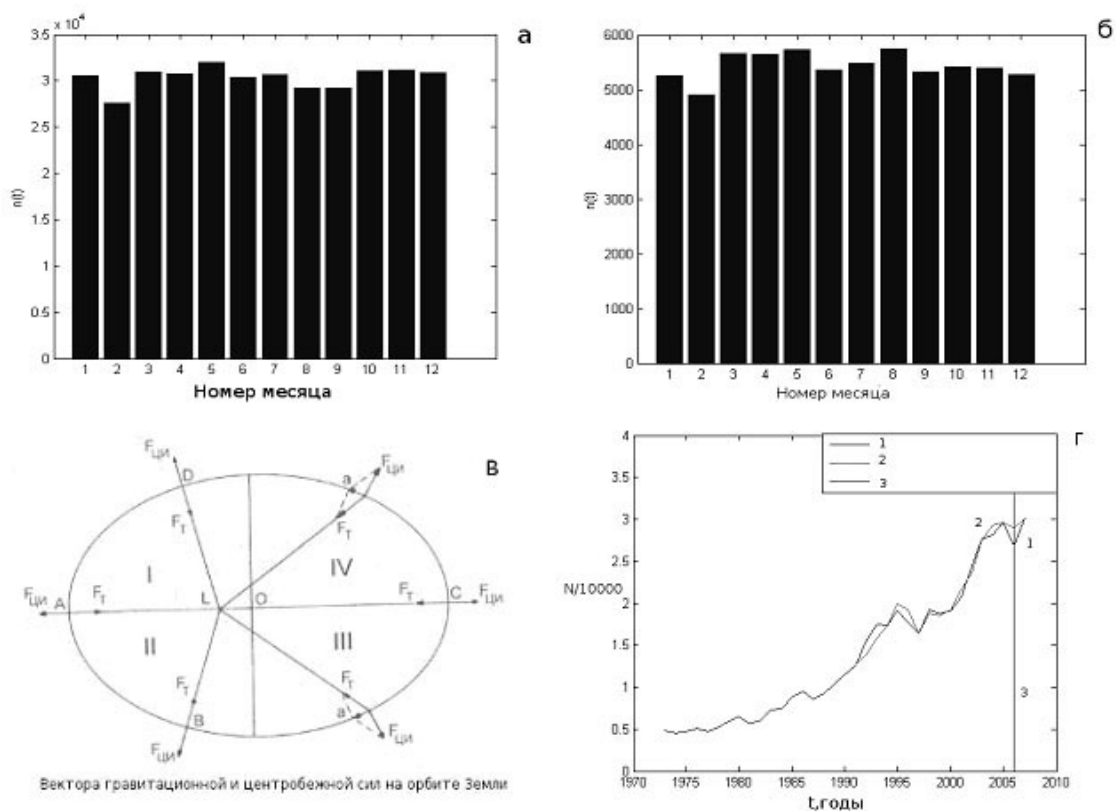


Рис. 5. Распределения событий по месяцам для глубин гипоцентров не более 100 км (а) и более 100 км (б), вектора гравитационной и центробежной сил на орбите Земли, определяющих тангенциальное ускорение [8] (в) и экстраполяция распределения сейсмических событий по времени методом «Гусеница» (г)

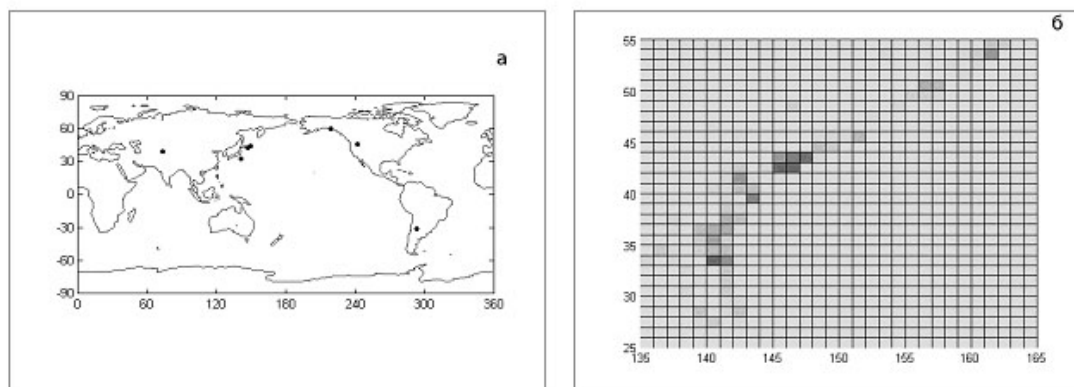


Рис. 6. Центры сейсмической активности на поверхности Земли 1973–1980 гг. (а) и распределение эпицентров землетрясений вблизи центра 1973 г. (б)

отражает общую трансформацию энергии напряжений, указывает на коррелированность событий в пространстве и времени, а, следовательно, на самосогласованность и самоорганизацию сейсмического процесса. Циклические изменения проявляются во временных зависимостях сейсмичности. Для коровых землетрясений основными являются периоды 10,5 и 9,7 лет, у мантийных событий – периоды 11,5, 8,4, 13,6 лет. Активность коровых землетрясений по отношению к мантийным (см. рис. 3) можно понять из представлений о гравитационной экранировке мантии земной корой. Благодаря такой экранировке устраняется влияние экзогенных факторов на тектонические процессы в мантии, а нестационарность мантийного сейсмического процесса выражена слабо в связи с меньшим числом точек бифуркации. Обнаруженная во временной зависимости коровых землетрясений бифуркация 1998 г. в аналогичной зависимости мантийных землетрясений отсутствует (рис. 4). Соответствие этой бифуркации началу роста экваториального радиуса Земли подтверждено лазерными измерениями формы Земли [8]. Присутствующую в обоих распределениях событий бифуркацию 2004 г., по-видимому, определяют эндогенные процессы. Интерпретация экстремумов распределения коровых событий по месяцам проведена с использованием эллиптической формы орбиты движения Земли вокруг Солнца [4]. В системе отсчета, связанной с Землей, в апреле-мае, октябре-ноябре тангенциальное ускорение Земли отлично от нуля, способно вызывать дополнительные упругие напряжения в земной коре. В июне и декабре тангенциальное ускорение близко к нулю. Соответственно и число землетрясений в эти месяцы меньше. Максимум распределения мантийных событий по месяцам приходится на август. С учетом влияния эндогенных механизмов это позволяет высказать гипотезу о связи землетрясений и повышенного числа, происходящих в августе природных и техногенных катастроф, с режимами дефлюидизации мантии и воздействием ее продуктов на атмосферу, биосферу, психофизическое состояние и здоровье людей. Выявленные центры сейсмической активности в основном связаны с Тихоокеанским сейсмическим кольцом. В этих центрах возбужденные состояния недр быстро формируются и быстро сбрасываются. Имеют место смещения гипоцентров.

ЗАКЛЮЧЕНИЕ

Исследование показало, что распределения сейсмических событий на основе каталога «NEIC» являются информативными по срав-

нению с первичными данными. Они не используют априорных сведений и позволяют делать выводы относительно самых разных сторон современного сейсмического процесса.

Выводы

1. Пространственные, временные, спектральные и энергетические параметры сейсмичности имеют особенности, присущие земной коре, мантии и тектоническим доменам литосферы.
2. Различия в распределениях коровых и мантийных сейсмических событий могут быть объяснены гравитационной экранировкой мантии земной корой, снижающей влияние экзогенных факторов на сейсмический процесс.
3. Центры сейсмической активности являются долгоживущими образованиями, могут включаться в тектонические структуры и направленно эволюционировать.
4. Современная сейсмичность является самосогласованным, самоорганизующимся процессом.

ЛИТЕРАТУРА

1. Говорухин В., Цибулин В. Компьютер в математическом исследовании. Maple. Matlab. LaTeX. СПб.: Питер, 2001. 624 с.
2. Голяндина Н. Метод «Гусеница»-SSA: анализ временных рядов: Учебное пособие. СПб.: Изд-во СПбГУ, 2004. 76 с.
3. Дьяконов В. П. Вейвлеты. От теории к практике. М.: Солон-Р, 2002. С. 176–179.
4. Ротационные процессы в геологии и физике / Под ред. Е. Е. Милановского. М.: КомКнига, 2007. 528 с.
5. Хаин В. Е., Халилов Э. Н. Цикличность геодинамических процессов. Ее возможная природа. М.: Научный мир, 2009. 520 с.
6. Belashev B. Methods to reveal Hidden Structures of Signals and their Applications. Вестник РУДН. 2010. N 3, вып. 3(2). Математика. Физика. Информатика. С. 132–135.
7. Burg J. P. Maximum entropy Spectral Analysis. Ph. D. Thesis, Department of Geophysics, Stanford University, Stanford, Calif., 1975.
8. Cox C., Chao B. F. Detection of large-scale mass redistribution in terrestrial system science 1998. Science. Vol. 297, N 2. August, 2002. P. 831–833.
9. NEIC National Earthquake Information Center NEIC - <http://earthquake.usgs.gov/regional/neic/index.php>.

СВЕДЕНИЯ ОБ АВТОРЕ:

Белашев Борис Залманович

старший научный сотрудник, д. т. н.

Институт геологии Карельского научного центра РАН

ул. Пушкинская, 11, Петрозаводск, Республика Каре-

лия, Россия, 185910

профессор кафедры информационно-измерительных

систем и физической электроники

Петрозаводский государственный университет

пр. Ленина, 33, Петрозаводск,

Республика Карелия, Россия, 185910

эл. почта: belashev@krc.karelia.ru

тел.: (8142) 782753

Belashev, Boris

Institute of Geology, Karelian Research Centre, Russian

Academy of Sciences

11 Pushkinskaya St., 185610 Petrozavodsk, Karelia,

Russia

Petrozavodsk State University

33 Lenina St., 185910, Petrozavodsk, Karelia, Russia

e-mail: belashev@krc.karelia.ru

tel.: (8142) 782753

УДК 338.45:621.31(470.22)

МЕТОДЫ МИНИМИЗАЦИИ ПОТЕРЬ ЭНЕРГИИ И МОЩНОСТИ В ТОПЛИВНО-ЭНЕРГЕТИЧЕСКОМ ХОЗЯЙСТВЕ РЕГИОНА

Г. А. Борисов, Т. П. Тихомирова

*Институт прикладных математических исследований
Карельского научного центра РАН*

В статье показано, что при параллельной работе одинаковых переделов энергии выравнивание нагрузок позволяет достичь условного минимума суммарных потерь мощности. При неодинаковых переделах условный минимум потерь мощности достигается при выравнивании относительных приростов их потерь. В случаях работы отдельного передела или последовательном соединении их выравнивание графика нагрузки на уровне средней приводит к условному минимуму динамических потерь энергии.

Ключевые слова: технологические переделы энергии, потери мощности, потери энергии, расходная характеристика, неравномерность графика нагрузки, стабилизация нагрузки.

G. A. Borisov, T. P. Tikhomirova. METHODS TO MINIMIZE ENERGY AND POWER LOSSES IN THE REGIONAL FUEL-AND-ENERGY SECTOR

We demonstrate that equalization of the load in identical process stages in parallel operation yields the constrained minimum of combined power losses. The constrained minimum of power losses in non-identical process stages is achieved through equalization of relative rises in power losses from the stages. For process stages operating independently or sequentially stabilization of the demand curve at an average level would result in a constrained minimum of dynamic energy losses.

Key words: process transitions, power losses, energy losses, metering characteristic, irregularity of the load curve, load stabilization.

Целью энергетической политики России является максимально эффективное использование природных энергетических ресурсов и потенциала энергетического сектора [1]. Статья посвящена методам технологического управления нормальными установившимися режимами топливно-энергетического хозяйства региона, минимизирующими потери энергии и мощности. В качестве примера рассматривается электроэнергетическое хозяйство Респуб-

лики Карелия: от природных источников энергии до получения ее конечных видов.

Известные методы технологического управления нормальными установившимися режимами энергосистем позволяют достигать не абсолютного, а условного минимума расхода или стоимости топлива в системе, диктуемого требующейся для потребителей суммарной мощностью генерации в системе (P_c). Минимум получается путем наивыгоднейшего

распределения суммарной мощности генерации между параллельно работающими электростанциями. Для оптимизации используются методы проекции градиента, обобщенного метода приведенного градиента. При использовании штрафных функций для учета ограничений в форме неравенств применяется метод безусловной оптимизации [2]. Минимум расхода топлива определяется методом множителей Лагранжа (при любом значении суммарной мощности генерации в системе P_c у m параллельно работающих источников мощности), когда равны численные значения относительных приростов [3]:

$$\frac{\partial B_1(P_1)}{\partial P_1} = \dots = \frac{\partial B_j(P_j)}{\partial P_j} = \dots = \frac{\partial B_m(P_m)}{\partial P_m} = \lambda \quad (1)$$

с учетом

$$\sum_{j=1}^m P_j = P_c, \quad (2)$$

$$0 \leq P_j \leq P_{jH}. \quad (3)$$

Здесь $B_j(P_j)$ – расходная характеристика j -го источника (зависимость расхода топлива от производимой выходной мощности P_j); P_c – суммарная мощность генерации; $\partial B_j(P_j)/\partial P_j$ – характеристика относительных приростов (топлива) у j -го источника; P_{jH} – номинальная мощность j -го источника.

При работе источников на потребителей через электрическую сеть и наличии потерь активной мощности в ней равенства (1) заменяются на следующие [3]:

$$\frac{1}{1 - \sigma_1} \cdot \frac{\partial B_1(P_1)}{\partial P_1} = \dots = \frac{1}{1 - \sigma_j} \cdot \frac{\partial B_j(P_j)}{\partial P_j} = \dots = \frac{1}{1 - \sigma_m} \cdot \frac{\partial B_m(P_m)}{\partial P_m} = \lambda, \quad (4)$$

где σ_j – производные потерь активной мощности в сетях по нагрузкам (мощностям) j -го источника.

Такой метод позволяет за счет наивыгоднейшего распределения мощности между параллельно работающими электростанциями, т. е. на одном технологическом переделе топливно-энергетического хозяйства, получать до 1 % экономии топлива [6]. Однако генерация дает сравнительно небольшую часть потерь энергии топливно-энергетического хозяйства. Так, например, по данным топливно-энергетического баланса Республики Карелия, из общей доли потерь в 75 % от первичной

энергии на ТЭЦ приходится 14,1 % общих потерь энергии на ее пути от первичных до конечных видов [4]. Подавляющая часть потерь энергии приходится на другие технологические переделы и элементы системы, соединяемые не параллельно, как электрические станции, а последовательно по ветвящимся и замкнутым (контурным) схемам, для которых упомянутые методы минимизации потерь активной мощности и энергии при нормальных режимах отсутствуют.

В нашей работе [5] показано, что технологические переделы системы имеют расходные характеристики и характеристики потерь мощности достаточно точно аппроксимируемые полиномами от первой до третьей степени

$$P_{i-1} = f(P_i) = P_i + \Delta P_i(P_i), \quad (5)$$

где P_{i-1} – входная мощность i -го элемента или передела; P_i – его выходная мощность; ΔP_i – потери мощности.

Минимум входной энергии и ее потерь в заданное время T и при заданной величине выходной энергии W_i получается при стабилизации выходной энергии, когда в каждый малый промежуток времени Δt выходная мощность передела равняется ее среднему значению

$$P_i = \frac{W_i}{T} = \frac{\Delta t \sum_{k=1}^n P_{ik}}{T} = P_{i\text{ср}},$$

где P_{ik} – выходная мощность i -го передела в k -й интервал времени; Δt – одинаковые по величине интервалы времени, $\Delta t = T/n$; n – количество интервалов времени.

При отклонении на i -м технологическом переделе его выходной мощности P_i от среднего значения $P_{i\text{ср}}$ за время T на величину ΔP_i появляются дополнительные потери потребляемой энергии ΔW_{i-1} (рис. 1).

Пусть расходная характеристика i -го передела достаточно точно аппроксимирована полиномом:

$$P_{i-1} = a_0 + a_1 P_i + a_2 P_i^2 + a_3 P_i^3.$$

Тогда энергия, потребляемая переделом за период времени T при постоянной выходной мощности $P_{i\text{ср}}$, равна

$$W_{i-1}(P_{i\text{ср}}) = T(a_0 + a_1 P_{i\text{ср}} + a_2 P_{i\text{ср}}^2 + a_3 P_{i\text{ср}}^3). \quad (6)$$

Если в первый полупериод $T/2$ выходная мощность скачкообразно увеличивается до значения $(P_{i\text{ср}} + \Delta P_{i\text{ср}})$, а во второй уменьшается до $(P_{i\text{ср}} - \Delta P_{i\text{ср}})$, то выдаваемая переделом энергия за период остается неизменной, а потребляемая энергия в первый полупериод времени $T/2$ составит

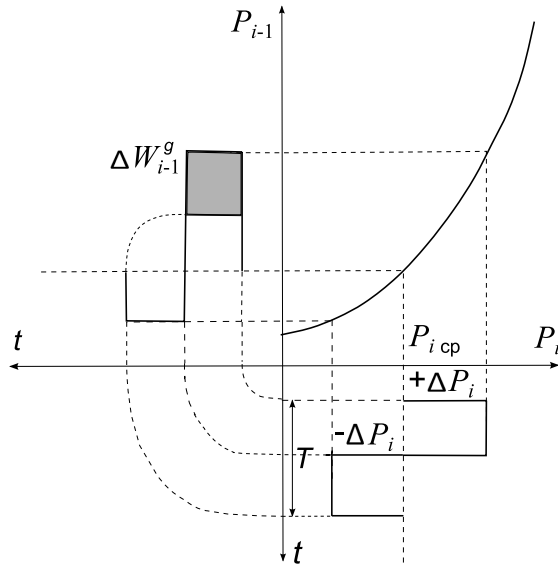


Рис. 1. Потери входной мощности передела при отклонении от средней выходной мощности

$$W_{i-1}(P_{i cp} + \Delta P_{i cp}) = 0.5 T [a_0 + a_1(P_{i cp} + \Delta P_{i cp}) + a_2(P_{i cp} + \Delta P_{i cp})^2 + a_3(P_{i cp} + \Delta P_{i cp})^3], \quad (7)$$

во второй полупериод

$$W_{i-1}(P_{i cp} - \Delta P_{i cp}) = 0.5 T [a_0 + a_1(P_{i cp} - \Delta P_{i cp}) + a_2(P_{i cp} - \Delta P_{i cp})^2 + a_3(P_{i cp} - \Delta P_{i cp})^3]. \quad (8)$$

Перерасход энергии за счет скачкообразной неравномерности графика выходной мощности за период T составит величину

$$\begin{aligned} \Delta W_{i-1}^g &= W_{i-1}(P_{i cp} + \Delta P_{i cp}) + \\ &+ W_{i-1}(P_{i cp} - \Delta P_{i cp}) - W_{i-1}(P_{i cp}) = \\ &= T(a_2 \Delta P_{i cp}^2 + 3a_3 P_{i cp} \cdot \Delta P_{i cp}^2) = \\ &= \frac{T}{2} \Delta P_{i cp}^2 (2a_2 + 6a_3 P_{i cp}) = \\ &= \frac{T}{2} \Delta P_{i cp}^2 \frac{\partial^2 P_{i-1}}{\partial P_{i cp}^2}. \end{aligned} \quad (9)$$

Этот перерасход можно назвать динамическими потерями энергии. В соответствии с полученной формулой (9) $\Delta P_{cp} = 0$ при стабильной выходной мощности, когда $P_{ik} = P_{i cp}$ и динамические потери энергии отсутствуют. Также они будут отсутствовать, если расходная характеристика или характеристика потерь мощности передела линейны и вторые производные их равны нулю, что получается при наличии только потерь холостого хода, независимых от нагрузки или потерь, линейно зависящих от нагрузки, к примеру, в теплопроводах с качественным регулированием отпуска тепла.

Вторая составляющая потерь энергии, независимая от изменений выходной мощности, может быть названа статическими потерями (ΔW_i^c). Она определяется только средним значением выходной мощности ($P_{i cp}$) за период T

$$\begin{aligned} \Delta W_{i-1}^c &= T(P_{i-1}(P_{i cp}) - P_{i cp}) = \\ &= T(a_0 + (a_1 - 1)P_{i cp} + \\ &+ a_2 P_{i cp}^2 + a_3 P_{i cp}^3). \end{aligned} \quad (10)$$

При произвольном графике выходной мощности i -го передела за период времени T потери энергии определяются формулой

$$\begin{aligned} \Delta W_{i-1} &= (P_{i-1}(P_{i cp}) - P_{i cp}) \cdot T + \\ &+ \frac{1}{2} \int_0^T (P_i(t) - P_{i cp})^2 \frac{\partial^2 P_{i-1}}{\partial P_i^2} dt, \end{aligned} \quad (11)$$

где первая составляющая — статические потери энергии, а вторая — динамические. При последовательном соединении технологических переделов в единичной технологической линии (рис. 2)

$$P_{i-1} = P_l + \sum_{\nu=i}^l \Delta P_{\nu}, \quad (12)$$

т. е. мощность, потребляемая i -м переделом, равняется сумме конечной мощности P_l и потерь мощности на последующих переделах.

Соответственно, при выравнивании или стабилизации нагрузки на единичной технологической линии, состоящей из последовательно соединенных переделов, увеличивается в сравнении с единичным переделом величина динамической составляющей потерь энергии. При этом первые переделы наиболее чувствительны по снижению динамических потерь энергии к выравниванию графика конечной мощности.

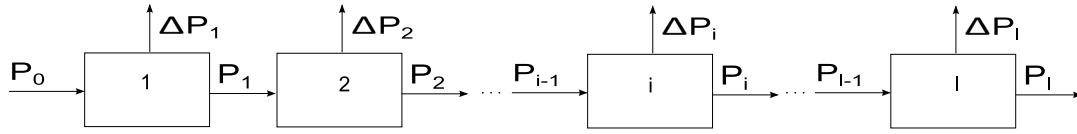


Рис. 2. Схема единичной технологической линии

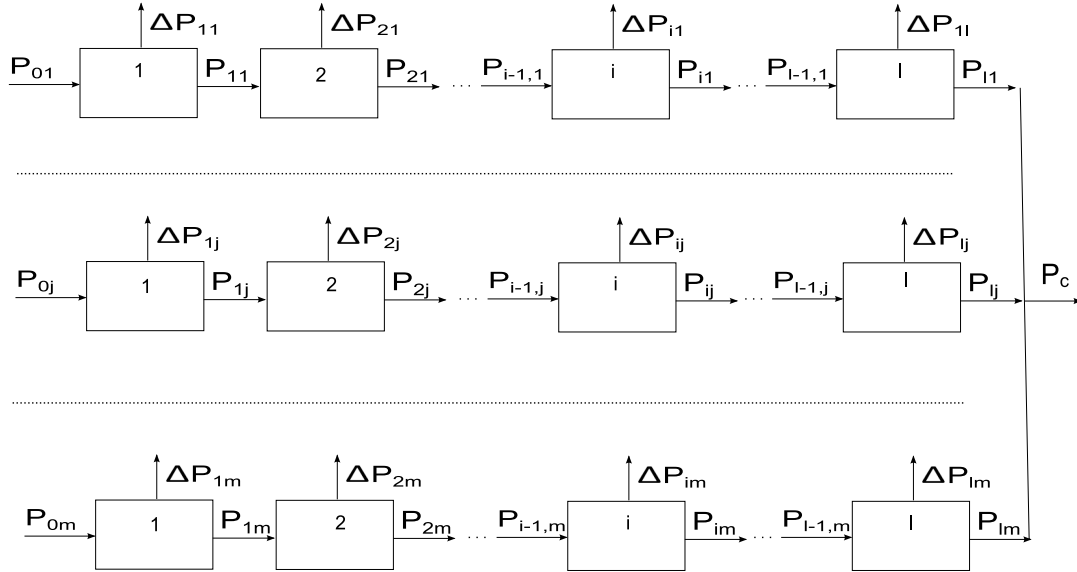


Рис. 3. Параллельная работа технологических линий

Таким образом, выравнивание графика конечной мощности снижает требования к установленной мощности предшествующих переделов и снижает на них динамическую составляющую потерь энергии. Такой способ используется в автономных энергетических установках – на гибридных автомобилях, устройствах непрерывного питания индивидуальных домов, в системах электроснабжения на базе возобновляемых источников энергии.

В «большой» энергетике метод стабилизации или сглаживания конечных нагрузок в целях энергосбережения не нашел применения. Он находит применение как способ ограничения максимума нагрузки при дефиците располагаемых генерирующих мощностей, при современной тенденции разуплотнения графиков нагрузки энергосистем в условиях дефицита пиковых и полупиковых мощностей [1].

При параллельной работе на общую нагрузку или соединении последовательных цепочек переделов (рис. 3) минимум суммарной первичной мощности будет получаться при заданной суммарной конечной мощности системы $P_c = P_{l1} + P_{l2} + \dots + P_{lm}$ при условии

$$\frac{\partial P_{01}}{\partial P_{l1}} = \dots = \frac{\partial P_{0j}}{\partial P_{lj}} = \dots = \frac{\partial P_{0m}}{\partial P_{lm}} = \lambda, \quad (13)$$

где P_{0j} – первичная мощность j -й технологической линии; P_{lj} – конечная мощность j -й технологической линии.

В соответствии с формулой (12)

$$\frac{\partial P_{0j}}{\partial P_{lj}} = 1 + \frac{\partial \sum_{i=1}^l \Delta P_{ij}}{\partial P_{lj}} \quad (14)$$

и

$$\begin{aligned} \frac{\partial \sum_{i=1}^l \Delta P_{i1}}{\partial P_{l1}} &= \dots = \frac{\partial \sum_{i=1}^l \Delta P_{ij}}{\partial P_{lj}} = \dots \\ &= \frac{\partial \sum_{i=1}^l \Delta P_{im}}{\partial P_{lm}} = \lambda - 1. \end{aligned} \quad (15)$$

При одинаковых расходных характеристиках параллельно работающих переделов или параллельно работающих единичных технологических линиях, когда

$$f(P_{01}) = f(P_{02}) = \dots = f(P_{0j}) = \dots = f(P_{0m}),$$

условие (1) приводит к равенствам

$$P_{l1} = P_{l2} = \dots = P_{lj} = \dots = P_{lm},$$

$$P_{01} = P_{02} = \dots = P_{0j} = \dots = P_{0m},$$

т. е. к необходимости выравнивания конечных и первичных мощностей параллельно работающих переделов или их технологических линий.

Выводы

1. У технологических переделов с нелинейными расходными характеристиками потери энергии разделяются на статическую и динамическую составляющие.
2. При автономной работе переделов или их группы, соединенных последовательно, условный минимум потерь энергии за период времени T достигается, когда выходная мощность неизменна и динамические потери энергии пренебрежимо малы.
3. При параллельной работе одинаковых переделов или единичных технологических линий условный минимум потерь мощности получается при равенстве их выходных мощностей, при неодинаковых переделах — при равенстве их относительных приростов потерь.

СВЕДЕНИЯ ОБ АВТОРАХ:

Борисов Георгий Александрович

старший научный сотрудник
Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: borisov@krc.karelia.ru
тел.: (8142) 766312

Тихомирова Тамара Петровна

ученый секретарь
Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: tihomiro@krc.karelia.ru
тел.: (8142) 785520

ЛИТЕРАТУРА

1. Энергетическая стратегия России на период до 2030 г.
2. *Баринов В. А., Савалов С. А.* Режимы энергосистем: Методы анализа и управления. М.: Энергоатомиздат, 1990. 440 с.
3. *Болотов В. В.* Теоретические основы выбора экономического режима электроэнергетической системы. М.; Л.: Изд-во АН СССР, 1947. 278 с.
4. *Борисов Г. А., Тихомирова Т. П.* Структурный анализ потерь энергии в электроэнергетическом хозяйстве Карелии // Учен. зап. Петр ГУ. 2009. № 9 (103). С. 93–97.
5. *Борисов Г. А., Тихомирова Т. П.* Исследование математической модели единичной технологической линии топливно-энергетического хозяйства региона // Труды Карельского научного центра РАН. Сер. Математическое моделирование и информационные технологии. Вып. 2. 2011. № 5. С. 10–17.
6. *Горништейн В. М.* Наивыгоднейшие режимы работы гидростанций в энергетических системах. М.; Л.: Госэнергоиздат, 1954. 248 с.

Borisov, George

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: borisov@krc.karelia.ru
tel.: (8142) 766312

Tikhomirova, Tamara

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: tihomiro@krc.karelia.ru
tel.: (8142) 785520

УДК 368:519

ОПТИМАЛЬНЫЕ СТРАТЕГИИ СТРАХОВАНИЯ ИНДИВИДУАЛЬНЫХ РИСКОВ И СУММАРНОГО УЩЕРБА ДЛЯ СТАТИЧЕСКОЙ МОДЕЛИ

В. Н. Гридин¹, А. Ю. Голубин², М. Н. Петрова²

¹ Центр информационных технологий в проектировании РАН,
Московская обл., г. Одинцово

² Московский институт электроники и математики Национального
исследовательского университета «Высшая школа экономики»

Статья посвящена решению задачи оптимального управления риском в статической модели, где разрешен выбор как стратегии страхования риска отдельного клиента, так и стратегии перестрахования суммарного риска. Интересы клиентов и перестраховщика учтены введением дополнительных ограничений с вероятностью единица на их остаточные риски. Оптимальным с точки зрения полезности страховщика оказывается stop loss перестрахование с верхним пределом и страхование, представляющее собой комбинацию stop loss стратегии и франшизы. Приведен пример, иллюстрирующий доказанные результаты в случае экспоненциальной функции полезности страховщика.

Ключевые слова: оптимальный дележ риска, страхование, перестрахование, функция полезности.

V. N. Gridin, A. Yu. Golubin, M. N. Petrova. OPTIMAL PER-CLAIM INSURANCE AND OPTIMAL REINSURANCE OF SUMMARY RISK IN A STATIC SETTING

The paper studies an individual risk model where both a policy of per-claim insurance and a policy of reinsurance of aggregated losses are chosen by the insurer in order to maximize their expected utility. The problem is solved under additional constraints on the reinsurer's risk and the residual risk of the insured that should be met with probability one. We show that the only solution to the problem is the following: the optimal reinsurance is a modification of the stop loss reinsurance policy, so-called stop loss reinsurance with an upper limit; the optimal insurer's indemnity is a combination of the stop loss and the deductible policies. The results are illustrated by a numerical example for the case of exponential utility function.

Key words: optimal risk sharing, insurance, reinsurance, utility function.

ВВЕДЕНИЕ

Объектом исследования в данной работе является так называемая модель индивидуального риска (или статическая модель

страхования), описанная, например в [1], где предусмотрена возможность перестрахования. Страховая компания может одновременно выбирать как долю своего возмещения (дележ)

каждому страхователю, так и дележ своего суммарного риска с перестраховщиком. Критерием оптимальности страховщика служит ожидаемая полезность его финального капитала. Желание клиентов и перестраховщика не иметь «больших» значений ущербов после заключения сделки отражено введением дополнительных ограничений сверху на их остаточные риски.

Работой, заложившей основы теории дележа риска в страховании, была статья Эрроу [4], где было показано, что при использовании принципа среднего значения для начисления премии оптимальным с точки зрения страхователя является дележ формы франшиза. В статье [6] были найдены Парето-оптимальные дележи между страховщиком и единственным страхователем, было показано, в частности, что максимизация полезности приводит к stop loss дележу, если решение о выборе функции дележа принимает страховщик, и к страхованию с франшизой, если выбирающей стороной является страхователь. В близкой постановке эти результаты были модифицированы на случай принципа среднего значения для вычисления премии в [9].

Оптимальность дележей страхования с франшизой изучалась в [10] с использованием разных типов распределений суммарного риска. В [5] были представлены условия на стоимость страхования, при которых оптимальными оказывались различные типы дележей страхования, такие как франшиза, сострахование и др. Исследование влияния различных типов функции полезности на уровень франшизы при ограничении сверху на риск страховщика проведено в [7]. Еще одним направлением исследований в статических моделях стал поиск оптимальных дележей при так называемом "Value at Risk" ограничении, которое означает установление нижней границы на вероятность разорения. Но, как было показано, например, в [11], учет этого ограничения ведет к тому, что оптимальные стратегии страхования оказываются разрывными функциями. Это означает наличие стимула к искажению действительного значения ущерба (moral hazard), что неприемлемо в страховой практике на развитых рынках. Работа [2] посвящена оптимизации страхования и индивидуального перестрахования при ограничении на среднее значение остаточного риска страхователя. Было показано, что решение такой задачи не единственно, и оптимальные дележи страхования образуют целый класс функций.

В отличие от предыдущих исследований, где разрешен выбор только стратегии стра-

хования или/и перестрахования индивидуальных рисков, в настоящей работе изучается задача одновременной оптимизации дележей страхования и перестрахования суммарного риска страховщика от целой группы клиентов. Интерес именно к суммарному перестрахованию берет начало с работы [6], где было установлено, что лучшим поведением страховщика будет не индивидуальное перестрахование, а создание пула из рисков клиентов с дальнейшим обращением за его перестрахованием. Еще один момент, отличающий предложенную модель, состоит в наличии естественных ограничений, наложенных с вероятностью единица на остаточные риски клиентов и перестраховщика.

Параграф 2 содержит формальное описание модели. В п. 3 рассмотрен частный случай, когда страховщик выбирает лишь стратегию (или, другими словами, дележ) перестрахования суммарного риска, целиком принимая на себя риски страхователей. Показано, что оптимум достигается на stop loss перестраховании с верхним пределом. В п. 4 рассмотрен общий случай оптимизации страхования и перестрахования при ограничениях на остаточные риски. Доказано (теорема 2), что оптимальное перестрахование по-прежнему имеет вид stop loss стратегии с верхним пределом, а оптимальное страхование есть комбинация stop loss страхования и франшизы. Выведена система уравнений оптимальности для двух параметров, входящих в выражения для оптимальных стратегий, а также проведено сравнение нашей модели с перестрахованием суммарного риска и модели с индивидуальным перестрахованием.

2. ОПИСАНИЕ МОДЕЛИ

Рассматривается модель страхового рынка, состоящего из страховщика, перестраховщика и n клиентов ($n \geq 2$). Потенциальные ущербы (риски) клиентов – независимые неотрицательные случайные величины X_j , $j = 1, \dots, n$, определенные на одном вероятностном пространстве $(\Omega, \mathcal{F}, P)$. В дальнейшем эта группа клиентов предполагается однородной: все X_j имеют одинаковое распределение $F(x) \stackrel{\text{def}}{=} P\{X_j \leq x\}$. Через $T = \sup\{\text{supp } F\} \leq \infty$ будем обозначать верхнюю грань носителя распределения F , т. е. множества возможных значений риска X_1 . Страховщик выбирает функцию дележа страхования $I(x)$ и функцию дележа перестрахования $A(x)$ из класса борелевских функций на $[0, \infty)$, удовлетворяющих неравенствам $0 \leq I(x) \leq x$ и $0 \leq A(x) \leq x$,

которые означают, что возмещение не может быть отрицательным и не может превосходить величины ущерба. Случайная величина $I(X_j)$ есть возмещаемая j -му клиенту часть ущерба, $A(\sum_{j=1}^n I(X_j))$ – доля, оплачиваемая страховщиком после перестрахования его суммарного риска $X^I = \sum_{j=1}^n I(X_j)$. Остаток $X^I - A(X^I)$ возмещается перестраховщиком.

Наложим дополнительные ограничения на дележи (I, A) , которые отражали бы желание других участников рынка – страхователей и перестраховщика – защититься от больших потерь, т. е. ввести ограничения сверху на максимальные возможные значения их рисков

(i) остаточный риск клиента после страхования $X_1 - I(X_1) \leq q$,

(ii) риск перестраховщика $X^I - A(X^I) \leq Q$. Здесь q и Q – заданные положительные константы, неравенства понимаются выполненными с вероятностью единица (п.н.). Легко видеть, что эти ограничения можно эквивалентно переписать в форме ограничений снизу на функции дележа страхования и перестрахования: $I(x) \geq x - q$ и $A(x) \geq x - Q$ для $x \in [0, \infty)$.

Премия P , полученная от клиента, и премия P_1 от страховщика за перестрахование определяются по формуле среднего значения, широко используемой в актуарной литературе (см., например, [1]): $P = (1 + \alpha)E I(X_1)$ и $P_1 = (1 + \alpha_1)E [X^I - A(X^I)]$. Здесь суммарный риск страховщика (до перестрахования) $X^I = \sum_{j=1}^n I(X_j)$, а α и α_1 – заданные коэффициенты нагрузки, соответственно, страховщика и перестраховщика, удовлетворяющие стандартным неравенствам $0 < \alpha < \alpha_1$.

Предпочтения страховщика описываются его функцией полезности: риск Y_1 предпочтительнее риска Y_2 тогда и только тогда, когда $E u(Y_1) > E u(Y_2)$, где $u(x)$ – заданная дважды дифференцируемая вогнутая функция полезности: $u' > 0$ и $u'' < 0$. Объектом изучения в данной работе является задача выбора страховщиком оптимальных дележей страхования и перестрахования, где целевым функционалом выступает ожидаемая полезность финального капитала страховщика

$$J[I, A] \stackrel{\text{def}}{=} E u \left(nP - P_1 - A \left(\sum_{j=1}^n I(X_j) \right) \right). \quad (1)$$

С учетом ограничений, введенных на допустимые дележи, эта задача имеет вид

$$\max_{I, A} J[I, A] \quad (2)$$

при ограничениях $(x - q)_+ \leq I(x) \leq x$ и $(x - Q)_+ \leq A(x) \leq x$, где мы обозначили через $(y)_+ = \max\{0, y\}$.

Ниже будем предполагать конечность всех математических ожиданий, указанных в тексте. Это условие не является слишком ограничительным, поскольку существование $E X_1 < \infty$ и $E |u(-n(1 + \alpha_1)E X_1 - \sum_{j=1}^n X_j)| < \infty$ влечет конечность ожидаемой полезности $J[I, A]$ при любых дележах риска. Отметим в заключение, что совпадение дележей $(I, A) = (I', A')$ означает $I(X_1) = I'(X_1)$ и $A(\sum_{j=1}^n I(X_j)) = A'(\sum_{j=1}^n I'(X_j))$ с вероятностью единица (п.н.).

3. ОПТИМИЗАЦИЯ ПЕРЕСТРАХОВАНИЯ

Сначала рассмотрим частный случай описанной выше модели, в котором не предполагается дележ риска между страховщиком и клиентом или, в наших обозначениях, $I(x) \equiv x$ – каждый индивидуальный ущерб страхуется целиком. Страховщик может выбирать лишь дележ перестрахования, и оптимизационная задача (2) сводится к следующей

$$J[A] \equiv E u(nP - P_1 - A(X)) \rightarrow \max \quad (3)$$

при ограничении $(x - Q)_+ \leq A(x) \leq x$,

где

$$\begin{aligned} P &= (1 + \alpha)E X_1, \\ P_1 &= (1 + \alpha_1)\{nE X_1 - E A(X)\} \text{ и} \\ X &= \sum_{j=1}^n X_j. \end{aligned}$$

Теорема 1. Задача (3) имеет единственное решение, stop loss перестрахование с верхним пределом $A^*(x) = (x \wedge a^*) \vee (x - Q)$, где $x \wedge y \stackrel{\text{def}}{=} \min\{x, y\}$ и $x \vee y \stackrel{\text{def}}{=} \max\{x, y\}$. Уровень

$$a^* = \begin{cases} a^0 & \text{при } a^0 < nT, \\ nT & \text{иначе,} \end{cases}$$

здесь $a^0 > 0$ есть минимальный корень уравнения $\psi(a) = 0$,

$$\begin{aligned} \psi(a) &\stackrel{\text{def}}{=} (1 + \alpha_1)E u'(P_a - A_a(X)) \\ &\quad - u'(P_a - a), \end{aligned} \quad (4)$$

$$\text{где } P_a \stackrel{\text{def}}{=} (1 + \alpha_1)E A_a(X) - n(\alpha_1 - \alpha)E X_1,$$

$$A_a(x) \stackrel{\text{def}}{=} (x \wedge a) \vee (x - Q).$$

Доказательство: В силу вогнутости функции полезности $u(\cdot)$, функционал $J[A]$ также вогнут по A . Поэтому дележ A^* оптимален тогда

и только тогда, когда производная по направлению

$$\left. \frac{d}{d\lambda} J[\lambda A^* + (1 - \lambda)A] \right|_{\lambda=1} \geq 0$$

для любого допустимого A . После вычисления производной (см. (1)) получаем, что последнее неравенство означает

$$\int_0^{nT} \eta(x)(A^*(x) - A(x))dF_\Sigma(x) \geq 0,$$

где функция

$$\eta(x) = (1 + \alpha_1)Eu'(nP - P_1^* - A^*(X)) - u'(nP - P_1^* - A^*(x)) \quad (5)$$

и $F_\Sigma(x)$ обозначает функцию распределения суммарного риска X . Таким образом, A^* – решение задачи

$$\max_A \int_0^{nT} \eta(x)A(x)dF_\Sigma(x)$$

на множестве борелевских функций $\{A(x) : (x - Q)_+ \leq A(x) \leq x\}$. Для решения такой задачи применим лемму Неймана-Пирсона (см., например, [3], применения в задачах оптимизации страхования можно найти в [9]): допустимый дележ A^* оптимален тогда и только тогда, когда

$$A^*(x) = \begin{cases} (x - Q)_+ & \text{если } \eta(x) < 0, \\ x, & \text{если } \eta(x) > 0 \end{cases} \quad (6)$$

для всех x с точностью до множества нулевой F_Σ -меры. Поскольку $u'(\cdot)$ убывает, то легко видеть, что $\eta(0) > 0$. При возрастании x от $x = 0$ имеем, в силу (6), что $A^*(x)$ остается равной x и $\eta(x)$ убывает вплоть до точки a^0 , в которой $\eta(x)$ впервые становится равной нулю. На интервале $[a^0, a^0 + Q]$ функция $\eta(x)$ не может ни возрастать ни убывать, поскольку в противном случае мы получили бы противоречие с (6).

Таким образом, $\eta(x)$ убывает на $[0, a^0]$ при этом $A^*(x) = x$, равна нулю на интервале $[a^0, a^0 + Q]$, где $A^*(x) = a^0$. Когда x возрастает от точки $a^0 + Q$, значения $A^*(x)$ лежат на нижней границе $x - Q$. Действительно, если $A^*(x) > x - Q$, то $\eta(x)$ убывает от $0 = \eta(a^0 + Q)$ и, согласно (6), мы имели бы $A^*(x) = x - Q$. Окончательно получаем: оптимальная функция дележа есть stop loss перестрахование с верхним пределом, $(x \wedge a^0) \vee (x - Q)$, на $[0, \infty)$. Для нахождения a^0 , т. е. точки, в которой $\eta(x)$ впервые обращается в ноль, подставим в $\eta(x)$

выражение $(X \wedge a) \vee (X - Q)$ вместо $A^*(X)$ и a вместо $A^*(x)$. Уравнение, определяющее a^0 , примет тогда вид $\psi(a) = 0$, где функция $\psi(a)$ указана в (4). Принимая во внимание вырожденный случай $\psi(a) > 0$ на $[0, nT]$, где правая граница $nT = \sup\{\text{supp } F_\Sigma\}$, приходим к окончательной формуле $A^*(x) = (x \wedge a^*) \vee (x - Q)$ с a^* , определенной в теореме 1. $\square$

Оптимальное перестрахование $A^*(x)$ суммарного риска, найденное в теореме 1, представляет собой двухпараметрическую функцию, которую в литературе иногда называют перестрахованием эксцедента убыточности. Если суммарный ущерб X не превосходит уровень удержания страховщика a^* , то страховщик оплачивает весь ущерб; далее ущерб оплачивается перестраховщиком за вычетом a^* и, наконец, если $X > a^* + Q$, то перестраховщик платит предельную сумму Q , а остаток погашается страховщиком. Таким образом, перестраховщику достается "средняя" доля риска $X - A^*(X) = (X - a^*)_+ \wedge Q$. Константа Q имеет смысл максимальной суммы, которую перестраховщик готов заплатить, поэтому функцию дележа $A^*(x)$ часто называют stop loss перестрахованием с верхним пределом (т. е. верхним пределом Q ответственности перестраховщика).

Замечание 1: В вырожденном случае уравнение $\psi(a) = 0$ может, вообще говоря, не иметь корней на $[0, nT]$, т. е. $\psi(a) > 0$ на $[0, nT]$, где nT верхняя граница носителя F_Σ . Тогда уровень a^* полагаем равным nT , что означает отказ страховщика от перестрахования.

4. ОБЩИЙ СЛУЧАЙ: ОПТИМАЛЬНЫЙ ВЫБОР СТРАХОВАНИЯ И ПЕРЕСТРАХОВАНИЯ

В этом параграфе будет найдено решение задачи (2), которая формулируется здесь в полном объеме с выбором стратегий и страхования индивидуальных рисков, и перестрахования суммарного ущерба

$$J[I, A] \equiv Eu(nP - P_1 - A(X^I)) \rightarrow \max \quad (7)$$

при ограничениях $(x - q)_+ \leq I(x) \leq x$ и $(x - Q)_+ \leq A(x) \leq x$.

Ниже нам понадобится понятие двухпараметрической функции дележа страхования $I(x) = (x \wedge k) \vee (x - q)$, которая является комбинацией stop loss страхования $x \wedge k$ и франшизы $(x - q)_+$. При таком выборе функции $I(x)$ дележа страхования страховщику достается "хвост" распределения риска клиента X_1 , самому же страхователю остается "средняя" доля его риска $X_1 - I(X_1) = (X_1 - k)_+ \wedge q$.

Как будет показано, функция дележа такого типа вместе со stop loss перестрахованием с верхним пределом оптимальны в задаче (7).

Теорема 2. Оптимальные дележи (I^*, A^*) в задаче (7) имеют следующий вид: комбинация stop loss страхования и франшизы $I^*(x) = (x \wedge k^*) \vee (x - q)$, stop loss перестрахование с верхним пределом $A^*(x) = (x \wedge a^*) \vee (x - Q)$. Уровни $k^* = k^0 \wedge T$ и $a^* = a^0 \wedge nT$ удовлетворяют неравенству $0 < k^* < a^*$, параметры k^0 и a^0 являются решением пары уравнений $\psi_1(k, a) = 0$ и $\psi_2(k, a) = 0$, в которых

$$\begin{aligned} \psi_1(k, a) &\stackrel{def}{=} (1 + \alpha) E u'(P_{k,a} - A_a(X_k^I)) \\ &- E [u'(P_{k,a} - A_a(X_k^I)) | X_1 = k], \end{aligned} \quad (8)$$

$$\begin{aligned} \psi_2(k, a) &\stackrel{def}{=} (1 + \alpha_1) E u'(P_{k,a} - A_a(X_k^I)) \\ &- u'(P_{k,a} - a), \end{aligned} \quad (9)$$

где обозначены

$$P_{k,a} = (1 + \alpha_1) E A_a(X_k^I) - n(\alpha_1 - \alpha) E I_k(X_1), \quad A_a(x) = (x \wedge a) \vee (x - Q),$$

$$I_k(x) = (x \wedge k) \vee (x - q), \quad \text{и} \quad X_k^I = \sum_{j=1}^n I_k(X_j).$$

Доказательство: Покажем сначала, что исследуемая задача имеет решение.

Лемма 1. В задаче (7) существует оптимальная пара дележей (I^*, A^*) .

Доказательство: Обозначим через $\{I_n, A_n\}$ максимизирующую последовательность в (7), т. е., $\lim_{n \rightarrow \infty} J[I_n, A_n] = J^* \stackrel{def}{=} \sup_{I, A} J[I, A]$. По

теореме Хелли, существует подпоследовательность $\{I_m(X_1), A_m(X^{I_m})\}$ с $X^{I_m} = \sum_{j=1}^n I_m(X_j)$,

слабо сходящаяся к (δ, ε) . Для доказательства того, что δ и ε есть собственные случайные величины, заметим: $I_m(X_1) \leq X_1$ и $A_m(X^{I_m}) \leq \sum_{j=1}^n X_j$ п.в. Поэтому соответствующие функции распределения $F_m^I(x) \geq F(x)$ и $F_m^A(x) \geq F_\Sigma(x)$, и тогда пределы δ и ε удовлетворяют требуемому неравенству $P\{\delta < \infty, \varepsilon < \infty\} = 1$. Так как δ измерима относительно сигма-алгебры $\sigma(X_1)$, порожденной X_1 , и $(X_1 - q)_+ \leq \delta \leq X_1$ п.н., то предел δ может быть представлен как $\delta = I^*(X_1)$ для некоторой допустимой функции дележа страхования $I^*(x)$. Аналогично, другой предел ε представим в виде $\varepsilon = A^*(\sum_{j=1}^n I^*(X_j))$ для неко-

торой допустимой функции дележа перестрахования $A^*(x)$. Наконец, тот факт, что $J^* =$

$J[I^*, A^*]$, следует из доказанной слабой сходимости, непрерывности функции полезности u , и конечности ожидаемой полезности $J[I^*, A^*]$. $\square$

Обозначим через (I^*, A^*) решение (7), существующее в силу леммы 1. Так как A^* решает задачу $\max_A J[I^*, A]$, то, в силу теоремы 1 (где X заменен на $X^* = \sum_{j=1}^n I^*(X_j)$),

$A^*(x) = (x \wedge a^*) \vee (x - Q)$ для подходящего уровня a^* .

Рассмотрим задачу $\max_I J[I, A^*]$. Так же, как в доказательстве теоремы 1, определим $I_\rho = \rho I^* + (1 - \rho)I$. Тогда необходимым условием оптимальности I^* будет

$$\frac{d}{d\rho} J[I_\rho, A^*]|_{\rho=1} \geq 0$$

для любого допустимого I . Выражение для $J[I, A^*]$ задано в (1), и после дифференцирования будем иметь

$$\begin{aligned} &\int_0^T \dots \int_0^T \xi(x_1, \dots, x_n) \left[\sum_1^n I^*(x_j) \right. \\ &\left. - \sum_1^n I(x_j) \right] dF(x_1) \dots dF(x_n) \geq 0, \end{aligned}$$

где функция

$$\begin{aligned} \xi(\bar{x}) &= \mathbf{1}(\bar{x}) [(1 + \alpha_1) E u'(nP^* - P_1^* - A^*(X^*)) \\ &- u'(nP^* - P_1^* - A^*(\sum_1^n I^*(x_j)))] \\ &- (\alpha_1 - \alpha) E u'(nP^* - P_1^* - A^*(X^*)). \end{aligned} \quad (10)$$

Здесь $\mathbf{1}(\bar{x}) = \mathbf{1}\{\sum_1^n I^*(x_j) \in [0, a^*] \cup [a^* + Q, \infty)\}$

и $\mathbf{1}\{\cdot\}$ обозначает индикаторную функцию.

Для анализа выведенного условия оптимальности введем новую функцию $\xi_\Sigma(\cdot)$ от скалярного аргумента соотношением $\xi(\bar{x}) \equiv \xi_\Sigma(\sum_1^n I^*(x_j))$. В определенном смысле она соответствует модели с единственным "коллективным" страхователем. Теперь мы можем записать приведенное выше неравенство в виде $\int_0^T \theta(x) (I^*(x) - I(x)) dF(x) \geq 0$, где

$$\theta(x) = E \xi_\Sigma(I^*(x) + \sum_{j=2}^n I^*(X_j)), \quad (11)$$

и математическое ожидание берется по случайным переменным $X_2, \dots, X_n$. Другими сло-

вами, оптимальный дележ I^* есть решение задачи

$$\max_0^T \int \theta(x) I(x) dF(x)$$

при ограничении $(x - q)_+ \leq I(x) \leq x$. (12)

Докажем теперь, что единственное решение в (12) имеет вид $I^*(x) = (x \wedge k^*) \vee (x - q)$, где уровень $0 < k^* < a^*$.

Применяя лемму Неймана-Пирсона, получаем, что допустимая функция I^* решает (12) тогда и только тогда, когда

$$I^*(x) = \begin{cases} (x - q)_+ & \theta(x) < 0, \\ x, & \theta(x) > 0 \end{cases} \quad (13)$$

с точностью до множества нулевой F -меры. Функция $\theta(x)$, определенная в (11), есть усреднение $\xi_\Sigma(I^*(x) + z)$ по распределению $G(z)$ случайной величины $Z = \sum_{j=1}^n I^*(X_j)$. После этого усреднения по z получаем, что $\theta(0) = (1 + \alpha)Eu'(nP^* - P_1^* - A^*(X^*)) -$

$$-E \left[u'(nP^* - P_1^* - A^*(x + \sum_{j=1}^n I^*(X_j))) \right] \Big|_{x=0} > 0,$$

поскольку $u'(\cdot)$ убывает и $A^*(\cdot)$ есть неубывающая функция (вида $A^*(x) = (x \wedge a^*) \vee (x - Q)$). При возрастании x от точки $x = 0$, функция $\theta(x)$ убывает от положительного значения $\theta(0) > 0$, поскольку для этих x значение $I^*(x) = x$, как следует из (13).

Докажем, что точка k^0 , в которой $\theta(x)$ первый раз попадает в ноль, лежит левее a^* . В оптимальном перестраховании $A^*(x)$ уровень $a^* = a^0 \wedge nT$, где (см. доказательство теоремы 1) a^0 есть минимальный нуль функции $\eta(x)$, определенной в (5), при $A^*(x) = x$ для $x \leq a^0$. Поскольку коэффициент нагрузки перестраховщика $\alpha_1 > \alpha$, легко видеть, что $\eta(x) > \xi_\Sigma(x) = (1 + \alpha)Eu'(nP^* - P_1^* - A^*(X^*)) - u'(nP^* - P_1^* - A^*(x))$ для x от 0 до точки K_Σ , нуля функции $\xi_\Sigma(x)$. Следовательно, $K_\Sigma < a^0$ и тогда наименьший нуль функции $\xi_\Sigma(x + z)$ при фиксированном $z > 0$ есть $K_\Sigma - z < a^0$. После усреднения по z и получения $\theta(x) = E \xi_\Sigma(I^*(x) + Z)$ такое положение вещей сохранится: k^0 – наименьший нуль $\theta(x)$ – лежит левее точки $K_\Sigma < a^0$.

Когда x пробегает интервал $[k^0, k^0 + q)$, функция $\theta(x)$ не убывает и не возрастает, оставаясь нулевой, при этом $I^*(x)$ равна константе k^0 . Действительно, если $I^*(x)$ убывает, то из определения (10)–(11) следует, что $\theta(x)$ возрастает от нуля. Тогда, согласно (13), $I^*(x) = x$ – мы имеем противоречие. Пусть $I^*(x)$ возрастает, тогда $\theta(x)$ убывает от $\theta(k^0) = 0$, становясь отрицательной. В силу (13), $I^*(x) =$

$(x - q)_+ < I^*(k^0) = k^0$, и мы получаем противоречие с предположением о возрастании $I^*(x)$.

При возрастании x от $k^0 + q$, значения $I^*(x)$ равны $x - q$, нижней границе допустимых значений. Предположим противное: $I^*(x) > x - q$. Как было показано выше, это возрастание от $I^*(k^0 + q) = k^0$ ведет к тому, что $\theta(x) < 0$. Тогда из (13) получаем, что $I^*(x)$ должна совпадать с $x - q$.

Таким образом, единственный оптимальный дележ в задаче (12) имеет вид $I^*(x) = (x \wedge k^0) \vee (x - q)$ на $[0, \infty)$, где $0 < k^0 < a^* = a^0 \wedge nT$.

Для того чтобы найти точку k^0 , в которой значение $\theta(x)$ первый раз становится нулевым, подставим в (11) найденное выражение $(X_j \wedge k) \vee (X_j - q)$ вместо $I^*(X_j)$, $j = 2, \dots, n$ и k вместо $I^*(x)$. Тогда k^0 будет минимальным корнем уравнения $\psi_1(k, a^*) = 0$, где функция $\psi_1(k, a)$ задана выражением (8). Осталось рассмотреть вырожденный случай: если $\psi_1(k, a^*) > 0$ на $[0, T)$, то полагаем $k^* = T$. Это означает, что $I^*(X_1) = X_1$ п.н. – страховщик оплачивает весь ущерб клиента.

Для нахождения уровня a^* в оптимальном перестраховании $A^*(x) = (x \wedge a^*) \vee (x - Q)$, обратимся к теореме 1. Было показано, что в задаче оптимизации перестрахования уровень $a^* = a^0 \wedge nT$, где a^0 является корнем уравнения $\psi(a) = 0$ с функцией $\psi(a)$, определенной в (4). В нашем случае при использовании дележа страхования вида $I_k(x) = (x \wedge k) \vee (x - q)$, суммарный риск $X = \sum_{j=1}^n X_j$ заменяется в (4)

на $X_k^I = \sum_{j=1}^n I_k(X_j)$. В результате приходим к

уравнению $\psi_2(k, a) = 0$, где $\psi_2(k, a)$ определена в (9). $\square$

Найденные в теореме 2 стратегии I^* и A^* применяются, соответственно, к каждому индивидуальному риску X_j , $j = 1, \dots, n$ и к суммарному ущербу $X^* = \sum_{j=1}^n I^*(X_j)$, погасить

который обязался перед клиентами страховщик. В этом принципиальное отличие от модели с выбором индивидуального страхования и индивидуального перестрахования в [2], где, как и здесь, принятый страховщиком от клиента риск есть с. в. $I^*(X_j)$, но суммарная доля страховщика после перестрахования равна $\sum_{j=1}^n A^*(I^*(X_j))$, в то время как в нашем случае

это $A^*(\sum_{j=1}^n I^*(X_j))$. Если выбрать подходящую

аппроксимацию для $\sum_{j=1}^n I^*(X_j)$ (пусть, для

определенности, нормальную), то доля страховщика окажется кусочно-линейным преобразованием $A^*(\cdot)$ от нормальной с.в. В случае же индивидуального перестрахования, к доле страховщика $\sum_{j=1}^n A^*(I^*(X_j))$ непосредственно применяется нормальная аппроксимация.

Замечание 2. Найденные в теореме 2 оптимальные дележи (I^*, A^*) допускают, вообще говоря, два вырожденных случая: Если уровень удержания страховщика $k^* \geq T = \sup \text{supp} F$ тогда, очевидно, $I^*(X_1) = X_1$ п.н., т. е. страховщик берет ответственность за весь риск клиента – дележ риска ему не выгоден. После дележа страхования верхняя грань носителя суммарного риска $X^* = \sum_{j=1}^n I^*(X_j)$, принятого страховщиком, становится равной $nT^* = n(T \wedge k^*) \vee (T - q) \leq nT$. Если в дележе перестрахования найденный уровень a^* оказался не меньше, чем эта граница, а именно, $a^* \geq nT^*$, то $A^*(X^*) = (X^* \wedge a^*) \vee (X^* - Q) = X^*$ п.н. – страховщик не обращается за перестрахованием.

Представляется интересным сравнить оптимальные дележи в случае индивидуального перестрахования, где финальный капитал страховщика $nP - nP^1 - \sum_{j=1}^n A(I(X_j))$ с премией перестраховщику $P^1 = (1 + \alpha_1)E[I(X_1) - A(I(X_1))]$, и в нашем случае суммарного перестрахования, где финальный капитал $nP - P_1 - A(\sum_{j=1}^n I(X_j))$. Для простоты рассмотрим модели без дополнительных ограничений, где допустимыми считаются все дележи, такие что $0 \leq I(x) \leq x$ и $0 \leq A(x) \leq x$. Как было показано в [2], оптимальными дележами в первом случае будут stop loss стратегии $I^*(x) = x \wedge k^*$ и $A^*(x) = x \wedge a^*$ причем, в силу неравенства $\alpha < \alpha_1$, всегда выполнено $k^* \leq a^*$. Это означает, что риск страховщика после перестрахования $A^*(I^*(X_1)) = I^*(X_1)$ п.н. – перестрахование не востребовано. Однако в случае перестрахования суммарного риска, такой отказ от перестрахования необязательно имеет место. Из теоремы 2, формально полагая $q = Q = \infty$, получаем $I^*(x) = x \wedge k^*$ и $A^*(x) = x \wedge a^*$, где $k^* < a^*$. Но a^* – уровень удержания не в индивидуальном перестраховании, а в перестраховании суммарного риска $X^* = \sum_{j=1}^n X_j \wedge k^*$, верхняя граница возможных значений которого равна nk^* . Поэтому для $n > 1$ возможна ситуация, когда $nk^* > a^*$ и, следовательно, доля страховщи-

ка после перестрахования $A^*(X^*) = X^* \wedge a^*$ не совпадает с X^* – перестрахование оказывается выгодным. Следующее ниже предложение устанавливает, что такая ситуация имеет место, когда разница между коэффициентами нагрузки страховщика и перестраховщика, где $\alpha < \alpha_1$, "не слишком большая" т. е. когда перестрахование "не слишком" дорого.

Предложение 1. Для задачи $\max_{I, A} J[I, A]$ без дополнительных ограничений в невырожденном случае, когда $k^* < T$, существует $\alpha' > \alpha$ такой, что для любого $\alpha_1 \in [\alpha, \alpha']$ неравенство $nk^* > a^*$ выполняется.

Доказательство: Как следует из доказательства теоремы 2 (где теперь $q = Q = \infty$), уровень $k^* < T$ является наименьшим корнем уравнения $\theta(x) = 0$, в котором (см. (11)) $\theta(x) = E \xi_\Sigma(I^*(x) + Z)$, $I^*(x) = x \wedge k^*$ и $Z = \sum_{j=1}^n I^*(X_j)$. Функцию распределения с.в. Z обозначим через $G(z)$. Было показано, что $k^* < K_\Sigma < a^*$, где K_Σ есть точка первого обнуления $\xi_\Sigma(x)$. С другой стороны, k^* может быть также определена как наименьший нуль функции $E \xi_\Sigma(x + Z)$.

Докажем, что $nk^* > K_\Sigma$. Предположим противное, $nk^* \leq K_\Sigma$. Так как $I^*(x) = x \wedge k^*$, имеем $\sup\{\text{supp} G\} = (n-1)k^*$ и, по предположению, $\sup\{\text{supp} G\} \leq K_\Sigma - k^*$. Как было отмечено, k^* – наименьший нуль $E \xi_\Sigma(x + Z)$, поэтому $k^* > K_\Sigma - \sup\{\text{supp} G\} \geq K_\Sigma - (K_\Sigma - k^*) = k^*$. Полученное противоречие доказывает требуемое неравенство $nk^* > K_\Sigma$.

Принимая во внимание то, что по определению функции $\xi_\Sigma(x)$ точка K_Σ совпадает с a^* , если $\alpha_1 = \alpha$, из непрерывности $\xi_\Sigma(\cdot)$ получаем существование $\alpha' > \alpha$ такого, что $nk^* > a^*$ при любом $\alpha_1 \in [\alpha, \alpha']$. $\square$

Обоснование такого эффекта выгоды перестрахования именно суммарного риска состоит, как было отмечено выше, в том, что лучшая стратегия страховщика – это не перестрахование отдельных рисков, а образование пула $\sum_{j=1}^n I^*(X_j)$ из них и затем обращение за перестрахованием этого суммарного риска. Отметим в заключение, что в случае, когда дополнительное ограничение $0 \leq A(x) \leq (x - Q)_+$ наложено на дележи перестрахования (допустимые I по-прежнему такие, что $0 \leq I(x) \leq x$), утверждение предложения 1 остается в силе. Доказательство получается повторением рассуждений в доказательстве предложения 1. В последнем случае, однако, сужение множества допустимых дележей A ведет к меньшему, вообще говоря, значению α'

правой границы интервала $[\alpha, \alpha']$ значений коэффициента α_1 нагрузки перестрахования.

Пример. Рассмотрим задачу нахождения пары оптимальных дележей (I^*, A^*) в случае экспоненциальной функции полезности страховщика $u(x) = c^{-1}(1 - \exp(-cx))$ с заданным коэффициентом неприятия риска $c = 0, 1$. Пусть численность группы клиентов $n = 15$, распределение $F(x)$ индивидуального ущерба равномерно на $[0, 10]$, верхние границы рисков $q = 7$ и $Q = 50$, коэффициент нагрузки страховщика $\alpha = 0, 5$. Для моделирования распределения суммарного риска страховщика $\sum_{j=1}^n I(X_j)$ мы используем усеченное нормальное распределение, которое довольно часто применяется в моделях страхования (см., например, [1]). Его плотность есть

$$1\{x \geq 0\} \frac{\exp(-\frac{(x-m)^2}{2\sigma^2})}{\int_0^\infty \exp(-\frac{(x-m)^2}{2\sigma^2})dx}.$$

По теореме 2, оптимальные дележи в задаче (7) максимизации полезности страховщика равны $I^*(x) = (x \wedge k^*) \vee (x - q)$ и $A^*(x) = (x \wedge a^*) \vee (x - Q)$. Уровни k^* и a^* определяются из системы двух уравнений: $\psi_1(k, a) = 0$ и $\psi_2(k, a) = 0$ (см. (8)–(9)). В данном случае, когда $u'(x) = e^{-cx}$, эти уравнения имеют вид

$$(1 + \alpha)E \exp[cA_a(X_k^I)] - E \exp[cA_a(k + \sum_{j=2}^n I_k(X_j))] = 0 \quad (14)$$

$$(1 + \alpha_1)E \exp[cA_a(X_k^I)] - \exp[ca] = 0, \quad (15)$$

где $A_a(x) = (x \wedge a) \vee (x - Q)$, $I_k(x) = (x \wedge k) \vee (x - q)$, и $X_k^I = \sum_{j=1}^n I_k(X_j)$. По предполо-

жению, случайные величины X_k^I и $\sum_{j=2}^n I_k(X_j)$ имеют, соответственно, усеченные нормальные распределения $\Phi_{m, \sigma}^{tr}(x)$ и $\Phi_{m_1, \sigma_1}^{tr}(x)$ с параметрами $m = nE I_k(X_1)$, $\sigma^2 = nVar I_k(X_1)$, и $m_1 = (n - 1)E I_k(X_1)$, $\sigma_1^2 = (n - 1)Var I_k(X_1)$. Формулы для первых двух моментов $I_k(X_1) = (X_1 \wedge k) \vee (X_1 - q)$ легко выводятся:

$$E I_k(X_1) = \int_0^k \bar{F}(x)dx + \int_{k+q}^\infty \bar{F}(x)dx,$$

$$E I_k^2(X_1) = 2[\int_0^k \bar{F}(x)xdx + \int_{k+q}^\infty \bar{F}(x)(x-q)dx].$$

Численное решение уравнений оптимальности (14)–(15) после подстановки равномерного распределения $F(x)$ дает: при значениях коэффициента нагрузки перестраховщика $\alpha_1 = 0, 6$

и $\alpha_1 = 2$ оптимальные уровни равны, соответственно, $(k^*, a^*) = (3, 446, 12, 011)$ и $(k^*, a^*) = (1, 149, 52, 175)$ — удорожание перестрахования приводит к уменьшению доли страхуемого риска $I^*(X_j)$ и увеличению риска страховщика $A(X^*)$ после перестрахования.

ЗАКЛЮЧЕНИЕ

В статье рассмотрена задача одновременного выбора оптимального дележа риска между страховщиком и отдельным клиентом и оптимального дележа суммарного риска страховщика между ним и перестраховочной компанией. В аналитическом виде получены функции дележа — ими оказались кусочно-линейные функции специального вида — и найдены уравнения для определения параметров этих дележей. В качестве критерия была использована ожидаемая полезность страховщика. Особое внимание уделено сравнению оптимальных вариантов в рамках схем индивидуального и суммарного перестрахования.

Работа выполнена при поддержке РФФИ, проект 12-01-00078.

ЛИТЕРАТУРА

1. Бауэрс Н., Гербер Х., Джонс Д. и др. Актуарная математика. М.: Янус-К, 2001. 599 с.
2. Голубин А. Ю., Гридин В. Н., Газов А. И. Оптимизация дележа риска в статической модели с перестрахованием // Автоматика и телемеханика. 2009. Т. 70, № 8. С. 133–144.
3. Леман Э. Проверка статистических гипотез. М.: Наука, 1964. 369 с.
4. Arrow K. J. Essays in the Theory of Risk Bearing. Chicago: Wyley and Sons, 1971. 278 p.
5. Blazenko G. Optimal Indemnity Contracts // Insurance: Mathematics and Economics. 1985. Vol. 4. P. 267–278.
6. Raviv A. The Design of an Optimal Insurance Policy // Amer. Economic Review. 1979. P. 84–96.
7. Cummins J., Mahul O. The Demand for Insurance with an Upper Limit on Coverage // Journal of Risk and Insurance. 2004. Vol. 71(2). P. 253–264.
8. Golubin A. Y. Pareto-optimal Insurance Policies in the Models with a Premium Based on the Actuarial Value // Journal of Risk and Insurance. 2006. Vol. 73, N 3. P. 469–487.
9. Murray M. L. A Deductible Selection Model: Development and Applications // Journal of Risk and Insurance. 1971. Vol. 38. P. 423–436.

10. *Schlesinger H.* The Optimal Level of Deductibility in Insurance Contracts // Journal of Risk and Insurance. 1981. Vol. 48. P. 465–481.

11. *Zhou C., Wu C.* Optimal Insurance Under the Insurer's VaR Constraint // The Geneva Risk and Insurance Review. 2009. Vol. 34. P. 140–154.

СВЕДЕНИЯ ОБ АВТОРАХ:

Гридин Владимир Николаевич

профессор, д. т. н.

Центр информационных технологий в проектировании

РАН, Московская обл., г. Одинцово

эл. почта: info@ditc.ras.ru

тел.: (495) 596 02 19

Gridin, Vladimir

Russian Academy of Sciences, Design Information Technologies Center

7-A Marshala Biryuzova St., Odintsovo, Moscow region, 143000, Russia

e-mail: info@ditc.ras.ru

tel.: (495) 596 02 19

Голубин Алексей Юрьевич

доцент, к. ф.-м. н.

Московский институт электроники и математики Национального исследовательского университета «Высшая школа экономики»

Б. Трехсвятительский пер., 3, Москва, 109028

эл. почта: io@miem.edu.ru

тел.: (495) 916 88 13

Golubin, Alexey

National Research University Higher School of Economics

3 B. Trechsvjatitsky St., Moscow, 109028, Russia

e-mail: io@miem.edu.ru

tel.: (495) 916 88 13

Петрова Мария Николаевна

аспирантка

Московский институт электроники и математики Национального исследовательского университета «Высшая школа экономики»

Б. Трехсвятительский пер., 3, Москва, 109028

эл. почта: io@miem.edu.ru

тел.: (495) 916 88 13

Petrova, Maria

National Research University Higher School of Economics

3 B. Trechsvjatitsky St., Moscow, 109028, Russia

e-mail: io@miem.edu.ru

tel.: (495) 916 88 13

УДК 004.9

КРИТЕРИЙ ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАТЕЛЬНЫХ МУЛЬТИМЕДИЙНЫХ ПОТОКОВ, ТРАНСЛИРУЕМЫХ В РЕАЛЬНОМ РЕЖИМЕ ВРЕМЕНИ

А. Л. Забровский

Петрозаводский государственный университет

В данной статье представлен критерий оценки качества образовательных мультимедийных потоков, транслируемых в реальном режиме времени в IP сети. Данный критерий используется для определения величины битового потока (битрейта), удовлетворяющего конечных пользователей.

Ключевые слова: критерий оценки качества, мультимедийный поток, битрейт.

A. L. Zabrovskiy. THE CRITERION OF QUALITY ASSESSMENT OF EDUCATIONAL MULTIMEDIA STREAMS TRANSMITTED IN REAL TIME MODE

The criterion of quality assessment of educational multimedia streams transmitted in real time mode is presented. It gives us information about the ability of each client to receive a certain multimedia stream (bit rate) within a definite time.

Key words: criterion of quality assessment, multimedia stream, bit rate.

ВВЕДЕНИЕ

Трансляция образовательных мультимедийных потоков в реальном режиме времени на сегодняшний день активно используется во многих университетах мира. Потокковая передача осуществляется как в локальных сетях учебных заведений, сетях уровня города, так и между территориально разделенными образовательными учреждениями по сети Интернет. Образовательное потоковое видео становится одним из основных видов дистанционного взаимодействия участников образовательного процесса.

Основным отличием образовательного видео от других типов видео зачастую является его относительная статичность сюжета.

Данное видео обычно включает в себя такие объекты, как изображение выступающего и информационные слайды. Образовательное видео должно обеспечивать хорошую читабельность символов и букв, не упускать произносимые фразы и слова. Для организации подобного рода взаимодействия, как трансляция потокового видео, используются следующие три основных компонента: кодер, потоковый сервер и клиент [1].

В свою очередь, во время передачи мультимедиа потоков по сети в режиме реального времени (например, трансляций лекций или выступлений с конференций) медиа потоки подвергаются влиянию различных факторов, которые могут ухудшать качество воспроизведения на стороне клиента [2].

Поэтому основная задача, которая стоит перед администратором системы трансляций, – это отправка на сервер потоков с выбранными битовыми скоростями, которые называются битовый поток или битрейт и определяются как величина потока данных, передаваемого в реальном режиме времени. Битрейт выражается в битах в секунду.

На сегодняшний день тестирование способности удаленных клиентов получить тот или иной мультимедийный поток либо совсем не производится, либо происходит по следующему сценарию: удаленные пользователи получают от администратора ссылку на web-страницу, где осуществляется тестовая трансляция мультимедийного потока, с каким-то одним битрейтом. Пользователи могут запустить плеер, посмотреть изображение и сделать самостоятельно вывод о качестве принимаемого потока. При этом время, в которое пользователи заходят на web-страничку, является произвольным и не может точно характеризовать состояние загрузки Интернет канала. То есть, тестирование в данном случае не может правильно оценить возможности удаленных клиентов получить тот или иной медиа поток в заданное время.

Существует технология динамического переключения битрейтов медиа потоков, при возникновении, например, проблемы с сетью (большая задержка, потери пакетов и др.). Данная технология поддерживается различными потоковыми медиа серверами. Недостатком данного метода является то, что администратор должен создавать множество потоков с разными битовыми скоростями, при этом в некоторых случаях мультимедийные потоки с малыми битовыми скоростями могут быть неприемлемы для трансляции определенных мероприятий [3, 4].

В свою очередь, во многих случаях важно заранее получить достоверную информацию о возможностях клиентов получить тот или иной поток, чтобы при необходимости можно было заблаговременно исправить проблемы, вызывающие ухудшения качества мультимедийного потока. Предлагаемый критерий оценки видеопотока как раз предназначен для решения задачи определения величины битового потока, удовлетворяющего конечных пользователей. Предполагается использование данного критерия оценки в тех случаях, когда необходимо обеспечить приемлемое качество передаваемого потока. Программное обеспечение, использующее данный критерий оценки, может автоматически и заранее предоставить администратору информацию о способностях

всех клиентов получать те или иные потоки. Его можно использовать для оценки качества в технологиях динамического переключения битрейтов [5].

ОПИСАНИЕ ВХОДНЫХ ДАННЫХ

Определение величины потоков в данной работе основано на анализе параметров, полученных от плееров удаленных пользователей [6].

Были выбраны четыре характеристики T_{start} , F_{min} , F_{drop} , B_{min} , которые представлены ниже.

- T_{start} – время начала воспроизведения. Время, при котором первое значение количества воспроизводимых видеок кадров в секунду больше 24.

- F_{min} – минимальное количество кадров в секунду, которое было зафиксировано в течение воспроизведения мультимедиа потока. Стандартное значение количества кадров в секунду FPS равняется 25 кадрам в секунду для оригинального потока, но в некоторых случаях оно может уменьшаться, например, если процессор компьютера загружен и не успевает отображать все кадры, соответственно, происходит потеря кадров.

- F_{drop} – максимальный скачок потери кадров, который был зафиксирован в течение воспроизведения мультимедийного потока.

- B_{min} – минимальный размер буфера в секундах, который был зафиксирован в течение всего воспроизведения мультимедийного потока.

Возникает задача: можно ли по этим четырем характеристикам определить качество мультимедийного потока.

Для решения этой задачи было решено использовать дискриминантный анализ. На основании опытных данных требуется найти функцию

$$y = G(T_{start}, F_{min}, F_{drop}, B_{min}),$$

которая принимает положительные значения, когда качество видеопотока хорошее и отрицательные значения в противоположном случае.

ЭКСПЕРИМЕНТЫ

Проводя натурные эксперименты, было получено всего 15 значений переменных, что явно недостаточно для построения дискриминантной функции, поэтому было принято решение построить систему имитационного моделирования (плеер) для определения ха-

рактических характеристик T_{start} , F_{min} , F_{drop} , B_{min} . На качество мультимедийного потока оказывают существенное влияние различные параметры состояния сети, такие как потери пакетов, задержки сети и др. Большое значение также имеет текущее состояние компьютера на стороне клиента.

Разработанный плеер запрограммирован таким образом, что способен отправлять различные характеристики состояния воспроизводимого мультимедийного потока (потери видеокадров, уровень буфера и др.) в реальном режиме времени в удаленную базу данных (рис. 1).

В табл. 1 представлены шесть влияющих факторов, воздействие которых было исследовано в ходе проведенных в данной работе экспериментов. Третья колонка таблицы указывает на место, в котором происходило эмулирование того или иного воздействия на мультимедийный видеопоток.

Представленные в табл. 1 параметры оказывают влияние на значения состояния мультимедийного потока, воспроизводимого плеером, такие как:

- *currentFPS* – количество воспроизводимых кадров в секунду;
- *videoBufferByteLength* – количество байт видеоданных в буфере плеера;
- *droppedFrames* – количество отброшенных кадров. На основе значений состояния мультимедийного потока определялись характеристики: T_{start} , F_{min} , F_{drop} , B_{min} . То есть:

$$T_{start} = f_1(x_1, x_2, x_3, x_4, x_5, x_6);$$

$$F_{min} = f_2(x_1, x_2, x_3, x_4, x_5, x_6);$$

$$F_{drop} = f_3(x_1, x_2, x_3, x_4, x_5, x_6);$$

$$B_{min} = f_4(x_1, x_2, x_3, x_4, x_5, x_6).$$

На основании проведенных экспериментов были сформированы две выборки: обучающая и контрольная. Обучающая выборка включала данные по 129 экспериментам, а контрольная выборка по 66. Обучающая выборка содержит данные, полученные как в реальных условиях, так и при эмулировании.

В ходе проведения экспериментов производилась визуальная оценка качества воспроизводимого мультимедийного видеопотока на стороне клиента. Данная оценка отражает мнение эксперта о качестве мультимедийного потока и имеет 2 уровня: *Good* (хорошее качество) и *Bad* (плохое качество). Если во время

просмотра потока появлялись какие-либо задержки или длительные замирания, качество потока считалось как “плохое”, в ином случае “хорошее”. Отрывок исходных данных приведен в табл. 2.

Полученные на обучающей и контрольной выборке результаты обобщены в табл. 3. Два первых столбца табл. 3 характеризуют отдельные изменения влияющих факторов.

Количество задержек и их продолжительность фиксировалась экспериментатором. Поток считается хорошего качества (помечается как “хорошее”), если в течение $T_{exp} = 182$ секунд не было никаких задержек и остановок видео. Кроме визуальной оценки осуществлялась автоматическая передача характеристик состояния воспроизводимого потока самим плеером в удаленную базу данных каждые две секунды. С учетом того, что продолжительность эксперимента T_{exp} равна или чуть больше 182 секунд, то для каждого эксперимента была автоматически создана минимум девяносто одна запись в таблице базы данных. Также для экспериментов плеер был запрограммирован таким образом, чтобы перед началом воспроизведения видеопотока он накапливал буфер данных $B_{start} = 10$ секунд и затем начинал отображение видеопотока. Если буфер плеера не успевал заполниться полностью за это время, то время начала воспроизведения T_{start} увеличивалось. Параметры F_{min} , F_{drop} , B_{min} вычисляются с двадцатой секунды, после нажатия кнопки проигрывания плеера, с интервалом в две секунды.

Для организации трансляций мультимедийных потоков, было установлено и настроено следующее программное обеспечение.

- Кодер: программное обеспечение WireCast 4.1.2
- Поточный медиа сервер: Adobe Flash Media Streaming Server 4.0.
- Клиент: браузер Internet Explorer 9.0 с Flash player 10.3.183 (плеер Strobe Media Playback).

В качестве эмулятора сетевых характеристик применялось программное обеспечение Wide Area Network Emulator (WANEM), которое находилось между потоковым сервером и клиентом. В качестве источника сигнала на входе кодера использовалось образовательное видео, специально созданное для экспериментов. Характеристики образовательного видеопотока представлены в табл. 4.



Рис. 1. Медиа плеер посылает различные характеристики в удаленную базу данных

Таблица 1. Влияющие параметры

Параметр	Название параметра	Место воздействия
x_1	задержка	сеть
x_2	потеря пакетов	сеть
x_3	джиттер	сеть
x_4	пропускная способность	сеть
x_5	загрузка процессора	компьютер клиента
x_6	загрузка оперативной памяти	компьютер клиента

Таблица 2. Отрывок таблицы исходных данных

<i>Experiment</i>	<i>Bit_{rate}</i>	<i>T_{start}</i>	<i>F_{min}</i>	<i>F_{drop}</i>	<i>B_{min}</i>	<i>Quality</i>
Experiment 40	1500	30	0	2	0	BAD
Experiment 41	100	12	25	0	6,41	GOOD
Experiment 42	100	12	25	0	6,17	GOOD
Experiment 43	100	12	24	0	6,43	GOOD
Experiment 44	100	12	25	0	4,61	GOOD
Experiment 45	100	20	25	0	5,81	GOOD
Experiment 46	200	12	24	0	6,33	GOOD
Experiment 47	200	10	24	0	5,59	GOOD
Experiment 48	200	10	19	0	6,34	BAD
Experiment 49	200	14	0	0	0	BAD
Experiment 50	200	14	0	0	0	BAD

Таблица 3. Обучающая и контрольная выборки

Условия эксперимента	Значения параметров	Количество экспериментов	Число совпадений
Обучающая выборка			
Идеальные условия	—	15	15
Эмуляция потери пакетов	2 %, 4 %, 8 %, 12 %, 16 %	25	25
Эмуляция сетевой задержки	200 мс, 400 мс, 600 мс, 800 мс, 1000 мс	25	24
Загрузка процессора	—	15	15
Загрузка процессора и памяти	—	15	15
Эмуляция сетевой задержки, потери пакетов и джиттера	Задержка: 60 мс. Джиттер: 5 % Потери пакетов: 1–15 %	34	34
Контрольная выборка			
Эмуляция сетевой задержки и потери пакетов	Задержка: 50 мс, 100 мс, 150 мс Потери пакетов: 0 %, 1 %, 2 %, 3 %	60	60
Использование критерия в реальных условиях	—	6	6

Таблица 4. Характеристики образовательного видеофайла

Параметр	Значение
Разрешение	640 x 360
Кодек	H.264
Частота ключевых кадров	5 секунд
Длительность	182 секунды
Количество кадров в секунду	25 кадров

Были проведены эксперименты двух типов: 1. Без каких-либо воздействий на передаваемый поток (рис. 2). При проведении данного типа экспериментов исследовались значения, которые описывают систему в условиях, максимально приближенных к идеальным условиям. Влиянием параметров x_1, x_2, x_3, x_4, x_5 и x_6 в данном случае можно пренебречь.

2. С эмуляцией различных сетевых воздействий на поток с помощью программного обеспечения WANEM. Сюда также включаются эксперименты по исследованию влияния загрузки процессора и памяти на стороне клиента (рис. 3). Именно влияющие факторы x_1, x_2, x_3, x_4, x_5 и x_6 используются в экспериментах второго типа.

Все эксперименты проводились для пяти потоков, с разными битовыми скоростями, которые представлены ниже. При этом максимальная полоса пропускания сетевого канала

для всех экспериментов была ограничена величиной 2048 Кбит/с.

- 148 Кбит/с (100 Кбит/с видео и 48 Кбит/с аудио);
- 248 Кбит/с (200 Кбит/с видео и 48 Кбит/с аудио);
- 548 Кбит/с (500 Кбит/с видео и 48 Кбит/с аудио);
- 1048 Кбит/с (1000 Кбит/с видео и 48 Кбит/с аудио);
- 1548 Кбит/с (1500 Кбит/с видео и 48 Кбит/с аудио).

С математической точки зрения задачу можно описать следующим образом. Существует два класса принадлежности объектов K_1 и K_2 . Класс K_1 содержит объекты хорошего качества, а класс K_2 плохого качества. В качестве дискриминантной функции будем использовать функцию вот такого вида:

$$y = G(T_{start}, F_{min}, F_{drop}, B_{min})$$

$$= \begin{cases} \geq 0, & G(T_{start}, F_{min}, F_{drop}, B_{min}) \in K_1 \\ < 0, & G(T_{start}, F_{min}, F_{drop}, B_{min}) \in K_2 \end{cases};$$

Функция подбиралась исходя из физических соображений исследуемого процесса. Было обнаружено, что дискриминантная функция вида:

$$y = \left(\frac{F_{min} \cdot B_{min}}{T_{start} \cdot F_{drop} + T_{start} + 2^{(25 - F_{min})}} - 5 \right) \quad (1)$$

обладает требуемым качеством.

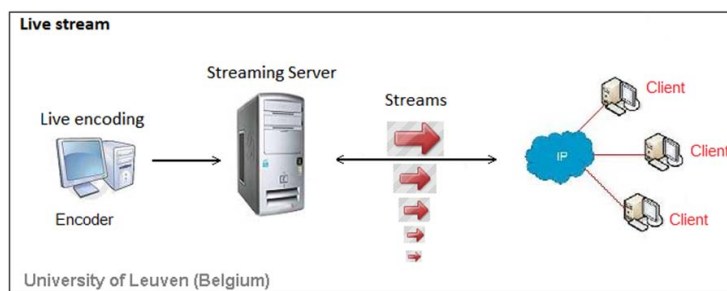


Рис. 2. Воздействия на передаваемый поток отсутствуют

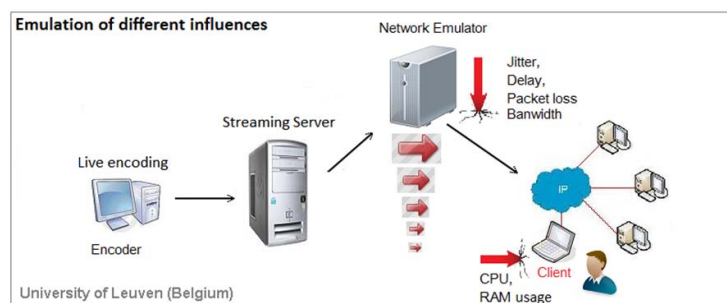


Рис. 3. Эмуляции сетевых воздействий на поток

Константа $FPS = 25$ означает количество видеок кадров в секунду в оригинальном источнике мультимедийного потока.

С помощью представленной формулы были оценены 129 экспериментов обучающей выборки. Полученные значения с помощью данного критерия оценки видеопотока были сравнены с визуальной оценкой качества мультимедийных потоков проведенных экспериментов. Из 129 экспериментов в 128 сравнение показало совпадение результатов обоих используемых подходов оценки качества.

Значения предлагаемого критерия, полученные с помощью экспериментов, принадлежат отрезку $[-5, 19]$. В табл. 3 в четвертом столбце приведены подробные данные о качестве предлагаемого критерия.

В свою очередь, из числителя данной формулы видно, что если одно из значений F_{min} или B_{min} принимают значение, равное нулю, то функция принимает значение, равное -5 , что означает плохое качество потока. Это можно объяснить тем, что чем меньше минимальный размер буфера плеера B_{min} , тем больше вероятность того, что будет ухудшение качества, то же самое касается и минимального количества кадров F_{min} , воспроизводимого в секунду плеером. Если загрузка процессора или памяти на приемном оборудовании имеют значения, близкие к максимальным величинам, то неизбежно будут возникать потери кадров при декодировании, а, как следствие,

уменьшение качества воспроизводимого мультимедийного потока.

В знаменатель формулы входят три параметра: T_{start} , F_{drop} и F_{min} , которые влияют на результирующее значение формулы в разной степени. Чем больше значения параметров T_{start} и F_{drop} , тем больше вероятность ухудшения мультимедийного потока.

Один эксперимент из 129, оцененный с помощью созданного критерия, в ходе обучающей выборки был определен как плохой, а визуальная оценка показала, что данный поток имеет хорошее качество. В то же время в данном эксперименте наблюдается значение $F_{drop} = 12$, поэтому можно предположить, что ухудшение потока было пропущено экспериментатором в ходе проведения эксперимента.

В свою очередь, в ходе проведения 66 экспериментов контрольной выборки результаты, полученные с помощью критерия оценки, совпали с визуальным наблюдением.

Выводы

Разработан критерий оценки качества образовательных мультимедийных потоков, транслируемых в реальном режиме времени в IP сети.

Разработано программное обеспечение для проведения и отображения результатов экспериментов, использующее предложенный критерий оценки качества. Проведены тестовые

расчеты (контрольная выборка), которые подтвердили адекватность выбранного критерия оценки.

Полученные результаты и созданное экспериментальное программное обеспечение в дальнейшем будут использованы для создания автоматизированной системы тестирования удаленных пользователей. Предложенный критерий оценки можно использовать в реализации алгоритмов динамического переключения битрейтов.

ЛИТЕРАТУРА

1. *Austerberry D.* The Technology of Video and Audio Streaming. Burlington, Oxford: Elsevier, 2005. P. 134–144.

2. *Lervold M. G., Xing L., Perkis A.* [Электронный ресурс.] Режим доступа: http://enpub.fulton.asu.edu/resp/vpqm/vpqm10/Proceedings_VPQM2010/vpqm_p34.pdf

3. *Jan L Ozer.* Video Compression for Flash, Apple Devices and HTML5., Doceo Publishing, Inc. 2011. P. 115–140.

4. *David Hassoun.* 2010. [Электронный ресурс.] Режим доступа: http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_advanced_pt1.html

5. *Abhinav Kapoor.* 2009. [Электронный ресурс.] Режим доступа: http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_live.html

6. *Abhinav Kapoor.* 2009. [Электронный ресурс.] Режим доступа: http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_actionscript.html

СВЕДЕНИЯ ОБ АВТОРЕ:

Забровский Анатолий Леонидович
аспирант, ведущий программист
Петрозаводский государственный университет
пр. Ленина, 33, Петрозаводск, Республика Карелия,
Россия, 185910
эл. почта: z_anatoliy@petrsu.ru
тел.: (8142) 711069

Zabrovskiy, Anatoliy
Petrozavodsk State University
33 Lenina St., Petrozavodsk, Karelia, Russia, 185910
e-mail: z_anatoliy@petrsu.ru
tel.: (8142) 711069

УДК 519.6:539.2

МОДЕЛЬ ТЕРМОДЕСОРБЦИИ ВОДОРОДА ИЗ МЕТАЛЛООРГАНИЧЕСКОЙ СТРУКТУРЫ

Ю. В. Заика

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассматривается дегазация методом термодесорбции (ТДС) металлоорганического образца, предварительно насыщенного водородом. Система мелких пор является каналом переноса, крупные поры играют роль ловушек с обратимым захватом. Представлена математическая модель ТДС-эксперимента.

Ключевые слова: водородопроницаемость, моделирование, краевые задачи.

Yu. V. Zaika. MODEL OF HYDROGEN THERMO- DESORPTION FROM METAL-ORGANIC FRAMEWORKS

Degassing by the method of thermal desorption (TDS) of an organometallic sample pre-saturated with hydrogen is considered. The system of small pores is the channel for migration, and large pores act as traps with reversible capture. The mathematical model of the TDS-experiment is presented.

Key words: hydrogen permeability, modeling, boundary-value problems.

Введение

Исследования в области водородного материаловедения инициированы прежде всего перспективами водородной энергетики. Рассмотрим прикладную задачу, связанную с проблемой накопления и хранения водорода. Одними из перспективных материалов в этом отношении являются металлоорганические структуры (metal-organic frameworks, MOF) [4, 1, 3], имеющие пористую структуру. В [2] представлена модель переноса водорода в пористом материале. В качестве приложения рассмотрена водородопроницаемость пористого вольфрама. Поры считаются настолько большими, что накапливающийся в них водород является

газообразным и приходится оперировать параметром давления H_2 в порах. В MOF характерные размеры пор в диаметре значительно меньше (исчисляются ангстремами — АА), что требует корректировки модели. Приведем подходящую математическую модель для конкретного эксперимента: после предварительного насыщения водородом материал дегазируется методом термодесорбции. Канал переноса (диффузии) молекул H_2 — система мелких пор, соизмеримых с размерами молекулы. Большие поры изолированы в том смысле, что молекула H_2 не может попасть сразу из одной большой поры в другую. Поэтому систему больших пор считаем ловушками-накопителями. Вместе с тем, они относительно малы (несколько десятков Å в диаметре).

Образец является двухслойной пластиной из подложки и исследуемого материала. Образец насыщается водородом при комнатной температуре и медленно охлаждается до криогенной температуры. После установления равновесия происходит в режиме вакуумирования медленный нагрев со стороны подложки, с помощью масс-спектрометра регистрируется выходной десорбционный поток водорода.

Математическая модель

Пренебрегаем торцами и считаем подложку практически непроницаемой для молекул H_2 со стороны слоя MOF. Обозначим через S площадь одной стороны поверхности. По толщине слоя MOF выделим отрезок $[x, x + dx]$ и рассмотрим элемент объема $dV = Sdx$. В дальнейшем по контексту под S , V понимаем как сами геометрические фигуры, так и численные значения их площади и объема. Введем обозначения, следуя [2]: n_p — концентрация больших пор, $[n_p] = 1/\text{см}^3$; V_p — объем поры; S_p — ее площадь поверхности. В силу регулярности структуры MOF эти величины являются константами в образце. Тогда $n_p V_p$ — объем пор в единичном кубике, т. е. численно это объемная доля пор в материале. Эта безразмерная величина называется пористостью. Обозначим ее $\eta = n_p V_p$. Величина $n_p S_p$ имеет смысл общей поверхности пор в единице объема. Эту объемную плотность поверхности обозначим $\tilde{S}$.

Перейдем к уравнениям материального баланса. Объем dV разбивается на две части: ηdV — объем пор, $(1 - \eta)dV$ — объем, в котором осуществляется диффузионный перенос (система мелких пор). Обозначим через J_p^+ плотность потока молекул H_2 из поры: столько молекул пересекает единичную площадку сферы большой поры изнутри в единицу времени. Аналогично определяем величину J_p^- — это плотность встречного потока диффундирующего водорода, попавшего в пору-накопитель. Через $c(t, x)$ обозначим концентрацию диффундирующего H_2 в системе мелких пор (в канале переноса) в момент времени t на глубине x по толщине слоя MOF.

Далее в выкладках опускаем бесконечно малые $o(\Delta t)$, $o(\Delta x)$ в силу последующего предельного перехода $\Delta t \rightarrow 0$, $\Delta x \rightarrow 0$. Запишем изменение количества молекул H_2 в объ-

еме $(1 - \eta)dV$ в единицу времени (скорость):

$$(1 - \eta)dV \Delta c / \Delta t \rightarrow \partial_t c(t, x)(1 - \eta)dV \quad (\Delta t \rightarrow 0).$$

Найдем выражение этой величины через баланс потоков. Площадь плоского сечения равна S . Саму геометрическую фигуру и ее площадь для простоты обозначаем одним символом. Сечение делится на две части: S_d — через эту площадь идет диффузия, $(S - S_d)$ — общая площадь кругов, являющихся сечениями крупных пор. Для численного значения S_d имеем $S_d = (1 - \eta)dV/dx = (1 - \eta)S$. Через плоское сечение с координатой x (по толщине слоя MOF) входит диффузионный поток H_2 плотности $-(D\partial c/\partial x)|_x$, а через сечение $x + \Delta x$ выходной поток имеет плотность $-(D\partial c/\partial x)|_{x+\Delta x}$. Умножив упомянутые плотности на S_d , получим потоки H_2 по каналу мелких пор сквозь сечения x и $x + \Delta x$. Разность $-D\partial c/\partial x|_x S_d - (-D\partial c/\partial x|_{x+\Delta x}) S_d$ определяет изменение количества молекул H_2 в объеме dV в единицу времени за счет диффузии. Кроме того, молекулы водорода попадают в диффузионный канал переноса из больших пор и захватываются обратно из канала мелких пор. Дисбаланс этих двух встречных потоков равен $(J_p^+ - J_p^-)S_p n_p S dx$. Здесь $n_p S dx$ — общее количество больших пор в объеме $dV = S dx$, так что $S_p n_p S dx = \tilde{S} dV$ — это общая поверхность больших пор в объеме dV . Для MOF $\tilde{S} \gg 1$.

Составляем баланс молекул H_2 в объеме канала диффузии $(1 - \eta)dV = (1 - \eta)S dx$ (с учетом $\Delta t = dt \rightarrow 0$, $\Delta x = dx \rightarrow 0$):

$$\begin{aligned} \frac{\partial c}{\partial t} S_d dx &= \frac{\partial c}{\partial t}(t, x)(1 - \eta)dV = \\ &= \left[\left(D \frac{\partial c}{\partial x} \right) \Big|_{x+\Delta x} - \left(D \frac{\partial c}{\partial x} \right) \Big|_x \right] S_d + \\ &+ (J_p^+ - J_p^-) \underbrace{\tilde{S}}_{S_p n_p} \cdot \underbrace{dV}_{S dx}. \end{aligned}$$

Подставим $S_d = (1 - \eta)S$, поделим на $dV = S dx$ и устремим dx к нулю: $\tilde{S} \equiv n_p S_p$,

$$(1 - \eta) \frac{\partial c}{\partial t} = (1 - \eta) \frac{\partial}{\partial x} \left(D \frac{\partial c}{\partial x} \right) + (J_p^+ - J_p^-) \tilde{S}.$$

Поделив дополнительно на величину $(1 - \eta)$, окончательно получаем: $T = T(t, x)$,

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D(T) \frac{\partial c}{\partial x} \right) + (J_p^+ - J_p^-) \frac{\tilde{S}}{1 - \eta}.$$

Для определенности зависимость «физико-химических» коэффициентов модели от температуры T считаем аррениусовской. В частности, в приведенном уравнении $D(T) = D_0 \exp\{-E_D/[RT]\}$, где E_D — энергия активации, R — универсальная газовая постоянная.

Аналогично составляем материальный баланс для объема ηdV , занятого большими порами (w — концентрация H_2 в них):

$$\eta \frac{\partial w}{\partial t} = -(J_p^+ - J_p^-) \tilde{S} \Rightarrow \frac{\partial w}{\partial t} = -(J_p^+ - J_p^-) \frac{\tilde{S}}{\eta}.$$

В качестве модели плотностей потоков $J_p^\pm$ сквозь поверхность больших пор примем

$$J_p^+ = k^+(T)w(t, x)[1 - c(t, x)c_{\max}^{-1}],$$

$$J_p^- = k^-(T)c(t, x)[1 - w(t, x)w_{\max}^{-1}].$$

Величина $\theta^+ = c/c_{\max}$ имеет смысл степени насыщения канала диффузии. Чем больше концентрация в больших порах и меньше степень насыщения малых пор, тем больше плотность выходного потока молекул H_2 из поры-ловушки в канал диффузии. Соответствующий коэффициент пропорциональности k^+ зависит от текущей температуры T . Аналогично интерпретируется выражение J_p^- для плотности встречного потока в поры-накопители.

Подчеркнем, что в уравнения переноса входит не только пористость материала η , но и объемная плотность поверхности больших пор $\tilde{S}$. При одной и той же пористости величина $\tilde{S}$ для MOF может варьироваться в широких пределах. Перенос тепла диффундирующими молекулами (эффект Дюфо) и термодиффузию (эффект Соре) считаем второстепенными.

Перейдем к *граничным условиям*. Подложка практически непроницаема со стороны MOF: $\partial_x c(t, 0) = 0$, $t \geq 0$. Значение $x = 0$ соответствует началу слоя MOF, $t = 0$ — началу нагрева образца в вакууме. Выходная сторона слоя MOF обращена в вакуум и вследствие пористости материала имеет изрезанный вид. Можно ввести параметр σ шероховатости поверхности [2], равный отношению «истинной» площади поверхности к геометрической. Но численно значение σ оценить трудно. Поступим по-другому. Выделим тонкий приповерхностный слой толщины ℓ , который в силу изрезанности существенно отличается по физико-химическим свойствам от

регулярных внутренних слоев общей толщины L ($\ell \ll L$). Кроме того, в процессе изготовления материала, подготовки и проведения эксперимента неизбежно появление микропримесей и окислов на поверхности. Фактически этот слой является несколько иным материалом по сравнению с объемным. Внешний слой обладает накопительным потенциалом: молекула H_2 имеет дополнительную возможность зацепиться за неровности поверхности. Рассмотрим сечение $x = L$. Площадь этого сечения равна S . Из объема подходит диффузионный поток $-S_d D(T) \partial_x c|_L$. Молекулы H_2 могут не сразу десорбироваться, а предварительно «растекаться» в поперечном направлении. Таким образом, десорбция в вакуум идет «сквозь окно» площади S . Выделенный иррегулярный внешний слой настолько мал, что в его пределах объемная концентрация молекул водорода практически не меняется по толщине $x \in [L, L + \ell]$. Обозначим ее $c_\ell(t)$. Поскольку имеем дело с физической адсорбцией (абсорбцией в ℓ -слой), то плотность потока десорбции моделируем реакцией первого порядка: $b(T)c_\ell(t)$. Эффективный коэффициент десорбции b имеет размерность скорости (см/с). Дисбаланс потоков определяет изменение концентрации в приповерхностном слое:

$$S \ell \dot{c}_\ell(t) = -S_d D(T) \partial_x c(t, L) - S b(T) c_\ell(t).$$

Остается поделить на S и заменить отношение S_d/S на $(1 - \eta)$. В [2] в аналогичной ситуации (но в больших порах-полостях находится газ) принято $c_\ell(t) = c(t, L)$, что при $\ell \ll 1$ соответствует относительно малой поверхностной концентрации $q(t) = \ell c_\ell(t)$. Можно расширить возможности модели, полагая $c_\ell(t) = h(T)c(t, L)$ и допуская $h \gg 1$: к диффундирующему водороду добавляется водород из открытых приповерхностных пор. Обозначая $\ell h = 1/g$, формально приходим к динамическому граничному условию, аналогично условиям баланса потоков

$$c_0(t) = g(T)q_0(t), \quad c_\ell(t) = g(T)q_\ell(t),$$

$$\dot{q}_0(t) = \mu s(T)p_0(t) - b(T)q_0^2(t) + D(T)c_x(t, 0),$$

$$\dot{q}_\ell(t) = \mu s(T)p_\ell(t) - b(T)q_\ell^2(t) - D(T)c_x(t, \ell)$$

для экспериментального метода проницаемости H сквозь пластину толщины ℓ (μ — кинетический коэффициент, s — коэффициент

прилипания, p — давление газообразного водорода). Такая форма пригодна и в ситуации, когда поверхностная концентрация больше объемной (при относительно низких температурах). В любом случае вместе с начальными условиями краевые задачи приобретают замкнутый вид. Если накопительный эффект пренебрежимо мал, то полагаем левую часть, равной нулю, и получаем граничное условие $(\eta - 1)D\partial_x c(t, L) = bc(t, L)$: сколько молекул H_2 диффузионно подошло к выходной поверхности, столько и десорбировалось за единицу времени. Можно подразумевать, что в эффективный коэффициент десорбции b «включены» как параметр h , так и шероховатость выходной поверхности.

Кратко остановимся на *начальных условиях*. В режиме насыщения образца следует учесть наличие внешнего давления газа (H_2):

$$\ell \dot{c}_\ell(t) = \mu s(T)p(t) - (1 - \eta)D(T)\partial_x c(t, L) - b(T)c_\ell(t).$$

При начальной температуре $T = \bar{T}$ эксперимента установилось равновесие, поэтому все производные равны нулю: $\mu s\bar{p} = b\bar{c}_\ell$. В объеме получаем: $\bar{c}g = \bar{c}_\ell$,

$$J_p^+ = J_p^- \Rightarrow k^+ \bar{w} [1 - \bar{c}c_{\max}^{-1}] = k^- \bar{c} [1 - \bar{w}w_{\max}^{-1}].$$

Модель позволяет однозначно определить начальные данные по цепочке $\bar{p} \rightarrow \bar{c}_\ell \rightarrow \bar{c} \rightarrow \bar{w}$. Как оценить значения $c_{\max}$, $w_{\max}$? Рассмотрим единичный кубик с максимальным насыщением. Количество молекул H_2 в нем равно $A = \eta w_{\max} + [1 - \eta]c_{\max}$. Учтем соотношения $\eta = n_p V_p = n_p 4\pi r^3/3 = n_p S_p r/3 = \tilde{S}r/3$, где n_p — количество пор в единичном кубике, r — радиус большой поры, S_p — площадь ее поверхности, V_p — объем, $\tilde{S} = n_p S_p$ — объемная плотность общей поверхности больших пор. Тогда

$$\begin{aligned} A &= w_{\max} \tilde{S}r/3 + c_{\max} - c_{\max} \tilde{S}r/3 = \\ &= c_{\max} + \tilde{S}[w_{\max} - c_{\max}]r/3. \end{aligned}$$

Зафиксируем параметр $\tilde{S}$ и будем менять r в физически реальном диапазоне. Если уменьшить r , то, чтобы сохранить значение $\tilde{S}$, нужно соответствующим образом увеличить количество пор. Кроме того, в более мелких порах молекуле легче удержаться, так что имеется тенденция увеличения $w_{\max}$ с уменьшением r . Но тогда коэффициент $r[w_{\max} - c_{\max}]$ при $\tilde{S}$

имеет тенденцию к стабилизации, слабой зависимости от r . Практически линейная зависимость $A = k_1 \tilde{S} + k_2$ ($k_i = \text{const}$) при варьировании $\tilde{S}$ обнаружена экспериментально [3]. Тогда $k_2 = c_{\max}$, $(k_1, A) \leftrightarrow (w_{\max}, r)$. Можно ограничиться некоторыми средними значениями $\tilde{c}_{\max}$, $\tilde{w}_{\max}$ или грубыми оценками по результатам полной дегазации с учетом того, что из геометрических соображений количество малых шаров (молекул H_2) по объему в большой поре радиуса r не должно превысить $4\pi r^3/3$. Если в условиях эксперимента $c \ll c_{\max}$, $w \ll w_{\max}$, что согласуется с $D \neq D(c)$, то исключаем параметры $c_{\max}$, $w_{\max}$ из модели: $c/c_{\max} \approx 0$, $w/w_{\max} \approx 0$.

Теплопередача в металлоорганических каркасах происходит значительно быстрее диффузии. Поэтому вместо распределенного уравнения теплопроводности целесообразно оперировать квазистационарами. Если в уравнении теплопроводности $\nu \partial_t T = \lambda \partial_{xx} T$ считать изменения по времени медленными, то в относительном масштабе получаем $\partial_t T \approx 0$ и линейную зависимость $T = \alpha x + \beta$. С течением времени квазистационар перестраивается: $T(t, x) = \alpha(t)x + \beta(t)$. На границе слоев два линейных квазистационара стыкуются с изломом из-за различных коэффициентов теплопроводности. Температура эксперимента низка, так что тепловым излучением в вакуум, по-видимому, можно пренебречь. Дальнейшая детализация зависит от набора теплофизических измерений в эксперименте. Для достаточно тонкого образца для начала можно принять, что прогрев равномерный.

Изложенное следует воспринимать лишь как один из вариантов модели. Различных МОФ синтезировано уже более тысячи и говорить о достаточно общей модели не приходится. Приведем одну из модификаций представленной модели с учетом необходимости решать обратную задачу параметрической идентификации. Это приводит к необходимости учета лимитирующих факторов минимальным набором неизвестных априори параметров. Коэффициенты имеют смысл обобщенных (интегральных, эффективных) показателей переноса водорода в слое МОФ. В частности, диффузию не разделяем на молекулярную и кнудсеновскую, считая $D = D_{\text{eff}}$. Рассмотрим МОФ со следующими свойствами. Имеется две системы пор: большие ($\sim 10\text{\AA}$)

и малые ($\sim 5\text{\AA}$). Грубо диаметр молекулы водорода $\sim 3\text{\AA}$. Система больших пор взаимосвязана и является каналом диффузионного переноса (диффузионного в формальном «фииковском» смысле: поток пропорционален градиенту концентрации). Большие поры имеют кубическую форму, малые являются «треугольными карманами» в вершинах куба. Геометрическая и химическая структура каркаса малых пор позволяет считать их ловушками водорода. В «щели» при наличии ионов металла молекула H_2 связана с каркасом значительно сильнее, чем в большой поре. Пор огромное число, так что для анализа макропереноса следует ввести подходящие макропараметры. Ограничимся лишь двумя геометрическими характеристиками: пористость η и удельная поверхность $\tilde{S}$ ($[\cdot] = \text{cm}^2/\text{cm}^3 = 1/\text{cm}$). Общий баланс подкачки водорода из ловушек в процессе равномерного монотонного (обычно линейного) нагрева $T = T(t)$ моделируем соотношением $J_p^+ - J_p^- = a(T)w(t)$ ($[a] = \text{cm/s}$), где $w(t)$ — концентрация в объеме малых пор. Из материального баланса получаем:

$$(1 - \eta) \frac{\partial c}{\partial t} = (1 - \eta) D(T) \frac{\partial^2 c}{\partial x^2} + a(T) w(t) \tilde{S},$$

$$\eta dw/dt = -a(T) w(t) \tilde{S}.$$

После интегрирования второго линейного обыкновенного уравнения получаем одно дифференциальное уравнение для $c(t, x)$. Начальное насыщение по постановке ТДС-равномерно: $c(0, x) = \bar{c}$, $w(0) = \bar{w}$, $x \in [0, L]$. Пренебрегая накоплением на выходе, принимаем граничные условия

маем граничные условия

$$\partial_x c(t, 0) = 0, \quad (\eta - 1) D \partial_x c(t, L) = bc(t, L).$$

Ловушки, обращенные в вакуум ($x = L$), опустошились при $t < t_0 \ll 1$. Неизвестные значения D_0 , E_D , a_0 , E_a , b_0 , E_b оцениваются по плотности десорбции $J(t) = b(T(t))c(t, L)$, часто известной лишь в относительных единицах ($J/J_{\max}$). Для первого приближения (оценка значений D_0 , E_D , $\bar{a}$) можно принять

$$a = 0, \quad T < T_{\text{crit}}, \quad a = \bar{a} > 0, \quad T \geq T_{\text{crit}},$$

$$c(t, L) = 0, \quad J = (\eta - 1) D \partial_x c(t, L).$$

При необходимости нетрудно учесть в форме ловушки определенной емкости накопление водорода на границе с подложкой.

ЛИТЕРАТУРА

1. Исаева В. И., Кустов Л. М. Металлоорганические каркасы — новые материалы для хранения водорода // Российский химический журнал. 2006. № 6. С. 56–72.
2. Пусарев А. А., Цветков И. В., Маренков Е. Д., Ярмо С. С. Проницаемость водорода через металлы. М.: МИФИ, 2008. 144 с.
3. Hirscher M., Panella B., Schmitz B. Metal-organic frameworks for hydrogen storage // Microporous and mesoporous materials. 2010. Vol. 129. P. 335–339.
4. Varin R. A., Czujko T., Wronski Z. S. Nanomaterials for solid state hydrogen storage. Springer, 2009. P. 338.

СВЕДЕНИЯ ОБ АВТОРЕ:

Заика Юрий Васильевич

зав. лаб. моделирования природно-технических систем, д. ф.-м. н.
Институт прикладных математических исследований Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: zaika@krc.karelia.ru
тел.: (8142) 766312

Zaika, Yury

Institute of Applied Mathematical Research,
Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
Russia
e-mail: zaika@krc.karelia.ru
tel.: (8142) 766312

УДК 004.01:006.72 (470.22)

МОДЕЛЬ РЕГУЛИРОВАНИЯ ЭКОНОМИКИ САНКТ-ПЕТЕРБУРГА ЗА СЧЕТ ИНВЕСТИЦИОННОГО ПОТОКА

К. А. Замураев, А. В. Прасолов

Санкт-Петербургский государственный университет

В работе строится модель управления инвестициями в основной капитал с целью достижения определенных уровней оборота предприятий. Моделирование процесса производится с использованием статистических данных по Санкт-Петербургу в предположении, что инвестиции в основной капитал и оборот организаций образуют регрессионную зависимость и одновременно с этим имеют эндогенную природу. Параметры модели определяются с помощью МНК. Учитывается инфляционная составляющая статистических данных, предлагается экономически интерпретируемая схема сглаживания, строится прогноз инвестиций и сравнивается с реальными данными.

Ключевые слова: оборот организаций, инвестиции в основной капитал, эндогенные параметры, регрессионная модель, управление инвестициями.

K. A. Zamuraev, A. V. Prasolov. MODEL OF REGULATING THE ST. PETERSBURG ECONOMY BY THE INVESTMENT FLOW

A model of managing investments into fixed capital in order to reach certain levels of company turnover is considered. The process is modeled with the assumption that investments into fixed capital and the company turnover are related by a regression function, simultaneously having endogenous nature. The model is based on statistical data from St. Petersburg. The model parameters are found using OLS. The inflationary component is taken into account, a special smoothing scheme for statistical data is suggested, the investment forecast is compared with actual data.

Key words: company turnover, investments into fixed capital, endogenous parameters, regression model, investment management.

ВВЕДЕНИЕ

Одной из важнейших задач, решение которых важно для развития субъектов Российской Федерации, является ведение оправданной и эффективной инвестиционной политики региона. В частности, одной из подзадач является достижение определенных уровней развития экономики, промышленности, отраслей.

Основное средство достижения этих целей – эффективное управление инвестициями.

В данной статье представлена модель оценки объемов инвестиций, требуемых для достижения определенных уровней оборота предприятий города.

В основе модели рассматривается регрессионная связь оборота с инвестициями в основной капитал. Для оборота делается предположение о его эндогенной природе, а

относительно значений инвестиций предполагаем, что они образуют авторегрессию.

Применение нового подхода к управлению инвестициями рассматривается на статистических данных оборота организаций¹ Санкт-Петербурга и инвестиций в основной капитал², параметров, сбором которых занимается территориальный орган Федеральной службы государственной статистики по Санкт-Петербургу и Ленинградской области (Петростат). Кроме того, учитывается инфляционное влияние на рассматриваемые параметры, также для более адекватного прогнозирования, при построении модели практического примера предлагается схема сглаживания данных.

Исследованию значимости влияния инвестиционных процессов на изменение роста экономики уделяется достаточное внимание уже несколько десятилетий. Так, в работах Роберта Солоу [7] и Роберта Лукаса [6] рассматриваются неоклассические модели эндогенного экономического роста, в которых с использованием производственных функций определена связь чистых инвестиций с общим объемом производства.

Кроме того, известны модели, в которых значения инвестиций образуют авторегрессию, несколько из них представлено в книге Э. Берндта [1].

В свою очередь, в данной статье одновременно учитываются оба вышеизложенных элемента. Как уже отмечалось ранее, строится модель взаимосвязи оборота организаций региона и инвестиций в основной капитал, с предположением об эндогенной природе оборота и авторегрессии инвестиций.

Еще одним нововведением статьи является управляющая составляющая в модели взаимодействия оборота и инвестиций. Сама идея ввода параметра, за счет которого можно производить регулирование экономического роста, не нова, например, в статье Томаса Сарджента и Нила Уоллеса [8], авторы рассматривают пример стохастического уравнения регулирования ВВП, кстати, который имеет эндогенную природу, за счет изменения объемов денежной массы, однако основной упор в их работе делается на исследовании оптимального уровня ВВП. В данной статье рас-

сматривается предположение о том, что регулирующий параметр является обособленным показателем, для которого есть собственные статистические данные, но который входит в систему с коэффициентом, определенным для текущего значения оборота.

Для предложенной в статье модели рассматривается пример, опирающийся на реальные данные оборота организаций Санкт-Петербурга и инвестиций в основной капитал. Исследование взаимосвязи инвестиций в основной капитал и индекса развития регионов Российской Федерации уже делались, например, в статьях Ю. А. Шипунова [5] и А. С. Коротченко [2], авторами приводятся регрессионные модели связи индекса ВРП и индекса инвестиций федеральных округов Российской Федерации, однако, их модели имеют простой вид и не содержат предположений о эндогенности роста оборота, авторегрессионных инвестициях, управляющего параметра.

ОПИСАНИЕ МОДЕЛИ

Поставим перед собой задачу оценки инвестиций в основной капитал, которые будут способствовать достижению определенных уровней оборота в течение последующего периода времени.

Обозначим за x_k статистические данные оборота организаций (в момент k), символом u_k объем инвестиций в основной капитал (в момент k), предположим, что нам известны m значений обоих параметров и требуется построить прогноз u_{k+1} .

Как уже отмечалось ранее, сделаем предположение о регрессионной зависимости оборота организаций и инвестиций в основной капитал, примем также, что текущее значение оборота и инвестиций зависят от некоторого набора своих предыдущих значений, например от n параметров.

Кроме того, для более точного прогнозирования экономических параметров необходимо учесть инфляционную составляющую. Для этого необходимо подготовить данные – привести их к определенному заданному уровню, извлекая накапливающуюся инфляцию. Величина накопленной инфляции определяется по

¹Оборот организаций – это стоимость отгруженных товаров собственного производства, выполненных собственными силами работ и услуг, а также выручка от продажи приобретенных на стороне товаров (без налога на добавленную стоимость, акцизов и других аналогичных обязательных платежей). Экономический доклад «Социально-экономическое положение Санкт-Петербурга и Ленинградской области в январе-ноябре 2011 года» Петростат, 2011 г.

²Инвестиции в основной капитал – это совокупность затрат на новое строительство, расширение, а также реконструкцию и модернизацию объектов, которые приводят к увеличению первоначальной стоимости объекта, на приобретение машин, оборудования, транспортных средств, на формирование основного стада, многолетние насаждения и т. д. Экономический доклад «Социально-экономическое положение Санкт-Петербурга и Ленинградской области в январе-ноябре 2011 года» Петростат, 2011 г.

формуле (1).

$$Z_i = \sum_{j=1}^i Z_j, j \leq i, \quad (1)$$

где Z – параметр, характеризующий инфляционный процесс, а $j = 1$ – начало отсчета накопления инфляции.

Формула приведения параметров x_k и u_k к очищенной от инфляции форме имеет вид:

$$\varphi_{i-Z} = \varphi_i - \frac{\varphi_i Z_i}{100\%}, \quad (2)$$

где φ_i, φ_{i-Z} – значение временного ряда (x_k или u_k) с инфляционной составляющей $Z(\varphi_i)$ и без нее (φ_{i-Z}), соответственно.

Сформулируем общий вид новой модели (3) вычисления объемов инвестиций, требуемых для достижения определенных значений оборота организаций (3).

$$u_{k+1} = a_{k+1} \widetilde{x_{k+1}} + \sum_{i=k-n}^k a_i x_i + \sum_{j=k-n}^k b_j u_j, \quad (3)$$

где $k = \overline{n+1, m}$. Здесь параметр ($\widetilde{x_{k+1}}$) – это плановые значения оборота, достижение которых возможно за счет привлечения объемов инвестиций u_{k+1} .

Коэффициенты a_i, b_j из (3) принимаются равными коэффициентам из уравнения (4), которые в свою очередь находятся методом наименьших квадратов по известным значениям x_k и u_k

$$u_{k+1} = \sum_{i=k-n}^{k+1} a_i x_i + \sum_{j=k-n}^k b_j u_j, \quad (4)$$

где $k = \overline{n+1, m}$.

А расчетные значения x_i определяются по формуле, где коэффициенты α_i, β_j определяются методом наименьших квадратов по известным значениям x_k и u_k [3]

$$x_{k+1} = \sum_{i=k-n}^k \alpha_i x_i + \sum_{j=k-n}^k \beta_j u_j, \quad (5)$$

где $k = \overline{n+1, m}$.

ПРАВИЛО ПРИНЯТИЯ РЕШЕНИЯ

Строя прогноз значений инвестиций в момент $k+1$, т. е. u_{k+1} , мы не знаем значение x_{k+1} , но имеем цель – достичь оборота, равного (x_{k+1}), который будем использовать в модели.

Модель (3)–(5), используя статистические данные параметров x_k и u_k , модель вычисляет объемы инвестиций u_{k+1} , которые будут способствовать достижению уровней оборота организаций $\widetilde{x_{k+1}}$. Таким образом, при построении модели мы учитываем значение оборота, которое хотели бы достигнуть, тем самым вводим управляемую составляющую. При этом модель вычисляет получаемый за счет инвестиций u_{k+1} оборот организаций x_{k+1} . Причем значения $\widetilde{x_{k+1}}$ и x_{k+1} не равны из-за случайных составляющих параметров x_k и u_k (ε_i) для $i = \overline{1, m}$, а равенство $\widetilde{x_{k+1}}$ и x_{k+1} наступило бы при $\varepsilon_i = 0$ для $i = \overline{1, m}$.

После построения прогноза необходимо учесть для него инфляцию, для этого воспользуемся формулой:

$$\varphi_{i+Z} = \frac{100\% \varphi_{i-Z}}{100\%}, \quad (6)$$

где φ_{i+Z} – i -ое прогнозное значение временного ряда (x_k или u_k) с возвращенной инфляционной составляющей.

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МОДЕЛИ: ОБРАБОТКА ФАКТИЧЕСКИХ ДАННЫХ

Приведем численный пример. Рассмотрим помесечные данные оборота организаций Санкт-Петербурга, а также объем инвестиций в основной капитал в период с января 2005 г. и по декабрь 2010 г. (рис. 1).

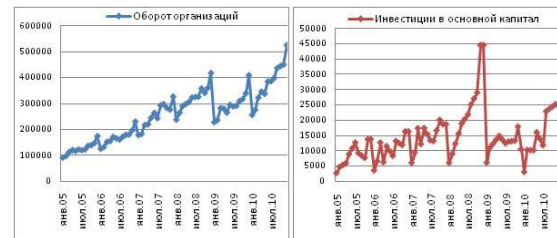


Рис. 1. Оборот организаций (в действующих ценах) и инвестиции в основной капитал, в млн руб. за текущий месяц

Для реальных данных экономических параметров естественно наличие инфляционной составляющей. Удобнее работать с "чистыми"

данными, без инфляции, так как в таком случае прослеживается реальное изменение параметра во времени, не искаженное влиянием инфляционных процессов.

Для "очистения" данных от инфляции будем использовать Индекс потребительских цен (далее – ИПЦ), величину, характеризующую эти процессы. Для этого рассмотрим данные ИПЦ в Санкт-Петербурге, определенные Петростатом в период с января 2005 г. по декабрь 2010 г. и представленные в процентах за текущий месяц.

Вспоминая формулу (1), выберем базовый месяц (декабрь 2004 г.) и вычислим для каждого последующего месяца накопленное значение ИПЦ:

$$CPI_i = \sum_{j=1}^i CPI_j, j \leq i. \quad (7)$$

Далее применим формулу (2) к значениям оборота и инвестиций:

$$\varphi_{i-CPI} = \varphi_i - \frac{\varphi_i CPI_i}{100\%}, \quad (8)$$

где $\varphi_i, \varphi_{i-CPI}$ – значение оборота или инвестиций с ИПЦ и без ИПЦ соответственно.

Получим значения рядов без ИПЦ для оборота организации и инвестиций (рис. 2).

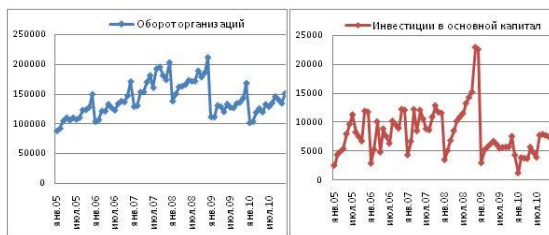


Рис. 2. Оборот организаций и инвестиции в основной капитал, в млн руб. за текущий месяц, с исключенной инфляционной составляющей

На рис. 2 видно, что ежегодная тенденция динамики оборота организаций имеет следующую особенность: каждый январь происходит сильное падение по сравнению со значениями за предыдущий год, далее оборот начинает поступательно расти, а затем в декабре месяц происходит резкий рост, образуются декабрьские пики. Сложившаяся ситуация, возможно, объясняется тем, что в январе в связи с продолжительными выходными меньшее по сравнению с другими месяцами количество рабочих дней (примерно 2/3 от среднего количества), и, следовательно, значения оборота организаций будут заметно меньше среднемесячного значения. В свою очередь стремительный

рост к концу года и декабрьские пики можно объяснить желанием производителя улучшить свою статистику, для этого уже произведенную или готовую до определенной стадии продукцию можно учитывать в нужные месяцы, например в конце года, чтобы сформировать общую картину роста.

Рассматривая рис. 2, также замечаем, что для инвестиций в основной капитал в январе и декабре ежегодно повторяются падения и пики роста соответственно.

Для временных рядов, имеющих неоднородную структуру, часто применяют различные виды сглаживания. В данном случае для того чтобы предлагаемая модель производила более точное вычисление прогноза, мы также предложим алгоритм сглаживания для оборота организаций и инвестиций в основной капитал, но это сглаживание носит экономически интерпретируемый смысл.

Пусть x_i – месячные значения оборота организаций за год, т. е. $i = \overline{1; 12}$. Введем величину $h = x_{12} - x_{11}$, т. е. – это разница между декабрьским и ноябрьским значениями оборота.

Преобразуем значения с января по ноябрь по следующей форме:

$$\overline{x_j} = x_j + \frac{(12-j)}{66}h, j = \overline{1, 11}, \overline{x_{12}} = x_{12} - h. \quad (9)$$

Данные преобразования применяются для значений каждого имеющегося года.

Согласно формуле (9) к значениям будут прибавляться некоторые величины, причем в январе, так как именно в этот месяц наблюдается наименьшее значение оборота, эта величина будет наибольшей, а затем она равномерно уменьшается и в ноябре будет минимальной, одновременно с этим из декабрьского значения вычитается величина h . Сумма всех прибавляемых значений равна h , т. е. мы не меняем суммарный оборот за год, а изменяем его "отчетные" месячные значения. К тому же данную схему сглаживания отличает то, что не изменяется количество элементов временных рядов.

Сглаженный ряд оборота организаций Санкт-Петербурга представлен на рис. 3.

Относительно значений инвестиций в основной капитал (см. рис. 2) также можно заметить сезонность падений в январе с последующим ростом до пиков графика в значениях, соответствующих ноябрю и декабрю. Чтобы изменить большую неоднородность данных, предложим схему сглаживания (10).

$$\begin{aligned}\bar{u}_j &= u_j + \frac{(10-j)}{45}(h_1 + h_2), j = \overline{1,9}, \\ \bar{u}_{11} &= h_1 = \frac{u_{11}}{2}, \bar{u}_{12} = h_2 = \frac{u_{11}}{2}.\end{aligned}\quad (10)$$

Согласно формуле (10) к значениям с января по сентябрь будут прибавляться некоторые величины, причем в январе, так как именно в этот месяц наблюдается наименьшее значение оборота, эта величина будет наибольшей, а затем она равномерно уменьшается и в сентябре будет минимальной, в октябре значение не изменяется, и одновременно с этим из ноябрьского и декабрьского значений вычитается величина h_1 и h_2 соответственно. Сумма всех прибавляемых значений равна $h_1 + h_2$, т. е. мы не меняем суммарный оборот за год, а изменяем его "планируемые" месячные значения.

Сглаженный ряд инвестиций в основной капитал представлен на рис. 3.

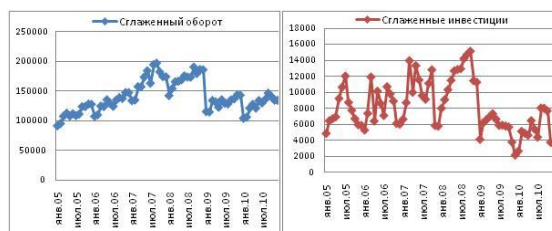


Рис. 3. Сглаженные оборот и инвестиции в основной капитал, в млн руб. за текущий месяц

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ: ПОСТРОЕНИЕ МОДЕЛИ

Приступим к построению модели. Согласно сделанным ранее предположениям считаем, что временные ряды, характеризующие оборот организаций, и инвестиции в основной капитал зависят от n предыдущих значений, например, от 12 (год).

Как отмечалось ранее, одним из нововведений данной модели служит введение в формулу расчета инвестиций плановой составляющей x_{k+1} . Ежегодно органами исполнительной власти Санкт-Петербурга выпускаются прогнозы общеэкономических показателей, в числе которых есть выпуск товаров и услуг. Прогнозные значения с 2005 по 2011 г., в том числе с извлеченным ИПЦ по аналогичным обороту и инвестициям формулам, представлены на рис. 4.



Рис. 4. Сглаженные оборот и инвестиции в основной капитал, в млн руб. за текущий месяц

Будем считать, что задача состоит в том, чтобы посчитать инвестиции в основной капитал u_{k+1} , требуемые в 2011 г. для достижения значений x_{k+1} из рис. 4 (2011 г.).

Сформулируем в общем виде модель для данного числового примера:

$$u_{k+1} = a_{k+1}\widetilde{x_{k+1}} + \sum_{i=k-12}^k a_i x_i + \sum_{j=k-12}^k b_j u_j, \quad (11)$$

$$k = \overline{13; 72}$$

$$u_{k+1} = \sum_{i=k-12}^{k+1} a_i x_i + \sum_{j=k-12}^k b_j u_j, \quad k = \overline{13; 72} \quad (12)$$

$$x_{k+1} = \sum_{i=k-12}^k \alpha_i x_i + \sum_{j=k-12}^k \beta_j u_j, \quad k = \overline{13; 72}. \quad (13)$$

Коэффициенты в уравнениях вычисляются по методу наименьших квадратов. Чтобы отсеять статистически незначимые параметры, исследуем их значимость с помощью t -критерия Стьюдента.

Сперва рассчитаем фактические значения t -критерия по формулам

$$t_x = \frac{x_i}{m_{x_i}}, t_u = \frac{u_i}{m_{u_i}}. \quad (14)$$

Сравнивая фактические значения t -критерия с табличным значением $t \approx 2$ при 142 степенях свободы, по одному удаляем из модели незначимые коэффициенты и заново оцениваем остальные, в итоге приходим к модели, в которой останутся только значимые коэффициенты. В формулах (15) и (16) значимость оставшихся коэффициентов указана в скобках под ними.

$$x_{k+1} = \underbrace{0,64x_k}_{(6,53)} + \underbrace{0,35x_{k-1}}_{(3,87)} + \underbrace{2,01u_k}_{(3,1)} - \underbrace{1,77u_{k-2}}_{(2,69)} \quad (15)$$

$$u_{k+1} = \underbrace{0,07\widetilde{x_{k+1}}}_{(4,02)} - \underbrace{0,05x_k}_{(2,83)} + \underbrace{0,62u_k}_{(6,34)}. \quad (16)$$

При идентификации коэффициентов для формулы (15) получили значение коэффициента детерминации $R^2 = 0,99135$, для формулы (16) $R^2 = 0,94943$, что говорит о достаточно высокой степени адекватности модели.

Перед началом построения прогноза построим оптимальную начальную точку прогнозирования [4].

Применяя к системе (15), (16) значения $\widetilde{x}_{k+1}$ за 2011 г., получим искомые значения, построим их графики (рис. 5). Можем сравнить плановые и расчетные обороты (рис. 6).

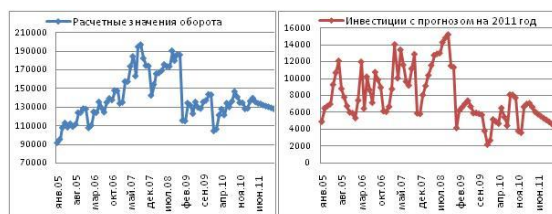


Рис. 5. Расчетные значения оборота и инвестиции в основной капитал, в том числе требуемые для достижения заданных уровней оборота в 2011 г. в млн руб.



Рис. 6. Расчетные и плановые обороты организаций в 2011 г. в млн руб.

Чтобы быть более точными в прогнозировании, нам необходимо учесть индекс потребительских цен (ИПЦ), который ранее мы выделили из данных. Имея временной ряд ИПЦ, можем построить, например, линейный тренд.

Продолжив тренд на интересующий будущий интервал времени, получим прогноз ИПЦ (рис. 7). Но ранее мы решили, что выделять будем ИПЦ, приведенный к базовому периоду, поэтому прогнозные значения необходимо привести к базовому месяцу, которым у нас является декабрь 2004.

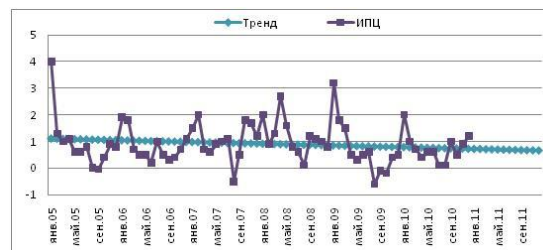


Рис. 7. Значения и тренд ИПЦ

Прогнозные данные ИПЦ к базовому месяцу приведенные в табл. 1.

Таблица 1. Прогноз значений накопленного ИПЦ в 2011 г.

январь.	фев.	март	апр.
71,97 %	72,67 %	73,37 %	74,06 %
май	июнь	июль	авг.
74,74 %	75,42 %	76,10 %	76,77 %
сентяб.	окт.	ноябрь	дек.
77,43 %	78,09 %	78,75 %	79,39 %

Преобразовываем прогнозные данные без ИПЦ в прогнозные данные с учетом ИПЦ по формуле:

$$\varphi_{i+CPI} = \frac{100\% \varphi_i - CPI}{100\% - CPI_i} \cdot \quad (17)$$

Строим окончательный график оборота организаций с построенным прогнозом на год и расчетными значениями, и инвестиций в основной капитал с построенной моделью значений на год (рис. 8).



Рис. 8. Расчетные и плановые обороты организаций, инвестиции в основной капитал, в том числе в 2011 г., с включенным ИПЦ, в млн руб.

Построив на одном графике значения оборота организаций: реальные, полученные моделью, и прогнозные (рис. 9), видно, что реальные значения достигнутого оборота организаций примерно совпадают с плановыми значениями. В свою очередь значения оборота, рассчитанные моделью, оказались большими,

нежели реальные или прогнозные. При этом рассчитанные моделью инвестиции больше реальных, примерно на такой же порядок, что и расчетный оборот больше прогнозного, но в реальных значениях инвестиций наблюдается неоднородность и пиковость.

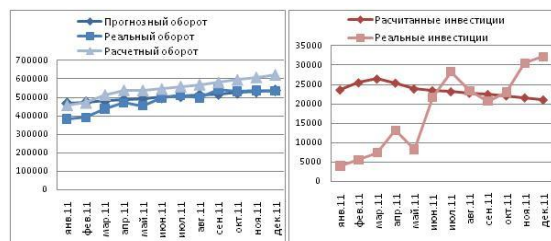


Рис. 9. Реальный оборот организаций в 2011 г., планируемый оборот организаций, оборот, рассчитанный моделью, рассчитанные и реальные инвестиции, в млн руб.

5. РЕЗУЛЬТАТ ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ

Просуммировав смоделированные значения инвестиций в основной капитал, мы получим сумму в 281 575 млн руб. необходимую для закладки в бюджет на следующий год, которая будет способствовать достижению заранее заданному росту оборота организаций Санкт-Петербурга в 2011 г.

Для того чтобы проверить эффективность модели, подберем помесечные значения прогноза оборота организаций так, чтобы расчетные значения оборота в сумме за год были примерно равны реальным значениям оборота, а затем сравним рассчитанный моделью набор потока инвестиций с реальными инвестициями.

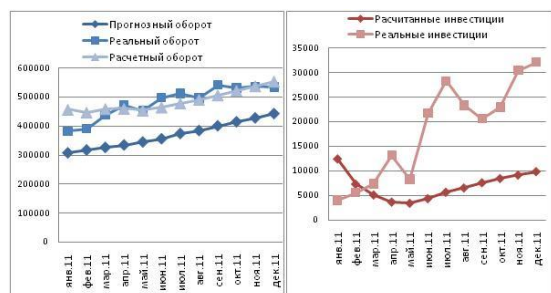


Рис. 10. Выбор планируемого оборота: реальный оборот организаций в 2011 г., совпадающий с оборотом, рассчитанный моделью, рассчитанные и реальные инвестиции, в млн руб.

На рис. 10 представлен набор планового (прогнозного) оборота, за счет которого модель будет давать оборот за год, равный реальным значениям. Расчетный оборот за 2011 г. $x_{2011} = 5\,791\,367$ млн руб., $x_{2011} = 5\,791\,391$ млн руб.

Значения расчетных и реальных значений оборота также представлены на рис. 10. Сравним потребовавшиеся для модели инвестиции с реальными (рис. 10). Расчетные инвестиции за 2011 г. $u_{2011} = 83\,928$ млн руб., реальные $u_{2011} = 218\,120$ млн руб.

Видно, что при одинаковых объемах оборота, модель показывает необходимость привлечения объемов инвестиций более чем в 2 раза меньших, потраченных в реальности.

Таким образом, представлен метод, который позволяет учесть в прогнозировании наперед заданный оборот городских предприятий и который можно достичь за счет управления инвестициями.

ЛИТЕРАТУРА

1. Берндт Э. Р. Практика эконометрики: классика и современность. М.: Юнити, 2005. С. 259–329.
2. Коротченко А. С. Корреляционно-регрессионный анализ влияния инвестиционной активности на рост регионального валового продукта Калининградской области // Вестник Балтийского федерального университета им. И. Канта. К., 2010. № 3. С. 108–115.
3. Мхитарян В. С., Архипова М. Ю., Балаш В. А. и др. Эконометрика. М.: Проспект, 2010. 384 с.
4. Прасолов А. В., Хованов Н. В. О прогнозировании с использованием статистических и экспертных методов // Автоматика и телемеханика. М., 2008. № 6. С. 129–142.
5. Шипунов Е. А. Бюджетные инвестиции как фактор социально-экономического развития российских регионов // Транспортное дело России. М., 2009. № 5. С. 49–51.
6. Lucas R. E. On the Mechanics of Economic Development // Journal of Monetary Economics. 1988. N 22. P. 3–42.
7. Solow R. M. A Contribution to the Theory of Economic Growth // Quarterly Journal of Economics. 1956. N 22. P. 65–94.
8. Wallace N., Sargent T. Rational expectations and the theory of economic policy // Journal of econometrics. 1976. N 2. P. 169–183.

СВЕДЕНИЯ ОБ АВТОРАХ:

Замураев Константин Александрович

аспирант

Санкт-Петербургский государственный университет
Университетская набережная, 7/9, Санкт-Петербург,
Россия, 199034

эл. почта: konstantin_zamurayev@mail.ru

тел.: (812) 5765282

Прасолов Александр Витальевич

профессор

Санкт-Петербургский государственный университет
Университетская набережная, 7/9, Санкт-Петербург,
Россия, 199034

эл. почта: alexander.prasolov@gmail.com

тел.: (812) 7349441

Zamurayev, Konstantin

Saint-Petersburg State University

7/9 Universitetskaja nab., Saint-Petersburg, Russia,
199034

e-mail: konstantin_zamurayev@mail.ru

tel.: (8142) 5765282

Prasolov, Alexandr

Saint-Petersburg State University

7/9 Universitetskaja nab., Saint-Petersburg, Russia,
199034

e-mail: alexander.prasolov@gmail.com

tel.: (8142) 7349441

УДК 004.01:006.72(470.22)

МАТЕМАТИЧЕСКИЙ АНАЛИЗ ОДНОГО СПОСОБА ПРЕДСТАВЛЕНИЯ ДВУХ FIFO-ОЧЕРЕДЕЙ В ОБЩЕЙ ПАМЯТИ

Н. В. Каблукова¹, А. В. Соколов²

¹ *Петрозаводский государственный университет*

² *Институт прикладных математических исследований
Карельского научного центра РАН*

Во многих приложениях требуется работа с несколькими FIFO-очередями, расположенными в общем пространстве памяти. Для этого применяют различные программные или аппаратные решения [4, 7, 8].

В работе [8] поставлена задача построения математической модели процесса работы с несколькими FIFO-очередями в общей памяти, когда на нечетном шаге дискретного времени с известными вероятностями допускаются операции включения элементов в очереди, а на четном – операции исключения.

В [5] предложена математическая модель этого процесса для двух FIFO-очередей, и решается задача оптимального разбиения общей памяти для очередей в случае их последовательного циклического представления.

В данной работе построены математическая и имитационная модели процесса работы с двумя очередями, когда они двигаются по кругу друг за другом [9].

Ключевые слова: FIFO-очередь, случайное блуждание, регулярные цепи Маркова.

N. V. Kablukova, A. V. Sokolov. MATHEMATICAL ANALYSIS OF A WAY OF PRESENTING TWO FIFO QUEUES IN SHARED MEMORY

Many applications need to work with several FIFO queues located in a common memory space. Various software or hardware solutions are used to this end [4, 7, 8]. In [8] the problem was formulated for building a mathematical model of the process of working with several FIFO queues in shared memory, where insertion of an item in the queue is allowed on an odd step of the discrete time, and deletion operations – on an even step. The probabilities of the operations are known. In [5] a mathematical model of this process for the number of queues $n = 2$ was proposed, and the problem of optimal partitioning of the shared memory between the two FIFO queues in the case of successive cyclic representation was dealt with.

In this paper we construct a mathematical and a simulation models of the process of working with two queues, when they move around a circle, one after another [9].

Key words: FIFO queue, casual walk, regular Markov chains.

ВВЕДЕНИЕ

Эффективные алгоритмы работы с несколькими FIFO-очередями в общем пространстве памяти необходимы при разработке различных сетевых устройств и встроенных операционных систем, управляющих потоками пакетов Internet, таких, например, как Cisco IOS, где требования на время обработки пакетов маршрутизатором очень жесткие. Механизм страничной виртуальной памяти здесь не используется, и вся работа происходит в нескольких пулах оперативной памяти. Количество очередей в таких устройствах может достигать нескольких сотен и тысяч, а в будущем, по экспертным оценкам, может достигнуть нескольких миллионов [14]. Для представления FIFO-очередей применяют различные программные или аппаратные решения [4, 7, 8].

В данной статье анализируются математические модели для некоторых способов представления нескольких FIFO-очередей в памяти одного уровня [7]. В качестве критерия оптимальности рассмотрена минимальная доля потерянных элементов при бесконечном времени работы очередей. Эту величину разумно минимизировать, когда переполнение очереди является не аварийной, а стандартной ситуацией (здесь мы подчеркиваем, что в некоторых приложениях при переполнении очереди работа программы заканчивается, и тогда в качестве критерия оптимальности надо рассматривать максимальное среднее время до переполнения памяти). Т. е., если очередь занимает всю предоставленную ей память, то все последующие элементы, поступающие в нее, отбрасываются до тех пор, пока не появится свободная память (т. е. до тех пор, пока не произойдет исключение элемента из очереди). Такая схема работы применяется, например, в работе сетевых маршрутизаторов [4] в том случае, когда по мере увеличения трафика очередь на исходящем интерфейсе маршрутизатора заполняется пакетами. Такое поведение маршрутизатора называется "сбросом хвоста". Потери пакетов приводят к нежелательному результату, поэтому число таких ситуаций необходимо свести к минимуму. Мы в этой работе строим математическую модель в виде случайных блужданий по целочисленной пирамиде. Первоначально такие модели в виде случайного блуждания в треугольнике [10, 12, 13] были построены для решения задачи анализа процесса работы с двумя стеками, растущими навстречу друг другу, поставленной в [7]. В этих моделях предполагается, что на

каждом шаге дискретного времени с заданными вероятностями происходят некоторые операции со структурами данных. Время выполнения операций это не случайная величина, а константа, поэтому фиксированным является и шаг времени. В [2, 3] предлагались модели работы со стеками в двухуровневой памяти. В [1, 11] предлагались модели для последовательного, связанного и страничного способов представления нескольких FIFO-очередей в памяти одного уровня.

В работе [8] приведены результаты имитационных экспериментов и поставлена задача построить математическую модель процесса работы с несколькими FIFO-очередями в общей памяти, когда операции с очередями выполняются по несколько другому принципу. В данной схеме работы на нечетном шаге допускаются операции включения элементов в одну из n очередей с равными вероятностями, а на четном шаге – операции исключения элементов из очередей с равными вероятностями. Исключение из пустой очереди не приводит к завершению работы. В [8] ставилась задача определить вероятность (как функцию от n и j) того, что очередь, выбранная для операции на j -ом шаге, будет пустой, а также вычислить математическое ожидание количества элементов в очередях после j операций. В данной задаче не рассматривался конкретный способ представления очередей в памяти, т. е. предполагалось, что очереди могут быть неограниченной длины, что на практике невыполнимо.

В [5] предложена математическая модель этого процесса для числа очередей $n = 2$, и решается задача оптимального разбиения общей памяти для двух FIFO-очередей в случае их последовательного циклического представления в предположении, что операции с очередями выполняются по этому принципу, но с неравными вероятностями.

В данной работе построены математическая и имитационная модели процесса работы с двумя очередями, когда они двигаются по кругу друг за другом [9]. В случае, когда одна из очередей была пустой, но ожила, она начинает работу с середины пустого места.

Для решения поставленных задач используется аппарат управляемых случайных блужданий, регулярных цепей Маркова, система Intel® Math Kernel Library PARDISO*. Анализируются результаты численных экспериментов, в которых производится сравнение двух методов представления очередей. Вычисления производились с помощью кластера КарНЦ РАН.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть в памяти размера m единиц мы работаем с двумя циклическими FIFO очередями, которые двигаются друг за другом по кругу. Возобновление движения пустой очереди начинается с середины промежутка между очередями.

Операции, производимые с очередями, выполняются по следующей схеме: на нечетном шаге происходит операция включения в одну из очередей, на четном шаге – операция исключения из какой-либо очереди, причем известны некоторые вероятностные характеристики операций, производимых с очередями. Пусть p_1 и p_2 – вероятности включения информации в первую и вторую очереди соответственно, q_1 и q_2 – вероятности исключения информации из первой и второй очередей соответственно.

Поскольку построенная на основе такой постановки задачи марковская цепь не будет являться регулярной и однородной, два последовательных шага объединяем в один, а также вводим следующие вероятности операций, не изменяющих длины очередей (например, чтение): r_1 – на нечетном шаге и r_2 – на четном шаге, при этом $r_1 \neq 0$, $r_2 \neq 0$.

Соответственно, $p_1 + p_2 + r_1 = 1$, $q_1 + q_2 + r_2 = 1$.

Тогда на каждом шаге могут быть выполнены следующие действия:

1. Включение в первую, исключение из второй очереди с вероятностью $p_1 q_2$,
2. Включение во вторую, исключение из первой очереди с вероятностью $p_2 q_1$,
3. Включение в первую очередь с вероятностью $p_1 r_2$,
4. Включение во вторую очередь с вероятностью $p_2 r_2$,
5. Исключение из первой очереди с вероятностью $q_1 r_1$,
6. Исключение из второй очереди с вероятностью $q_2 r_1$,
7. Остаться на месте с вероятностью $p_1 q_1 + p_2 q_2 + r_1 r_2$,

где $p_1 q_2 + p_2 q_1 + p_1 r_2 + p_2 r_2 + q_1 r_1 + q_2 r_1 + p_1 q_1 + p_2 q_2 + r_1 r_2 = 1$.

Предполагается, что в очередях хранятся данные фиксированного размера. При исключении информации из пустой очереди не происходит завершения работы. Область блуждания показана на рис. 1.

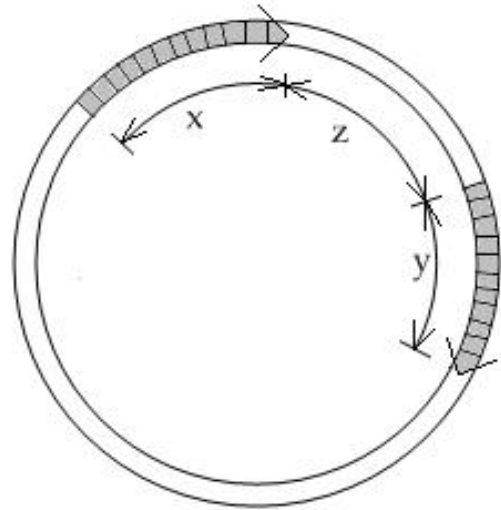


Рис. 1. Схема движения очередей

Целью исследования является определение средней доли потерянных элементов для сравнения со средней долей потерянных элементов при последовательном способе организации очередей в случае оптимального разбиения общей памяти [5].

Обозначим через x и y – текущие длины очередей, через z – расстояние между концом первой очереди и началом второй [7]. В качестве математической модели рассматриваем блуждание по целочисленной трехмерной пирамиде с вершиной $(0, 0, 0)$ и основанием $x + y + z = m$.

Для учета потерянных элементов вводим два фиктивных экрана:

1. (а) $z = -1$ и $x + y' + z = m + 1$ ($y' = y + 1$), где блуждание переходит на экран $z = -1$, если происходит включение в первую очередь в то время, когда $z = 0$, и на экран $x + y' + z = m + 1$, если происходит включение во вторую очередь в то время, когда $x + y + z = m$. В данном случае элемент будет потерян, а очереди будут находиться на экранах до тех пор, пока поступают новые элементы, в случае исключения элемента очереди вернутся в состояние, принадлежащие пирамиде.
- (б) В случае, когда происходит включение во вторую очередь и исключение из первой очереди, в то время когда $x + y + z = m$, происходит потеря элемента и одновременно увеличение расстояния между концом второй и началом первой очереди. Поэтому для этого также необходимо

ввести экран, который будет учитывать данную ситуацию ($x' = x - 1$).

2. Если происходят операции включения и исключения из первой очереди при $z = 0$ либо операции включения и исключения из второй очереди при $x + y + z = m$, происходит потеря элемента, а очереди возвращаются в состояния, принадлежащие пирамиде. Для учета потери элемента вводим экраны $z = -2$ и $x + y' + z = m + 2$ ($y' = y + 2$) соответственно.

Определим схему переходов между состояниями. Пусть (x, y, z) – текущее состояние процесса, тогда блуждание по пирамиде можно описать следующим образом (1). Величина *distance* означает расстояние между концом первой очереди и началом второй. Если оживает первая очередь, то $distance = \left\lceil \frac{m-y-1}{2} \right\rceil$, и $distance = \left\lceil \frac{m-x-1}{2} \right\rceil$, если оживает вторая очередь. Черта сверху состояния указывает на то, что состояние принадлежит одному из фиктивных экранов.

$$\begin{aligned}
 (x, y, z) &\xrightarrow{r_1 r_2} (x, y, z) \\
 (x, y, z) &\xrightarrow{p_1 r_2} (x', y', z') = \begin{cases} (x+1, y, z-1) & x, y, z > 0 \\ (x+1, 0, 0) & y = 0, x < m \\ (1, y, distance-1) & x = 0, y < m \\ \overline{(x, y, -1)} & z = 0, y \neq 0 \text{ либо } x = m \\ \overline{(x, y+1, z)} & x = 0, y = m \end{cases} \\
 (x, y, z) &\xrightarrow{p_2 r_2} (x', y', z') = \begin{cases} (x, y+1, z) & x, y, z > 0 \\ (x, 1, distance) & y = 0, x < m \\ \overline{(0, y, 0)} & x = 0, y < m \\ \overline{(x, y+1, z)} & x + y + z = m \end{cases} \\
 (x, y, z) &\xrightarrow{q_1 r_1} (x', y', z') = \begin{cases} (0, y, 0) & x = 0 \text{ либо } x = 1 \\ (x-1, 0, 0) & y = 0 \\ (x-1, y, z) & \text{иначе} \end{cases} \\
 (x, y, z) &\xrightarrow{q_2 r_1} (x', y', z') = \begin{cases} (x, 0, 0) & y = 0 \text{ либо } y = 1 \\ (0, y-1, 0) & x = 0 \\ (x, y-1, z+1) & \text{иначе} \end{cases} \\
 (x, y, z) &\xrightarrow{p_1 q_2} (x', y', z') = \begin{cases} (x+1, y-1, z) & x, y, z > 0 \\ (x+1, 0, 0) & x < m, y = 0 \text{ либо } y = 1 \\ (1, y-1, distance) & x = 0, y < m \\ \overline{(x, y-1, -1)} & z = 0, y \neq 0 \text{ либо } x = m \\ \overline{(x, y-1, z+1)} & x = 0, y = m \end{cases} \quad (1) \\
 (x, y, z) &\xrightarrow{p_2 q_1} (x', y', z') = \begin{cases} (x-1, y+1, z) & x, y, z > 0 \\ (x-1, 1, distance) & y = 0, x < m \\ \overline{(0, y+1, 0)} & y < m, x = 0 \text{ либо } x = 1 \\ \overline{(x', y, z)} & x + y + z = m \end{cases} \\
 (x, y, z) &\xrightarrow{p_1 q_1} (x', y', z') = \begin{cases} (x, y, z-1) & x, y, z > 0 \\ (0, y, 0) & x = 0 \\ (x, 0, 0) & y = 0 \\ \overline{(x, y, -2)} & z = 0, x \neq 0 \\ \overline{(x, y+1, 0)} & z = 0, y = m \end{cases} \\
 (x, y, z) &\xrightarrow{p_2 q_2} (x', y', z') = \begin{cases} (x, y, z+1) & x, y, z > 0 \\ (x, 0, 0) & y = 0 \\ \overline{(0, y, 0)} & x = 0 \\ \overline{(x, y+2, z+1)} & x + y + z = m \end{cases}
 \end{aligned}$$

Переходы с первого фиктивного экрана:

1) переходы с плоскости $z = -1$ (2):

$$\begin{aligned}
& \overline{(x, y, -1)} \xrightarrow{r_1 r_2} (x, y, 0) \\
& \overline{(x, y, -1)} \xrightarrow{p_1 r_2} \overline{(x, y, -1)} \\
& \overline{(x, y, -1)} \xrightarrow{p_2 r_2} (x', y', z') = \begin{cases} \overline{(x, y + 1, 0)} & x + y = m \\ (x, y + 1, 0) & \text{иначе} \end{cases} \\
& \overline{(x, y, -1)} \xrightarrow{q_1 r_1} (x - 1, y, 0) \\
& \overline{(x, y, -1)} \xrightarrow{q_2 r_1} (x', y', z') = \begin{cases} (x, y, 0) & y = 0 \text{ либо } y = 1 \\ (x, y - 1, 1) & y > 1 \end{cases} \\
& \overline{(x, y, -1)} \xrightarrow{p_1 q_1} \overline{(x, y, -2)} \\
& \overline{(x, y, -1)} \xrightarrow{p_2 q_2} (x', y', z') = \begin{cases} \overline{(x, y + 2, 1)} & x + y = m \\ (x, y, 1) & \text{иначе} \end{cases} \\
& \overline{(x, y, -1)} \xrightarrow{p_1 q_2} (x', y', z') = \begin{cases} \overline{(x, y, -1)} & y = 0 \\ (x, y - 1, -1) & \text{иначе} \end{cases} \\
& \overline{(x, y, -1)} \xrightarrow{p_2 q_1} (x', y', z') = \begin{cases} \overline{(x', y, 0)} & x + y = m \\ (x - 1, y + 1, 0) & \text{иначе} \end{cases}
\end{aligned} \tag{2}$$

2) переходы из состояний $x + y' + z = m + 1$
 $(y' = y + 1)$ (3):

$$\begin{aligned}
& \overline{(x, y + 1, z)} \xrightarrow{r_1 r_2} (x, y, z) \\
& \overline{(x, y + 1, z)} \xrightarrow{p_1 r_2} (x', y', z') = \begin{cases} (x + 1, y, z - 1) & x, y, z > 0 \\ (x + 1, 0, 0) & y = 0, x < m \\ (1, y, \text{distance} - 1) & x = 0, y < m \\ \overline{(x, y, -1)} & z = 0, y \neq 0 \text{ либо } x = m \\ (x, y + 1, z) & x = 0, y = m \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{p_2 r_2} (x, y + 1, z) \\
& \overline{(x, y + 1, z)} \xrightarrow{q_1 r_1} (x', y', z') = \begin{cases} (0, y, 0) & x = 0 \text{ либо } x = 1 \\ (x - 1, 0, 0) & y = 0 \\ (x - 1, y, z) & \text{иначе} \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{q_2 r_1} (x', y', z') = \begin{cases} (x, 0, 0) & y = 0 \text{ либо } y = 1 \\ (0, y - 1, 0) & x = 0 \\ (x, y - 1, z + 1) & \text{иначе} \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{p_1 q_2} (x', y', z') = \begin{cases} (x + 1, y - 1, z) & x, y, z > 0 \\ (x + 1, 0, 0) & x < m, y = 0 \text{ либо } y = 1 \\ (1, y - 1, \text{distance}) & x = 0, y < m \\ \overline{(x, y - 1, -1)} & z = 0, y \neq 0 \text{ либо } x = m \\ (x, y - 1, z + 1) & x = 0, y = m \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{p_2 q_1} (x', y', z') = \begin{cases} \overline{(x - 1, y + 1, z + 1)} & x = 0 \\ \overline{(x, y + 1, z)} & y = 0 \\ (x', y, z) & x \neq 0, y \neq 0 \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{p_1 q_1} (x', y', z') = \begin{cases} (x, y, z - 1) & x, y, z > 0 \\ (0, y, 0) & x = 0 \\ (x, 0, 0) & y = 0 \\ \overline{(x, y, -2)} & z = 0, x \neq 0 \\ (x, y + 1, 0) & z = 0, y = m \end{cases} \\
& \overline{(x, y + 1, z)} \xrightarrow{p_2 q_2} \overline{(x, y + 2, z)}
\end{aligned} \tag{3}$$

3) переходы из состояний (x', y, z) ($x' = x - 1$) повторяют переходы из соответствующих состояний, принадлежащих пирамиде (4):

Переходы из состояний второго фиктивного экрана повторяют переходы из соответствующих состояний, принадлежащих пирамиде (5):

$$\overline{(x', y, z)} \Leftrightarrow \begin{cases} (0, y, 0) & x' = 0 \\ (x', y, z) & \text{иначе.} \end{cases} \tag{4}$$

$$\overline{(x, y, z)} \Leftrightarrow (x - 1, y, z + 2) \text{ для } z = -2$$

$$\overline{(x, y', z)} \Leftrightarrow \begin{cases} (x, 0, 0) & y = 0 \\ (x - 1, 0, 0) & y = 1 \\ (0, y - 1, 0) & y + z = m \\ (x, y, z + 1) & \text{иначе.} \end{cases} \text{ для } x + y' + z = m + 2 \quad (5)$$

Случайное блуждание будем рассматривать в виде регулярной конечной цепи Маркова с переходной матрицей P . Зададим нумерацию так, как показано на рис. 2: сначала нумеруем состояния, принадлежащие основанию пирамиды ($z = 0$), далее поперечным сечениям ($z = 1, 2, \dots, m - 1$) и вершоте пирамиды ($z = m$), затем состояния первого фиктивного экрана, последними нумеруем состояния второго фиктивного экрана.

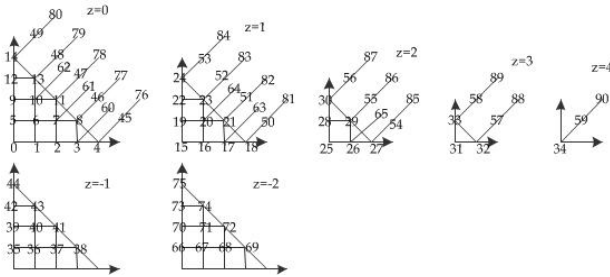


Рис. 2. Нумерация состояний при $m = 4$

Нумерацию начинаем с нуля, общее количество состояний в цепи будет

$$\sum_{i=0}^m \frac{(i+1) \cdot (i+2)}{2} + \frac{5m^2 + 7m + 4}{2}.$$

Далее согласно введенной нумерации состояний и вышеуказанной схеме переходов составляется матрица переходных вероятностей P . В данном исследовании был установлен вид матрицы P для произвольных значений параметра m . При составлении матрицы для каждого конкретного состояния определяются те состояния, в которые процесс переходит при выполнении допустимых операций, и вычисляются соответствующие вероятности переходов. Данный процесс был автоматизирован при помощи программы на языке Си.

Следующим шагом является решение уравнения $\alpha \cdot P = \alpha$, где α – предельный вектор для полученной марковской цепи. Для решения этого уравнения применялось средство решения Intel® Math Kernel Library PARDISO* библиотеки Intel® MKL. Данный инстру-

мент используется для прямого решения систем линейных алгебраических уравнений с разреженной матрицей¹, причем разработаны функции как для симметричных, так и для несимметричных матриц.

Элемент вектора α_i – это средняя доля времени, которое процесс проводит в состоянии i [6]. Для вычисления времени, проводимого процессом в состояниях, где происходят потери элементов очередей, нужно просуммировать элементы вектора α , соответствующие состояниям на экранах. При введенной нумерации это будут последние $\left(\frac{5m^2 + 7m + 4}{2}\right)$ элементов вектора α .

ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

После завершения работы над моделью было проведено несколько экспериментов для возможного выявления присущих ей свойств и сравнения с ранее полученными данными и графиками.

1. Для выяснения ситуаций, когда выгоднее применить случай движения очередей по кругу, был проведен ряд вычислений с различными наборами исходных данных. Аналогичные вычисления были проведены для последовательного способа организации очередей в общей памяти при оптимальном разбиении памяти [5].

Также был рассмотрен случай, когда память заранее делится пополам при последовательном способе организации очередей. Это может быть полезно, если заранее неизвестны вероятностные характеристики операций, производимых с очередями. Некоторые полученные данные указаны в таблице, где столбцы (1) и (2) означают последовательное расположение очередей в памяти при оптимальном делении памяти и делении памяти пополам соответственно, столбец (3) – движение очередей по кругу друг за другом. Указанные результаты были подтверждены имитационными экспериментами.

¹Матрица с большим количеством нулей.

Таблица 1. Величина потерь при различном представлении очередей при $m = 16$

Входные данные	Величина потерь Т при переполнении		
	(1)	(2)	(3)
$p_1 = 0,4, p_2 = 0,45, q_1 = 0,4, q_2 = 0,45$	0,0568543452 ($s = 8$)	0,0568543452	0,0319323619
$p_1 = 0,3, p_2 = 0,45, q_1 = 0,35, q_2 = 0,5$	0,0190375345 ($s = 8$)	0,0190375345	0,0125850917
$p_1 = 0,8, p_2 = 0,1, q_1 = 0,1, q_2 = 0,8$	0,7000000000 ($s = 10$)	0,7000000040	0,6958703274
$p_1 = 0,75, p_2 = 0,1, q_1 = 0,15, q_2 = 0,7$	0,6000000000 ($s = 10$)	0,6000000099	0,5749273399
$p_1 = 0,7, p_2 = 0,15, q_1 = 0,2, q_2 = 0,65$	0,5000000008 ($s = 9$)	0,5000000032	0,4557056302
$p_1 = 0,6, p_2 = 0,3, q_1 = 0,15, q_2 = 0,75$	0,4500000039 ($s = 9$)	0,4500000354	0,4455273285
$p_1 = 0,4, p_2 = 0,45, q_1 = 0,6, q_2 = 0,35$	0,1022326838 ($s = 6$)	0,1030173486	0,0889535997
$p_1 = 0,3, p_2 = 0,6, q_1 = 0,45, q_2 = 0,45$	0,1502357136 ($s = 6$)	0,1514556020	0,1200671464
$p_1 = 0,25, p_2 = 0,65, q_1 = 0,45, q_2 = 0,45$	0,1901271731 ($s = 5$)	0,2002810614	0,1943398747
$p_1 = 0,2, p_2 = 0,3, q_1 = 0,45, q_2 = 0,05$	0,2500009312 ($s = 7$)	0,2500844509	0,3931085764
$p_1 = 0,3, p_2 = 0,5, q_1 = 0,1, q_2 = 0,8$	0,2000042218 ($s = 8$)	0,2000042218	0,3319594498

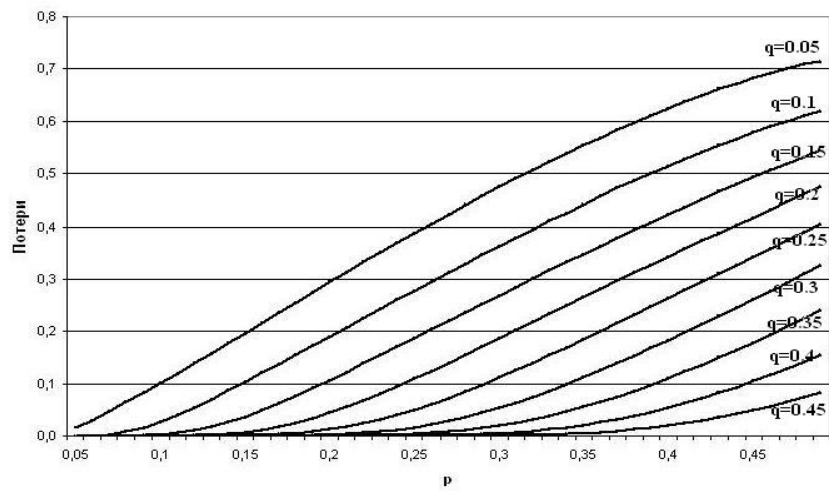


Рис. 3. Потери при постоянной величине q

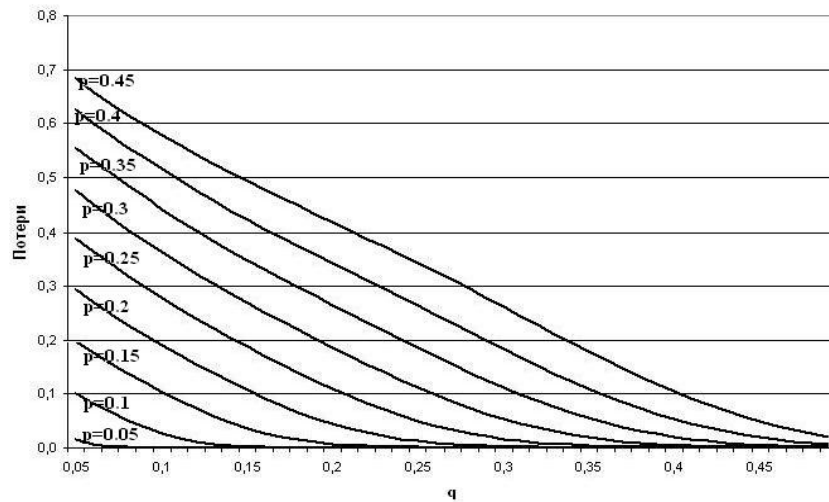


Рис. 4. Потери при постоянной величине p

По этим данным видно, что при движении очередей по кругу потери меньше в двух случаях:

- (а) вероятности включения и исключения в очереди равны (либо практически равны). Это можно прокомментировать тем, что в такой ситуации обе очереди равномерно двигаются друг за другом по кругу, не переполняясь;
- (б) вероятность включения в одну из очередей намного больше, чем вероятность включения в другую очередь, а вероятность исключения из этой очереди – намного меньше, чем из другой. Данный случай объясняется тем, что при данных условиях та из очередей, вероятность включения в которую меньше, практически не занимает места в памяти, соответственно, теряется меньше включаемых элементов.

В других случаях лучше использовать оптимальное разбиение памяти в последовательном способе организации очередей. Если же вероятностные характеристики вообще неизвестны, предпочтительнее делить память пополам при последовательном расположении очередей.

- 2. Пусть $p_1 = p_2 = p$, $q_1 = q_2 = q$, т. е. рассматриваем ситуацию равновероятного включения и исключения информации из очередей, тогда $r_1 = 1 - 2p$, $r_2 = 1 - 2q$.

В данных условиях были построены графики зависимости суммарных потерь от изменения p при постоянной величине q (рис. 3), и от изменения q при постоянной величине p (рис. 4).

Подобные графики были также представлены в работе [5].

Работа выполнена при финансовой поддержке гранта РФФИ (проект 12-01-00253) и Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

ЛИТЕРАТУРА

1. Аксенова Е. А., Драц А. В., Соколов А. В. Оптимальное управление n FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2009. № 6. С. 46–54.
2. Аксенова Е. А., Лазутина А. А., Соколов А. В. Исследование немарковской модели управления стеком в двухуровневой памяти // Программирование. 2004. № 1. С. 1–10.
3. Аксенова Е. А., Соколов А. В. Оптимальное управление двумя параллельными стеками // Дискретная математика. 2007. № 1. С. 67–75.
4. Боллапрагада В., Мэрфи К., Расс У. Структура операционной системы Cisco IOS. М.: Вильямс, 2002. 208 с.
5. Каблукова Н. В., Соколов А. В. Оптимальное разбиение общей памяти для двух последовательных циклических FIFO-очередей // Прикладная информатика. 2012. № (40). С. 113–125.
6. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1960. 272 с.
7. Кнут Д. Искусство программирования для ЭВМ. Т. 1. М.: Вильямс, 2001. 736 с.
8. Седжвик Р. Фундаментальные алгоритмы на C++. К.: Издательство «Диасофт», 2001. 688 с.
9. Соколов А. В. Математические модели и алгоритмы оптимального управления динамическими структурами данных. Петрозаводск: ПетрГУ, 2002. 216 с.
10. Соколов А. В. О распределении памяти для двух стеков // Автоматизация эксперимента и обработки данных. Петрозаводск: изд-во Карельского филиала АН СССР, 1980. С. 65–71.
11. Aksenova E. A., Sokolov A. V. The optimal implementation of two FIFO-queues in single-level memory // Applied Mathematics. 2011. Vol. 2, N 10. P. 1297–1302.
12. Flajolet P. The evolution of two stacks in bounded space and random walks in a triangle // Lecture Notes in Computer Science. 1986. Vol. 223. P. 325–340.
13. Louchard G., Schott R., Tolley M., Zimmermann P. Random walks, heat equation and distributed algorithms // Journal of Computational and Applied Mathematics. 1994. N 53. P. 243–274.
14. Nikolgiannis A., Katevenis M. Multi-Queue Management for Advanced QoS in High-Speed Communication Systems Computer Architecture and VLSI Systems Lab, Institute of Computer Science (ICS) Head.
<http://archvlsi.ics.forth.gr/muqpro/queueMgt.html>

СВЕДЕНИЯ ОБ АВТОРАХ:

Каблукова Наталья Вениаминовна

студентка

Петрозаводский государственный университет
ул. Ленина, 33, Петрозаводск, Республика Карелия,
Россия, 185910

эл. почта: xitushka@mail.ru

Соколов Андрей Владимирович

ведущий научный сотрудник

Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Каре-
лия, Россия, 185910

эл. почта: avs@krc.karelia.ru

тел.: (8142) 766312

Kablukova, Natalia

Petrozavodsk State University

33 Lenina St., 185910 Petrozavodsk, Karelia, Russia

e-mail: xitushka@mail.ru

Sokolov, Andrey

Institute of Applied Mathematical Research, Karelian

Research Centre, Russian Academy of Sciences

11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,

Russia

e-mail: avs@krc.karelia.ru

tel.: (8142) 766312

УДК 519.218

О МИНИМАКСНЫХ ПОДХОДАХ В ЗАДАЧАХ БЕЗОПАСНОСТИ

В. А. Каштанов¹, О. Б. Зайцева²

¹Московский институт электроники и математики, Национальный исследовательский университет «Высшая школа экономики»

²Армавирская государственная педагогическая академия

В статье исследуется модель управляемого полумарковского процесса с катастрофами применительно к проблеме безопасности. Вводятся характеристики (показатели) безопасности. Математическая модель используется для анализа характеристик безопасности технической системы, которая обеспечивает защиту объекта (информации, территории и т. п.). Устанавливается связь характеристик надежности (безотказности и работоспособности) и характеристик безопасности. Анализируется ситуация выбора оптимальной стратегии управления в условиях неполной информации о характеристиках надежности системы.

Ключевые слова: безопасность, управляемый полумарковский процесс с катастрофами, однородная Марковская рандомизированная стратегия управления, надежность, безотказность, ремонтпригодность, оптимальное управление, неполная информация.

V. A. Kashtanov, O. B. Zaytseva. ON MINIMAX APPROACHES TO PROBLEMS OF SAFETY

The model of a controlled semi-Markov process with accidents is investigated as applied to a safety problem. The characteristics (indicators) of safety are introduced. A mathematical model is used to analyze the safety characteristics of the technical system responsible for the security of the facility (information, premises, etc.) The relationship between the reliability characteristics (non-failure operation and operability) and the safety characteristics is investigated. The situation of choice of the optimal control strategy given incomplete information about the reliability characteristics of a system is analysed.

Key words: safety, controlled semi-Markov process with accidents, homogeneous Markov randomized control strategy, reliability, non-failure operation, maintainability, incomplete information.

ВВЕДЕНИЕ

Практическая значимость проблемы безопасности во всех сферах деятельности отдельного человека и общества в целом трудно переоценить. В связи с этим возникает острая проблема создания математических

моделей анализа безопасности, которые позволили бы оценивать количественные характеристики и показатели безопасности, осуществлять прогноз развития опасных ситуаций, принимать аргументированные решения, обеспечивающие безопасное развитие процессов функционирования различных систем (технических, экономических, политических и

т. д.). В настоящей работе мы понимаем безопасность как свойство процесса функционирования системы, которая является объектом исследования и управления. Именно создание математических моделей дает возможность прогнозирования развития опасных ситуаций, выработки стратегии управления, которая обеспечит безопасное течение прогнозируемых прецессов, построить количественные оценки последствий опасного и безопасного развития реальных процессов. Основу нормативных документов в России составляет Федеральный закон "О безопасности" от 28.12.2010 №390-ФЗ. В этом законе безопасность определяется как состояние защищенности жизненно важных интересов личности, общества и государства от внутренних и внешних угроз. Другими словами, есть объект, который должен находиться во множестве безопасных состояний, и есть угрозы, выводящие процесс из этого множества. Основная проблема заключается в выработке стратегии управления процессами функционирования и существования (эволюции) субъектов, которая обеспечивала бы оптимальное в каком-то смысле течение этих процессов. Когда говорят о смысле оптимизации, то тем самым определяют количественный показатель (один или несколько), характеризующий качество управления. В условиях неопределенности (в частности, стохастической неопределенности) эффективной моделью для исследования характеристик безопасности может служить класс управляемых полумарковских процессов с катастрофами, введенный и исследованный в [7, 9]. Ранее аналогичные математические проблемы исследовались А. Д. Соловьевым для процессов восстановления [3, 4], причем в своих работах автор использует терминологию "неблагоприятное событие" вместо термина "катастрофа".

ОПРЕДЕЛЕНИЕ УПРАВЛЯЕМОГО ПОЛУМАРКОВСКОГО ПРОЦЕССА С КАТАСТРОФАМИ

Исходным объектом для конструктивного построения управляемого полумарковского процесса с катастрофами является однородная четырехмерная Марковская цепь или *однородный управляемый процесс Марковского восстановления с катастрофами*

$$(\xi_n, \theta_n, u_n, \eta_n), n \geq 0, \xi_n \in E, \theta_n, \eta_n \in R^+ = [0, \infty), u_n \in U. \quad (1)$$

В этих обозначениях считаем:

- $E = \{1, 2, \dots, N\}$, $N < \infty$, — конечное множество состояний, первая компонента ξ_n однородного управляемого процесса Марковского восстановления с катастрофами принимает дискретные значения из этого множества (ограничимся случаем конечного множества);

- $R^+ = [0, \infty)$ — множество положительных действительных чисел, поэтому вторую компоненту θ_n и последнюю компоненту η_n однородного управляемого процесса Марковского восстановления с катастрофами отождествляем со временем, на пространстве $R^+ = [0, \infty)$ задаем борелевскую алгебру;

- U есть некоторое пространство управлений с σ -алгеброй $\mathbf{A}$ подмножеств этого пространства;

Однородная Марковская цепь $(\xi_n, \theta_n, u_n, \eta_n)$ определяется переходными вероятностями и начальным распределением

$$\mathbf{P}\{\xi_0 = i, \theta_0 < t, u_0 \in B, \eta_0 < x\}.$$

В рассматриваемом случае предполагаем, что Марковская цепь задается переходными вероятностями специального вида

$$\begin{aligned} &\mathbf{P}\{\xi_{n+l} = j, \theta_{n+l} < t, u_{n+l} \in B, \eta_{n+l} < x | \xi_n = i, \\ &\theta_n = \tau, u_n = u, \eta_n = \tau, u_n = y\} \\ &= \mathbf{P}\{\xi_{n+l} = j, \theta_{n+l} < t, u_{n+l} \in B, \eta_{n+l} < x | \xi_n = i\}, i, j \in E, \quad t, \tau, x, y \in R^+, \\ &B \in \mathbf{A}, u \in U, \end{aligned}$$

в которых нет зависимости от параметров τ , u и y — значений второй, третьей и четвертой компонент на предыдущем шаге и номера шага n . Следовательно, будущее поведение однородного управляемого процесса Марковского восстановления с катастрофами зависит только от значения первой компоненты и имеет место однородность этого управляемого Марковского процесса восстановления. При этих дополнительных ограничениях в качестве начального распределения можно задавать константы

$$p_i = \mathbf{P}\{\xi_0 = j, \theta_0 < \infty, u_0 \in U, \eta_0 < \infty\}, \quad \sum_{j \in E} p_j = 1, \quad (2)$$

поскольку событие $\{u_0 \in U\}$ является достоверным, а относительно θ_0 и η_n далее полагаем $\mathbf{P}\{\theta_0 = 0\} = 1, \mathbf{P}\{\eta_0 > 0\} = 1$.

В дальнейшем будем использовать обозначения

$$\begin{aligned} &\tilde{Q}_{i,j}(t, B, x) \\ &= \mathbf{P}\{\xi_{n+l} = j, \theta_{n+l}, t, u_{n+l} \in B, \eta_{n+l} < x | \xi_n = i\}. \end{aligned} \quad (3)$$

Из равенства (3) следует, что модель можно усложнить и считать, что область определения функции $\tilde{Q}_{ij}(t, B, x)$ по переменной B зависит от состояния i . Это значит, что для каждого $i \in E$ задано множество управлений U_i и σ -алгебра $\mathbf{A}_i$ подмножеств этого пространства U_i . Функция $\tilde{Q}_{ij}(t, B, x)$, определяемая равенством (3), задана для $i, j \in E$, $t, x \in R^+$, $B \in \mathbf{A}_j$.

Семейство функций $\tilde{Q}_{ij}(t, B, x)$ порождает на измеримом пространстве $(U_i, \mathbf{A}_i)$ вероятностную меру

$$G_i(B) = \mathbf{P}\{u_{n+l} \in B | \xi_n = i\} \\ = \sum_{j \in E} \tilde{Q}_{ij}(\infty, B, \infty), \quad i \in E, B \in \mathbf{A}_i, \quad (4)$$

где обозначено

$$\tilde{Q}_{ij}(t, B) = \tilde{Q}_{ij}(t, B, \infty) = \lim_{x \rightarrow \infty} \tilde{Q}_{ij}(t, B, x), \\ \tilde{Q}_{ij}(\infty, B, \infty) = \lim_{t \rightarrow \infty} \tilde{Q}_{ij}(t, B, \infty).$$

Так как для любых $j \in E$, $t, x \in R^+$, $B \in \mathbf{A}_j$ справедливо неравенство $G_i(B) \leq \tilde{Q}_{ij}$, то мера $\tilde{Q}_{ij}(t, B, x)$ абсолютно непрерывна относительно меры $G_i(B)$, и на основании теоремы Радона-Никодима, существуют непрерывные измеримые функции $Q_{ij}(t, u, x)$, $u \in U_i$, для которых имеет место равенство

$$\tilde{Q}_{ij}(t, B, x) = \int_B Q_{ij}(t, u, x) G_i(du). \quad (5)$$

Функции $Q_{ij}(t, u, x)$ есть условные вероятности

$$Q_{ij}(t, u, x) = \mathbf{P}\{\xi_{n+l} = j, \theta_{n+l} < t, \eta_{n+l}, x | \xi_n = i, u_{n+l} = u\}. \quad (6)$$

Таким образом, однородный управляемый процесс Марковского восстановления с катастрофами может быть задан семейством матриц

$$\{Q_{ij}(t, u, x)\}, t, x \in R^+, u \in U_i, i, j \in E,$$

множеством вероятностных мер $G + i(B)$, $i \in E$, $B \in \mathbf{A}_i$ и начальным распределением вероятностей состояний p_i , $i \in E$.

Семейство матриц $\{Q_{ij}(t, u, x)\}$, определяемое равенствами (6), будем называть полумарковским ядром управляемого полумарковского процесса с катастрофами, а семейство вероятностных мер $G_i(B)$, определяемое равенствами (4), будем называть семейством управляющих мер или Марковской однородной стратегией управления.

Для введенных выше характеристик справедливо равенство

$$Q_{ij}(t, u) = Q_{ij}(t, u, \infty) = \lim_{x \rightarrow \infty} Q_{ij}(t, u, x),$$

которое устанавливает связь полумарковского ядра $Q_{ij}(t, u, x)$ управляемого полумарковского процесса с катастрофами со стандартным полумарковским ядром $Q_{ij}(t, u)$ управляемого полумарковского процесса [3, 7].

В дальнейшем будем использовать общепринятые обозначения [7, 10]:

$$p_{ij}(u) = Q_{ij}(\infty, u) = \lim_{n \rightarrow \infty} Q_{ij}(t, u),$$

$$F_{ij}(t, u) = \frac{Q_{ij}(t, u)}{p_{ij}(u)},$$

$$Q_{ij}(t) = \int_{u \in U_i} Q_{ij}(t, u) G_i(du), \quad (7)$$

$$p_{ij} = Q_{ij}(\infty) = \lim_{t \rightarrow \infty} Q_{ij}(t),$$

$$F_{ij}(t) = Q_{ij}(t)/p_{ij},$$

если $p_{ij} > 0$, $p_{ij}(u) > 0$, в противном случае функции $F_{ij}(t)$, $f_{ij}(t, u)$ можно доопределить произвольным образом.

Вероятностный смысл введенных обозначений вытекает из определения полумарковского ядра (6), в частности, матрицы

$$P = \begin{pmatrix} p_{11} & \dots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \dots & p_{NN} \end{pmatrix},$$

$$P(u) = \begin{pmatrix} p_{11}(u) & \dots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1}(u) & \dots & p_{NN}(u) \end{pmatrix}$$

определяют переходные вероятности вложения цепи Маркова, характеризующей эволюцию первой компоненты исследуемого процесса, а функции $F_{ij}(t)$, $F_{ij}(t, u)$ определяют условное распределение второй компоненты этого процесса

$$F_{ij}(t) = \mathbf{P}\{\theta_{n+l}, t | \xi_n = i, \xi_{n+l} = j\},$$

$$F_{ij}(t, u) = \mathbf{P}\{\theta_{n+l}, t | \xi_n = i, \xi_{n+l} = j, u_{n+l} = u\}.$$

Из равенств (6) и (7) получаем условное совместное распределение второй и четвертой компонент управляемого процесса Марковского восстановления с катастрофами:

$$\mathbf{P}\{\theta_{n+l} < t, \eta_{n+l} < x | \xi_n = i, \xi_{n+l} = j, u_{n+l} = u\} \\ = \frac{Q_{i,j}(t, u, x)}{p_{i,j}(u)} \leq F_{ij}(t, u). \quad (8)$$

В силу монотонности функции (8) по переменной x получаем неравенство

$$\begin{aligned} \mathbf{P}\{\theta_{n+l}, t, \eta_{n+l}, x | \xi_n = i, u_{n+l} = u\} \\ = \frac{Q_{ij}(t, u, x)}{P_{ij}(u)}. \end{aligned} \quad (9)$$

Следовательно, мера

$$\mathbf{P}\{\theta_{n+l} < t, \eta_{n+l} < x | \xi_n = i, \xi_{n+l} = j, u_{n+l} = u\}$$

абсолютно непрерывна относительно меры $F_{ij}(t, u)$, и на основании теоремы Радона-Никодима существуют измеримые функции

$$\begin{aligned} F_{ij}(x, t, u) = \mathbf{P}\{\eta_{n+l} | \xi_n = i, \xi_{n+l} = j, \\ \theta_{n+l} = t, u_{n+l} = u\}, \end{aligned} \quad (10)$$

для которых имеет место равенство

$$\begin{aligned} \mathbf{P}\{\theta_{n+l} < t, \eta_{n+l} < x | \xi_n = i, \xi_{n+l} = j, u_{n+l} = u\} \\ = \int_0^t F_{ij}(x, t, u) d_y F_{ij}(y, u). \end{aligned} \quad (11)$$

Приведенные рассуждения позволяют сделать вывод о том, что однородный управляемый процесс Марковского восстановления с катастрофами можно задавать стандартными характеристиками управления процесса Марковского восстановления [3, 7] – равенствами (4) и (7) и условными распределениями момента катастрофы, которое определяется равенством (10).

Далее определим *управляемый полумарковский процесс с катастрофами* $Y(t)$ как случайный процесс с четырьмя компонентами:

$$Y(t) = (\xi(t), u(t), \theta(t), \eta(t)), \quad (12)$$

где $\xi(t) = \xi_{\nu(t)}$, $u(t) = u_{\nu(t)+l}$, $\theta(t) = \theta_{\nu(t)+l}$, $\eta(t) = \eta_{\nu(t)+l}$, а считающий процесс $\nu(t)$ определяется равенством $\nu(t) = \sup\{n : \sum_{k \leq n} \theta_k \leq t\}$, $\theta_0 = 0$.

Заметим, что процесс $\xi(t)$ совпадает со стандартным полумарковским процессом [10], вторая компонента управляемого полумарковского процесса $u(t)$ определяет траекторию принимаемых решений.

Эта пара совпадает с управляемым полумарковским процессом $X(t) = (\xi(t), u(t))$ [3]. Третья и четвертая компоненты $\theta(t)$ и $\eta(t)$ принимают значения из пространства $R^+ = [0, \infty)$ и определяют длительность периода между соседними моментами изменения состояния первой компоненты и характеристику развития катастрофы на данном периоде.

Компоненты управляемого полумарковского процесса с катастрофами $Y(t)$ и введенный

выше считающий процесс $\nu(t)$ имеют ступенчатые траектории, для которых совпадают моменты изменения состояний (изменения состояний происходят в моменты $t_n = \sum_{k \leq n} \theta_k$, $n \geq 1$).

ПОСТРОЕНИЕ ФУНКЦИОНАЛА И ЕГО АНАЛИЗ. ПОСТРОЕНИЕ ОПТИМАЛЬНОЙ СТРАТЕГИИ

Компоненты $\theta(t)$ и $\eta(t)$ увяжем с моментами появления некоторого события A , называемого катастрофой (неблагоприятное событие). Если для некоторого $t > 0$ выполняется неравенство

$$\theta(t) = \theta_{\nu(t)+l} > \eta(t) = \eta_{\nu(t)+l},$$

то считаем, что на периоде $[t_{\nu(t)}, t_{\nu(t)+l}]$ произошла катастрофа в момент $t_{\nu(t)} + \eta(t)$. Значение процесса $\nu(t) + l$ определяет номер периода, на котором произошла катастрофа.

Таким образом, получаем последовательность (поток) моментов катастроф при $t > 0$. В частности, если $\zeta = \inf\{t : \theta(t) > \eta(t)\}$, то $\tau_l = \zeta + \eta(\xi)$ есть момент первой катастрофы.

Объектом исследования в моделях безопасности является математическое ожидание момента первой катастрофы $M_i = \mathbf{M}(\tau_l | \xi(0) = i)$. При исследовании этого функционала, построенного на траекториях управляемого полумарковского процесса, возникают следующие проблемы:

- Определение условий, при которых это математическое ожидание существует;
- Определение зависимости этого математического ожидания от вероятностных мер, определяющих Марковскую однородную рандомизированную стратегию;
- Определение оптимальной стратегии, обеспечивающей максимум математического ожидания времени до катастрофы.

Используя принятые выше обозначения, поставим задачу исследовать зависимость условного математического ожидания $M_i = \mathbf{M}(\tau_l | \xi(0) = i)$ от исходных характеристик, определяющих управляемый полумарковский процесс с катастрофами. Решение этих вопросов зависит от свойств поведения вложенной цепи Маркова, определяющей эволюцию первой компоненты $\xi(t)$.

Известно [11], что конечное множество состояний марковской цепи разбивается на непересекающиеся подмножества: подмножество невозвратных состояний, для которых вероятность возвращения строго меньше единицы, и конечный набор неразложимых классов, попав

в любой из которых цепь не может из него выйти.

Обозначим через E_{3s} множество состояний, составляющее один неразложимый класс, s — номер класса, $s = 1, 2, \dots, n$, n — число классов, E_{30} — множество невозвратных состояний. Очевидны равенства

$$E = E_{30} \cup \bigcup_{s=1}^n E_{3s},$$

$$E_{3i} \cap E_{3j} = \emptyset, \quad i \neq j, i, j \in \{0, 1, \dots, n\}.$$

Если цепь Маркова стартует из состояния, принадлежащего некоторому неразложимому множеству, то блуждание будет только по этому множеству и отдельно можно изучать цепь. Если цепь стартует из невозвратного состояния, то некоторое конечное число шагов цепь проведет в подмножестве невозвратных состояний, затем перейдет в одно из неразложимых подмножеств и там останется. Эта эволюция вложенной цепи Маркова будет использована при исследовании проблемы существования условного математического ожидания $M_i = \mathbf{M}(\tau_1 | \xi(0) = i)$.

В исследуемой модели с катастрофами введем другую классификацию состояний вложенной цепи Маркова.

Обозначим через β_{ij} условную вероятность того, что процесс $\xi(t)$ перешел в состояние j (следующим состоянием будет состояние j), и за время этого перехода не произошло катастрофы при условии, что на этом переходе процесс находился в состоянии i . Тогда, используя введенные выше обозначения, получим равенство для введенной вероятности:

$$\beta_{ij} = \int_{u \in U_i} \int_0^\infty \{1 - F_{ij}(t, t, u)\} d_i Q_{ij}(t, u) G_i(du). \quad (13)$$

Из определения вероятности (13) следует, что сумма

$$\beta_i = \sum_{j \in E} \beta_{ij} \quad (14)$$

есть вероятность того, что на периоде не произойдет катастрофы при условии, что процесс этот период проводит в состоянии i .

Если для состояния i выполняется равенство $\beta_i = 0$, то на этом периоде катастрофа произойдет с вероятностью единица. Назовем такое состояние особо опасным и обозначим множество таких особо опасных состояний через E_0 .

Если для состояния i выполняется равенство $\beta_i = 1$, то на этом периоде катастрофа

произойдет с вероятностью ноль. Назовем такое состояние безопасным и обозначим множество таких безопасных состояний через E_2 .

Если для состояния i выполняется неравенство $0 < \beta_i < 1$, то на этом периоде катастрофа произойдет с положительной вероятностью $1 - \beta_i > 0$, отличной от единицы. Назовем такое состояние опасным и обозначим множество таких опасных состояний через E_1 . Величина вероятности $0 < \beta_i < 1$ характеризует степень опасности.

Очевидны соотношения

$$E = E_0 \cup E_1 \cup E_2, \quad E_i E_j = \emptyset,$$

$$i \neq j, i, j = 0, 1, 2.$$

Ответ на вопрос о существовании математического ожидания

$$M_i = \mathbf{M}(\tau_1 | \xi(0) = i)$$

дает теорема, доказательство которой приведено в статье [6].

Теорема 1. Если в каждом неразложимом классе состояний вложенной цепи Маркова управляемого полумарковского процесса с катастрофами и конечным множеством состояний есть хотя бы одно опасное или особо опасное состояние $j \in E_0 \cup E_1$, то математическое ожидание $M_i = \mathbf{M}(\tau_1 | \xi(0) = i)$ существует и представляется как решение алгебраической системы уравнений

$$(1 - \beta_{ii})M_i - \sum_{j \in E, j \neq i} \beta_{ij}M_j = b_i, \quad (15)$$

где

$$b_i = \sum_{j \in E} \int_{u \in U_i} \int_0^i n f t y \left[\int_0^t x d_x F_{ij}(t, x, u) + \int_t^\infty t d_x F_{ij}(t, x, u) \right] d_t Q_{ij}(t, u) G_i(du).$$

Таким образом, получаем, что искомые математические ожидания отыскиваются как решения неоднородной системы линейных алгебраических уравнений с коэффициентами, зависящими от исходных характеристик. Заметим, что коэффициент b_i есть условное математическое ожидание минимума двух случайных величин — длительности периода и случайной величины, определяющей момент катастрофы на периоде, при условии, что процесс провел период в состоянии i .

Решение системы (15) представляется отношением определителей $M_i = \Delta_i / \Delta$, где Δ — определитель матрицы $(I - B)$, (здесь

I — единичная матрица, B — матрица с элементами $\beta_{ij}, i, j \in E$, $\Delta_i, i \in E$, — определитель матрицы $(I - B)$, в которой i -й столбец заменяется столбцом из свободных членов $b = (b_1, b_2, \dots, b_N)^T$, где N — число состояний полумарковского процесса. В теореме 1 сформулированы достаточные условия того, что определитель Δ отличен от нуля. Известно [11], что матрица переходных вероятностей имеет клеточную (блочную) структуру: диагональные клетки (блоки), соответствующие переходам в эргодических множествах и переходам в множестве невозвратных состояний, клетка, соответствующая переходам из невозвратных состояний в эргодические, и, наконец, клетка, соответствующая переходам из эргодических состояний в невозвратные. Так как переходы из эргодических состояний в невозвратные состояния невозможны, то последняя клетка состоит из нулевых элементов.

Из определения вероятностей p_{ij} и β_{ij} следует естественное неравенство $p_{ij} \geq \beta_{ij}$. Поэтому делаем вывод, что матрицы B и $(I - B)$ также имеют блочную (клеточную) структуру. Блочная структура матрицы $(I - B)$ показывает, что система (15) разбивается на блоки, соответствующие эргодическим подмножествам и невозвратным состояниям. Для эргодического подмножества справедливо равенство

$$M_i = \sum_{j \in E_{3k}} \beta_{ij} M_j + b_i, \quad i \in E_{3k},$$

или, в матричном виде, при $k = 1, 2, \dots, n$,

$$(I - B_k)M(k) = b(k), \quad (16)$$

$$I(\bar{G}) = \frac{\int_{U(N)} A(u_1, u_2, \dots, u_N) dG_1(U_1) d(G_2(u_2) \dots dG_n(u_N))}{\int_{U(N)} B(u_1, u_2, \dots, u_N) dG_1(U_1) d(G_2(u_2) \dots dG_n(u_N))}$$

достигается на вырожденных распределениях, если этот максимум существует и пространство Ω , по которому определяется экстремум, содержит все вырожденные распределения. В последнем соотношении

$$U^{(N)} = \prod_{i \in E} U_i -$$

прямое произведение пространств решений в каждом состоянии, $\bar{G} = \{G_1, \dots, G_N\}$ — вероятностные меры, определяющие Марковскую однородную рандомизированную стратегию управления.

где $M(k)$ — вектор-столбец элементов $M_i, i \in E_{3k}$, $b(k)$ — вектор-столбец элементов $b_i, i \in E_{3k}$.

Справедливы следующие два утверждения.

Следствие 1. Для состояний из фиксированного неразложимого класса $i \in E_{3k}$ решение системы (16) определяется отношением

$$M_i = \Delta_i^{(k)} / \Delta^{(k)}, \quad (17)$$

где $\Delta_i^{(k)}$ — определитель матрицы B_k , в котором i -й столбец заменен столбцом $b(k)$.

Следствие 2. Если состояние $i \in E_0$ — особо опасное состояние, то

$$M_i = b_i$$

$$= \sum_{j \in E} \int_{u \in U_i} \int_0^\infty \int_0^t x d_t Q_{ij}(t, u) G_i(du). \quad (18)$$

Элементы матрицы $I - B$, определяемые равенством (13), являются линейными функционалами относительно распределений $G_i(B)$, $i \in E$, $B \in \mathbf{A}_i$, где i — номер строки. Тогда [3] определители $\Delta, \Delta_i, i \in E$, являются линейными функционалами относительно этих распределений. Поэтому решения (15) и (16) — это дробно-линейные функционалы, а решение (17) есть линейный функционал относительно вероятностных мер $G_i(B)$, $i \in E$, $D \in \mathbf{A}_i$, определяющих марковскую рандомизированную стратегию управления.

Исследованная структура характеристики безопасности позволяет построить оптимальную стратегию управления. Известно [9], что максимум дробно-линейного функционала

Тогда задача сводится к поиску максимума отношения подинтегральных функций

$$\begin{aligned} \max_{\bar{G} \in \Omega} I(\bar{G}) &= \max_{u_i, U_i, i \in E} \frac{A(u_1, u_2, \dots, u_N)}{B(u_1, u_2, \dots, u_N)} \\ &= \frac{A(u_1^0, u_2^0, \dots, u_N^0)}{B(u_1^0, u_2^0, \dots, u_N^0)}. \end{aligned}$$

МОДЕЛЬ ЗАЩИТЫ

Далее используем изложенную выше модель управляемого полумарковского процесса с катастрофами для анализа и оценки характеристик безопасности и увязки этих характеристик с характеристиками надежности.

Связь проблем надежности и безопасности была очевидна специалистам давно. При низких характеристиках надежности (безотказности и ремонтпригодности) нельзя обеспечить высокую безопасность функционирования различных технических систем. Примером могут служить транспортные системы, энергетические системы, вычислительные системы и т. п. В ряде работ исследовалась зависимость показателей безопасности от характеристик надежности и стратегии технического обслуживания, поддерживающей систему в работоспособном состоянии [8]. При построении таких математических моделей учитывались характерные особенности ситуации, такие как наличие противоборствующих сторон, случайный характер возникновения отказов, наличие управления (управляющих воздействий), позволяющего улучшать показатель безопасности, ставить и решать оптимизационную задачу. При построении случайного процесса, описывающего эволюцию системы, и постановке задачи оптимизации предполагается, что характеристики надежности известны.

Описание процесса атак на систему защиты

При описании этого процесса, прежде всего, заметим, что он носит дискретный характер, т. е. атаки или попытки пройти систему защиты осуществляются периодически через некоторые, возможно, случайные, интервалы времени. В настоящей работе будем предполагать, что этот процесс описывается процессом Пуассона с параметром λ . У пуассоновского процесса интервалы η между соседними атаками независимы в совокупности и распределены по экспоненциальному закону с параметром λ :

$$\mathbf{P}\{\eta < x\} = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Пуассоновский процесс является Марковским процессом. Для произвольного момента T время ожидания момента следующей попытки (атаки) распределено также по экспоненциальному закону с тем же параметром. Это свойство будет использовано при выводе основных соотношений.

Описание процесса эволюции системы защиты

Пусть задана система, у которой время безотказной работы ξ распределено по закону $F(x) = \mathbf{P}\{\xi < x\}$, $\bar{F}\{x\} = 1 - F(x) = \mathbf{P}\{\xi \geq x\}$. Предположим, что появившийся при функционировании системы отказ самостоятельно обнуляется (проявляется) мгновенно.

В начальный момент $t_0 = 0$ начинается эксплуатация системы защиты и начинается плановое предупредительное обновление (профилактика) системы через время $v \geq 0$, распределенное по закону

$$G(x) = \mathbf{P}\{v < x\}, G(0) = 0.$$

Назначение плановых предупредительных обновлений системы через случайное время $v \geq 0$ означает введение рандомизации в процесс принятия решений, т. е. в тот момент, когда нужно принимать решение, строится реализация τ случайной величины v , $\{v = \tau\}$, распределенной по закону $G(x)$, и плановое предупредительное обновление системы производится через время τ . Если к назначенному моменту $v \geq 0$ система не отказала (произошло событие $\{v < \tau\}$), то в момент $v \geq 0$ начинается плановое предупредительное обновление системы, которое по предположению полностью обновляет систему. Обозначим длительность этого планового предупредительного (профилактического) обновления γ_1 , а $F_1(x) = \mathbf{P}\{\gamma_1 < x\}$ обозначим функцию распределения этой длительности, $\bar{F}_1(x) = \mathbf{P}\{\gamma_1 \geq x\}$.

Наконец, если отказ системы наступил до назначенного момента времени $v \geq 0$ (произошло событие $\{v \geq \xi\}$), то в момент обнаружения отказа ξ начинается внеплановое аварийное обновление системы. Длительность этой восстановительной работы обозначим γ_2 , а закон распределения обозначим $F_2(x) = \mathbf{P}\{\gamma_2 < x\}$, $\bar{F}_2 = \mathbf{P}\{\gamma_2 \geq x\}$.

После проведения возможных восстановительных работ, когда по предположению система полностью обновляется, осуществляется перепланирование момента проведения следующей предупредительной восстановительной работы независимо от прошлого течения процесса и весь процесс обслуживания повторяется заново.

Описанная выше физическая модель полностью укладывается в математическую модель управляемого полумарковского процесса с катастрофами.

Построим этот управляемый полумарковский процесс с катастрофами. Алгоритм построения этого процесса включает следующие этапы: определение Марковских моментов и пространства состояний, построение пространства управлений и стратегий управления, определение полумарковского ядра управляемого полумарковского процесса, определение условных распределений моментов катастроф и, наконец, построение математического ожидания момента катастрофы.

В рассматриваемом случае Марковские моменты — это моменты начала и окончания восстановительных работ. По определению первая компонента управляемого полумарковского процесса с катастрофами $\xi(t)$ между Марковскими моментами не изменяется.

Будем считать, что $\xi(t) = 1$, если в ближайший Марковский момент, предшествующий t , началась плановая предупредительная профилактика системы.

Будем считать, что $\xi(t) = 2$, если в ближайший Марковский момент, предшествующий t , началось внеплановое аварийное восстановление системы.

Наконец, полагаем $\xi(t) = 0$, если ближайший Марковский момент, предшествующий t , является моментом обновления системы. Следовательно, первая компонента управляемого полумарковского процесса с катастрофами $\xi(t)$ принимает значения из конечного множества $E = \{0; 1, 2\}$.

Определение множества управлений и стратегии управления. Из физического описания процесса функционирования системы следует, что решение о назначении периода проведения плановых предупредительных профилактик принимается только в моменты окончания восстановительных работ (процесс $\xi(t)$ находится в состоянии ноль, $\xi(t) = 0$). В этом состоянии можно назначить проведение плановой предупредительной профилактики через любое положительное время. Множество управлений U_0 совпадает с множеством положительных чисел, $U_0 = [0, \infty)$.

Следовательно, по определению, $G_0(u) = G(u)$.

Построение полумарковского ядра. Из описания процесса функционирования системы следует, что при $i = 1, 2$ справедливы равенства

$$\begin{aligned} Q_{i0}(t, u) &= \mathbf{P}\{\gamma_i < t\} = F_i(t), \\ Q_{ij}(t, u) &= 0, \quad j = 1, 2. \end{aligned} \quad (19)$$

Для перехода в состояние ноль необходимо и достаточно, чтобы до момента t закончилась проводимая восстановительная работа, а последние равенства в (19) справедливы, поскольку Марковские моменты начала и окончания восстановительных работ чередуются.

При $i = 0$ справедливы равенства:

- если $j = 1$, то

$$\begin{aligned} Q_{01}(t, u) &= \mathbf{P}\{v < t, v \leq \xi | v = u\} \\ &= \begin{cases} 0, & u > t, \\ \bar{F}(u), & u \leq t, \end{cases} \end{aligned} \quad (20)$$

поскольку при $u > t$ не выполняется первое неравенство и, следовательно, вероятность такого события равна нулю, а при $u \leq t$ необходимо и достаточно выполнения неравенства $\xi \geq U$, вероятность которого равна $\bar{F}(u)$.

Заметим, что при указанных условиях перехода решение совпадает с интервалом между Марковскими моментами с вероятностью единица, т. е.

$$\mathbf{P}\{\theta_{n+1} = u_{n+1} | \xi_n = 0, \xi_{n+1} = 1\} = 1; \quad (21)$$

- если $j = 2$, то

$$\begin{aligned} Q_{02}(t, u) &= \mathbf{P}\{v > \xi, t > \xi | v = u\} \\ &= \begin{cases} F(t), & u > t, \\ F(u), & u \leq t. \end{cases} \end{aligned} \quad (22)$$

Пояснения аналогичны предыдущим: при $u > t$ необходимо и достаточно выполнения неравенства $t > \xi$, вероятность которого равна $F(t)$, а при $u \leq t$ необходимо и достаточно выполнения неравенства $t > \xi$, вероятность которого равна $F(u)$.

Заметим, что при указанных условиях перехода решение также совпадает с интервалом между Марковскими моментами с вероятностью единица, т. е. справедливо равенство, аналогичное равенству (21):

$$\mathbf{P}\{\theta_{n+1} = u_{n+1} | \xi_n = 0, \xi_{n+1} = 2\} = 1. \quad (23)$$

- если $j = 0$, то

$$Q_{00}(t, u) = 0, \quad (24)$$

поскольку переход в состояние ноль невозможен.

Нетрудно проверить очевидное равенство, справедливое при любом $u > 0$:

$$\lim_{t \rightarrow \infty} \sum_{j \in E} Q_{ij}(t, u) = 1.$$

Из равенств (19)–(24) получаем интегрированием по мере $G_0(u) = G(u)$ полумарковское ядро стандартного полумарковского процесса

$$Q_{i0} = F_i(t), \quad Q_{ij}(t) = 0, \quad j = 1, 2, \quad (25)$$

поскольку нет зависимости от управления и функций (19);

$$Q_{00}(t) = 0, \quad Q_{0,1}(t) = \int_0^t \bar{F}(u) dG(u), \quad (26)$$

$$Q_{02}(t) = \int_0^t F(u) dG(u) + \bar{F}(t)[1 - G(t)].$$

Предельным переходом при $t \rightarrow \infty$ получаем переходные вероятности состояний вложенной цепи Маркова

$$p_{ij} = \lim_{t \rightarrow \infty} Q_{ij}(t), \quad i, j \in E = \{0, 1, 2\}.$$

Для исследуемой модели из равенств (25) и (26) находим, что

$$\begin{aligned} p_{00} &= 0, \quad p_{0,1} = \int_0^\infty \bar{F}(u) dG(u), \\ p_{02} &= \int_0^\infty F(u) dG(u), \\ p_{i0} &= 1, \quad p_{ij} = 0, \quad i, j = 1, 2. \end{aligned} \quad (27)$$

Нетрудно проверить очевидное равенство при $i = 0, 1, 2$

$$\lim_{t \rightarrow \infty} \sum_{j \in E} Q_{ij}(t) = \sum_{j \in E} p_{ij} = 1.$$

Отметим, что вложенная цепь Маркова имеет один замкнутый класс сообщающихся состояний с периодом $d = 2$.

Распределение моментов катастроф. Характеристики катастрофы задаются условными вероятностями (10).

Для рассматриваемой модели при $i = 0, j = 1$ справедливо равенство

$$\begin{aligned} F_{01}(x, t, u) &= \mathbf{P}\{\eta_{n+1} < x | \xi_n = 0, \quad \xi_{n+1} = 1, \\ &\quad \theta_{n+l} = t, \quad u_{n+1} = u\} \\ &= \begin{cases} 0, & x \leq t, \\ 1 - e^{-\lambda(x-t)}, & x \geq t. \end{cases} \end{aligned} \quad (28)$$

Поясним соотношение (28). При переходе из состояния $i = 0$ в состояние $j = 1$ система защиты в течение времени t исправно функционирует и, следовательно, по предположению успешно парирует угрозы. Поэтому момент возникновения угрозы равен сумме постоянной величины t или u и некоторой случайной величины ζ , распределенной по экспоненциальному закону (свойство процесса Пуассона). Следовательно,

$$\begin{aligned} F_{01}(x, t, u) &= \mathbf{P}\{\zeta + t < x\} \\ &= \begin{cases} 0, & x \leq t, \\ 1 - e^{-\lambda(x-t)}, & x \geq t. \end{cases} \end{aligned} \quad (29)$$

Таким образом, получается равенство (28).

Для $i = 0, j = 2$ система защиты также в течение всего периода t находится в работоспособном состоянии и способна парировать возникающие угрозы. Переход из состояния $i = 0$

в состояние $j = 2$ означает выполнение неравенства $\xi \geq v$. В этом случае интервал между Марковскими моментами совпадает с моментом появления отказа ξ . Поэтому из определения (10) условного распределения $F_{03}(t, x, u)$ следует, что справедливы соотношения $\xi < v$; $\xi = t$; $v = u$. Следовательно, получаем равенство

$$\begin{aligned} F_{02}(x, t, u) &= F_{01}(x, t, u) \\ &= \begin{cases} 0, & x \leq t, \\ 1 - e^{-\lambda(x-t)}, & x \geq t. \end{cases} \end{aligned} \quad (30)$$

Наконец, если учесть, что в состояниях, когда проводятся ремонты, $i = 1, 2$, система защиты не способна парировать угрозы, получаем для этих состояний, что любое возникновение угрозы означает наличие катастрофы.

Поэтому при $i = 1, 2$

$$F_{i0}(t, x, u) = 1 - e^{-\lambda x}, \quad x \geq 0. \quad (31)$$

Далее покажем, что состояние $i = 0$ — безопасное состояние, а состояния $i = 1, 2$ — опасные состояния. Из определений (13) и (14) и равенства (19), определяющего полумарковское ядро, следует, что при $i = 1, 2$

$$0 < \beta_{i0} = \int_0^\infty e^{-\lambda t} dF_i(t) < 1, \quad (32)$$

так как $\lambda > 0$.

Для $i = 0$ покажем, что $\beta_0 = 1$, т. е. состояние $i = 0$ безопасное.

Для вероятности β_{01} справедливо равенство

$$\beta_{01} = p_{01} = \int_0^\infty \bar{F}(u) dG(u). \quad (33)$$

Для вероятности β_{02} справедливо равенство

$$\beta_{02} = p_{02} = \int_0^\infty F(u) dG(u). \quad (34)$$

Таким образом, доказано, что для $i = 0$ справедливо равенство $\beta_0 = 1$ и, следовательно, состояние $i = 0$ безопасное.

Выше приведенными рассуждениями доказано выполнение всех условий теорем и следствий, приведенных ранее.

Тогда получаем выражение математического ожидания времени до катастрофы через исходные характеристики:

$$M_0(G) = \frac{\int_0^\infty [\int_0^u \bar{F}(y) dy + \bar{F}(u) \int_0^\infty e^{-\lambda t} \bar{F}_1(t) dt + F(u) \int_0^\infty e^{-\lambda t} \bar{F}_2(t) dt] dG(u)}{\int_0^\infty [1 - \bar{F}(u) \int_0^\infty e^{-\lambda t} dF_1(t) - F(u) \int_0^\infty e^{-\lambda t} dF_2(t)] dG(u)} \quad (35)$$

Перечислим основные выводы, которыми мы воспользуемся при последующем анализе модели защиты:

1. Математическое ожидание времени до катастрофы есть дробно-линейный функционал.

2. Оптимальную стратегию управления можно искать в классе детерминированных

стратегий

$$G(x) = \begin{cases} 0, & x \leq u, \\ 1, & x > u. \end{cases} \quad (36)$$

Предположим, что процесс стартует из состояния ноль, когда система защиты новая.

Тогда из равенств (35) и (36) получаем, что соотношение

$$M_0(u) = \frac{\int_0^u \bar{F}(y)dy + \bar{F}(u) \int_0^\infty e^{-\lambda t} \bar{F}_1(t)dt + F(u) \int_0^\infty e^{-\lambda t} \bar{F}_2(t)dt}{1 - \bar{F}(u) \int_0^\infty e^{-\lambda t} dF_1(t) - F(u) \int_0^\infty e^{-\lambda t} dF_2(t)} \quad (37)$$

определяет математическое ожидание времени до катастрофы при условии, что процесс стартует из состояния $i = 0$ и предупредительные профилактики начинаются через время u .

Математическая задача сводится к определению максимума функции (37) и точки u_0 , в которой этот максимум достигается,

$$V_0(u_0) = \max_{u \geq 0} M_0(u). \quad (38)$$

ВЫВОД. Нужно назначать проведение предупредительных профилактик через время u_0 , тогда получим максимальное математическое ожидание времени до катастрофы.

УПРАВЛЕНИЕ ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ

Однако реальная ситуация такова, что характеристики надежности, как правило, точно не известны, а получаются в результате статистических испытаний и обычно на практике имеются только оценки неизвестных характеристик. Следовательно, точно не известны ве-

роятностные характеристики случайного процесса, описывающего эволюцию исследуемой системы. Поэтому в реальной ситуации управлять приходится по неполным данным, что меняет постановку математической задачи.

Сначала исследуем структуру функционала (35) относительно распределения $F(x) = \mathbf{P}\{\xi < x\}$.

Величины

$$a_i = \int_0^\infty e^{-\lambda t} dF_i(t), \quad i = 1, 2.$$

определяют вероятность того, что во время восстановительной работы i -го вида не произойдет катастрофы. Естественно считаем, что справедливо неравенство $a_1 \geq a_2$, так как аварийный ремонт длится дольше предупредительного. Тогда

$$\int_0^\infty e^{-\lambda t} \bar{F}_i(t)dt = \frac{1 - a_i}{\lambda}$$

и из (1) получаем, что

$$\begin{aligned} M_0(F, G) &= \frac{\int_0^\infty [\lambda \int_0^u \bar{F}(y)dy + (1 - a_1) - (a_2 - a_1) \int_0^u dF(y)]dG(u)}{\lambda \int_0^\infty [(1 - a_1) - (a_2 - a_1) \int_0^u dF(y)]dG(u)} \\ &= \frac{\int_0^\infty \int_0^\infty \min(y, u) dF(y) dG(u)}{\int_0^\infty \int_0^\infty B(y, u) dF(y) dG(u)} + \frac{1}{\lambda}, \end{aligned} \quad (39)$$

где

$$\begin{aligned} B(y, u) &= (1 - a_2)I(y \leq u) + (1 - a_1)I(y > u) \\ &= \begin{cases} 1 - a_2, & y \leq u, \\ 1 - a_1, & y > u, \end{cases} \end{aligned} \quad (40)$$

$I(A)$ обозначает индикатор события A .

Равенство (39) показывает, что математическое ожидание $M_0(F, G)$ есть дробно-линейный функционал относительно распределения G , определяющего периодичность

проведения плановых восстановительных работ, и распределения F — распределения времени безотказной работы:

$$M_0(F, G) = \frac{\int_0^\infty \int_0^\infty A(y, u) dF(y) dG(u)}{\int_0^\infty \int_0^\infty B(y, u) dF(y) dG(u)}.$$

Как указывалось выше, точные значения распределений F неизвестны, а известны оценки каких-то характеристик этого распределения. Опишем некоторые конкретные ситуации:

• В результате статистических испытаний определяются значения функции распределения в отдельных точках. Пусть $W = \Omega(n, \vec{y}, \vec{\pi})$ есть множество распределений, которые в n заданных точках

$$\vec{y} = (y_0 = 0, y_1, y_2, \dots, y_n, y_{n+1} = \infty)$$

принимают заданные значения $\vec{\pi} = (\pi_0 = 0, \pi_1, \pi_2, \dots, \pi_n, \pi_{n+1} = 1)$, $F(y_i) = \pi_i$. Для введенных параметров справедливы соотношения

$$y_0 = 0 < y_1 < y_2 < \dots < y_n < y_{n+1} = \infty, \\ \pi_0 = 0 \leq \pi_1 \leq \pi_2 \leq \dots \leq \pi_n \leq \pi_{n+1} = 1.$$

Тогда при постановке задачи можно считать, что $F \in W$.

• В результате статистических испытаний определяются оценки математического ожидания. Пусть $W = \Omega(\mu)$ есть множество распределений с фиксированным математическим ожиданием μ

$$\Omega(\mu) = \{F : \mu = \int_0^\infty [1 - F(x)]dx\}.$$

Тогда можно считать, что $F \in W$.

Теперь сформулируем математическую задачу.

Так как характеристики времени безотказной работы определяются по результатам статистических испытаний и после их обработки получают оценки этих характеристик, то считаем, что функция распределения времени известна неточно, а известно множество распределений W , которому она принадлежит, $F \in W$. Если обозначить через Ω множество распределений положительных случайных величин, то задачу можно сформулировать так.

Найти

$$\max_{G \in \Omega} \inf_{F \in W} M_0(F, G) \quad (41)$$

и распределения F_0, G_0 , на которых этот максимум достигается.

Решение задачи (41) зависит от структуры множества распределений W и структуры исследуемого функционала $M_0(F, G)$. Как показано выше, исследуемый функционал дробно-линеен относительно распределений F, G .

Пусть $F \in W = \Omega(n, \vec{y}, \vec{\pi})$. При поиске внутреннего экстремума (минимума) при фиксированном $G \in \Omega$ воспользуемся доказанным в [1] утверждением: если существует минимум дробно-линейного функционала

$$I(F) = \frac{\int_0^\infty A(y) dF(y)}{\int_0^\infty B(y) dF(y)} \quad (42)$$

по множеству функций распределения $F \in W = \Omega(n, \vec{y}, \vec{\pi})$, то он достигается на ступенчатой функции, имеющей в каждом интервале

$$y \in [y_k, y_{k+1}), \quad k = 0, 1, 2, \dots, n,$$

не более одной точки роста

$$\begin{aligned} \min_{F \in \Omega(n, \vec{y}, \vec{\pi})} I(F) &= \min_{F \in \Omega^{(0)}(n, \vec{y}, \vec{\pi})} I(F) \\ &= \min_{z_k \in [y_k, y_{k+1}), k=0,1,\dots,n} \frac{\sum_{k=0}^n A(z_k)(\pi_{k+1} - \pi_k)}{\sum_{k=0}^n B(z_k)(\pi_{k+1} - \pi_k)} \\ &= \frac{\sum_{k=0}^n A(z_k^0)(\pi_{k+1} - \pi_k)}{\sum_{k=0}^n B(z_k^0)(\pi_{k+1} - \pi_k)}. \end{aligned}$$

Если использовать этот результат для рассматриваемого случая, когда $F \in W = \Omega(n, \vec{y}, \vec{\pi})$, то

$$A(y) = \int_0^\infty A(y, u) dG(u),$$

$$B(y) = \int_0^\infty B(y, u) dG(u)$$

и, следовательно, точка $(z_1^0, z_2^0, \dots, z_n^0)$ зависит от распределения $G(u)$, поэтому поиск внешнего экстремума затруднен, поскольку неизвестна структура исследуемого функционала.

При решении этой проблемы воспользуемся результатом, изложенным в [1].

Пусть задан функционал

$$I(F, G) = \frac{\int_0^\infty A(y, u) dF(y) dG(u)}{\int_0^\infty B(y, u) dF(y) dG(u)}.$$

Если функция $A(y, u)$ не убывает по переменной y при любом u , а функция $B(y, u)$ не возрастает по переменной y при любом u , то при любом распределении G справедливо неравенство

$$\begin{aligned} I(F_1, G) &= \frac{\int_0^\infty \int_0^\infty A(y, u) dF_1(y) dG(u)}{\int_0^\infty \int_0^\infty B(y, u) dF_1(y) dG(u)} \\ &\leq \frac{\int_0^\infty \int_0^\infty A(y, u) dF_2(y) dG(u)}{\int_0^\infty \int_0^\infty B(y, u) dF_2(y) dG(u)} \\ &= I(F_2, G), \end{aligned}$$

функция $F_1(y)$ мажорирует функцию $F_2(y)$, $F_1(y) \geq F_2(y)$, $y \in [0, \infty)$.

Это утверждение позволяет сделать общий вывод: если в пространстве W существует распределение F_0 , мажорирующее любое распределение F , $F_0 \geq F \in W$, то для значений функционала справедливо неравенство $I(F_0, G) \leq I(F, G)$. Значит на мажорирующем распределении достигается минимум.

Таким образом, получаем, что при выполнении сформулированных выше условий на подынтегральные функции, функционал

$$I(F_0, G) = \min_{F \in W} I(F, G),$$

дробно-линеен относительно распределений $G \in \Omega$. Поэтому, если максимум этого функционала существует, то оптимальную стратегию управления можно искать в классе детерминированных стратегий $\Omega^{(0)}$ [3]

$$G(x) = \begin{cases} 0, & x \leq u, \\ 1, & x > u. \end{cases} \quad (43)$$

Таким образом, получаем равенство

$$\begin{aligned} \max_{G \in \Omega} \min_{F \in W} I(F, G) &= \max_{G \in \Omega^{(0)}} I(F_0, G) \\ &= \max_{u \geq 0} \frac{\sum_{k=0}^n A(z_k^0, u)(\pi_{k+1} - \pi_k)}{\sum_{k=0}^n B(z_k^0, u)(\pi_{k+1} - \pi_k)} \\ &= \frac{\sum_{k=0}^n A(z_k^0, u^0)(\pi_{k+1} - \pi_k)}{\sum_{k=0}^n B(z_k^0, u^0)(\pi_{k+1} - \pi_k)}. \end{aligned} \quad (44)$$

В пространстве $W = \Omega(n, \vec{y}, \vec{\pi})$ существует мажорирующее распределение

$$F_0(x) = \begin{cases} 0, & x \leq 0, \\ \pi_k, & x \in (y_k, y_{k+1}], \\ 1, & x > y_n, \end{cases} \quad (45)$$

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

у которого скачки расположены в крайних левых точках интервала $(y_k, y_{k+1}]$.

Для функционала (39) подынтегральная функция числителя определяется равенством $A(y, u) = \min(y, u)$, а подынтегральная функция знаменателя определяется равенством (40). Подынтегральная функция числителя — неубывающая функция по переменной y при любом u , а функция $B(y, u)$ — невозрастающая функция по переменной y при любом u , так как естественно предполагать выполнение неравенства $a_1 \geq a_2$. Таким образом, доказано, что выполняются все условия, при которых внутренний минимум задачи (41) достигается на мажорирующем распределении $F_0(x)$, определяемом равенством (45).

Следовательно,

$$\begin{aligned} M_0(F_0, G_0) \max_{G \in \Omega} \min_{F \in W} M_0(F, G) &= \max_{G \in \Omega} M_0(F_0, G) \\ &= \max_{x \in [0, \infty)} \frac{\sum_{k=0}^n \min(x, y_k)(\pi_{k+1} - \pi_k)}{\sum_{k=0}^n B(x, y_k)(\pi_{k+1} - \pi_k)} + \frac{1}{\lambda} \\ &= \frac{\sum_{k=0}^n \min([0, y_k](\pi_{k+1} - \pi_k))}{\sum_{k=0}^n B(x_0, y_k)(\pi_{k+1} - \pi_k)} + \frac{1}{\lambda}. \end{aligned} \quad (46)$$

Профилактики надо проводить через время x_0 , и равенство (46) дает значение гарантированного уровня безопасности в этих условиях неопределенности.

Рассмотрим другой пример, когда управлять приходится при наличии неполной информации. Часто по результатам статистических испытаний получают довольно точные оценки моментов распределений, в частности, математического ожидания.

Пусть $W = \Omega(\mu)$ есть множество распределений положительных случайных величин с фиксированным математическим ожиданием μ ,

$$W = \Omega(\mu) = \{F : \mu = \int_0^\infty [1 - F(x)]dx\}.$$

Для технических систем естественно считать, что распределение стареющее, т. е. интенсивность отказов $\lambda(x) = f(x)/(q - F(x))$ — неубывающая функция, где $f(x)$ — плотность распределения.

Тогда известно [2], что существует распределение

$$F_0(x) = \begin{cases} 1 - e^{-x/\mu}, & 0 \leq x \leq \mu; \\ 1, & x > \mu, \end{cases} \quad (47)$$

которое мажорирует все распределения множества

$$\Omega(\mu) = \{F : \mu = \int_0^\infty [1 - f(x)]dx\}.$$

Следует заметить, что распределение $F_0(x)$, определяемое соотношением (47), не принадлежит множеству $W = \Omega(\mu)$. Следовательно,

$$\begin{aligned} \inf_{f \in \Omega(\mu)} M_0(F, G) &\geq M_0(F_0, G) \\ &= \frac{1}{\lambda} + \frac{[\mu(1 - e^{-\min(u, \mu)/\mu})]}{(1 - \alpha) - (\alpha_2 - \alpha_1)F_0(u)}. \end{aligned} \quad (48)$$

Отсюда получаем, что

$$\max_{G \in \Omega} \inf_{F \in \Omega(\mu)} M_0(f, g) \geq \max_{G \in \Omega} M_0(F_0, G)$$

или, с учетом равенства (48),

$$\begin{aligned} \max_{G \in \Omega} \inf_{F \in \Omega(\mu)} M_0(F, G) &\geq \max_{u \geq 0} \left\{ \frac{1}{\lambda} + \frac{[\mu(1 - e^{-\min(u, \mu)/\mu})]}{(1 - \alpha_1) - (\alpha_2 - \alpha_1)F_0(u)} \right\} \\ &= \frac{1}{\lambda} + \frac{[\mu(1 - e^{-1})]}{(1 - \alpha_1) - (\alpha_2 - \alpha_1)F_0(u)} \end{aligned} \quad (49)$$

и достигается этот максимум при $u_0 = \mu$.

Таким образом, получаем следующий вывод: если распределение времени безотказной работы системы защиты — стареющее распределение и известно только его математическое ожидание μ , то предупредительные профилактики целесообразно начинать через $u_0 = \mu$, при этом гарантированное значение математического ожидания времени до катастрофы равно

$$\max_{G \in \Omega} \inf_{F \in \Omega(\mu)} M_0(F, G) \\ \geq \frac{1}{\lambda} + \frac{[\mu(1 - e^{-1})]}{(1 - \alpha_1) - (\alpha_2 - \alpha_i)F_0(u)}.$$

ЛИТЕРАТУРА

1. Барзилович Е. Ю., Каштанов В. А. Организация и обслуживание при ограниченной информации о надежности системы. М.: Советское радио, 1975. 136 с.
2. Барлоу Р., Прошан Ф. Математическая теория надежности. М.: Советское радио, 1969. 448 с.
3. Вопросы математической теории надежности / Под ред. Б. В. Гнеденко. М.: Радио и связь, 1983. 376 с.

СВЕДЕНИЯ ОБ АВТОРАХ:

Каштанов Виктор Алексеевич

зав. каф. исследования операций, д. ф.-м. н., профессор
Московский институт электроники и математики, национальный исследовательский университет «Высшая школа экономики»
Б. Трехсвятительский пер., д. 3, Москва, 109028
эл. почта: vakashtan@yandex.ru
тел.: (8495) 9168813

Зайцева Ольга Борисовна

доцент кафедры «Информатики и информационных технологий обучения», к. ф.-м. н.
Армавирская государственная педагогическая академия
Кирова, 50, Армавир, Россия, 352900
эл. почта: o_zaitseva@mail.ru
тел.: (86137) 39756

4. Гнеденко Д. Б., Соловьев А. Д. Одна общая модель резервирования с восстановлением // Известия АН СССР // Техническая кибернетика. 1974. № 6. С. 113–118.

5. Зайцева О. Б. Анализ безопасности функционирования технических систем // Обзорные прикладной и промышленной математики. 2011. Т. 18, вып. 1. С. 94–95.

6. Зайцева О. Б. Анализ полумарковской модели безопасности // Обзорные прикладной и промышленной математики. 2011. Т. 18, вып. 2. С. 223–225.

7. Каштанов В. А. Элементы теории случайных процессов. М.: МИЭМ, 2010. 113 с.

8. Каштанов В. А., Зайцева О. Б. О построении математических моделей безопасности // Труды IV Всерос. научно-практической конференции, Армавир, 30–31 марта 2010. С. 45–49.

9. Каштанов В. А., Янишевский И. М. Исследование функционалов на траекториях процесса с конечным множеством состояний // Кибернетика и системный анализ. АН Украины. 1998. № 1. С. 145–156.

10. Королюк В. С., Турбин А. Ф. Полумарковские процессы и их приложения. Киев: Наукова думка, 1976. 182 с.

11. Ширяев А. Н. Вероятность – 2. М.: МЦНМО, 2004. 408 с.

Kashtanov, Victor

Moscow State Institute of Electronics and Mathematics,
Higher School of Economics
3 B. Trehsvyatitsky St., 109028 Moscow, Russia
e-mail: vakashtan@yandex.ru
tel.: (8495) 9168813

Zaytseva, Olga

Armavir State Pedagogical Academy
50 Kirova St., 352900 Armavir, Russia
e-mail: o_zaitseva@mail.ru
tel.: (86137) 39756

УДК 517.977

СТАБИЛИЗАЦИЯ УПРАВЛЯЕМЫХ ДИНАМИЧЕСКИХ СИСТЕМ ЗА КОНЕЧНОЕ ВРЕМЯ

А. Н. Кириллов

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассмотрена задача стабилизации управляемых динамических систем за конечное время. Предложен метод стабилизации линейной системы с периодическими коэффициентами.

Ключевые слова: стабилизация, конечное время, линейная система.

A. N. Kirillov. FINITE TIME STABILIZATION OF CONTROLLABLE DYNAMICAL SYSTEMS

The problem of finite time stabilization of controllable dynamical systems is considered. A method for stabilization of a linear system with periodical coefficients is proposed.

Key words: stabilization, finite time, linear system.

ОСНОВНЫЕ ПОНЯТИЯ

В конце 1960-х гг. Н. Н. Петров [4] предложил понятие нормальной локальной управляемости. Рассмотрим динамическую систему управления

$$\dot{x} = f(x, u), \quad (1)$$

где x – состояние, $x \in \mathbb{R}^n$, u – управление, $u \in \mathbb{R}^m$. Пусть допустимое управление – кусочно постоянная функция времени t со значениями в ограниченном множестве $U \in \mathbb{R}^m$.

Определение 1. Система (1) называется нормально локально или N -локально управляемой в окрестности начала $x = 0$, если для любого $T > 0$ существует окрестность начала, из каждой точки которой с помощью допустимого управления можно попасть в 0 за время, меньшее T .

Для аналитической динамической системы (1) при $n = 2$, $m = 1$ получены достаточные

условия N -управляемости в терминах коэффициентов степенных рядов, в которые раскладывается f [5]. В [1] автор ввел понятие T -стабилизируемости.

Определение 2. Система (1) называется T -стабилизируемой в точке x^* , если для любого $T > 0$ существует такая окрестность $U(x^*)$ точки x^* , что все траектории системы (1), выходящие из $U(x^*)$, за время, меньшее чем T , с помощью допустимого управления попадут в любую сколь угодно малую окрестность точки x^* и в дальнейшем в ней останутся.

Это понятие мотивировано необходимостью решать задачи стабилизации систем с изменяющейся структурой, когда время стабилизации ограничено временем существования структуры. Введем понятие T -синхронизации. Пусть $x(t, x_0, u)$ – траектория системы (1), соответствующая управлению u и удовлетворяющая условию $x(0, x_0, u) = x_0$.

Определение 3. Система (1) называется T -синхронизуемой в точке x^* , если для любого $T > 0$ существует такая окрестность $U(x^*)$ точки x^* , что для любой точки $x_0 \in U(x^*)$ найдется такое допустимое управление u , что траектория $x(t, x_0, u)$ системы (1) за время $t(x_0) < T$ попадет в точку x^* и в дальнейшем не будет ее покидать, т. е. $x(t, x_0, u) = x^*$ при $\forall t \geq t(x_0)$.

Понятие T -синхронизуемости введено для решения задач управления объектами с большой инерционностью, например, в задачах управления процессами атомной энергетики. Очевидно, T -синхронизуемость является частным случаем T -стабилизируемости и нормальной локальной управляемости.

N -ЛОКАЛЬНАЯ УПРАВЛЯЕМОСТЬ И T -СТАБИЛИЗИРУЕМОСТЬ

Установим связь между понятиями нормальной локальной управляемости и T -стабилизируемости.

Теорема 1. Пусть f, f_x непрерывны в некоторой окрестности начала координат. Если система (1) N -локально управляема, то она T -стабилизируема.

Доказательство. Пусть для определенности система (1) N -локально управляема в начале $x = 0$. Рассмотрим две окрестности начала $U_r(0), U_R(0)$, $r < R$. В [4] получена оценка для времени Δt прохождения траекторией системы (1) слоя $U_R(0) \setminus U_r(0)$

$$\Delta t \geq \frac{R^2 - r^2}{c},$$

где $0 < c(R^2, r^2, K)$ – некоторая постоянная, причем в окрестности начала координат $\|f(x, u)\| \leq K$. Возьмем произвольное $\varepsilon > 0$. Покажем, что найдется окрестность $U_\delta(0)$ такая, что для любой точки $x_0 \in U_\delta(0)$ траектория, начинающаяся в точке x_0 , при соответствующем допустимом (кусочно-постоянном) управлении не покинет $U_\varepsilon(0)$. Возьмем ε_1 такое, что $\varepsilon > \varepsilon_1 > 0$. Время Δt_1 прохождения траекторией слоя удовлетворяет оценке

$$\Delta t_1 \geq \frac{\varepsilon^2 - \varepsilon_1^2}{c}.$$

Поскольку система N -локально управляема в точке $x = 0$, то для любого $\tau > 0$ такого, что

$$0 < \tau < \frac{\varepsilon^2 - \varepsilon_1^2}{c},$$

существует окрестность нуля $U(< \tau)$ такая, что все ее точки за время, меньшее τ , можно перевести с помощью допустимого управления в начало. Если хотя бы одна траектория, начинающаяся в $U(< \tau)$, выйдет на границу окрестности $U_\varepsilon(0)$ до попадания в начало, то для попадания в начало ей придется пройти слой $U_\varepsilon(0) \setminus U_{\varepsilon_1}(0)$, на что потребуется время $\Delta t_1 > \tau$. Это значит, что система не является N -локально управляемой в точке 0, что противоречит предположению. Применяя соответствующее управление для траектории, уже прошедшей начало и находящейся в $U(< \tau)$, получим, что она также не может достичь границы окрестности $U_\varepsilon(0)$. Итак, все траектории, начинающиеся в $U(< \tau)$, остаются в $U_\varepsilon(0)$. Выберем $\delta > 0$ такое, чтобы $U_\delta(0) \subset U(< \tau)$. Тогда получаем, что точка 0 T -стабилизируема. $\square$

Замечание 1. Понятие T -стабилизируемости является более широким, чем понятие нормальной локальной управляемости. Рассмотрим систему

$$\dot{x} = -\frac{x}{\sqrt{|x| - u}},$$

где $x \in \mathbb{R}$, u – кусочно-постоянная функция времени, принимающая значения из множества $\{\frac{1}{n}, n \in \mathbb{N}\} \ni u(t)$. Несложно показать, что точка $x = 0$ T -стабилизируема. Для попадания в любую ε -окрестность начала достаточно положить $u = \frac{1}{n} < \varepsilon$. При этом ни одна траектория не попадает в начало, т. е. N -локальной управляемости в начале нет.

T -СТАБИЛИЗИРУЕМОСТЬ НЕСТАЦИОНАРНЫХ ЛИНЕЙНЫХ СИСТЕМ

Рассмотрим сначала линейную стационарную систему управления

$$\dot{x} = Ax + Bu := f(x, u), \quad (2)$$

где $x \in \mathbb{R}^n$, $u \in \mathbb{R}^k$, A, B – постоянные матрицы соответствующих размерностей. Тогда векторы $f(0, u^i) = Bu^i$, $1 \leq i \leq m$, не могут образовывать положительный базис [4] в $\mathbb{R}^n$ при $m > n$ ни при одном наборе постоянных векторов u^i , $1 \leq i \leq m$, $u^i \in \mathbb{R}^k$, если размерность вектора управлений меньше размерности вектора состояний, т. е. $k < n$. Действительно

$$Bu = b^1 u_1 + \dots + b^k u_k,$$

где b^j – столбцы матрицы B , u_j – компоненты вектора u , $1 \leq j \leq k$, а тогда вектор

Bu принадлежит линейной оболочке векторов $b^1, \dots, b^k$, т. е. собственному подпространству пространства $\mathbb{R}^n$. Таким образом, векторы $f(0, u^i) = Bu^i$, $1 \leq i \leq m$ не образуют положительный базис, и поэтому нельзя применять теорему о T -стабилизации, которая утверждает, что существование положительного базиса $f(0, u^i)$, $1 \leq i \leq n+1$ является достаточным условием T -стабилизируемости системы (1) в точке $x = 0$ [2]. Более того, справедлива

Лемма 1. Пусть $k < n$, управление u — кусочно-постоянно, $u \in \{u^1, \dots, u^m\}$. Тогда точка $x = 0$ не является T -стабилизируемой для системы (2).

Доказательство. Доказательство повторяет рассуждение из теоремы работы [4] о том, что, если векторы $f(0, u^i)$, $i = 1, \dots, m$ образуют одностороннюю совокупность, то система (2) не является локально управляемой. $\square$

Рассмотрим теперь нестационарную линейную систему с постоянной матрицей коэффициентов A

$$\dot{x} = Ax + B(t)u := f(x, u, t), \quad (3)$$

где $B(t)$ — кусочно-непрерывная матрица. Предположим, что по-прежнему, $k < n$. Покажем, что при определенных условиях, накладываемых на $B(t)$, точку $x = 0$ можно стабилизировать за конечное, но не любое время с помощью кусочно-постоянного управления.

Пример 1. Рассмотрим систему

$$\dot{x} = g_1(t)u, \quad \dot{y} = g_2(t)u,$$

где $x, y, u \in \mathbb{R}$,

$$g_1(t) = \begin{cases} 1, & 0 \leq t < 0,25T_0, \\ -1, & 0,25T_0 \leq t < 0,75T_0, \\ 1, & 0,75T_0 \leq t \leq T_0, \end{cases}$$

$$g_2(t) = \begin{cases} 1, & 0 \leq t < 0,5T_0, \\ -1, & 0,5T_0 \leq t \leq T_0, \end{cases}$$

$$g_i(t + T_0) = g_i(t), \quad i = 1, 2,$$

где T_0 — положительная постоянная. Пусть управление u кусочно-постоянно и может принимать два значения: 0 или 1. Покажем, что данная система стабилизируема в нуле за конечное, но не любое время T . При $u = 1$ участки траекторий системы принадлежат прямым $x + y = c$ или $x - y = c$, c — постоянная. Траектории системы периодические, с периодом

T_0 , геометрически представляют собой стороны равных квадратов, параллельные указанным прямым. Покажем, что любую начальную точку $M_0(x_0, y_0)$ можно перевести в начало. Для этого, например, сначала переведем точку M_0 на одну из прямых $x + y = 0$ или $x - y = 0$, а потом по этой прямой переведем точку в начало и там оставляем, полагая $u = 0$ (тем самым решая даже задачу синхронизации за конечное время). Построим соответствующее управление. Предположим, что M_0 принадлежит прямой $x + y = c_0$ (или $x - y = c_0$), которая пересекает $x - y = 0$ (или $x + y = 0$). Полагаем $u = 0$ до того момента времени, когда $g_1 = 1$, $g_2 = -1$, если M_0 лежит выше прямой $x - y = 0$, или когда $g_1 = -1$, $g_2 = 1$, если M_0 лежит ниже прямой $x - y = 0$. В этот момент времени полагает $u = 1$ до тех пор, пока g_1, g_2 не поменяют знаки. Тогда останавливаемся, положив $u = 0$, и возобновляем процедуру. И так до попадания на прямую $x - y = 0$. Потом возобновляем аналогичный процесс управления до попадания в начало. При этом, в силу периодичности правых частей системы, остановка фазовой точки до возобновления движения длится в течение времени, не превышающего $0,75T_0$. Расстояние d , которое надо пройти фазовой точке от M_0 до начала вдоль пути, описанного выше, как нетрудно показать, равно $d = \frac{1}{\sqrt{2}}(|x - y| + |x + y|)$. Модуль вектора фазовой скорости при $u = 1$ равен $\sqrt{\dot{x}^2 + \dot{y}^2} = \sqrt{2}$. Время движения не превосходит величины

$$\Delta T = 0,75pT_0 + 0,5(|x_0 - y_0| + |x_0 + y_0|),$$

где p — количество простоев. Слагаемое $0,75pT_0$ равно максимальному времени простоя. Итак, для любого $T > 0,75pT_0$ можно найти окрестность начала такую, что $\Delta T < T$. Для этого достаточно, чтобы выполнялось условие

$$0,5(|x_0 - y_0| + |x_0 + y_0|) < T - 0,75T_0p,$$

т.е. все начальные точки M_0 , для которых выполняется это условие, переводятся в начало за время, меньшее T , и остаются там. Множество, задаваемое последним неравенством, является кругом радиуса $R = \frac{1}{\sqrt{2}}\sqrt{(x_0 - y_0)^2 + (x_0 + y_0)^2} \leq \frac{1}{\sqrt{2}}(|x_0 - y_0| + |x_0 + y_0|) < \frac{1}{\sqrt{2}}(2T - 1,5T_0p)$. В построенном примере время попадания в произвольно малую окрестность начала ограничено снизу величиной $0,75T_0p$. Заметим, что p зависит от M_0 и значение $p(M_0)$ можно оценить.

Пример показывает, что система вида (3) может быть стабилизируема в нуле за ко-

нечное время T при условии $T > \Theta$, где $0 < \Theta$ – постоянная. Таким образом, введение нестационарной матрицы коэффициентов позволяет решить задачу стабилизации линейной системы за конечное, хотя и не любое время. Этот результат созвучен в некотором смысле результату Г. А. Леонова, решившему проблему Р. Брокетта: насколько введение нестационарности в матрицу коэффициентов линейной обратной связи расширяет возможность стационарной стабилизации системы? Г. А. Леонов дал конструкцию периодической кусочно-постоянной матрицы, расширяющей возможности стационарной стабилизации [3]. Рассмотрим систему

$$\dot{x} = B(t)u, \quad (4)$$

где $x \in \mathbb{R}^n$, $B(t)$ – кусочно-непрерывная матрица размерности $(n \times 1)$, u – управление, $u \in \mathbb{R}$.

Теорема 2. Пусть u – кусочно-постоянная функция, $u(t) \in \{0, 1\}$; $B(t)$ – кусочно-непрерывная, периодическая функция с периодом τ , предположим, что существуют различные постоянные t_i , $1 \leq i \leq n + 1$, удовлетворяющие условиям:

- а) $t_i \in (0, \tau)$;
- б) t_i – внутренние точки промежутков непрерывности $B(t)$;
- в) векторы $B(t_i)$, $1 \leq i \leq n + 1$, образуют положительный базис в $\mathbb{R}^n$.

Тогда точка $x = 0$ синхронизируема за любое время $T > \Theta \geq 0$, где постоянная Θ зависит от окрестности начала, переводимой в точку $x = 0$.

Доказательство. Опишем метод синхронизации. В силу условия б) существуют окрестности U_i точек t_i такие, что если $\tilde{t}_i \in U_i$, то векторы $B(\tilde{t}_i)$, $1 \leq i \leq n + 1$, образуют положительный базис. Если начальный момент времени $t_0 \in U_l$ при некотором $l \in \{1, \dots, n + 1\}$, и при этом вектор $B(t_0)$ направлен внутрь соответствующего цилиндра стабилизации [2], то полагаем $u = 1$, и в течение некоторого времени продвигаемся внутрь цилиндра стабилизации до тех пор, пока $B(t)$ направлен внутрь его. Далее полагаем $u = 0$ (останавливаем движение) и "ждем" наступления момента времени $\tilde{t}_{i_1} \in U_{i_1}$ такого, что вектор $B(\tilde{t}_{i_1})$ направлен внутрь очередного цилиндра стабилизации. В этот момент полагаем $u = 1$. Поскольку длины всех интервалов U_i больше некоторой по-

стоянной $\Delta > 0$, то траектория попадет на основание некоторого очередного цилиндра стабилизации. После этого процесс стабилизации продолжится аналогично тому, как это показано в теореме о T -стабилизации. При этом, если суммарное время ожидания равно Θ , то $T > \Theta$. $\square$

Пример 2. Рассмотрим систему вида

$$\dot{x} = g_1(t - 0, 25uT_0), \quad \dot{y} = g_2(t - 0, 25uT_0),$$

где функции g_1, g_2 – из примера, рассмотренного выше. Пусть управление кусочно-постоянно и может принимать значения $0, 1, 2, 3$. Тогда начало, $x = 0$ T -стабилизируемо. Для того, чтобы это показать, достаточно воспользоваться способом управления, представленном в первом примере, с той разницей, что теперь можно избежать "простоев".

ЗАКЛЮЧЕНИЕ

Представлен способ стабилизации за конечное время линейной нестационарной системы при ограничениях специального вида.

Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

ЛИТЕРАТУРА

1. Кириллов А. Н. Нелинейная стабилизация динамических систем управления // Мехатроника, автоматизация, управление. 2008. № 12. С. 6–11.
2. Кириллов А. Н. Некоторые методы кусочно-постоянной стабилизации нелинейных динамических систем // Вторая российская мультиконференция по проблемам управления. Доклады. СПб., 2008. С. 70–71.
3. Леонов Г. А. Проблема Брокетта в теории устойчивости линейных дифференциальных уравнений // Алгебра и анализ. 2001. № 4. С. 134–155.
4. Петров Н. Н. Локальная управляемость автономных систем // Дифференциальные уравнения. 1968. № 7. С. 1218–1232.
5. Петров Н. Н. Решение одной задачи теории управляемости // Дифференциальные уравнения. 1969. № 5. С. 962–963.

СВЕДЕНИЯ ОБ АВТОРЕ:

Кириллов Александр Николаевич

ведущий научный сотрудник, д. ф.-м. н.

Институт прикладных математических исследований
Карельского научного центра РАН

ул. Пушкинская, 11, Петрозаводск, Республика Каре-
лия, Россия, 185910

эл. почта: kirillov@krc.karelia.ru

тел.: (8142) 763370

Kirillov, Alexandr

Institute of Applied Mathematical Research, Karelian

Research Centre, Russian Academy of Sciences

11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,

Russia

e-mail: kirillov@krc.karelia.ru

tel.: (8142) 763370

УДК 519.865

ЧИСЛЕННАЯ ОПТИМИЗАЦИЯ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ В СТАЦИОНАРНОЙ СЛУЧАЙНОЙ СРЕДЕ

А. О. Олейников

Новгородский государственный университет имени Ярослава Мудрого

В задаче о поведении в случайной среде (также известной как задача о двуруком бандите) предложена стратегия, которая применяет варианты к группам данных на конечном заданном числе этапов. Предложен алгоритм оптимизации размеров групп обрабатываемых данных. Представлены результаты численной оптимизации.

Ключевые слова: поведение в случайной среде, задача о двуруком бандите, робастное управление, параллельная обработка.

A. O. Oleynikov. NUMERICAL OPTIMIZATION OF PARALLEL PROCESSING IN A STATIONARY ENVIRONMENT

A strategy which applies variants to data groups in a bounded number of stages is considered for the problem of control in a stationary environment (also known as the Two-armed bandit problem). An algorithm is suggested for group size optimization. The results of the algorithm application are presented.

Key words: behavior in random environment, two-armed bandit problem, robust control, parallel processing.

ВВЕДЕНИЕ

Рассматривается задача о поведении в случайной среде [5], также известная как задача о двуруком бандите [4], [6] с нормально распределенными доходами на конечном числе шагов N [1]. Доходы при использовании различных вариантов имеют распределения с плотностями

$$f(x|m_\ell) = (2\pi)^{-1/2} \exp \left\{ -(x - m_\ell)^2 / 2 \right\},$$

где ℓ – номер выбранного варианта, а m_ℓ – математическое ожидание этого варианта (в нашем случае $\ell = 1, 2$). Такой двурукий бандит описывается векторным параметром $\theta = (m_1, m_2)$.

Выбор варианта на следующем шаге осуществляется на основе информации о доходах на предыдущих шагах и описывается стратегией σ . Множество, состоящее из всех возможных стратегий, обозначим Σ . Целью управления является получение наибольшего дохода.

При известном параметре θ наилучшей стратегией является та, которая указывает всегда применять вариант, которому соответствует большая из величин m_1, m_2 . В таком случае полный ожидаемый доход равен $N(m_1 \vee m_2)$. Если же параметр неизвестен, то потери дохода вследствие неполноты информации описываются функцией

$$L_N(\sigma, \theta) = E_{\sigma, \theta} \left(\sum_{n=1}^N ((m_1 \vee m_2) - \xi_n) \right).$$

Здесь $E_{\sigma, \theta}$ – математическое ожидание по мере, порожденной стратегией σ и параметром θ . Множество допустимых значений параметра имеет вид $\Theta = \{(m_1, m_2) : |m_1 - m_2| \leq c\}$, где c – некоторая константа ($0 < c < \infty$) [1].

При использовании минимаксного подхода, цель управления состоит в минимизации максимальных ожидаемых потерь на множестве параметров Θ по множеству стратегий Σ , величина

$$R_N^M(\Theta) = \inf_{\Sigma} \sup_{\Theta} L_N(\sigma, \theta)$$

называется минимаксным риском, а соответствующая стратегия – минимаксной стратегией. Эта постановка рекомендована в работе [7].

Одним из возможных видоизменений стратегии является разбиение шагов на несколько групп, таким образом, чтобы каждый шаг принадлежал какой-либо группе и в одной группе могут оказаться только шаги, расположенные подряд. В дальнейшем будем называть шаги пакетами данных, а выбор одного из вариантов на каждом шаге – выбором варианта обработки для пакета.

В данной работе рассмотрен класс стратегий, применяющих одинаковый вариант для всех пакетов в группе и позволяющих вследствие этого вести их параллельную обработку. Для такого класса стратегий предложен алгоритм оптимизации размеров групп и представлены результаты численной оптимизации.

СТРАТЕГИЯ ПАРАЛЛЕЛЬНОГО УПРАВЛЕНИЯ

В [1] установлено, что для описанного случая применима основная теорема теории игр (минимаксные стратегия и риск совпадают с байесовскими на наихудшем априорном распределении). Также в этой работе установлено, что наихудшее априорное распределение является симметрическим и асимптотически однородным, и получено рекуррентное уравнение для вычисления соответствующих байесовских стратегии и риска.

Положим $m_1 = u + v$, $m_2 = u - v$, тогда $\theta = (u + v, u - v)$, $\Theta = \{\theta : |v| \leq c\}$. В новых переменных асимптотически наихудшая плотность распределения может быть выбрана в виде $\nu_a(u, v) = \kappa_a(u)\rho(v)$, где $\kappa_a(u)$ – постоянная плотность при $|u| \leq a$, $\rho(v) = \rho(-v)$ – симметрическая плотность и $a \rightarrow \infty$. Соответствующий байесовский риск равен [1]:

$$R_N^B(\nu_a(u, v)) \quad (1)$$

$$= \inf_{\Sigma} \iint_{\Theta} L_N(\sigma, (u + v, u - v)) \nu_a(u, v) du dv.$$

Обозначим через n_1 и n_2 количество шагов, на которых применены первый и второй варианты соответственно, а через X_1 , X_2 – полные доходы при их применении. При поиске байесовской стратегии на первых двух шагах оба варианта применяются по очереди, стратегию на следующих шагах можно найти, используя приведенные в [1] рекуррентные уравнения.

Далее рассмотрим стратегию, в которой используется разбиение шагов на $k + 2$ группы и управление ведется с тем ограничением, что для всех пакетов в группе применяется один вариант [2], [3]. В начале стратегия указывает применять каждый вариант по M_0 раз (первые две группы), а затем осуществляет оптимальное управление с описанным выше ограничением. Размеры групп обозначим M_0 для первых двух и $M_1, M_2, \dots, M_k$ для оставшихся (таким образом, первые две группы всегда имеют одинаковый размер). Считаем, что $2M_0 + M_1 + M_2 + \dots + M_k = N$.

Такой вид управления позволяет производить параллельную обработку данных. Действительно, если в байесовской стратегии для выбора варианта для следующего пакета нужно знать результат обработки предыдущего, то для описанной стратегии возможно применение выбранного варианта ко всем пакетам в одной группе параллельно. В этом случае полное время управления равно времени обработки $k + 1$ пакетов данных (поскольку первые две группы могут быть обработаны одновременно).

В работе [2] описывается похожая стратегия, но предписывающая разбивать обработку на группы равного размера. Разбиение же на группы разного размера вызвано желанием уменьшить максимальные потери при использовании параллельной стратегии. Как будет видно из результатов оптимизации, уменьшение размеров первых групп ведет к уменьшению максимальных рисков.

Обозначим через $R_{n_1, n_2}^B(Z)$ байесовский риск на последних $N - (n_1 + n_2)$ шагах относительно текущего апостериорного распределения, через $R_{n_1, n_2}^{B(\ell)}(Z)$ – аналогичный риск, вычисленный при условии, что сначала M_i раз выбирается ℓ -й вариант, а затем выполняется оптимальное управление, где $Z = X_1 n_2 - X_2 n_1$, $n_1 \geq M_0$, $n_2 \geq M_0$, $\ell = 1, 2$. Вычисления удобно выполнять для рисков $R_{n_1, n_2}(Z) = R_{n_1, n_2}^B(Z) p_{n_1, n_2}(Z)$, $R_{n_1, n_2}^{(\ell)}(Z) = R_{n_1, n_2}^{B(\ell)}(Z) p_{n_1, n_2}(Z)$, $\ell = 1, 2$, где

$p_{n_1, n_2}(Z)$ – плотность распределения Z при фиксированных n_1, n_2 [2].

При оптимизации использовалось следующее рекуррентное уравнение для вычисления байесовского риска, соответствующего стратегии параллельного управления [2], [3]:

$$R_{n_1, n_2}(\cdot) = \min(R_{n_1, n_2}^{(1)}(\cdot), R_{n_1, n_2}^{(2)}(\cdot)), \quad (2)$$

где $R_{n_1, n_2}^{(1)}(Z) = R_{n_1, n_2}^{(2)}(Z) = 0$ при $n_1 + n_2 = N$,

$$R_{n_1, n_2}^{(1)}(Z) = M_i g_{n_1, n_2}^{(1)}(Z) + n_2^{-1} \times \int_{-\infty}^{+\infty} R_{n_1 + M_i, n_2}(Z + z) h_{n_1, M_i} \left(\frac{Z M_i - n_1 z}{n_2} \right) dz, \quad (3)$$

$$R_{n_1, n_2}^{(2)}(Z) = M_i g_{n_1, n_2}^{(2)}(Z) + n_1^{-1} \times \int_{-\infty}^{+\infty} R_{n_1, n_2 + M_i}(Z + z) h_{n_2, M_i} \left(\frac{Z M_i - n_2 z}{n_1} \right) dz \quad (4)$$

при $n_1 + n_2 = 2M_0 + \dots + M_{i-1}$, $n_1 \geq M_0$, $n_2 \geq M_0$. Здесь

$$g_{n_1, n_2}^{(\ell)}(Z) = \int_0^\infty 2v g_{n_1, n_2}(Z, (-1)^{\ell+1} v) \rho(v) dv, \quad \ell = 1, 2, \quad (5)$$

$$g_{n_1, n_2}(Z, v) = (2\pi n_1 n_2 (n_1 + n_2))^{-1/2} \times \exp \left(-\frac{(Z + 2v n_1 n_2)^2}{2n_1 n_2 (n_1 + n_2)} \right), \quad (6)$$

$$h_{n, M}(z) = \left(\frac{n + M}{2\pi n M} \right)^{1/2} \times \exp \left(-\frac{z^2}{2n M (n + M)} \right). \quad (7)$$

При этом байесовский риск (1) вычисляется по формуле:

$$\lim_{a \rightarrow \infty} R_N^B(\nu_a(u, v)) = 4M_0 \int_0^\infty v \rho(v) dv + \int_{-\infty}^\infty R_{M_0, M_0}(z) dz. \quad (8)$$

Запоминая для каждой тройки Z, n_1, n_2 минимальный риск R^ℓ , мы находим соответствующую байесовскую стратегию.

Для произвольной стратегии параллельного управления $\{\sigma_\ell(z, n_1, n_2)\}$ и плотности $\nu_a(u, v)$ вычислить потери можно, используя следующее рекуррентное уравнение [2]:

$$L_{n_1, n_2}(Z) = \sigma_1(Z, n_1, n_2) L_{n_1, n_2}^{(1)}(Z) + \sigma_2(Z, n_1, n_2) L_{n_1, n_2}^{(2)}(Z), \quad (9)$$

где $L_{n_1, n_2}^{(1)}(Z) = L_{n_1, n_2}^{(2)}(Z) = 0$ при $n_1 + n_2 = N$, а функция $\sigma_\ell(Z, n_1, n_2)$ равна единице, если найденная байесовская стратегия предписывает на данном шаге при данных доходах выбрать ℓ -й вариант, и нулю – в противном случае.

$$L_{n_1, n_2}^{(1)}(Z) = M_i g_{n_1, n_2}^{(1)}(Z) + n_2^{-1} \times \int_{-\infty}^{+\infty} L_{n_1 + M_i, n_2}(Z + z) h_{n_1, M_i} \left(\frac{Z M_i - n_1 z}{n_2} \right) dz, \quad (10)$$

$$L_{n_1, n_2}^{(2)}(Z) = M_i g_{n_1, n_2}^{(2)}(Z) + n_1^{-1} \times \int_{-\infty}^{+\infty} L_{n_1, n_2 + M_i}(Z + z) h_{n_2, M_i} \left(\frac{Z M_i - n_2 z}{n_1} \right) dz \quad (11)$$

при $n_1 + n_2 = 2M_0 + \dots + M_{i-1}$, $n_1 \geq M_0$, $n_2 \geq M_0$. Тогда функция потерь вычисляется по формуле

$$\lim_{a \rightarrow \infty} L_N(\nu_a(u, v)) = 4M_0 \int_0^\infty v \rho(v) dv + \int_{-\infty}^\infty L_{M_0, M_0}(z) dz. \quad (12)$$

ОПТИМИЗАЦИЯ РАЗМЕРОВ ГРУПП ДАННЫХ

Для оптимизации размеров групп использовалась показанная выше методика нахождения минимаксного риска ((2) – (8)) и потерь ((9) – (12)) для заданного разбиения данных $M_0, M_1, \dots, M_k$. Этот риск зависит от c и от разбиения $M_0, M_1, \dots, M_k$. Задача состоит в поиске оптимального разбиения (для фиксированного c)

$$R(M_0^0, M_1^0, \dots, M_k^0) = \min_{M_0, M_1, \dots, M_k} R(M_0, M_1, \dots, M_k).$$

Минимизация может выполняться численными методами, например, методом покоординатного спуска. Однако необходимо учитывать, что в данном случае переменные связаны между собой. Поэтому будем использовать следующую модификацию метода покоординатного спуска: будем поочередно находить оптимальные значения для каждого M_i , начиная с M_0 . При этом все оставшиеся M_i будем считать равными (в случае, если оставшееся количество шагов не

делится на количество оставшихся групп, необходимое количество последних групп увеличивается на единицу). Получившуюся подзадачу однопараметрической оптимизации будем решать прямым перебором. Полученное значение фиксируется, и оптимизация продолжается для следующей группы данных. Минимальное значение для M_i равняется единице, максимальное значение определяется в зависимости от i : для $M_0 \leq (N - k)/2$, для остальных

$$M_i \leq N - (k + 1 - i) - 2M_0 - \sum_{j=1}^{j=i-1} M_j.$$

Прямой перебор возможен для одного параметра благодаря небольшим рассматриваемым значениям N , в дальнейшем он может быть заменен другим методом. Начало оптимизации с M_0 обусловлено тем, что данное значение сильно влияет на потери (так как каждый вариант, в том числе худший, должен быть применен M_0 раз в начале управления), а также тем, что на начальных этапах управления информированность о параметрах среды меняется быстрее, чем на последующих.

Для того, чтобы убедиться, что найденное разбиение действительно соответствует минимуму функции, выбранной в качестве критерия, проверяются разбиения, полученные из найденного путем изменения размеров двух групп на единицу. Если в процессе проверок находилось лучшее разбиение, оно в дальнейшем проверялось по тем же правилам.

Оптимизация проводилась по предложенному алгоритму для $N = 16, 30$ и $k = 6$. Для d были использованы значения от 0,9 до 5,5 включительно с шагом 0,2 (параметр d характеризует априорное распределение: считается, что плотность $\rho(v)$ сосредоточена в двух точках: $v = \pm dN^{-1/2}$). Критерием для сравнений разбиений являлся минимаксный риск (производилась минимизация данного критерия). В результате оптимизации были получены разбиения со следующими размерами: для

$N = 16$: $M_0^0 = 1, M_1^0 = 1, M_2^0 = 1, M_3^0 = 1, M_4^0 = 3, M_5^0 = 1$ и $M_6^0 = 7$. Для $N = 30$: $M_0^0 = 1, M_1^0 = 3, M_2^0 = 2, M_3^0 = 4, M_4^0 = 2, M_5^0 = 5$ и $M_6^0 = 12$.

Для получения результата при $N = 16$ потребовалось найти риски для 71 возможного разбиения, тогда как их общее количество для приведенного примера равняется:

$$\sum_{i=1}^{2i \leq N-k} C_{N-2i-1}^{k-1} = 1897.$$

В данном случае это количество равно количеству композиций числа N длины $k + 2$, у которых первые два слагаемых равны.

Для $N = 30$ общее число вариантов равняется 210574. При использовании предложенного алгоритма потребовалось перебрать 283 варианта (включая проверки).

Таким образом, нам удалось значительно сократить перебор. Однако из-за применения метода покоординатного спуска есть риск нахождения локального минимума.

На рис. 1 показаны приведенные риски r для $N = 16$ ($r = \frac{R(M_0, M_1, \dots, M_k)}{\sqrt{N}}$): линия 1 – для

разбиения с одинаковыми размерами групп $M_i = 2$, линия 2 – для найденного оптимального разбиения, линия 3 – для такого же числа шагов для стратегии, в которой выбор можно менять на каждом шаге. Как видно на рисунке, потери для полученного разбиения меньше, чем для разбиения с одинаковыми размерами групп. Однако они выше, чем потери для стратегии с последовательным управлением, что является платой за возможность параллельной обработки.

На рис. 2 показаны приведенные риски и потери для параллельной (линии 1 и 2) и обычной (линии 3 и 4) стратегий ($N = 30$). На данном рисунке также видно, что применение параллельной стратегии увеличивает риски.

ЗАКЛЮЧЕНИЕ

Предложена стратегия параллельной обработки данных, предписывающая применять разные варианты обработки к группам последовательно поступающих пакетов. Минимаксный риск ищется с помощью рекуррентного уравнения как байесовский, соответствующий наихудшему априорному распределению, и зависит от разбиения групп обрабатываемых данных (для фиксированного c). Рассмотрена задача оптимизации разбиения данных для минимизации минимаксного риска.

В дальнейшем стратегия разбиения может быть развита добавлением ограничения на максимальный размер группы (например, для случая, когда максимальное количество одновременно обрабатываемых групп данных ограничено).

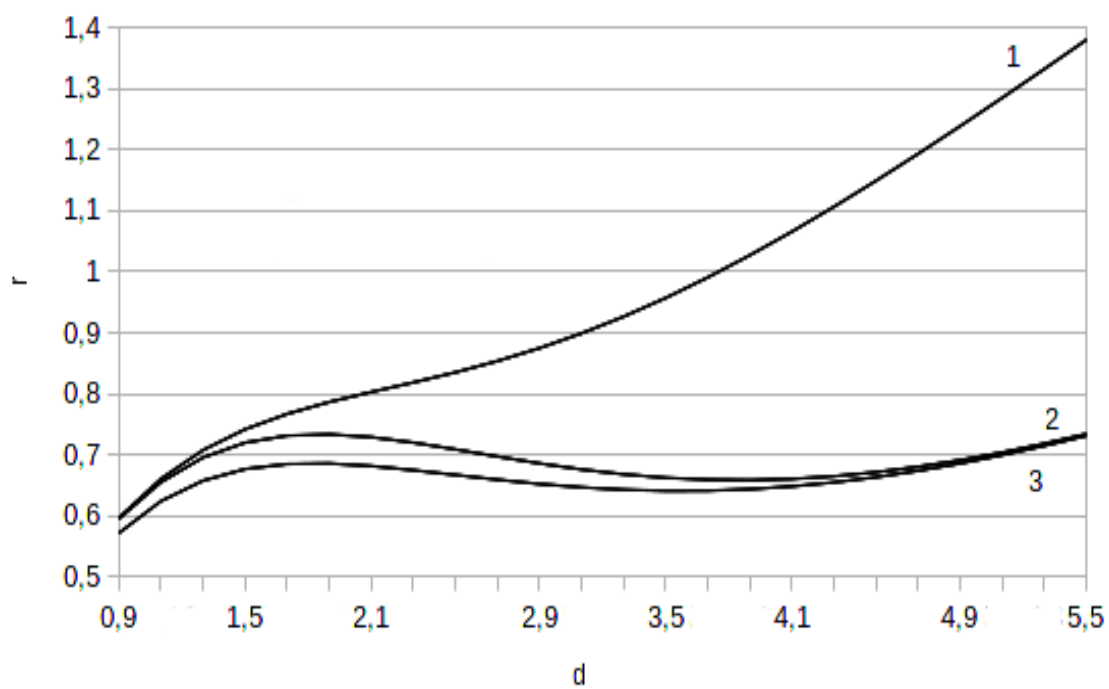


Рис. 1. Результаты численной оптимизации для 16 пакетов

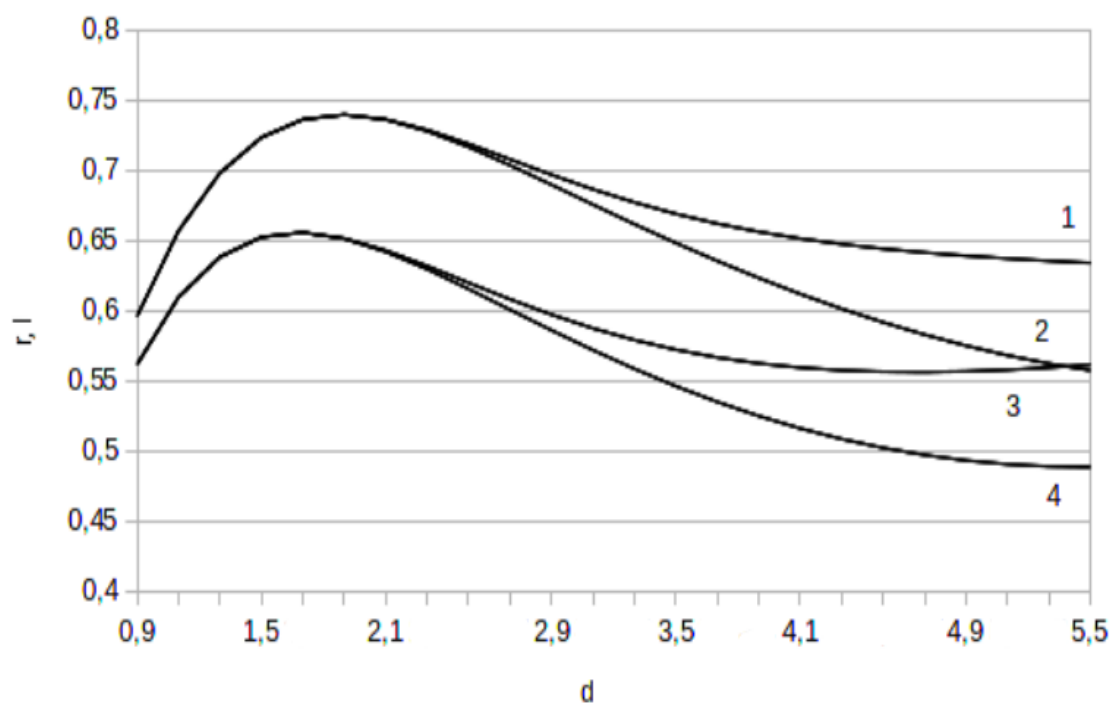


Рис. 2. Результаты численной оптимизации для 30 пакетов

ЛИТЕРАТУРА

1. Колногоров А. В. Нахождение минимаксных стратегии и риска в случайной среде (задача о двуруком бандите) // *АиТ*. 2011. № 5. С. 127–138.
2. Колногоров А. В. Робастное параллельное управление в случайной среде (задача о двуруком бандите) // *АиТ*. 2012. № 4. С. 114–130.
3. Колногоров А. В., Олейников А. О. Оптимизация параллельной многоэтапной обработки в случайной среде // *Обозрение прикладной и*

промышленной математики. 2012. Т. 19, вып. 2. С. 210–211.

4. Пресман Э. Л., Сонин И. М. Последовательное управление по неполным данным. Байесовский подход. М.: Наука, 1982. 286 с.

5. Цетлин М. Л. Исследования по теории автоматов и моделированию биологических систем. М.: Наука, 1969. 316 с.

6. Berry D. A., Fristedt B. Bandit problems. London, New York: Chapman and Hall, 1985. 275 p.

7. Robbins H. Some aspects of the sequential design of experiments // *Bulletin AMS*. 1952. V. 58(5). P. 527–535.

СВЕДЕНИЯ ОБ АВТОРЕ:

Олейников Андрей Олегович

аспирант

Новгородский государственный университет имени Ярослава Мудрого

ул. Большая Санкт-Петербургская, 41, Великий Новгород, Новгородская область, Россия, 173000

эл. почта: duke.gh@yandex.ru

тел.: (8953) 901 64 23

Oleynikov, Andrey

Novgorod State University

41 Bol. Sankt-Peterburgskaya St., Velikiy Novgorod, Russia, 173000

e-mail: duke.gh@yandex.ru

tel.: (8953) 901 64 23

УДК 519.2

О ПРЕДЕЛЬНОМ ПОВЕДЕНИИ ЧИСЛА ДЕРЕВЬЕВ ЗАДАННОГО ОБЪЕМА В СЛУЧАЙНОМ НЕПОМЕЧЕННОМ НЕКОРНЕВОМ ЛЕСЕ

Е. С. Петрова

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассматривается множество $F_{N,n}$ всех случайных лесов, состоящих из N упорядоченных некорневых деревьев и n незанумерованных вершин, на котором задано равномерное распределение вероятностей. Для таких лесов доказаны предельные теоремы о числе деревьев заданного объема при $N, n \rightarrow \infty$ так, что $n/N > 2$.

Ключевые слова: случайные леса, число деревьев заданного объема, обобщенная схема размещения, предельные теоремы.

E. S. Petrova. ON THE LIMIT BEHAVIOUR OF NUMBER OF TREES OF A GIVEN SIZE IN A RANDOM UNLABELLED UNROOTED FOREST

We consider the set $F_{N,n}$ of all random forests consisting of N ordered unrooted trees and n unlabelled vertices with uniform probability distribution on this set. We prove limit theorems for the number of trees of a given size as $N, n \rightarrow \infty$ so that $n/N > 2$.

Key words: random forests, number of trees of a given size, generalized allocation scheme, limit theorems.

ВВЕДЕНИЕ

Изучение случайных лесов, состоящих из некорневых деревьев с незанумерованными вершинами, началось в работах [1] и [2] с помощью обобщенной схемы размещения частиц по ячейкам, введенной и подробно изученной В. Ф. Колчиным (см., например, [4]). Были найдены предельные распределения таких важных характеристик случайного леса, как максимальный объем дерева и число деревьев заданного объема. В [1] рассматривалось предельное поведение случайной величины μ_r , равной числу де-

ревьев, содержащих r вершин. Для случая $(n - LN)/(L_1 N)^{2/3} \rightarrow -\infty$, где константы L и L_1 будут определены ниже в (4), доказано, что предельными распределениями для μ_r являются нормальное при фиксированном r и распределение Пуассона при $r \rightarrow \infty$. Настоящая статья завершает исследование предельного поведения μ_r в случае, когда $(n - LN)/(L_1 N)^{2/3} \geq \beta > -\infty$.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Обозначим $F_{N,n}$ – множество случайных непомеченных лесов, состоящих из N

некорневых деревьев, упорядоченных одним из $N!$ возможных способов, с общим числом вершин n . Зададим на этом объекте равномерное распределение вероятностей.

Рассмотрим производящую функцию

$$t(x) = \sum_{k=1}^{\infty} t_k x^k, \quad (1)$$

где t_k – число непомеченных некорневых деревьев, содержащих k вершин. Свойства функции (1) изучены в [5]. Показано, что начало этой суммы вглядит следующим образом:

$$t(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 6x^6 + 11x^7 + \dots,$$

ее радиус сходимости $R = 0,3383219\dots$. Кроме того, функцию $t(x)$ можно разложить в ряд по степеням $\sqrt{R-x}$:

$$t(x) = a_0 - a_1(R-x) + a_2(R-x)^{3/2} + \dots, \quad (2)$$

где $a_0 = t(R) = 0,5628769\dots$, $a_1 = t'(R) = 3,4127749\dots$, $a_2 = 6,4243753\dots$, а при $k \rightarrow \infty$

$$t_k = \alpha \left(R^k k^{5/2} \right)^{-1} + O \left(\left(R^k k^{7/2} \right)^{-1} \right), \quad (3)$$

$$\alpha = (3a_2/4\sqrt{\pi}) R^{3/2} = 0,5349485\dots$$

Введем обозначения:

$$\begin{aligned} L &= a_1 R / a_0 = 2,0512772\dots, \\ L_1 &= a_2 R^{3/2} / (\sqrt{2} a_0) = 1,5881723\dots \end{aligned} \quad (4)$$

Пусть независимые одинаково распределенные целочисленные случайные величины $\xi_1, \dots, \xi_N$ имеют распределение следующего вида:

$$p_k = \mathbf{P}\{\xi_1 = k\} = t_k R^k a_0^{-1}, \quad k = 1, 2, \dots \quad (5)$$

Обозначим $m = \mathbf{E} \xi_1$, $\sigma^2 = \mathbf{D} \xi_1$. Заметим, что $m = L$.

В работе [2] показано, что справедливо равенство:

$$\begin{aligned} \mathbf{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} \\ = \mathbf{P}\{\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n\}, \end{aligned} \quad (6)$$

где $\eta_1, \dots, \eta_N$ – случайные величины, равные объемам деревьев леса из $F_{N,n}$. Это равенство означает, что для рассматриваемых случайных лесов выполнены условия обобщенной схемы размещения частиц по ячейкам (см., например, [4]).

Справедливо следующее утверждение о предельном поведении числа вершин заданного объема μ_r .

Теорема. Пусть

$$u_r = (k - Np_r) / \sqrt{Np_r(1-p_r)},$$

а $N, n \rightarrow \infty$ так, что

$$(n - LN) / (L_1 N)^{2/3} \geq \beta > -\infty.$$

Тогда для целых неотрицательных k равномерно относительно u_r в любом конечном интервале:

1) при фиксированном r

$$\mathbf{P}\{\mu_r = k\} = \sqrt{2\pi Np_r(1-p_r)}^{-1} e^{-\frac{u_r^2}{2}} (1 + o(1));$$

2) при $r \rightarrow \infty$

$$\mathbf{P}\{\mu_r = k\} = \frac{1}{k!} (Np_r)^k \exp\{-Np_r\} (1 + o(1)).$$

Ниже приводятся леммы 1–3, с помощью которых будет доказана сформулированная теорема.

ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Пусть r – натуральное число. Введем независимые случайные величины $\xi_i^{(r)}$, $i = 1, \dots, N$, такие, что

$$\mathbf{P}\{\xi_i^{(r)} = k\} = \mathbf{P}\{\xi_i = k | \xi_i \neq r\}. \quad (7)$$

Положим

$$\zeta_N = \xi_1 + \dots + \xi_N, \quad \zeta_N^{(r)} = \xi_1^{(r)} + \dots + \xi_N^{(r)}.$$

Обозначим также $m_r = \mathbf{E} \xi_1^{(r)}$, $\sigma_r^2 = \mathbf{D} \xi_1^{(r)}$. Из (5) и (7) вытекают следующие соотношения:

$$\begin{aligned} m_r &= \frac{m - rp_r}{1 - p_r}, \\ \sigma_r^2 &= \frac{\sigma^2}{(1 - p_r)^2} \left(1 - p_r - p_r \frac{(m - r)^2}{\sigma^2} \right). \end{aligned} \quad (8)$$

Пусть $\varphi(t)$ и $\varphi_r(t)$ означают характеристические функции случайных величин ξ_1 и $\xi_1^{(r)}$ соответственно. Используя (1), (2), (4), (5) и (7), нетрудно получить, что

$$\begin{aligned} \varphi(t) &= 1 - L(1 - e^{it}) + \sqrt{2} L_1 (1 - e^{it})^{3/2} + \dots, \\ \varphi_r(t) &= \frac{\varphi(t) - p_r e^{itr}}{1 - p_r}. \end{aligned} \quad (9)$$

В условиях обобщенной схемы размещения, как показано в [4], удобно использовать следующую лемму, которая является следствием соотношения (6).

Лемма 1. *Справедливо равенство*

$$\mathbf{P}\{\mu_r = k\} = \binom{N}{k} p_r^k (1-p)^{N-k} \frac{\mathbf{P}\{\zeta_{N-k}^{(r)} = n - kr\}}{\mathbf{P}\{\zeta_N = n\}}.$$

Лемма 1 показывает, что для изучения асимптотики μ_r достаточно изучить предельное поведение сумм независимых случайных величин ζ_N и $\zeta_N^{(r)}$ и биномиальных вероятностей $\binom{N}{k} p_r^k (1-p)^{N-k}$.

Лемма 2. *Пусть*

$$N, n \rightarrow \infty, \quad S = N(1-p_r)(1+o(1)).$$

Тогда для натуральных h равномерно относительно $v = (h - m_r S)/(L_1 S)^{2/3}$ в любом конечном фиксированном интервале

$$(L_1 S)^{2/3} \mathbf{P}\left\{\frac{\zeta_S^{(r)} - m_r S}{(L_1 S)^{2/3}} = v\right\} = g(v) (1 + o(1)),$$

где $g(v)$ — плотность устойчивого распределения с характеристической функцией

$$f(t) = \exp\left\{-|t|^{3/2} \left(1 - i \frac{t}{|t|}\right)\right\}. \quad (10)$$

Доказательство. Обозначим $\psi_r(t)$ характеристическую функцию $(\zeta_S^{(r)} - m_r S)/(L_1 S)^{2/3}$. Нетрудно видеть, что

$$\psi_r(t) = \exp\left\{-\frac{im_r S t}{(L_1 S)^{2/3}}\right\} \varphi_r^S\left(\frac{t}{(L_1 S)^{2/3}}\right).$$

Используя разложение e^{it} в ряд Тейлора при $t \rightarrow 0$, из (9) получаем:

$$\varphi(t) = 1 - L_1 |t|^{3/2} + i \left(L t + L_1 |t|^{3/2} \frac{t}{|t|} \right) + O(t^2). \quad (11)$$

Отсюда, учитывая (8), получаем выражение

$$\begin{aligned} \ln \psi_r(t) &= -\frac{i L S t}{(L_1 S)^{2/3} (1-p_r)} + S \ln \left(1 - \frac{|t|^{3/2}}{S} \right. \\ &\quad \left. + \frac{i L t}{(L_1 S)^{2/3}} + \frac{i |t|^{3/2}}{S} \frac{t}{|t|} + O\left(\frac{t^2}{S^{4/3}}\right) \right), \end{aligned}$$

откуда

$$\psi_r(t) = \exp\left\{-|t|^{3/2} \left(1 - i \frac{t}{|t|}\right)\right\} (1 + o(1)) \quad (12)$$

По теореме 2.2.2 из [3] выражение $\exp\{-|t|^{3/2} (1 - i(t/|t|))\}$ является характеристической функцией устойчивого закона распределения с показателем $3/2$. Отсюда и из локальной предельной теоремы о сходимости распределений сумм независимых слагаемых (теорема 4.2.1 из [3]) при фиксированном r следует утверждение леммы.

При $r \rightarrow \infty$ $\xi_1^{(r)}, \dots, \xi_S^{(r)}$ образуют схему серий, поэтому применить известные достаточные условия сходимости (см., например, [3]) нельзя. Докажем локальную сходимость, следуя стандартному алгоритму доказательств предельных локальных теорем. Используя формулу обращения, получаем, что

$$\begin{aligned} \mathbf{P}\left\{\frac{\zeta_S^{(r)} - m_r S}{(L_1 S)^{2/3}} = v\right\} &= \frac{1}{2\pi (L_1 S)^{2/3}} \\ &\times \int_{-\pi (L_1 S)^{2/3}}^{\pi (L_1 S)^{2/3}} e^{-ivu} \psi_r(u) du. \end{aligned}$$

Отсюда и из (12) следует, что плотность распределения $f(y)$ функции $\psi_r(t)$ имеет вид

$$f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-iyu - |u|^{3/2} \left(1 - i \frac{u}{|u|}\right)\right\} du.$$

Таким образом, разность

$$R_S = 2\pi \left[(L_1 S)^{2/3} \mathbf{P}\left\{\frac{\zeta_S^{(r)} - m_r S}{(L_1 S)^{2/3}} = v\right\} - f(v) \right]$$

можно представить в виде суммы четырех интегралов:

$$R_S = I_1 + I_2 + I_3 + I_4, \quad (13)$$

где

$$\begin{aligned} I_1 &= \int_{-A}^A e^{-ivu} \left[\psi_r(u) - \exp\left\{-|u|^{3/2} \left(1 - i \frac{u}{|u|}\right)\right\} \right] du, \\ I_2 &= \int_{A < |u| \leq \varepsilon (L_1 S)^{2/3}} e^{-ivu} \psi_r(u) du, \\ I_3 &= \int_{\varepsilon (L_1 S)^{2/3} < |u| \leq \pi (L_1 S)^{2/3}} e^{-ivu} \psi_r(u) du, \\ I_4 &= - \int_{A < |u|} \exp\left\{-ivu - |u|^{3/2} \left(1 - i \frac{u}{|u|}\right)\right\} du. \end{aligned}$$

Покажем, что выбирая положительные постоянные A и ε , каждый из этих интегралов

можно сделать сколь угодно малым при достаточно больших S .

Поскольку

$$|I_4| \leq 2 \int_A^\infty e^{-u^{3/2}} du,$$

величина $|I_4|$ может быть сделана сколь угодно малой выбором достаточно большого A .

Из (12) следует, что при достаточно малом ε для $0 \leq |t| \leq \varepsilon$ справедливо неравенство $|\psi_r(t)| \leq e^{-b_1|t|^{3/2}}$, где $0 < b_1 < 1$. Отсюда получаем, что

$$\begin{aligned} |I_2| &\leq \int_{A < |u| \leq \varepsilon(L_1 S)^{2/3}} |\psi_r(u)| du \\ &\leq 2 \int_A^{\varepsilon(L_1 S)^{2/3}} e^{-b_1 u^{3/2}} du. \end{aligned}$$

Из сходимости последнего интеграла следует, что величину $|I_2|$ выбором достаточно большого A можно сделать сколь угодно малой.

При фиксированном A , согласно (12), интеграл $I_1 \rightarrow 0$. Запрещение случайным величинам $\xi_1^{(r)}, \dots, \xi_S^{(r)}$ принимать значение r не изменяет их максимального шага, а поскольку максимальный шаг равен единице, то при $\varepsilon \leq |t| \leq \pi$ справедлива оценка

$$\max |\varphi_r(t)| = q < 1.$$

Тогда

$$|I_3| \leq 2\pi q_1^S (L_1 S)^{2/3},$$

где величина $q_1 < 1$, следовательно, $I_3 \rightarrow 0$.

Итак, разность R_S из (13) при $S \rightarrow \infty$ стремится к нулю равномерно относительно v . Лемма доказана. $\square$

Лемма 3. Пусть $N, n \rightarrow \infty$ так, что

$$n/N > L, \quad (n - LN)/N^{2/3} \rightarrow \infty.$$

Тогда для $S = N(1 - p_r)(1 + o(1))$

$$\begin{aligned} \mathbf{P} \left\{ \zeta_S^{(r)} = n \right\} &= (\alpha/a_0) S(1 - p_r)^{-1} \\ &\times (n - m_r S)^{-5/2} (1 + o(1)). \end{aligned}$$

Доказательство. Обозначим

$$\gamma = \left(\frac{S^{2/3}}{n - m_r S} \right)^{1/3}, \quad (14)$$

$$\rho = \gamma(n - m_r S), \quad \delta = \gamma^{-1} S^{2/3}.$$

Рассмотрим независимые одинаково распределенные случайные величины $\xi_i' = \xi_i^{(r)} - m_r$, $i = 1, \dots, S$ и их сумму $\zeta_S' = \zeta_S^{(r)} - S m_r$. Представим искомую вероятность в следующем виде:

$$\mathbf{P} \left\{ \zeta_S' = n - m_r S \right\} = P_1 + S P_2 + P_3, \quad (15)$$

где

$$\begin{aligned} P_1 &= \mathbf{P} \left\{ \zeta_S' = n - m_r S; \xi_i' \leq \rho, i = 1, \dots, S \right\}, \\ P_2 &= \mathbf{P} \left\{ \zeta_S' = n - m_r S; \right. \\ &\quad \left. \xi_i' \leq \rho, i = 1, \dots, S-1, \xi_S' > \rho \right\}, \\ P_3 &= \mathbf{P} \left\{ \zeta_S' = n - m_r S; \bigcup_{i \neq j} \{ \xi_i' > \rho, \xi_j' > \rho \} \right\}. \end{aligned}$$

Покажем, оценивая последовательно эти вероятности, что основной вклад в сумму (15) дает второе слагаемое.

Рассмотрим P_1 . Положим

$$\begin{aligned} R(w) &= \\ &= \sum_{k \leq \rho + m_r} \exp\{(k - m_r)w\} \mathbf{P}\{\xi_1' = k - m_r\}. \end{aligned} \quad (16)$$

Из (3) и (5) следует, что при $k \rightarrow \infty$

$$p_k = (\alpha/a_0) k^{-5/2} (1 + o(1)). \quad (17)$$

Используя это соотношение, построим оценки для следующих сумм при достаточно большом l :

$$\sum_{k > l} p_k < c_1 \int_l^\infty y^{-5/2} dy = c_2 l^{-3/2}, \quad (18)$$

$$\sum_{k > l} (k - m_r) p_k < c_3 \int_l^\infty y^{-3/2} dy = c_4 l^{-1/2}, \quad (19)$$

$$\sum_{k \leq l} (k - m_r)^2 p_k < c_5 \int_1^l y^{-1/2} dy = c_6 l^{1/2}, \quad (20)$$

здесь и далее символы $c_1, c_2, \dots$ означают некоторые положительные постоянные.

Из (14) следует, что при выполнении условий леммы

$$\rho^{-3/2} = o(1/S). \quad (21)$$

Пусть $\omega = \rho^{-1}$. Учитывая, что при $0 < y \leq 1$ справедливо равенство $e^y = 1 + y + \delta(y)$, где $\delta(y) \leq y^2/2$, из (7), (16), (18) – (21) получаем соотношение

$$R(\rho^{-1}) = (1 - p_r)^{-1} (1 + o(S^{-1})). \quad (22)$$

Введем вспомогательные независимые одинаково распределенные случайные величины $\xi'_1(\gamma), \dots, \xi'_S(\gamma)$, имеющие распределение вероятностей вида:

$$\begin{aligned} \mathbf{P}\{\xi'_1(\gamma) = k - m_r\} \\ = \frac{\mathbf{P}\{\xi'_1 = k - m_r\} \exp\{(k - m_r)\rho^{-1}\}}{R(\rho^{-1})}, \end{aligned} \quad (23)$$

где $k \leq \rho + m_r$.

Положим $\zeta'_S(\gamma) = \xi'_1(\gamma) + \dots + \xi'_S(\gamma)$. Тогда P_1 имеет следующее представление:

$$\begin{aligned} P_1 = [R(\rho^{-1})]^S \exp\{-\gamma^{-1}\} \\ \times \mathbf{P}\{\zeta'_S(\gamma) = n - m_r S\}. \end{aligned} \quad (24)$$

Оценим последний множитель данного равенства. Обозначим $\varphi_\gamma(t)$ характеристическую функцию случайной величины $\xi'_1(\gamma)$. Получим для нее выражение, используя (16):

$$\varphi_\gamma(t) = \frac{R(\rho^{-1} + it)}{R(\rho^{-1})}. \quad (25)$$

По формуле обращения вероятность $\mathbf{P}\{\zeta'_S(\gamma) = n - m_r S\}$ имеет вид

$$\begin{aligned} \mathbf{P}\{\zeta'_S(\gamma) = n - m_r S\} \\ = \frac{1}{2\pi(L_1 S)^{2/3}} \int_{-\pi(L_1 S)^{2/3}}^{\pi(L_1 S)^{2/3}} \exp\left\{-\frac{it(n - m_r S)}{(L_1 S)^{2/3}}\right\} \\ \times \left[\varphi_\gamma\left(\frac{t}{(L_1 S)^{2/3}}\right)\right]^S dt. \end{aligned} \quad (26)$$

Если $k \leq \rho + m_r$, то

$$\exp\{(k - m_r)\rho^{-1}\} \leq 1 + 2(k - m_r)\rho^{-1},$$

поэтому из (16), (18), (19) и (21) получаем, что

$$|R(\rho^{-1} + it)| \leq (1 - p_r)|\varphi_{m_r}(t)| + o(S^{-1}), \quad (27)$$

где $\varphi_{m_r}(t)$ – характеристическая функция ξ'_1 .

Из определения величины ξ'_1 , следует, что $\varphi_{m_r}(t) = \exp\{-im_r t\}\varphi_r(t)$, где $\varphi_r(t)$ – характеристическая функция случайной величины $\xi_1^{(r)}$. Тогда из (9), (22), (25) и (27) получаем такую оценку:

$$\begin{aligned} \left|\varphi_\gamma\left(\frac{t}{(L_1 S)^{2/3}}\right)\right|^S \\ = c_8 \left|\varphi\left(\frac{t}{(L_1 S)^{2/3}}\right)\right|^S (1 + o(1)). \end{aligned} \quad (28)$$

Разобьем интеграл из равенства (26) на сумму двух интегралов по следующим областям: $|t| \leq \varepsilon(L_1 S)^{2/3}$ и $\varepsilon(L_1 S)^{2/3} < |t| \leq \pi(L_1 S)^{2/3}$, где ε достаточно мало. По свойству характеристических функций решетчатых распределений с максимальным шагом, равном единице, при $\varepsilon < |t| \leq \pi$ выполняется неравенство следующего вида:

$$|\varphi(t)| \leq \exp\{-c_9\}, \quad (29)$$

а из (9) при достаточно малом ε для $|t| \leq \varepsilon$ справедливо:

$$|\varphi(t)| \leq \exp\{-c_{10}|t|^{-3/2}\}. \quad (30)$$

Тогда из (26)–(30) следует, что при достаточно больших S и n :

$$\mathbf{P}\{\zeta'_S(\gamma) = n - m_r S\} \leq c_{11}S^{-2/3}.$$

Отсюда, из (14), (22) и (24):

$$\begin{aligned} P_1 &\leq c_{12}S^{-2/3} \exp\left\{-\frac{1}{\gamma}\right\} \\ &= o\left(S(n - m_r S)^{-5/2}\right). \end{aligned} \quad (31)$$

Оценим теперь вероятность P_2 . Очевидно, что

$$P_2 = \sum_M \mathbf{P}\{\xi'_S = n - m_r - k\} \quad (32)$$

$\times \mathbf{P}\{\zeta'_{S-1} = k - m_r(S-1), \xi'_i \leq \rho, i = 1, \dots, S-1\}$, где множество

$$M = S - 1 \leq k < (n - m_r) - \rho.$$

Введем вспомогательные независимые одинаково распределенные случайные величины $\xi'_i(u)$, $i = 1, \dots, S$, такие, что

$$\mathbf{P}\{\xi'_1(u) = k - m_r\} = \mathbf{P}\{\xi'_1 = k - m_r | \xi'_1 \leq u\}. \quad (33)$$

Пусть, кроме того, $\zeta'_S(u) = \xi'_1(u) + \dots + \xi'_S(u)$. Ниже будем рассматривать случаи, когда $u = \rho$ и $u = \delta$.

Используя (18), (21), (32) и (33), нетрудно получить, что

$$\begin{aligned} P_2 = (1 + o(1)) \sum_M \mathbf{P}\{\xi'_S = n - m_r - k\} \\ \times \mathbf{P}\{\zeta'_{S-1}(\rho) = k - m_r(S-1)\}. \end{aligned} \quad (34)$$

Обозначим через $\varphi_\rho(t)$ характеристическую функцию случайной величины $\xi'_1(\rho)$. Используя (3), (5), (7)–(9), (14), (21) и (33), найдем, что при любом фиксированном t

$$\varphi_\rho^S\left(\frac{t}{(L_1 S)^{2/3}}\right) = f(t)(1 + o(1)), \quad (35)$$

где $f(t)$ – характеристическая функция устойчивого закона с показателем $3/2$, заданная в (10). Отсюда

$$\mathbf{P}\{\zeta'_S(\rho) \leq y(L_1 S)^{2/3}\} \rightarrow \int_{-\infty}^y g(x)dx. \quad (36)$$

Далее покажем, что при достаточно больших S, n

$$\mathbf{P}\{\zeta'_{S-1}(\rho) > \delta\} \leq c_{13}\gamma^{3/2}(1-p_r)^{-1}. \quad (37)$$

Учитывая (33) и то, что $\rho > \delta$, для искомой вероятности находим следующее неравенство:

$$\begin{aligned} \mathbf{P}\{\zeta'_{S-1}(\rho) > \delta\} &\leq (S-1) \mathbf{P}\{\xi'_1(\rho) > \delta\} \\ &+ \left(\frac{1 - P\{\xi'_1 > \delta\}}{1 - P\{\xi'_1 > \rho\}} \right)^{S-1} \mathbf{P}\{\zeta'_{S-1}(\delta) > \delta\}. \end{aligned} \quad (38)$$

Из (18) и (33) нетрудно получить соотношение:

$$(S-1) \mathbf{P}\{\xi'_1(\rho) > \delta\} \leq c_{14}\gamma^{3/2}(1-p_r)^{-1}, \quad (39)$$

$$\mathbf{P}\{\zeta'_{S-1}(\delta) > \delta\} \leq c_{15}\gamma^{3/2}(1-p_r)^{-1}. \quad (40)$$

Также из (18) и (21) следует равенство:

$$\left(\frac{1 - P\{\xi'_1 > \delta\}}{1 - P\{\xi'_1 > \rho\}} \right)^{S-1} = 1 + o(1).$$

Поэтому из (38)–(40) получим оценку (37).

Учитывая (34), представим вероятность P_2 в виде суммы

$$P_2 = R_1 + R_2 + R_3 + R_4, \quad (41)$$

где

$$\begin{aligned} R_i &= (1 + o(1)) \sum_{K_i} \mathbf{P}\{\xi'_S = n - m_r - k\} \\ &\times \mathbf{P}\{\zeta'_{S-1}(\rho) = k - m_r(S-1)\}, \end{aligned}$$

$$\begin{aligned} K_1 &= \{k : S-1 \leq k \leq -\delta + m_r(S-1)\}, \\ K_2 &= \{k : -\delta + m_r(S-1) < k \leq \delta + m_r(S-1)\}, \\ K_3 &= \{k : \delta + m_r(S-1) < k \leq (n - m_r) - \gamma^{-1/2}\rho\}, \\ K_4 &= \{k : (n - m_r) - \gamma^{-1/2}\rho < k \leq (n - m_r) - \rho\}, \end{aligned}$$

при этом, как нетрудно видеть, область K_1 может быть пустой. Заметим, что

$$\begin{aligned} \mathbf{P}\{\xi'_S = n - m_r - k\} &= (1 - p_r)^{-1} \\ &\times \mathbf{P}\{\xi_S = (n - m_r S) + m_r S - k, \xi_S \neq r\}. \end{aligned}$$

Из (17) при $k \in K_2$ нетрудно получить, что

$$\begin{aligned} \mathbf{P}\{\xi'_S = n - m_r - k\} \\ = (\alpha/a_0)(1 - p_r)^{-1}(n - m_r S)^{-5/2}(1 + o(1)). \end{aligned}$$

Отсюда с помощью (14) и (36) получаем:

$$R_2 = (\alpha/a_0)(1 - p_r)^{-1}(n - m_r S)^{-5/2}(1 + o(1)). \quad (42)$$

Из (14) и (17) ясно, что при $k \in K_1$

$$\mathbf{P}\{\xi'_S = n - m_r - k\} \leq c_{15}(n + \delta - m_r S)^{-5/2}.$$

Отсюда, с учетом соотношения $\gamma \rightarrow 0$, из (36) видим, что

$$R_1 = o((n - m_r S)^{-5/2}). \quad (43)$$

Используя (14), (17) и (37), нетрудно получить выражения:

$$R_3, R_4 = o((n - m_r S)^{-5/2}). \quad (44)$$

Из представления (41) и оценок (42)–(44), делаем вывод, что

$$P_2 = (\alpha/a_0)(1 - p_r)^{-1}(n - m_r S)^{-5/2}(1 + o(1)). \quad (45)$$

Оценим вероятность P_3 . Заметим, что

$$\begin{aligned} P_3 &= \frac{S(S-1)}{2} \cdot \sum_{k < (n-2m_r)-2\rho} \mathbf{P}\{\zeta'_{S-2} = k - m_r(S-2)\} \\ &\times \mathbf{P}\{\xi'_{S-1} + \xi'_S = n - 2m_r - k, \xi'_{S-1} > \rho, \xi'_S > \rho\}. \end{aligned}$$

Отсюда следует, что

$$\begin{aligned} P_3 &\leq c_{16}S^2 \cdot \sum_{k < (n-2m_r)-2\rho} \mathbf{P}\{\zeta'_{S-2} = k - m_r(S-2)\} \\ &\times \sum_G \mathbf{P}\{\xi'_{S-1} = i\} \mathbf{P}\{\xi'_S = n - k - i\}, \end{aligned} \quad (46)$$

где $G = \{i : \rho < i < n - k - \rho\}$. Из (17) следует, что при $i > \rho$

$$\mathbf{P}\{\xi'_{S-1} = i\} = \mathbf{P}\{\xi_{S-1}^{(r)} = i + m_r\} \leq c_{16}\rho^{-5/2}.$$

Используя (17) и (18), получаем, что

$$\begin{aligned} \sum_G \mathbf{P}\{\xi'_{S-1} = i\} \mathbf{P}\{\xi'_S = n - k - i\} \\ \leq c_{17}\rho^{-5/2}(1-p_r)^{-1} \mathbf{P}\{\xi'_S > \rho\} < c_{18}(1-p_r)^{-1}\rho^{-4}, \end{aligned}$$

и в силу (46),

$$P_3 \leq c_{18}(1 - p_r)^{-1}S^2\rho^{-4},$$

а отсюда:

$$P_3 = o(S(n - m_r S)^{-5/2}). \quad (47)$$

Утверждение леммы следует из соотношений (15), (31), (45) и (47). $\square$

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ

В работе [2] было показано, что для сумм ζ_N при условиях, что $n/N > L$ и $(n - LN)/(L_1 N)^{2/3} \rightarrow \infty$

$$\mathbf{P}\{\zeta_N = n\} = (\alpha/a_0)N(n - LN)^{-5/2}(1 + o(1)) \quad (48)$$

равномерно относительно n , а для натуральных h равномерно относительно $v = (h - LN)/(L_1 N)^{2/3}$ в любом конечном фиксированном интервале

$$(L_1 N)^{2/3} \mathbf{P}\left\{\frac{\zeta_N - LN}{(L_1 N)^{2/3}} = v\right\} = g(v)(1 + o(1)). \quad (49)$$

При выполнении условия 2 теоремы $p_r \rightarrow 0$, поэтому, как известно, для целых положительных k

$$\binom{N}{k} p_r^k (1 - p_r)^{N-k} = \frac{(N p_r)^k}{k!} e^{-N p_r} (1 + o(1)) \quad (50)$$

равномерно относительно $(k - N p_r)/\sqrt{N p_r}$ в любом конечном интервале. Рассмотрим случай, когда $(n - LN)/(L_1 N)^{2/3} \rightarrow \tau$, где τ — некоторая постоянная. Для оценки вероятности $\mathbf{P}\{\zeta_{N-k}^{(r)} = n - kr\}$ используем лемму 2, полагая $S = N - k$ и $h = n - kr$. Заметим, что

$$k = N p_r + u_r \sqrt{N p_r}, \quad N - k = N(1 - p_r)(1 + o(1)), \quad (51)$$

где $u_r \leq C < \infty$. Тогда, используя (8) и соотношение (49), получаем, что

$$\frac{\mathbf{P}\{\zeta_{N-k}^{(r)} = n - kr\}}{\mathbf{P}\{\zeta_N = n\}} \rightarrow 1. \quad (52)$$

В случае, когда $n/N > L$, $(n - LN)/(L_1 N)^{2/3} \rightarrow \infty$, также выполняется (52) в силу леммы 3, замечания (51) и соотношения (48).

При выполнении условия 1 теоремы, мы используем нормальное приближение для биномиального распределения, справедливое при

$N p_r(1 - p_r) \rightarrow \infty$. Тогда

$$\binom{N}{k} p_r^k (1 - p_r)^{N-k} = \frac{1 + o(1)}{\sqrt{2\pi N p_r(1 - p_r)}} \times \exp\left\{-\frac{(k - N p_r)^2}{2N p_r(1 - p_r)}\right\} \quad (53)$$

равномерно относительно

$$(k - N p_r)/\sqrt{N p_r(1 - p_r)}$$

в любом фиксированном конечном интервале. Нетрудно видеть, что и в этом случае, как и выше, из лемм 2 и 3 вытекает (52).

Утверждения теоремы следуют теперь из лемм 1, соотношений (50), (52) и (53).

Работа выполнена при поддержке Программы стратегического развития на 2012–2016 гг. «Университетский комплекс ПетрГУ в научно-образовательном пространстве Европейского Севера: стратегия инновационного развития».

ЛИТЕРАТУРА

1. Берникович Е. С. О числе деревьев заданного объема в случайном непомеченном некорневом лесе // Труды Карельского научного центра РАН. Сер. Математическое моделирование и информационные технологии. 2011. Вып. 2, № 5. С. 4–9. С. 148–157.
2. Берникович Е. С., Павлов Ю. Л. О максимальном объеме дерева случайного непомеченного некорневого леса // Дискретная математика. 2011. Т. 23, № 1. С. 3–20.
3. Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. М.: Наука, 1965. 424 с.
4. Колчин В. Ф. Случайные графы. М.: Физматлит, 2000. 256 с.
5. Харари Ф., Палмер Э. Перечисление графов. М.: Мир, 1977. 326 с.

СВЕДЕНИЯ ОБ АВТОРЕ:

Петрова Елена Сергеевна

аспирант
лаб. теории вероятностей и компьютерной статистики
Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: bern@krc.karelia.ru
тел.: (8142) 781218

Petrova, Elena

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: bern@krc.karelia.ru
tel.: (8142) 781218

УДК 004.738.5:51-7

ПРИМЕНЕНИЕ ВЕБОМЕТРИЧЕСКИХ МЕТОДОВ ДЛЯ ИССЛЕДОВАНИЯ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА НАУЧНОЙ ОРГАНИЗАЦИИ (НА ПРИМЕРЕ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН)

А. А. Печников

*Институт прикладных математических исследований
Карельского научного центра РАН*

На примере Карельского научного центра РАН показано применение вебометрических методов для исследования веб-сайтов, составляющих информационное веб-пространство крупной научной организации для выработки рекомендаций, направленных на улучшение его присутствия в Вебе.

Ключевые слова: вебометрика, веб-пространство, веб-граф, индикаторы присутствия в Вебе.

**A. A. Pechnikov. APPLICATION OF WEBOMETRIC
TECHNIQUES FOR INVESTIGATION OF THE WEB SPACE
OF SCIENTIFIC ORGANIZATION (A CASE STUDY OF
KARELIAN RESEARCH CENTRE OF RAS)**

The use of webometric techniques for the study of web sites in the information web space of a large scientific organization in order to produce recommendations on improving its presence on the Web is shown with the example of Karelian Research Center of RAS.

Key words: webometrics, web space, web graph, indicators of presence on the Web.

ВВЕДЕНИЕ

К актуальным направлениям вебометрики, – одного из направлений информатики, – относятся исследования гиперссылок в Вебе [4]. Вследствие гигантской размерности Веба во многих случаях исследования проводятся на его достаточно узких фрагментах, таких, например, как множество сайтов университетов Великобритании [5] или Северной Европы [2].

Взаимодействие веб-сайтов в информационном веб-пространстве осуществляется посредством гиперссылок, связывающих сайты веб-пространства и определяющих его внутреннюю структуру. Наличие (или отсутствие) гиперссылок влияет на их присутствие в Вебе: известно, что увеличение количества гиперссылок на сайт связано с ростом индекса цитирования веб-сайта [11] и увеличивает (потенциально) количество переходов с одного сайта на другой, а значит, и количество посетителей.

Веб-ресурсы научных учреждений относятся к так называемым «регламентируемым

веб-ресурсам». Это означает, что существует (или, выразимся мягче, — «должен существовать») официальный документ, в котором изложены цели и задачи веб-ресурса, определена структура сайта, правила изменения информации и т. д. Под управляемостью процессами в Вебе понимается реализация управленческих решений в виде пунктов регламента, определяющих порядок их исполнения. В этом случае для сайтов информационного веб-пространства организации могут быть выработаны управленческие решения, направленные на улучшение их присутствия в Вебе. Однако прежде чем реализовывать управленческие решения, необходимо пройти путь от сбора информации о присутствии и взаимосвязях веб-сайтов до построения и анализа формальных моделей и разработки рекомендаций, сделанных на их основе. Целью данной работы является изучение взаимодействия веб-сайтов, составляющих информационное веб-пространство достаточно крупной научной организации для выработки рекомендаций, направленных на улучшения его присутствия в Вебе.

В качестве объекта исследования выбрано веб-пространство Карельского научного центра РАН (КарНЦ РАН), в состав которого входят 39 веб-сайтов, включая официальные сайты 7 институтов различной научной направленности. Представляется, что такой объект исследования является достаточно типичным для российской академической науки, и поэтому методы, апробированные здесь, могут быть перенесены на другие научные центры и крупные институты.

В работе используются методы, разработанные для исследований академического [8] и университетского фрагментов российского Веба [9]. Для получения, хранения и обработки вебометрической информации используется специализированный комплекс программ VeeBot, в состав которого входят поисковый робот для сканирования веб-сайтов с целью сбора информации о них, база данных, предназначенная для хранения и обработки собранной информации применительно к задачам вебометрических исследований и средства визуализации [10]. Кроме того, используется ряд расширенных возможностей поиска в Google.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Веб-сайт — совокупность html-страниц и веб-документов, связанных внутренними гиперссылками и обладающих единством содер-

жания, идентифицируемый в Вебе по уникальному доменному имени.

Уровень страницы сайта определяется следующим образом: начальная страница сайта, определяемая по уникальному доменному имени, имеет уровень 0; уровень любой другой страницы сайта — это минимальное количество внутренних гиперссылок, ведущих от начальной страницы к данной. Под контекстом гиперссылки в общем случае понимаются языковые выражения, окружающие гиперссылку в пределах веб-страницы [3], в более узком смысле рассматривается текст, сопровождающий гиперссылку на странице (так называемый «анкор»).

Уникальной прямой внешней гиперссылкой называется гиперссылка из множества всех гиперссылок, имеющих одинаковый контекст и адрес целевой страницы, сделанная с исходной страницы с наивысшим уровнем; если на наивысшем уровне таких ссылок несколько, то выбирается первая из них. Далее рассматриваются именно такие гиперссылки, поэтому вместо словосочетания «уникальная прямая внешняя гиперссылка» будут использоваться термины «гиперссылка» или просто «ссылка».

Официальный веб-сайт организации — веб-ресурс, обеспечивающий официальное представление информации в Вебе об организации, являющейся юридическим лицом, создаваемый и функционирующий в соответствии с нормативным актом (регламентом) организации-владельца сайта, определяющим цели и задачи официального веб-сайта, структуру и порядок размещения в Вебе информационных материалов, образующих веб-ресурс, права, обязанности и лиц, осуществляющих программно-техническую поддержку данного сайта, и лиц, осуществляющих предоставление информации для размещения на сайте.

Веб-сайт подразделения — веб-ресурс подразделения, входящего в состав организации и не являющегося юридическим лицом, обеспечивающий представление информации в Вебе о данном подразделении.

Тематический веб-сайт — веб-ресурс организации, не являющийся официальным веб-сайтом или веб-сайтом подразделения, создаваемый и функционирующий для представления в Вебе информации на заданную тему.

Неофициальный веб-сайт — веб-ресурс организации, обеспечивающий представление информации в Вебе об организации, являющейся юридическим лицом, не являющийся официальным сайтом организации.

Информационное веб-пространство – это множество взаимосвязанных посредством гиперссылок веб-сайтов перечисленных типов.

ИНФОРМАЦИОННОЕ ВЕБ-ПРОСТРАНСТВО КАРНЦ РАН

Веб-пространство КарНЦ РАН состоит из 39 взаимосвязанных сайтов (разобьем их на 7 групп, далее в скобках количество сайтов в группе):

1. Официальный сайт КарНЦ РАН (1),
2. Официальные сайты институтов, входящих в состав КарНЦ РАН (7),

3. Сайты подразделений КарНЦ РАН (2),
4. Тематические сайты КарНЦ РАН (10),
5. Тематические сайты институтов КарНЦ РАН (15),
6. Неофициальные сайты институтов КарНЦ РАН (3),
7. Сайт Карельской региональной организации профсоюза РАН (1).

Перечень веб-сайтов, входящих в веб-пространство КарНЦ РАН, приводится в табл. 1.

Таблица 1. Сайты веб-пространства Карельского научного центра РАН

№	Название организации/подразделения/сайта	Гр.	Доменное имя сайта	Стр.	Исх.	R
1	Карельский научный центр РАН	1	www.krc.karelia.ru	36000	1962	0,44
2	Институт биологии КарНЦ РАН	2	ib.krc.karelia.ru	1809	513	0,41
3	Институт водных проблем Севера КарНЦ РАН	2	water.krc.karelia.ru	1708	496	0,35
4	Институт геологии КарНЦ РАН	2	ig.krc.karelia.ru	1369	551	0,48
5	Институт леса КарНЦ РАН	2	forestry.krc.karelia.ru	2500	494	0,47
6	Институт прикладных математических исследований КарНЦ РАН	2	mathem.krc.karelia.ru	5141	1435	0,15
7	Институт экономики КарНЦ РАН	2	economy.krc.karelia.ru	3187	641	0,40
8	Институт языка, литературы и истории КарНЦ РАН	2	illhportal.krc.karelia.ru	2193	562	0,79
9	Вычислительный кластер КарНЦ РАН	3	cluster.krc.karelia.ru	24	24	0,27
10	Научная библиотека КарНЦ РАН	3	library.krc.karelia.ru	49	106	0,38
11	Электронная библиотека КарНЦ РАН	4	dl.krc.karelia.ru	43124	247	0,27
12	Труды КарНЦ РАН	4	transactions.krc.karelia.ru	927	1419	0,30
13	Зеленый пояс Фенноскандии	4	green-belt.krc.karelia.ru	462	763	0,27
14	ГЭП-анализ на территории Республики Карелия	4	gap-analysis.krc.karelia.ru	565	794	0,72
15	Проект HINTERLAND	4	hinterland.krc.karelia.ru	417	1	0,59
16	Экологический туризм на службе Муезерского района	4	muetacis.krc.karelia.ru	179	3	0,60
17	Развитие национального парка Койтайоки-Толвоярви	4	suotacis.krc.karelia.ru	111	5	0,54
18	Ладога	4	ladoga.krc.karelia.ru	652	1	0,23
19	Биоразнообразие Карелии	4	biodiv.krc.karelia.ru	68	133	3,43
20	Природные ресурсы Карелии	4	ias.krc.karelia.ru	15	40	0,44
21	Вебометрика	5(6)	webometrics.krc.karelia.ru	617	324	0,27
22	III Всероссийская научная конференция «Электронные библиотеки»	5(6)	rcdl2001.krc.karelia.ru	381	20	1,86

23	XI Всероссийская научная конференция «Электронные библиотеки»	5(6)	rcdl2009.krc.karelia.ru	579	51	0,32
24	Базы данных по топливно-энергетическому хозяйству Карелии	5(6)	energy.krc.karelia.ru	150	1	0,27
25	Разработка методов моделирования телекоммуникационных сетей	5(6)	morozov.krc.karelia.ru	52	7	0,75
26	Документация по программированию	5(6)	info.krc.karelia.ru	786	230	0,43
27	Журнал «МТИ& П»	5(6)	mgta.krc.karelia.ru	123	415	0,15
28	Российское ботаническое общество. Карельское отделение	5(2)	rbo.krc.karelia.ru	233	435	0,60
29	Геоботанический сайт	5(2)	geobotany.krc.karelia.ru	397	978	0,24
30	Топонимия Европейского Севера - TORIS	5(8)	toris.krc.karelia.ru	144	27	0,27
31	Фольклорный архив Института языка, литературы и истории КарНЦ РАН	5(8)	folk.krc.karelia.ru	89	4	0,32
32	Фонограммархив Института языка, литературы и истории КарНЦ РАН	5(8)	phonogr.krc.karelia.ru	70	30	0,38
33	Вепсский корпус	5(8)	vepsian.ru	745	2	0,27
34	Петроглифы Фенноскандии	5(8)	petroglyphs.ru	206	4	0,73
35	Отделение Русского географического общества в Республике Карелия	5(3)	rgo.karelia.ru	164	7	0,53
36	Институт водных проблем Севера КарНЦ РАН (неофиц.)	6(3)	nwpi.krc.karelia.ru	180	25	0,66
37	Институт геологии КарНЦ РАН	6(4)	geoserv.krc.karelia.ru	2567	29	0,19
38	Институт биологии КарНЦ РАН	6(2)	biology.krc.karelia.ru	476	45	0,52
39	Карельская организация Профсоюза РАН	7	profcom.krc.karelia.ru	24	16	0,35

В столбце «гр.» обозначена принадлежность сайта к одной из 7 указанных групп, цифра в скобках соответствует порядковому номеру сайта. Например, у сайта «Труды КарНЦ РАН», являющегося тематическим сайтом КарНЦ РАН, значение «гр.» равно 4. У сайта «Вепсский корпус», – тематического сайта Института языка, литературы и истории КарНЦ РАН, – значение «гр.» равно 5(8), где 5 обозначает принадлежность к группе, а 8 – принадлежность к институту (т. е. порядковый номер института в общем списке сайтов).

Из табл. 1 видно, что подавляющее большинство сайтов имеют доменные имена третьего уровня в зоне основного домена krc.karelia.ru, и лишь три сайта не ассоциированы с этим доменом.

Сайты сканировались на максимальную глубину до 5-го уровня, что считается достаточным для нахождения основных значимых гиперссылок, требуемых для построения

теоретико-графовых моделей. Вследствие этого, значения в столбцах «стр.» (количество страниц сайта) и «исх.» (количество исходящих гиперссылок), являются оценками снизу.

Всего на сайтах КарНЦ РАН отсканировано около 90 тысяч страниц, найдено почти 12 500 исходящих гиперссылок. Всего два сайта можно назвать большими – сайты КарНЦ РАН и Электронной библиотеки. Семь официальных сайтов институтов (плюс один неофициальный сайт института) имеют размеры в пределах от 1300 до 5150 страниц. Причем неофициальный сайт Института геологии почти в два раза больше официального. Остальные сайты имеют менее 1000 страниц. Корреляция между количеством страниц на сайте и количеством исходящих гиперссылок не обнаруживается.

Важным индикатором присутствия сайтов в Вебе считается количество документов, размещенных на них в виде файлов с расшире-

ниями .doc, .pdf, .ppt и др. (так называемые Rich Files) [6]. По указанной ссылке можно посмотреть ранжирование 120 российских научных учреждений РАН. Заметим, что КарНЦ в этот список не входит, а из региональных научных центров 20-е место занимает Коми научный центр Уральского отделения РАН (КомиНЦ УрО РАН).

Результаты сканирования сайтов веб-пространства КарНЦ РАН показывают, что основное количество Rich Files размещено на сайте КарНЦ РАН (www.krc.karelia.ru) — более 3500 файлов .doc и .pdf. Кроме того, активным участником информационного веб-пространства КарНЦ РАН является закрытый ресурс с адресом resources.krc.karelia.ru, используемый как хранилище файлов. По 4480 ссылкам на него, сделанным с других сайтов КарНЦ РАН, можно установить, что на нем содержится более 540 файлов .doc и 230 файлов .pdf. Например, сайт Зеленый пояс Фенноскандии (green-belt.krc.karelia.ru) отправляет к 300 файлам .doc, размещенным на resources.krc.karelia.ru. Однако более половины сайтов КарНЦ РАН не имеют Rich Files, в том числе и на resources.krc.karelia.ru.

Сказанное подтверждается замерами, сделанными в Google: команда «site:krc.karelia.ru filetype:doc» (количество файлов типа .doc на домене krc.karelia.ru) выдает 570 результатов, а «site:krc.karelia.ru filetype:pdf» — 1370. Та же команда, примененная к домену КомиНЦ УрО РАН, дает следующие результаты: «site:komisc.ru filetype:doc» — 310 результатов, а «site:komisc.ru filetype:pdf» — 5070.

Теперь о гиперссылках. Из более чем 12000 найденных гиперссылок, исходящих с сайтов КарНЦ РАН, половина являются гиперссылками, связывающими эти сайты. Ещё 4480 ссылок, как уже было сказано выше, сделано на закрытый ресурс resources.krc.karelia.ru. Остальные 1600 ссылок сделаны на 930 сайтов так называемого «сопутствующего множества» [9]. На рис. 1 представлен веб-граф, построенный на 969 вершинах, соответствующих 39 сайтам информационного веб-пространства КарНЦ РАН и 930 сайтам сопутствующего множества. Здесь нарисованы только исходящие дуги с вершин, соответствующих сайтам целевого множества (дуга изображена, если существует хотя бы одна гиперссылка, связывающая два соответствующих сайта).

Если на рис. 1 оставить только дуги, связывающие сайты целевого множества, то получим веб-граф, приведенный на рис. 2.

В овал заключен официальный сайт КарНЦ РАН, в верхнем прямоугольнике на-

ходятся сайты подразделений и тематические сайты КарНЦ РАН, в прямоугольниках в нижней части объединены все сайты, принадлежащие данному институту (официальный, и, если есть, то тематические и неофициальный).

К сайтам информационного веб-пространства КарНЦ РАН был применен подход, предложенный еще в [1]:

$$R(s) = (1-d) + d(R(t_1)/C(t_1) + \dots + R(t_n)/C(t_n)),$$

где $R(s)$ — значимость сайта s ,

d — коэффициент затухания (который обычно берут равным 0,85),

n — количество сайтов, ссылающихся на сайт s ,

$R(t_i)$ — значимость i -го сайта, имеющего ссылки на сайт s ,

$C(t_i)$ — количество ссылок с i -го сайта.

Значения $C(t_i)$ несложно посчитать для каждого сайта по матрице смежности веб-графа. Фрагмент матрицы смежности, показывающий количество гиперссылок, связывающих некоторые сайты, приведен на рис. 3. Как видно из него, веб-сайты имеют очень большое количество гиперссылок на «головной» сайт и обратно, и весьма незначительное между собой.

Содержательно значение $R(s)$ можно трактовать как значимость s -го сайта для других сайтов, входящих в заданное множество. Значения R для сайтов информационного веб-пространства КарНЦ РАН приведены в последней колонке табл. 1.

Достаточно предсказуемым результатом является высокая значимость официального сайта КарНЦ РАН как головной организации. Высокая позиция сайта Института прикладных математических исследований по сравнению с другими институтами также имеет свое объяснение (хотя и частичное) — многие сайты КарНЦ РАН имеют ссылки на него как на их разработчика. Официальные сайты институтов биологии, геологии и водных проблем Севера занимают последние места среди институтов, что может быть объяснено, в том числе, и наличием у них неофициальных сайтов, «отнимающим» ссылки у официальных.

Если в веб-графе, построенном на сайтах информационного веб-пространства КарНЦ, оставить только дуги, которым соответствует более двух гиперссылок, связывающих сайты, то получим веб-граф, изображенный на рис. 4 (изолированные вершины опущены). Максимальная компонента сильной связности этого веб-графа имеет диаметр, равный 3, и содержит 12 вершин, соответствующих сайту КарНЦ РАН, официальным сайтам всех

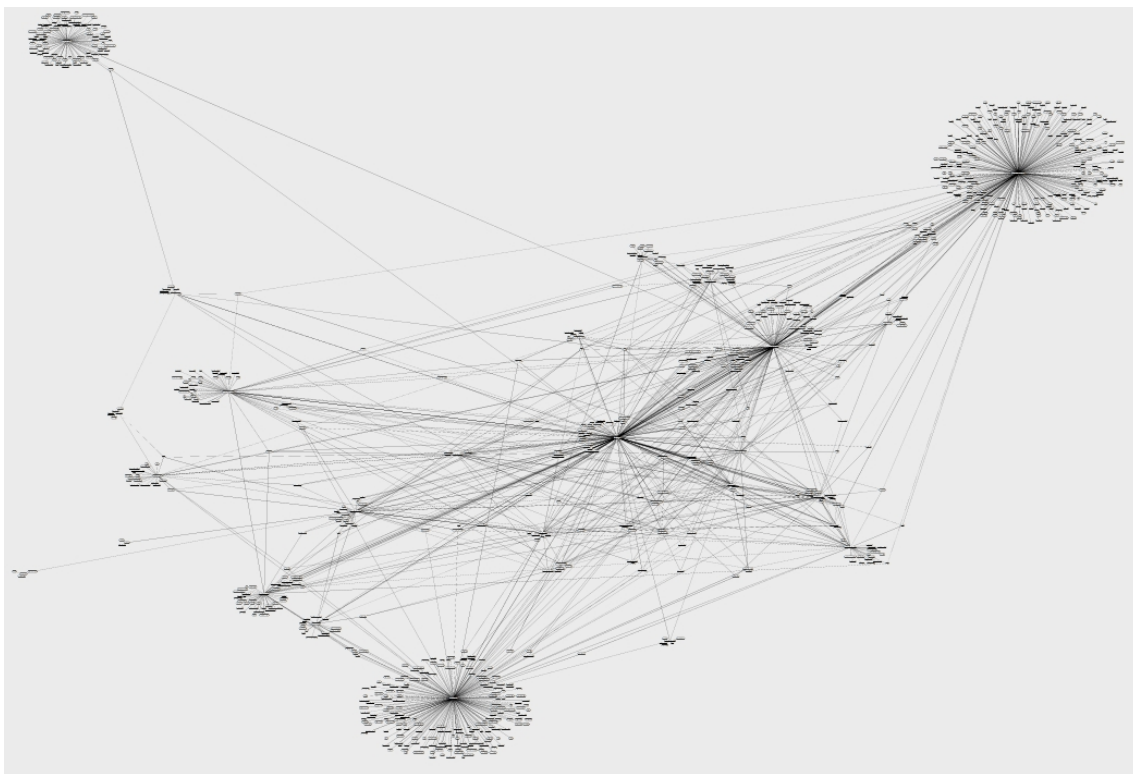


Рис. 1. Веб-граф, построенный на 969 вершинах

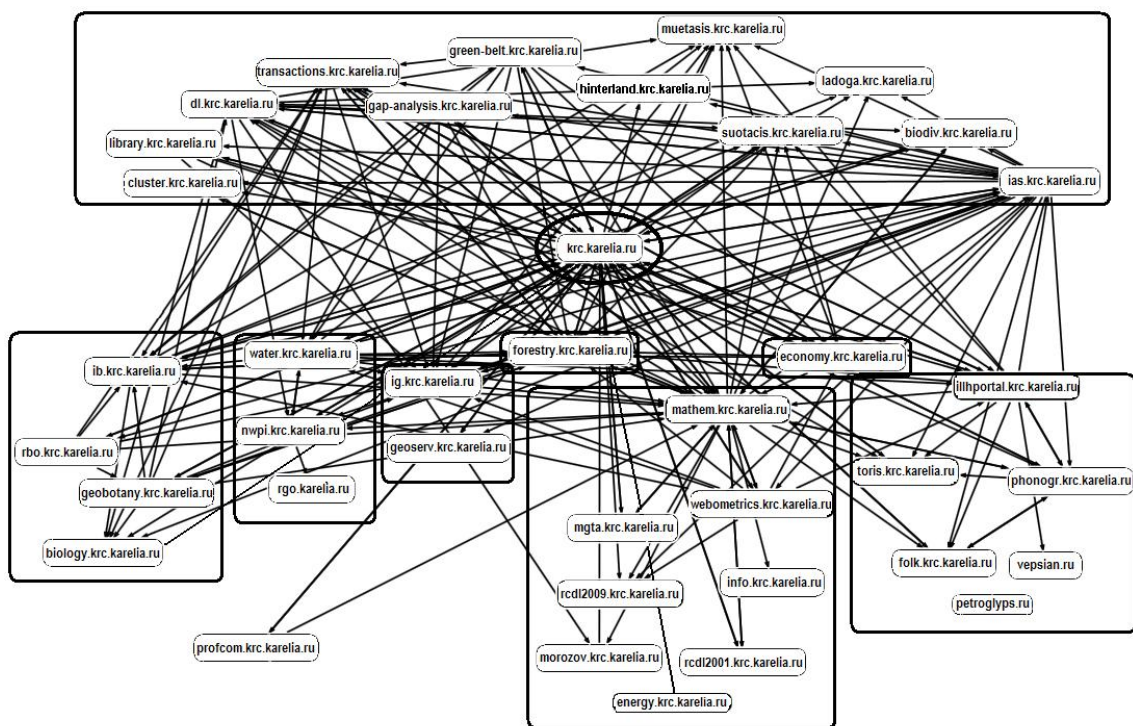


Рис. 2. Веб-граф, построенный на сайтах информационного веб-пространства КарНЦ (количество гиперссылок между сайтами больше 0)

	ias.krc.karelia.ru	ib.krc.karelia.ru	ig.krc.karelia.ru	illhportal.krc.karelia.ru	info.krc.karelia.ru	krc.karelia.ru
ias.krc.karelia.ru						4
ib.krc.karelia.ru	1					440
ig.krc.karelia.ru	1					517
illhportal.krc.karelia.ru	1					322
info.krc.karelia.ru						
krc.karelia.ru	1	19	12	19		

Рис. 3. Фрагмент матрицы смежности веб-графа КарНЦ РАН

семи институтов, и еще 4 сайтам — Электронная библиотека КарНЦ РАН, Зеленый пояс Фенноскандии, ГЭП-анализ на территории Республики Карелия и Труды КарНЦ РАН (на рис. 4 выделены овалами). Сайты, входящие в максимальную компоненту связности, можно назвать ядром информационного веб-пространства КарНЦ РАН: они образуют максимальную компоненту сильной связности, имея большое количество гиперссылок между собой (4834 из 6000).

ВЗАИМОСВЯЗИ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА КАРНЦ РАН С РОССИЙСКИМИ ВЕБ-САЙТАМИ

Сайты информационного веб-пространства КарНЦ содержат около 1150 гиперссылок, сделанных на 640 российских сайтов, относящихся

к сопутствующему множеству. Как оказалось, самое большое количество гиперссылок сделано на сайт фирмы, занимающейся компьютерными технологиями (правда, на него ссылаются лишь 2 сайта КарНЦ РАН). Вторым следует сайт РАН (32 ссылки с 15 сайтов КарНЦ РАН). На 490 сайтов сопутствующего множества сделано лишь по одной гиперссылке, из чего следует, что каждый из них интересен только одному сайту КарНЦ РАН.

Представляется существенным вопрос о том, как много гиперссылок сделано на сайты информационного веб-пространства КарНЦ с наиболее «популярных» у сайтов КарНЦ РАН российских сайтов сопутствующего множества. Для ответа на него были отобраны 15 сайтов, имеющих гиперссылки не менее чем с 4 сайтов КарНЦ РАН, перечисленные в табл. 2.

Таблица 2. 15 сайтов сопутствующего множества

№	Название организации/сайта	Доменное имя сайта	Сайты	Ссылки на сайт	Ссылки с сайта
1	РАН	www.ras.ru	15	32	8
2	LiveInternet.Ru	www.liveinternet.ru	13	13	0
3	Петрозаводский государственный университет	www.petrso.ru	8	26	4
4	РФФИ	www.rfbr.ru	8	14	1
5	Карельская государственная педагогическая академия	www.kspu.karelia.ru	7	13	1
6	Ботанический институт РАН	www.binran.ru	6	22	0
7	Учёные записки ПетрГУ	uchzap.petrso.ru	6	15	0
8	Портал органов власти Республики Карелия	gov.karelia.ru	5	6	0
9	WWW.KARELIA.RU	www.karelia.ru	5	6	0
10	Русское географическое общество	www.rgo.ru	5	5	0
11	РОМИП	www.romip.ru	4	15	5

12	Конференция «Электронные библиотеки»	www.rcdl.ru	4	10	4
13	Музей-заповедник «Кижы»	kizhi.karelia.ru	4	7	5
14	Некоммерческое партнерство «Север-Центр»	www.northcentre.ru	4	7	3
15	Теория управления организационными системами	www.mtas.ru	4	6	0

В колонке «сайты» указано количество сайтов КарНЦ РАН, имеющих ссылки на данный сайт, а в колонке «ссылки на сайт» – общее число ссылок с этих сайтов. В колонке «ссылки с сайта» указано количество ссылок, сделанных с данного сайта на сайты КарНЦ РАН.

Суммируя значения в двух последних колонках, получаем, что с сайтов КарНЦ РАН на 15 наиболее популярных сайтов сделано 197 ссылок, в то время как с этих сайтов на сайты КарНЦ РАН сделана всего 31 ссылка.

Можно сказать, что встречные ссылки между сайтом РАН и сайтами КарНЦ РАН во многом носят характер «начальник-подчиненный» и отражают реальное административное взаимодействие организаций РАН. Очевидное отражение в Вебе нашло взаимодействие КарНЦ РАН и Музея-заповедника «Кижы».

Однако вряд ли можно считать отражением реального взаимодействия наличие лишь одной ссылки с сайта Карельской государственной педагогической академии на сайты КарНЦ РАН. В меньшей степени, но это касается и сайта Петрозаводского государственного университета.

ВЗАИМОСВЯЗИ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА КАРНЦ РАН С ЗАРУБЕЖНЫМИ ВЕБ-САЙТАМИ

Сайты информационного веб-пространства КарНЦ содержат около 420 гиперссылок, сделанных на 290 зарубежных сайтов, относящихся к сопутствующему множеству (причем 130 гиперссылок сделано с сайта info.krc.karelia.ru). Наибольшее количество гиперссылок (а именно 6) сделано на сайт Electrical and Computer Engineering and Computer and Information Sciences networks (eecis.udel.edu) с одного сайта info.krc.karelia.ru. По три сайта КарНЦ РАН ссылаются на такие сайты, как Lake Model FLake (lakemodel.net) и Oulu University Library

(herkules.oulu.fi). Большинство остальных ссылок с сайтов КарНЦ РАН на зарубежные сайты можно охарактеризовать как «один сайт — одна-две ссылки». В целом можно говорить о невысокой активности взаимодействий с зарубежными сайтами.

Обратных ссылок с зарубежных сайтов, на которые ссылаются сайты КарНЦ РАН, по большому счету, не обнаружено. Исключениями являются сайт Lake Model FLake, имеющий 17 гиперссылок на официальный сайт Института водных проблем Севера и сайт Hinterland (hinterland-info.net), имеющий три ссылки на своего карельского коллегу Проект HINTERLAND.

Таким образом, реальное взаимодействие КарНЦ РАН с зарубежными организациями и коллегами отражения в Вебе практически не находит.

Выводы и рекомендации

Информационное веб-пространство КарНЦ РАН можно охарактеризовать как внутренне связанное, имеющее четко выраженное ядро, включающее наиболее значимые сайты. Однако высокая связность информационного пространства во многом обеспечивается одной-двумя гиперссылками, связывающими многие пары сайтов, что является отражением в Вебе скорее подчиненности, чем совместной деятельности. Обратим также внимание на то, что 9 сайтов КарНЦ вообще не имеют исходящих гиперссылок на другие сайты веб-пространства, а один сайт и вовсе изолированный.

Объемы сайтов КарНЦ РАН достаточно скромны, однако в целом соответствуют тенденциям, присущим институтам РАН. Например, сайт Института русской литературы РАН (www.pushkinskiydom.ru) содержит более 39000 страниц, а сайт Зоологического института РАН (www.zin.ru) — 1100. Более важной характеристикой присутствия в Вебе является

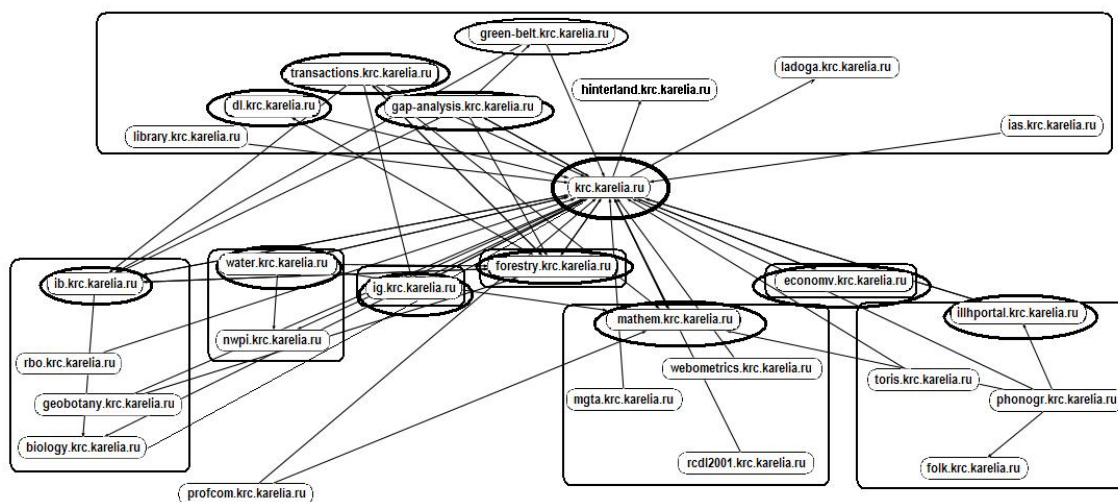


Рис. 4. Веб-граф, построенный на сайтах информационного веб-пространства КарНЦ (количество гиперссылок между сайтами больше 2)

ся количество выставленных на сайтах документов (так называемых Rich Files), а многие сайты КарНЦ РАН не имеют их вообще.

Взаимосвязи информационного веб-пространства КарНЦ РАН с российскими и зарубежными сайтами выглядят односторонне: большому количеству исходящих гиперссылок с сайтов КарНЦ РАН не соответствует большое количество ссылок, сделанных на эти сайты.

Возможно, это связано со слабой информированностью о наличии тех или иных веб-ресурсов КарНЦ РАН. Одним из характерных примеров слабой информированности коллег является ссылка с сайта Русского географического общества (www.rgo.ru) на веб-сайт Отделения РГО в Республике Карелия (kareliya.rgo.ru) и при этом отсутствие ссылки на другой сайт Отделения РГУ в Республике Карелия (rgo.karelia.ru), входящий в веб-пространство КарНЦ РАН. Сайт kareliya.rgo.ru является «дочерним» сайтом www.rgo.ru, поэтому о нем знают в Москве и делают на него ссылку, а сайт rgo.karelia.ru, по-видимому, неизвестен в Русском географическом обществе. Таких примеров недостаточной информированности при наличии реальной совместной деятельности можно привести много.

Из сделанных общих выводов можно сделать некоторые краткие рекомендации по усовершенствованию информационного веб-пространства КарНЦ РАН.

Необходимо обратить внимание на внутренние ссылки между сайтами веб-пространства КарНЦ РАН. Наверное, да-

леко не исчерпаны возможности веб-сайтов hinterland.krc.karelia.ru, ladoga.krc.karelia.ru и muetacis.krc.karelia.ru, которые могли сослаться хотя бы на официальный сайт КарНЦ РАН. Сайты petroglyphs.ru и vepsian.ru не имеют ссылок на сайт Института языка и литературы, biology.krc.karelia.ru – на сайт Института биологии, rgo.karelia.ru – на сайт Института водных проблем Севера, а rcdl2009.krc.karelia.ru и info.krc.karelia.ru – на сайт Института прикладных математических исследований.

Улучшению положения КарНЦ РАН в мировом рейтинге [6] (а точнее — его появлению в этом рейтинге), несомненно, должно помочь увеличение на сайтах веб-пространства КарНЦ РАН Rich Files. Во-первых, если у разработчиков есть такая возможность, то следует сделать открытым для поисковых машин хранилище файлов resources.krc.karelia.ru.

Во-вторых, на сайтах проектов вполне были бы уместны отчеты о проделанной работе в формате .doc или .pdf. Сюда же можно отнести и файлы различных презентаций, которые в большом числе делаются научными сотрудниками КарНЦ РАН. Пример может продемонстрировать сайт КарНЦ РАН, когда в новостной ленте информирует о научных докладах, сделанных на очередном заседании Президиума КарНЦ РАН: вместо фотографий (или вместе с фотографиями) полезно поставить и презентации научных докладов в формате .ppt.

Более сложной задачей представляется получение новых гиперссылок с внешних по отношению к веб-пространству КарНЦ РАН сай-

тов. Ясно, что покупка гиперссылок на так называемых «линкофермах» [7] является неприемлемой.

Очевидна рекомендация для научных сотрудников КарНЦ РАН: при создании персональных страниц на профессиональных веб-ресурсах (типа MathNet, www.mathnet.ru) обязательно указывайте официальный сайт своего института или КарНЦ РАН. То же самое относится к регистрации на сайтах конференций.

Научные сотрудники КарНЦ РАН ведут большую преподавательскую деятельность в Петрозаводском государственном университете, Карельской государственной педагогической академии, других вузах России, и информация о них размещается на сайтах соответствующих кафедр. Однако ссылки на основное место работы, а тем более на сайт организации, в ряде случаев отсутствуют.

Исследования научного и университетского Веба России часто убеждают в том, что ссылки на коллег и партнеров не выставляются не по злому умыслу, а по причине отсутствия точной информации о соответствующих веб-ресурсах. Здесь можно долго развивать тему PR-технологий и социальных коммуникаций, но одна из самых простых рекомендаций заключается в том, что на визитной карточке научного сотрудника должен быть адрес официального сайта его института. И эти карточки надо активно раздавать коллегам из других научных организаций.

Работа выполнена при частичной поддержке Программы стратегического развития Петрозаводского государственного университета на 2012–2016 гг. и гранта РГНФ № 12-03-12001.

ЛИТЕРАТУРА

1. Brin S., Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine // Seventh International World-Wide Web Conference (WWW 1998). April 14–18, 1998. Brisbane,

Australia. [Электронный ресурс] – Режим доступа: <http://ilpubs.stanford.edu:8090/361>.

2. Ortega J. L., Aguillo I. F. Visualization of the Nordic academic web: Link analysis using social network tools // Information Processing & Management. July 2008. Vol. 44, Iss. 4. P. 1624–1633.

3. Pant G., Srinivasan P., Menczer F. Crawling the Web // In Web Dynamics / M. Levene and A. Poulouvassilis, eds. Springer, 2004. P. 153–178.

4. Park H. W., Thelwall M. Hyperlink Analyses of the World Wide Web: A Review // Journal of Computer-Mediated Communication. 2003. N 8(4) [Электронный ресурс] – Режим доступа: <http://jcmc.indiana.edu/vol8/issue4/park.html?ref=ARKADASBUL.NET>.

5. Payne N., Thelwall V. A Statistical Analysis of UK Academic Web Links // Cybermetrics. International Journal of Scientometrics, Informetrics and Bibliometrics. 2004. Vol. 8, Iss. 1 [Электронный ресурс] – Режим доступа: <http://www.cindoc.csic.es/cybermetrics/articles/v8i1p2.html>.

6. Ranking Web of World Research Centers: CNR Institutes. RAS Institutes. <http://research.webometrics.info/RAS.asp> (дата обращения 18.04.2012).

7. Линкоферма — Википедия. <http://ru.wikipedia.org/wiki/Линкоферма> (дата обращения 18.05.2012).

8. Печников А. А., Луговая Н. Б., Чуйко Ю. В., Косинец И. Э. Разработка инструментов для вебметрических исследований гиперссылок научных сайтов // Вычислительные технологии. 2009. Т. 14, № 5. С. 66–78.

9. Печников А. А. Модель университетского Веба // Вестник Нижегородского университета им. Н. И. Лобачевского. 2010. № 6. С. 208–214.

10. Печников А. А., Чернобровкин Д. И. Адаптивный краулер для поиска и сбора внешних гиперссылок // Управление большими системами. Вып. 36. М.: ИПУ РАН, 2012. С. 301–315.

11. Яндекс. Как измеряется тИЦ. <http://help.yandex.ru/catalogue/?id=1111361> (дата обращения 14.05.2012).

СВЕДЕНИЯ ОБ АВТОРЕ:

Печников Андрей Анатольевич

вед. н. с. лаб. ТКС, к. ф.-м. н., д. т. н., доцент
Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: pechnikov@krc.karelia.ru
тел.: (8142) 763370

Pechnikov, Andrey

Institute of Applied Mathematical Research, Karelian
Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
Russia
e-mail: pechnikov@krc.karelia.ru
tel.: (8142) 763370

УДК 519.248.23:541.64:539.199

СУЩЕСТВЕННАЯ ВЫБОРКА ПРИ МОДЕЛИРОВАНИИ НЕПРЕРЫВНОГО СПЕКТРА КОНФОРМАЦИЙ МАКРОМОЛЕКУЛ МЕТОДОМ МОНТЕ-КАРЛО

А. Л. Рабинович¹, Д. В. Журкин²

¹ Институт биологии Карельского научного центра РАН,

² Петрозаводский государственный университет

Описан алгоритм, разработанный для моделирования молекул цепного строения методом Монте-Карло в предположении о непрерывном спектре конформаций, с использованием процедуры существенной выборки. Алгоритм учитывает взаимозависимость каждых трех углов внутреннего вращения вдоль по цепи. Он использован для изучения равновесных свойств углеводородных цепей различного строения.

Ключевые слова: метод Монте-Карло, существенная выборка, цепные молекулы, непрерывный спектр конформаций.

A. L. Rabinovich, D. V. Zhurkin. AN IMPORTANCE SAMPLING FOR MONTE CARLO SIMULATION OF A CONTINUOUS SPECTRUM OF CONFORMATIONS OF MACROMOLECULES

An importance sampling technique for Monte Carlo simulation of chain molecules assuming a continuous spectrum of their conformations is described. The algorithm takes into account the interdependence of each three torsion angles along the chain. It is used to study the equilibrium properties of hydrocarbon chains of different structure.

Key words: Monte Carlo simulation, importance sampling, chain molecules, continuous spectrum of conformations.

ВВЕДЕНИЕ

Для изучения свойств молекулярных систем широко используются методы компьютерного моделирования [10]. Наиболее распространенными являются методы Монте-Карло (МК) и молекулярной динамики (МД). Первый из них позволяет получить средние значения макроскопических наблюдаемых величин по статистическому ансамблю, а второй – средние по времени. Однако используемые в методе МК выборки конечны, а доступные

в методе МД времена моделирования сравнительно невелики, поэтому важно проводить сравнение между собой данных, полученных в рамках различных подходов. Актуальной задачей является также совершенствование методов компьютерного моделирования. В настоящей работе описана специальная вычислительная схема метода МК, разработанная для генерирования конформаций цепных молекул, в предположении о непрерывном

изменении углов внутреннего вращения в полном диапазоне от 0 до 2π . Учтена взаимозависимость каждых трех углов внутреннего вращения вдоль по цепи. При генерировании использована существенная выборка.

СРЕДНИЕ ЗНАЧЕНИЯ

Среднее значение $\langle H \rangle$ любой макроскопической наблюдаемой величины $H(x_1, x_2, \dots, x_n)$ системы частиц (атомов данной молекулы) в каноническом ансамбле можно вычислить следующим образом [1, 11]:

$$\langle H \rangle = \frac{\int \int \dots \int_{\Omega} H(x_1, x_2, \dots, x_n) \cdot \exp[-U(x_1, x_2, \dots, x_n)/k_B T] dx_1 dx_2 \dots dx_n}{\int \int \dots \int_{\Omega} \exp[-U(x_1, x_2, \dots, x_n)/k_B T] dx_1 dx_2 \dots dx_n}. \quad (1)$$

Здесь $(x_1, x_2, \dots, x_n)$ – точка в фазовом пространстве Ω этой системы, k_B – постоянная Больцмана, T – температура, $U(x_1, x_2, \dots, x_n)$ – гамильтониан системы, в котором опущены члены кинетической энергии, поскольку декартовы координаты и импульсы частиц входят в гамильтониан (в потенциальную и кинетическую энергию) по отдельности; переменные разделяются, интеграл по всем импульсам берется тривиальным образом и его можно сократить в числителе и знаменателе. В итоге под фазовым пространством Ω подразумевается конфигурационная его часть, а под $U(x_1, x_2, \dots, x_n)$ – потенциальная энергия, n – размерность пространства Ω . Если количество атомов в системе n_a , то n равно количеству декартовых координат ($3n_a$). Для молекул цепного строения, макромолекул конфигурационное пространство многомерно и функция $U(x_1, x_2, \dots, x_n)$, как правило, очень сложна и в аналитическом виде не представлена. Поэтому провести интегрирование в выражении (1) и выразить $\langle H \rangle$ в конечной аналитической форме в большинстве случаев не удастся. Но можно использовать один из численных методов решения, – например, метод МК; литература о нем и его приложениях весьма обширна (см., напр., [1–3, 5–7, 10–12, 14, 15, 19, 25, 26]).

Метод Монте-Карло

В методе МК задачу сводят к моделированию случайных величин и статистической оценке их характеристик. В простейшей формулировке это означает следующее. Для того чтобы приближенно вычислить некоторую скалярную величину a , предлагается придумать такую случайную величину ξ , у которой существует математическое ожидание $M\xi$, равное a . Тогда для оценки величины a выбирается K независимых реализаций $\xi_1, \xi_2, \dots, \xi_K$ случайной величины ξ и вычисляется среднее арифметическое $\bar{\xi}_K = (1/K) \cdot \sum_{i=1}^K \xi_i$. Последовательность одинаково распределенных независимых случайных величин, у которых существуют математические ожидания, под-

чиняется закону больших чисел (см., напр., [4, 6, 12], теорема Хинчина), поэтому среднее арифметическое этих величин сходится по вероятности к математическому ожиданию, т. е. при $K \rightarrow \infty$ имеет место $\bar{\xi}_K \xrightarrow{P} a$. При больших K величина $\bar{\xi}_K \approx a$, и вместо a можно использовать [6, 12] оценку $\bar{\xi}_K = (1/K) \cdot \sum_{i=1}^K \xi_i$. По определению, последовательность случайных величин $\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_K, \dots$ сходится по вероятности [12] к постоянной a , если для любого $h > 0$ вероятность $P\{|\bar{\xi}_K - a| \geq h\} \rightarrow 0$ при $K \rightarrow \infty$ (что равносильно $P\{|\bar{\xi}_K - a| < h\} \rightarrow 1$ при $K \rightarrow \infty$). Здесь $P\{|\bar{\xi}_K - a| \geq h\}$ есть вероятность того, что $\bar{\xi}_K$ удалена от a не менее чем на h . Смысл сходимости по вероятности в том, что вероятность значительного отклонения величины $\bar{\xi}_K$ от величины a при большом K небольшая; тем не менее, сильно отклоняющиеся значения $\bar{\xi}_K$ могут находиться сколь угодно далеко по K , но этих отклонений не очень много, вследствие чего вероятность их появления становится все меньше с ростом K .

Пусть $D\xi$ – (конечная) дисперсия случайной величины ξ . Можно использовать понятие *вероятной ошибки* [12] $r_K = 0,6745 \cdot (D\xi/K)^{1/2}$, для которой $P\{|\bar{\xi}_K - a| < r_K\} \approx P\{|\bar{\xi}_K - a| > r_K\} \approx 1/2$, т. е. одинаково вероятны ошибки, большие чем r_K , и ошибки, меньшие чем r_K , и охарактеризовать *порядок* ошибки метода. Вероятная ошибка $r_K \sim K^{-1/2}$. Критерий вероятной ошибки r_K полезен для сравнения степени "эффективности" [12] (или выгодности, удачности) разных алгоритмов, которые использованы (или можно использовать) для решения одной и той же задачи. Действительно, из выражения для r_K следует, что количества K осредняемых значений случайной величины $K = D\xi \cdot (0,6745/r_K)^2$. Пусть есть два алгоритма. Если задать вероятные ошибки r_K для обоих одинаковыми, то $K' = D\xi' \cdot (0,6745/r_K)^2$, $K'' = D\xi'' \cdot (0,6745/r_K)^2$. Пусть t' и t'' – время, затрачиваемое на расчет одного значения случайных величин ξ' и ξ'' , соответственно. Общее время расчета равно $t' \cdot K' = t' \cdot D\xi' \cdot (0,6745/r_K)^2$

и $t'' \cdot K'' = t'' \cdot D\xi'' \cdot (0,6745/r_K)^2$. Чем меньше время расчета ($t' \cdot K'$ или $t'' \cdot K''$), которое необходимо для достижения заданной вероятной ошибки, при прочих равных условиях, тем более эффективным является алгоритм. А если ввести понятие "трудоемкости" [12] алгоритма, которое используется для сравнения алгоритмов и равно произведению дисперсии осредняемой случайной величины $D\xi$ на время t расчета одного значения ξ (т. е. $t' \cdot D\xi'$ или $t'' \cdot D\xi''$), то можно сказать, что из двух алгоритмов более эффективным является менее трудоемкий.

Случайных величин ξ , – таких, что $M\xi = a$, существует бесконечно много, и теория методов МК отвечает на вопросы о том [12], каким образом выбрать удобную величину ξ для решения данной задачи и как находить значения $\xi_1, \xi_2, \dots$ произвольной случайной величины ξ .

Вычисление интегралов

Представляется целесообразным кратко привести сведения о двух известных схемах применения метода МК для вычисления интегралов; они важны для последующего описания разработанного алгоритма.

Схема 1 ([12], с. 93). Пусть требуется приближенно вычислить интеграл $J = \int_G f(x, y) \cdot p(x, y) dx dy$, где G – произвольная область плоскости x, y , а $p(x, y)$ – некоторая плотность вероятностей: $\int_G p(x, y) dx dy = 1$. Для оценки J методом МК вводится случайная точка Q в области G с плотностью вероятностей $p(x, y)$ и скалярная случайная величина $F = f(x, y)$, – такая, что ее математическое ожидание $MF = J$, т. е. равно искомому значению интеграла. Тогда, если $Q_1, \dots, Q_K$ – независимые реализации случайной точки Q , а $F_1 = f(Q_1), \dots, F_K = f(Q_K)$, то согласно приведенной выше схеме оценкой J будет величина [12] $\bar{F}_K = (1/K) \cdot \sum_{i=1}^K F_i$. При этом $\bar{F}_K \xrightarrow{P} J$.

Важно отметить [12], что если G – ограниченная область, то по этой же схеме можно получить оценку и интеграла вида $\int_G f(x, y) dx dy$. Действительно, пусть площадь области G есть S_G . Тогда плотность вероятностей случайной точки, равномерно распределенной в G , есть $1/S_G$. Если ввести $f_1(x, y) = S_G \cdot f(x, y)$ и обозначить $1/S_G \equiv p_1(x, y)$, то этот интеграл приводится к предыдущему виду: $\int_G f(x, y) dx dy = \int_G f_1(x, y) \cdot p_1(x, y) dx dy$.

Если искомую оценку получают усреднением по выборке значений функции с использованием плотности вероятностей случайной точки, распределенной в данной области равномерно (как в последнем примере), то такую

выборку в литературе по применению методов МК в физике (статистической, теоретической) принято называть "simple" sampling [15, 25], т. е. "простая" выборка [2] (иногда используется термин "прямая" выборка [14]).

Схема 2 ([12], с. 108). Пусть $J^0 = \int_G f(x, y) dx dy$, где G – ограниченная или неограниченная область, и $\int_G |f(x, y)| dx dy > 0$. В области G вводится плотность вероятности $p(x, y)$. Если $p(x, y) > 0$ в тех точках, в которых $f(x, y) \neq 0$, то $p(x, y)$ называется допустимой по отношению к $f(x, y)$; допустимая плотность $p(x, y)$ может обращаться в нуль только в тех точках, где $f(x, y) = 0$. Если же $p(x, y) > 0$ всюду в G , то она допустима по отношению к любым $f(x, y)$. Пусть G_0 – множество точек, в которых $f(x, y) = 0$, и пусть $G_+ = G - G_0$. Вводится функция $F^0(x, y)$, следующим образом: $F^0(x, y) = f(x, y)/p(x, y)$, если точка (x, y) принадлежит области G_+ , и $F^0(x, y) = 0$, если точка (x, y) принадлежит области G_0 . Здесь $p(x, y)$ – допустимая плотность. Для оценки J^0 методом МК, как и в *схеме 1*, вводится Q – случайная точка, определенная в области G с плотностью вероятностей $p(x, y)$. Математическое ожидание величины $F^0(Q)$ есть $MF^0(Q) = \int_G F^0(x, y) \cdot p(x, y) dx dy = \int_{G_+} f(x, y) dx dy = J^0$ ввиду того, что функция $f(x, y) = 0$ в точках (x, y) , не принадлежащих области G_+ . Итак, $MF^0(Q)$ равно искомому значению интеграла. Дальнейшие действия соответствуют *схеме 1*: если $Q_1, \dots, Q_K$ – независимые реализации случайной точки Q , а $F_1^0 = f(Q_1)/p(Q_1), \dots, F_K^0 = f(Q_K)/p(Q_K)$, то оценкой J^0 будет величина $\bar{F}_K^0 = (1/K) \cdot \sum_{i=1}^K F_i^0$.

Очевидно, плотность $p(x, y)$ желательно выбрать так, чтобы минимизировать дисперсию $DF^0(Q)$. Доказано [12], что минимальная дисперсия $DF^0(Q)$ реализуется в случае, когда плотность $p(x, y)$ пропорциональна $|f(x, y)|$; при этом она равна $|f(x, y)| \cdot [\int_G |f(x, y)| dx dy]^{-1}$. Ясно, однако, что несмотря на привлекательность достижения минимальной дисперсии, строго следовать этой рекомендации нельзя, поскольку задача вычисления $\int_G |f(x, y)| dx dy$ по трудности эквивалентна исходной задаче вычисления искомого интеграла. Но можно сделать и другой вывод: для получения величин F^0 с небольшими дисперсиями *желательно* выбирать плотность $p(x, y)$ *по возможности* пропорциональную $|f(x, y)|$ [12], – тогда в тех частях области G , в которых $|f(x, y)|$ больше и вклад которых в J^0 более существен, будет выбираться больше случайных точек [3, 12]. Иными словами,

чем меньше меняется функция $f(x, y)/p(x, y)$ (чем ближе она к постоянной), тем меньше окажется дисперсия $DF^0(Q)$. В этом состоит основная ценность данного приема; он предложен в работах [22, 23] и широко применяется [1, 2, 7, 12, 14]. Такая выборка называется "importance" sampling [15, 25] или "существенная" выборка [3, 7, 12], выборка "по значимости" [2, 5], выборка "по важности" [1, 6, 7], "предпочтительная" выборка [14].

Общий подход к оценке средних

Обе схемы, описанные выше для двух переменных, обобщаются на случай многих переменных. Возвращаясь к исходному выражению (1) для среднего значения $\langle H \rangle$, перейдем от точек (x, y) плоскости к n -мерным точкам $(x_1, x_2, \dots, x_n)$ конфигурационного пространства Ω . Интеграл в знаменателе выражения (1) есть так называемый

статистический интеграл Z (для дискретного случая – статистическая сумма [9]), $Z = \int \int \dots \int_{\Omega} \exp[-U(x_1, \dots, x_n)/k_B T] dx_1 \dots dx_n$, а величина $(1/Z) \cdot \exp[-U(x_1, \dots, x_n)/k_B T]$ есть плотность вероятности точки $(x_1, \dots, x_n)$. Для приближенного вычисления интегралов в выражении (1) можно применить как "простую", так и "существенную" выборку.

Реализуем "простую" выборку. n -мерная точка $(x_1, \dots, x_n)$ является аналогом рассмотренной выше 2-мерной точки Q . Независимыми реализациями Q_i случайной точки Q будут точки $(x_1^i, \dots, x_n^i)$, где $i = 1, 2, \dots, K$. Пусть они распределены равномерно в Ω . Скалярной случайной величиной F (из схемы 1) является выражение $(1/Z) \cdot H(x_1, \dots, x_n) \cdot \exp[-U(x_1, \dots, x_n)/k_B T]$, и его математическое ожидание, согласно (1), равно $\langle H \rangle$. Оценка $\bar{H}_K$ выражения (1) методом МК будет иметь вид [1, 2, 11]

$$\bar{H}_K = \frac{\sum_{i=1}^K H(x_1^i, x_2^i, \dots, x_n^i) \cdot \exp[-U(x_1^i, x_2^i, \dots, x_n^i)/k_B T]}{\sum_{i=1}^K \exp[-U(x_1^i, x_2^i, \dots, x_n^i)/k_B T]}. \quad (2)$$

При этом $\bar{H}_K \xrightarrow{P} \langle H \rangle$. Отметим, что выражение (2), в силу специфики исходного выражения (1), имеет вид формулы усреднения в каноническом ансамбле для случая, когда переменные имеют только дискретный спектр. Однако в данном случае суммирование производится не по дискретному спектру (всем состояниям), а по случайным точкам непрерывного спектра, и K является не количеством состояний, а размером выборки. В выражении (2) множитель $\exp[-U(x_1^i, \dots, x_n^i)/k_B T]$ может изменяться на многие порядки и, следовательно, более предпочтительной была бы "существенная" выборка.

Вариант "существенной" выборки, который можно было бы назвать наиболее эффективным, выгодным, т. е. наименее трудоемким (в том смысле, который описан выше), очевидно, был бы достигнут в случае использования функции $(1/Z) \cdot \exp[-U(x_1, x_2, \dots, x_n)/k_B T]$ в качестве плотности распределения случайной точки в области Ω . Пусть точки $(x_1^i, x_2^i, \dots, x_n^i)$ выбираются в Ω не равномерно, как в выражении (2), а именно с такой плотностью. Тогда оценка выражения

(1) сводится к среднему арифметическому $\bar{H}_K = (1/K) \cdot \sum_{i=1}^K H(x_1^i, x_2^i, \dots, x_n^i)$. Несмотря на столь простой вид итоговой оценки, реализовать эту идею невозможно, поскольку по трудности задача предварительного вычисления этой плотности сопоставима с задачей вычисления исходного выражения (1): речь идет о необходимости предварительного вычисления потенциальной энергии системы частиц $U(x_1, x_2, \dots, x_n)$. Тогда нужно подобрать такую плотность $p(x_1, x_2, \dots, x_n)$, которая была бы допустимой по отношению к подынтегральному выражению в (1) и, насколько возможно, пропорциональной ему, но при этом могла быть рассчитана предварительно, до проведения расчета методом МК. В итоге можно добиться дисперсии, значительно меньшей, чем в алгоритме простой выборки, и более эффективного (менее трудоемкого) алгоритма. Пусть некоторую плотность $p(x_1, x_2, \dots, x_n)$, удовлетворяющую этим условиям, найти удалось. Повторяя действия, описанные в схеме 2 выше, получим следующую оценку [1, 2, 11, 14]:

$$\bar{H}_K = \frac{\sum_{i=1}^K H(x_1^i, x_2^i, \dots, x_n^i) \cdot \exp[-U(x_1^i, x_2^i, \dots, x_n^i)/k_B T]/p(x_1^i, x_2^i, \dots, x_n^i)}{\sum_{i=1}^K \exp[-U(x_1^i, x_2^i, \dots, x_n^i)/k_B T]/p(x_1^i, x_2^i, \dots, x_n^i)}. \quad (3)$$

Различие между (3) и (2) обусловлено именно способом выбора точек в конфигурационном пространстве.

Если выбрать $p(x_1, x_2, \dots, x_n)$ в (3) пропорционально $\exp[-U(x_1, x_2, \dots, x_n)/k_B T]$, т. е. заранее вычислить какую-то часть энергии $U(x_1, x_2, \dots, x_n)$, то получим искомый результат. Действительно, энергию U можно представить, например, в виде суммы двух слагаемых (назовем их энергией ближних U_{short} и энергией дальних U_{long} взаимодействий), $U = U_{short} + U_{long}$. Тогда $\exp[-U(x_1, x_2, \dots, x_n)/k_B T]$ будет произведением $\exp[-U_{short}(x_1, x_2, \dots, x_n)/k_B T]$ и $\exp[-U_{long}(x_1, x_2, \dots, x_n)/k_B T]$. Выделение U_{short} из общей энергии U следует осуществить таким образом, чтобы была возможность вычислить ее на предварительном этапе,

до осуществления процедуры генерирования конформаций молекул методом МК. Детали реализации этой процедуры будут описаны ниже. Если U_{short} вычислена, то в качестве плотности $p(x_1, x_2, \dots, x_n)$ распределения случайных точек можно использовать $(1/Z_{short}) \cdot \exp[-U_{short}(x_1, x_2, \dots, x_n)/k_B T]$, где $Z_{short} = \int \int \dots \int_{\Omega} \exp[-U_{short}(x_1, \dots, x_n)/k_B T] dx_1 \dots dx_n$. Ясно, что это выражение является допустимой плотностью по отношению к подынтегральной функции: оно ≥ 0 во всей области конфигурационного пространства, а если обращается в нуль при $U_{short} \rightarrow \infty$, то в этих же точках обращается в нуль и экспонента, в которую входит полная потенциальная энергия системы U , поскольку тогда $U \rightarrow \infty$. В итоге вместо (3) получим следующую оценку

$$\bar{H}_K = \frac{\sum_{i=1}^K H(x_1^i, x_2^i, \dots, x_n^i) \cdot \exp[-U_{long}(x_1^i, x_2^i, \dots, x_n^i)/k_B T]}{\sum_{i=1}^K \exp[-U_{long}(x_1^i, x_2^i, \dots, x_n^i)/k_B T]}. \quad (4)$$

При компьютерной реализации эту плотность удастся воспроизвести лишь приблизительно, и итоговые формулы (см. ниже) отличаются от (4). Тем не менее, выборка оказывается в итоге "существенной".

Отметим, что при изучении свойств макромолекул идея вычисления или аппроксимации части энергии $U(x_1, x_2, \dots, x_n)$ в методе МК использовалась многократно, предложено много вариантов для разных молекулярных систем, и в каждом из них учтена специфика этих систем и характер задачи; ряд вариантов упомянут, например, в обзоре [11]. В настоящей работе эта идея использована для разработки алгоритма, который можно применить для исследования свойств молекул цепного строения определенных классов, и в нем учтена специфика этих цепей. В частности, ниже алгоритм расчета реализован для углеводородных цепей вида $CH_3 - (CH_2)_f - (CH = CH - CH_2)_d - (CH_2)_g - CH_3$, где f, d, g – целые.

ОПИСАНИЕ ЦЕПНЫХ МОЛЕКУЛ

Приведем основные сведения о принятом описании цепных молекул. Конформации мак-

ромолекулы могут быть охарактеризованы либо декартовыми координатами всех атомов, либо внутренними переменными: совокупностями валентных связей $\{l\}$, валентных углов $\{\theta\}$, углов внутреннего вращения $\{\varphi\}$ (т. е. торсионных углов). В рамках классической статистической механики существуют две схемы приближений [20], – "классическая гибкая модель" (согласно которой макромолекула – это система материальных точек, взаимодействие которых описывается некоторыми потенциалами) и "классическая жесткая модель" (согласно которой потенциалы действуют лишь между определенными участками макромолекулы, остальные же участки скреплены между собой геометрическими связями).

"Гибкая" модель при изучении равновесных свойств считается более предпочтительной [18, 20] (несмотря на ряд нерешенных вопросов [18, 20]), и она позволяет для среднего значения $\langle H \rangle$ вместо выражения (1) в декартовых координатах записать выражение, в котором все величины зависят от внутренних переменных:

$$\langle H \rangle = \frac{\int \int \dots \int H(\{\varphi\}, \{\theta\}, \{l\}) \cdot \exp[-U(\{\varphi\}, \{\theta\}, \{l\})/k_B T] d\{\varphi\} d\{\theta\} d\{l\}}{\int \int \dots \int \exp[-U(\{\varphi\}, \{\theta\}, \{l\})/k_B T] d\{\varphi\} d\{\theta\} d\{l\}}. \quad (5)$$

Здесь, как и в (1), H – некоторая макроскопическая наблюдаемая величина изучаемой системы частиц (молекулы), U – потенциальная энергия; $\{\varphi\}$, $\{\theta\}$ и $\{l\}$ – совокупности углов

внутреннего вращения, валентных углов и валентных связей, соответственно, которые образуют конфигурационное пространство.

Для получения оценки величины (5), аналогичной (3), методом МК можно использовать существенную выборку. Пусть случайные точки в конфигурационном пространстве есть $\{\varphi\}^\nu$, $\{\theta\}^\nu$ и $\{l\}^\nu$, а количество конформаций (состояний ν) молекулы в выборке есть

ω (для внутренних переменных обозначение ν использовано вместо i , а ω – вместо объема выборки K). Тогда, выбирая случайные точки с некоторой плотностью вероятностей $p(\{\varphi\}, \{\theta\}, \{l\})$, получим следующую оценку $\bar{H}_\omega$:

$$\bar{H}_\omega = \frac{\sum_{\nu=1}^{\omega} H(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu) \cdot \exp[-U(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu)/k_B T] / p(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu)}{\sum_{\nu=1}^{\omega} \exp[-U(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu)/k_B T] / p(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu)}. \quad (6)$$

Выражение (6) можно сделать более компактным, если убрать явное перечисление совокупностей случайных значений внутренних пере-

менных. Обозначая $H(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu) = H^\nu$; $U(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu) = U^\nu$; $p(\{\varphi\}^\nu, \{\theta\}^\nu, \{l\}^\nu) = p^\nu$, приведем (6) к виду

$$\bar{H}_\omega = \frac{\sum_{\nu=1}^{\omega} H^\nu \cdot \exp[-U^\nu/k_B T] / p^\nu}{\sum_{\nu=1}^{\omega} \exp[-U^\nu/k_B T] / p^\nu}. \quad (7)$$

Силовое поле

Рассмотрим вопрос об общем подходе к расчету энергии цепных молекул. Как правило, в расчетах используется метод атом-атомных потенциальных функций. В литературе существуют наборы таких функций и апробированные параметры (силовые поля) для цепных молекул разных типов. В частности, для молекул липидов и их компонентов – углеводородных цепей рассматриваемого типа предложено поле CHARMM27 [16, 17, 24, 28]; в настоящей работе использован модифицированный вариант этого поля [21]. Энергия U системы частиц вычисляется в виде суммы энергий валентных связей $U_l = \sum_{n=1}^{N_l} K_{l_n} \cdot (l_n - l_{0_n})^2$ и уг-

лов $U_\theta = \sum_{n=1}^{N_\theta} K_{\theta_n} \cdot (\theta_n - \theta_{0_n})^2$, энергии Юри-

Брэдли $U_u = \sum_{n=1}^{N_u} K_{u_n} \cdot (u_n - u_{0_n})^2$, энергии

неплоских отклонений атомов $U_\eta = \sum_{n=1}^{N_\eta} K_{\eta_n} \cdot$

$(\eta_n - \eta_{0_n})^2$, торсионной энергии $U_\varphi = \sum_{n=1}^{N_\varphi} K_{\varphi_n} \cdot$

$[1 + \cos(p_n \cdot \varphi_n - \delta_n)]$, энергии невалентных взаимодействий $U_{nb} = \sum_{i,j} \varepsilon_{ij} \cdot [(R_{min_{ij}}/r_{ij})^{12} -$

$2 \cdot (R_{min_{ij}}/r_{ij})^6]$ и электростатической энергии $U_{el} = \sum_{i,j} [(q_i \cdot q_j) / (\varepsilon_{el} \cdot r_{ij})]$.

Здесь N_l , N_θ , N_u , N_η , N_φ – соответственно, количество валентных связей; валентных углов; различных пар атомов, разделенных двумя валентными связями; углов неплоских отклонений в молекуле; торсионных углов. K_{l_n} , K_{θ_n} , K_{u_n} , K_{η_n} , K_{φ_n} – силовые постоянные для

расчета перечисленных компонентов энергии, соответственно. Среди параметров слагаемых энергии используются текущие и равновесные значения: длины n -й валентной связи (l_n и l_{0_n}), n -го валентного угла (θ_n и θ_{0_n}), n -го расстояния между атомами, разделенными двумя валентными связями (u_n и u_{0_n}), n -го угла неплоских отклонений (η_n и η_{0_n}). φ_n – это значение n -го торсионного угла, а p_n , δ_n – множитель и сдвиг фаз; r_{ij} – расстояние между атомами с номерами i и j , разделенными более чем двумя валентными связями; ε_{ij} и $R_{min_{ij}}$ – глубина и положение соответствующего минимума энергии взаимодействия; q_i , q_j – парциальные заряды на атомах, ε_{el} – диэлектрическая постоянная.

Сравнение внутренних переменных

Переменные $\{\varphi\}$, $\{\theta\}$, $\{l\}$ не являются, в определенном смысле, равноправными между собой. Действительно, при данной температуре, в условиях, близких к состоянию равновесия молекулы, каждый угол внутреннего вращения из совокупности $\{\varphi\}$ может с высокой вероятностью изменяться во всем диапазоне $[0, 2\pi]$, тогда как изменения валентных связей $\{l\}$ и валентных углов $\{\theta\}$ происходят лишь в малых областях вокруг их равновесных значений (это так называемые "малые колебания"). Данное различие использовано при разработке настоящей вычислительной схемы метода МК: генерирование каждой конформации макромолекулы осуществляли в 2 этапа.

На 1-м этапе генерирования конформации ν все величины $\{l\}$ и величины $\{\theta\}$ (или их большинство) фиксировали при равновесных значениях, т. е. соответствующие компонен-

ты энергии обращали в нуль. Конформацию номер ν генерировали только по торсионным углам $\{\varphi\}$. Последние изменялись в диапазонах $[0, 2\pi]$, но если рассматриваемая молекула содержала двойные связи *cis* (вращения вокруг которых сопряжены с преодолением большого потенциального барьера), то, аналогично валентным углам, соответствующие углы вращения φ фиксировали в положении равновесия. При генерировании конформации ν по совокупности углов $\{\varphi\}$ был предложен вариант существенной выборки, детали которого будут описаны ниже. В отличие от углов внутреннего вращения, независимое варьирование в методе МК всех валентных углов осуществить невозможно, – например, углов, имеющих вершину на одном и том же атоме. Поэтому на 2-м этапе осуществляли переход в полученной конформации от внутренних переменных (валентных связей, валентных углов, торсионных углов) к декартовым координатам атомов и проводили варьирование всех координат, – независимо для каждого атома. При этом было учтено, что, поскольку валентные связи $\{l\}$ и валентные углы $\{\theta\}$ претерпевают лишь малые колебания вокруг положений равновесия, изменения декартовых координат атомов также

могут быть лишь малыми. Оценка величины диапазона варьирования декартовых координат проведена из условия, чтобы максимальные изменения значений энергии валентных углов U_θ и валентных связей U_l , отвечающих этому диапазону, в сумме не превышали изменений торсионной энергии U_φ цепи и энергии невалентных взаимодействий U_{nb} . Расчеты показали, что этому условию отвечает диапазон $\pm 0,02 \text{ \AA}$ вокруг равновесных положений всех атомов. Ввиду малости этого диапазона полагали, что в его пределах значения координат распределены равномерно; процедуру генерирования координат осуществляли с использованием генератора (датчика) псевдослучайных чисел [27], равномерно распределенных в интервале $[0,1)$.

Итак, валентные связи l и валентные углы θ сначала фиксировали при равновесных значениях. В выражении (5) можно явно указать поэтому только торсионные углы. Пусть N – количество атомов основной цепи молекулы, $\varphi_1, \varphi_2, \dots, \varphi_{N-1}$ – углы внутреннего вращения вокруг связей основной цепи, $U(\varphi_1, \varphi_2, \dots, \varphi_{N-1})$ – потенциальная (конформационная) энергия. Тогда вместо (5) будем иметь

$$\langle H \rangle = \frac{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi} H(\varphi_1, \varphi_2, \dots, \varphi_{N-1}) \cdot \exp[-U(\varphi_1, \varphi_2, \dots, \varphi_{N-1})/k_B T] d\varphi_1 d\varphi_2 \dots d\varphi_{N-1}}{\int_0^{2\pi} \int_0^{2\pi} \dots \int_0^{2\pi} \exp[-U(\varphi_1, \varphi_2, \dots, \varphi_{N-1})/k_B T] d\varphi_1 d\varphi_2 \dots d\varphi_{N-1}}. \quad (8)$$

Энергия $U(\varphi_1, \varphi_2, \dots, \varphi_{N-1})$ при рассмотрении системы макромолекула – растворитель является энергией системы, состоящей из макромолекулы и окружающих молекул растворителя, усредненной (при любом данном состоянии $\varphi_1, \varphi_2, \dots, \varphi_{N-1}$ макромолекулы) по всем поло-

жениям и ориентациям молекул растворителя [18].

При использовании существенной выборки, выбирая случайные точки с некоторой допустимой плотностью $p(\varphi_1, \varphi_2, \dots, \varphi_{N-1})$, получим следующую оценку $\bar{H}_\omega$ методом МК:

$$\bar{H}_\omega = \frac{\sum_{\nu=1}^{\omega} H(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu) \cdot \exp[-U(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)/k_B T] / p(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)}{\sum_{\nu=1}^{\omega} \exp[-U(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)/k_B T] / p(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)}. \quad (9)$$

РАСЧЕТ ЭНЕРГИИ ЦЕПНОЙ МОЛЕКУЛЫ

Опишем более подробно процедуру расчета энергии U , а затем выбора допустимой плотности $p(\varphi_1, \varphi_2, \dots, \varphi_{N-1})$. Как упоминалось в разделе «Общий подход к оценке средних», потенциальная энергия всей молекулы U была

представлена суммой энергий ближних U_{short} и дальних U_{long} взаимодействий, $U = U_{short} + U_{long}$. Выделение U_{short} проведено так, чтобы ее можно было вычислить на предварительном этапе, до осуществления процедуры генерирования конформаций методом МК.

Исследуемая цепная молекула была условно разделена на небольшие фрагменты, – такие, что каждый из них содержал 3 последовательных угла внутреннего вращения вокруг связей основной цепи, $\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2}$, варьирование которых приводило к изменению взаимных положений атомов. При конструировании цепи принято следующее правило: фрагменты соединяются не "встык", а с наложением друг на друга, – так, чтобы последовательные (соседние, смежные) фрагменты имели по два общих варьируемых торсионных угла из трех. Этим учтена непрерывная взаимозависимость каждых трех углов вдоль по цепи.

Энергия U_{short} является суммой энергий U_{m_γ} этих фрагментов цепи, а энергия взаимодействий более удаленных вдоль по цепи атомов входит в U_{long} . В итоге $U_{short} =$

$\sum_{\gamma=1}^{N-3} U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2})$. Здесь m_γ – индекс фрагмента: m – его идентификационный номер, который характеризует химическое строение фрагмента ($m = 1, 2, \dots, m_f$, где m_f – общее количество разных фрагментов, необходимых для конструирования избранных молекул по указанному правилу), а γ – его номер в последовательности фрагментов вдоль по цепи. Номера m для некоторых γ в индексах m_γ в сумме U_{short} могут быть одинаковыми. Если молекула – это полимерная цепь, то номера m будут циклически повторяться для одинаковых мономерных единиц.

Как упоминалось, алгоритм расчета реализован для углеводородных цепей вида $CH_3 - (CH_2)_f - (CH = CH - CH_2)_d - (CH_2)_g - CH_3$, где f, d, g – целые. Анализ показал, что любая из цепей такого вида может быть сконструирована должным выбором нескольких фрагментов из следующей совокупности (три связи, углы вращения вокруг которых варьируются, ниже выделены дефисом):

$CH_3 - CH_2 - CH_2 - CH_2CH_2$;
 $CH_3CH_2 - CH_2 - CH_2 - CH_2CH_2$;
 $CH_2CH_2 - CH_2 - CH_2 - CH_2CH_2$;
 $CH_2CH_2 - CH_2 - CH_2 - CH_2CH$;
 $CH_2CH_2 - CH_2 - CH_2 - CHCHCH_2$;
 $CH_2CH_2 - CH_2 - CHCH - CH_2CH_2$;
 $CH_3CH_2 - CH_2 - CH_2 - CH_2CH$;
 $CH_2CH_2 - CH_2 - CHCH - CH_2CH$;
 $CH_2CH_2 - CHCH - CH_2 - CHCHCH_2$;
 $CH_3 - CH_2 - CHCH - CH_2CH$;
 $CH_3CH_2 - CHCH - CH_2 - CHCHCH_2$;
 $CH_2CHCH - CH_2 - CHCH - CH_2CH$;
 $CH_2CHCH - CH_2 - CH_2 - CH_2CH_3$;
 $CHCH_2 - CH_2 - CH_2 - CH_3$;
 $CHCH_2 - CHCH - CH_2 - CH_2CH_3$;
 $CH_2CHCH - CH_2 - CH_2 - CH_3$.

Один из этих фрагментов представлен в виде схемы на рис. 1.

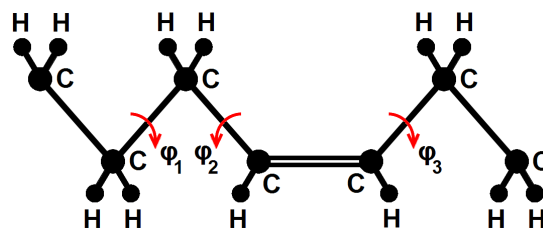


Рис. 1. Схема фрагмента $CH_2CH_2 - CH_2 - CHCH - CH_2CH_2$. Три варьируемых торсионных угла ($\varphi_1, \varphi_2, \varphi_3$) обозначены стрелками

Весовые множители

Энергию $U_m(\varphi_1, \varphi_2, \varphi_3)$ каждого из 16 представленных выше фрагментов вычисляли (с учетом только компонентов U_φ , U_{nb} и U_{el}) таким образом, чтобы сумма энергий любой совокупности фрагментов оказалась равной энергии ближних взаимодействий U_{short} той цепи, которая составлена из этих фрагментов. Для этого при вычислении каждой U_m вводили для трех перечисленных компонентов энергии весовые множители (1/3, 1/2 или 1). Множители были согласованы между смежными фрагментами и исключали возможность того, что одни и те же слагаемые окажутся дважды (или более раз) просуммированы в итоговой энергии U_{short} всей цепи. Избран принцип равноправного относительно смежных фрагментов учета каждого компонента энергии. А именно, слагаемые энергии, зависящие строго от одного торсионного угла, учитывали с множителем 1/3 в трех фрагментах (вдоль по цепи), в которых этот угол варьируется. В первом фрагменте из этих трех углов варьируется как φ_3 , во втором смежном фрагменте – как φ_2 , в третьем смежном фрагменте – как φ_1 . Слагаемые, зависящие строго от двух торсионных углов, учитывали с множителем 1/2 в каждом из двух смежных фрагментов, в которых эти углы варьируются (в первом фрагменте – как φ_2, φ_3 , а во втором смежном фрагменте – как φ_1, φ_2). Слагаемые, зависящие строго от трех торсионных углов, учитывали с множителем 1 только в том фрагменте, в котором эти углы варьируются. В энергии U_m каждого фрагмента из рассмотрения были исключены такие слагаемые, которые формально зависят от трех (или двух) углов вращения, но в данном фрагменте m один из них не варьируется. Эти слагаемые включали в энергию U_m того смежного фрагмента,

в котором варьируются все три (или, соответственно, два) торсионных угла.

В концевых фрагментах весовые множители некоторых компонентов энергии отличались от таковых для "срединных" фрагментов (для корректного расчета U_{short} всей цепи). Например, в концевых фрагментах все слагаемые энергии, которые зависели только от одного торсионного угла, учитывали полностью (с множителем 1, а не 1/3), если этот угол был первым (φ_1), поскольку ни в один из последующих фрагментов эти слагаемые больше не входят. Все слагаемые энергии, которые в концевых фрагментах зависели от второго угла (φ_2), учитывали не с множителем 1/3, а с 1/2, поскольку эти слагаемые входят не в три, а в два фрагмента: в концевой и смежный с ним фрагмент. Множители 1/3 для слагаемых энергии в концевых фрагментах были использованы только для третьего угла (φ_3). Табулирование значений энергии $U_m(\varphi_1, \varphi_2, \varphi_3)$ по трем углам вращения осуществлено с интервалом 1° .

СУЩЕСТВЕННАЯ ВЫБОРКА

Если энергия U_{short} исследуемой цепной молекулы рассчитана (по фрагментам), то алгоритм существенной выборки при генерировании конформаций этой молекулы методом МК будет эффективным, если организовать выборку совокупности случайных значений торсионных углов с плотностью вероятностей $p = (1/Z_{short}) \cdot \exp[-U_{short}(\varphi_1, \varphi_2, \dots, \varphi_{N-1})/k_B T]$.

$$\text{Здесь } Z_{short} = \int_0^{2\pi} \dots \int_0^{2\pi} \exp[-U_{short}(\varphi_1, \dots, \varphi_{N-1})/k_B T] \cdot d\varphi_1 \dots d\varphi_{N-1}.$$

$$\text{Поскольку } U_{short} = \sum_{\gamma=1}^{N-3} U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2}),$$

то числитель

$$\exp[-U_{short}(\varphi_1, \dots, \varphi_{N-1})/k_B T] / \\ = \prod_{\gamma=1}^{N-3} \exp[-U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2})/k_B T],$$

а

$$Z_{short} = \int_0^{2\pi} \dots \int_0^{2\pi} \prod_{\gamma=1}^{N-3} \exp[-U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2})/k_B T] \cdot d\varphi_1 \dots d\varphi_{N-1}.$$

Методика реализации этой идеи такова. Используя табулированные значения энергии U_m каждого фрагмента, при данной температуре

T рассчитывали, с интервалом 1° , функции $\exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T]$ и численными методами вычисляли интегралы

$$\int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T] d\varphi_1 d\varphi_2 d\varphi_3.$$

Пусть величина такого интеграла для фрагмента m равна I_m . Для каждого m области от 0 до 2π по каждому из трех углов ($\varphi_1, \varphi_2, \varphi_3$), т. е. "куб", разбивали на $100^3 = 1000000$ параллелепипедов (состояний) так, чтобы вероятности осуществления каждого из состояний были равны друг другу. Указанное разбиение осуществлено последовательно, численно, – с использованием метода прямоугольников [8]. Сначала были рассчитаны такие значения углов по оси φ_1 (положения плоскостей), которые разбивают трехмерное пространство ($\varphi_1, \varphi_2, \varphi_3$) на 100 "слоев", интегралы по которым от функции $\exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T]$ равны $I_m/100$. Каждый из слоев затем разделили по углу φ_2 на 100 равновероятных "подслоев" из условия, чтобы интегралы от функции $\exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T]$ по каждому подслою были равны $I_m/10000$, т. е. рассчитали значения углов по оси φ_2 , которые разделяют между собой подслои в каждом слое. Наконец, каждый подслои разделили по углу φ_3 на 100 состояний (параллелепипедов) так, чтобы интегралы от функции $\exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T]$ по каждому параллелепипеду были равны $I_m/1000000$ (рассчитали соответствующие граничные значения углов между параллелепипедами по оси φ_3 в каждом подслое). Очевидно, что при таком способе разбиения границы параллелепипедов в областях минимумов энергии каждого фрагмента m сгущаются, т. е. количество состояний – параллелепипедов в областях вблизи минимумов энергии оказывается больше, чем в областях вблизи максимумов. Массив чисел – границ между 1000000 параллелепипедов по всем трем углам $\varphi_1, \varphi_2, \varphi_3$ данного фрагмента используется далее для генерирования состояний с плотностью вероятностей, пропорциональной величине $\exp[-U_m(\varphi_1, \varphi_2, \varphi_3)/k_B T]$, которая входит в качестве сомножителя в плотность p .

Например, на рис. 2 схематично приведено несколько плоскостей, которые делят область изменения трех варьируемых углов вращения ("куб") некоторого фрагмента на слои по оси φ_1 , – это слои вблизи середины (180°) диапазона изменения угла φ_1 .

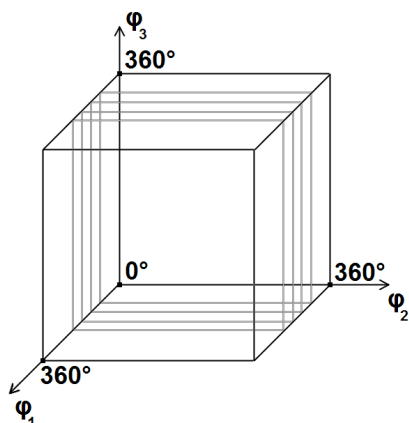


Рис. 2. Разбиение части области изменения трех варьируемых углов любого фрагмента на равновероятные слои по оси φ_1

Зависимость энергии фрагмента, изображенного на рис. 1, от углов φ_2 и φ_3 при $\varphi_1 = 180^\circ$, т. е. в одном из таких слоев, представлена в виде поверхности на рис. 3:

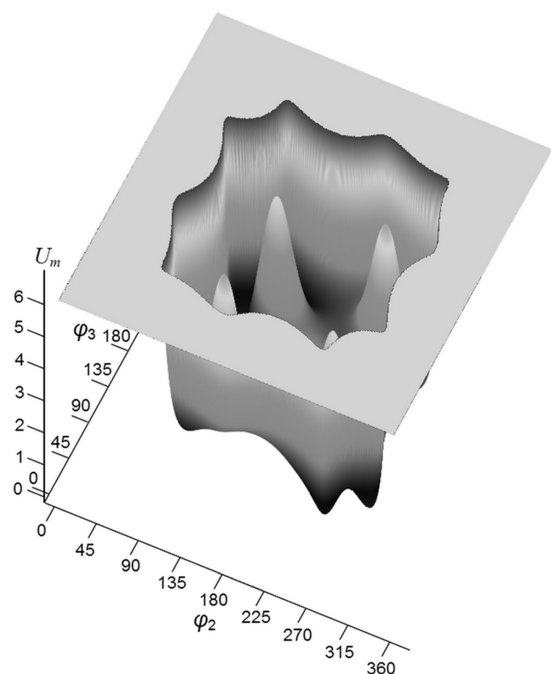


Рис. 3. Зависимость энергии (ккал/моль) фрагмента рис. 1 от углов φ_2 и φ_3 при $\varphi_1 = 180^\circ$. Отсчет энергии – от глобального минимума энергии данного фрагмента

Эту поверхность можно представить в виде эквиэнергетических линий, т. е. конформационной карты, рис. 4:

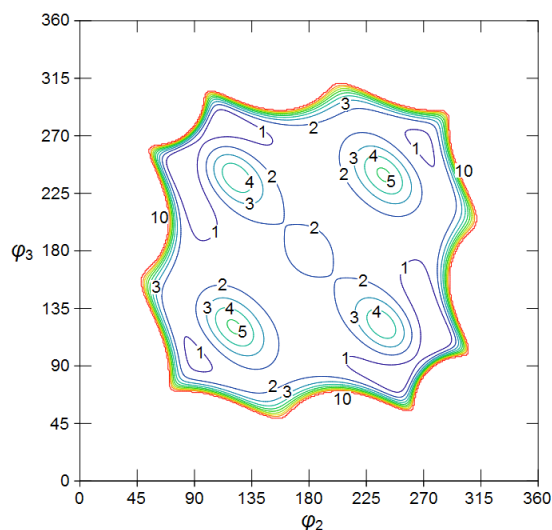


Рис. 4. Конформационная карта фрагмента рис. 1 при $\varphi_1 = 180^\circ$. Числа у кривых – энергия в ккал/моль, отсчет энергии – от глобального минимума

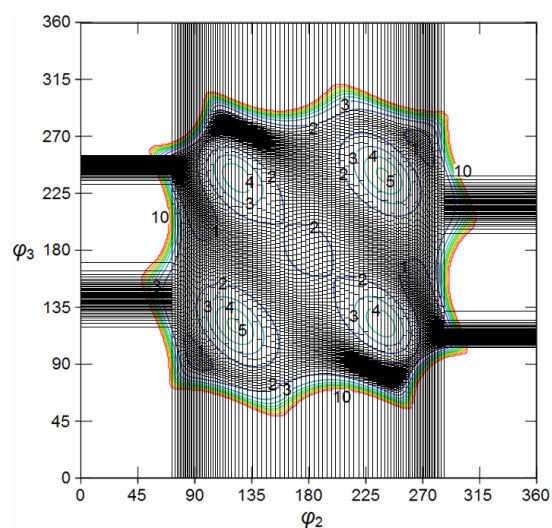


Рис. 5. Проекция границ параллелепипедов на плоскость (φ_2, φ_3) при $\varphi_1 = 180^\circ$ фрагмента рис. 1, совмещенная с соответствующей конформационной картой. Числа у кривых – энергия, ккал/моль

Слой, в который попадает значение $\varphi_1 = 180^\circ$, как и другие слои, разделен на подслои по оси φ_2 ('столбики'), а каждый подслой – на параллелепипеды по оси φ_3 . Проекция границ итоговых параллелепипедов на плоскость φ_2, φ_3 , в которой изображена также конформационная карта, имеет вид, представленный на рис. 5.

Проекция границ параллелепипедов на плоскость φ_2, φ_3 представлена на рис. 6:

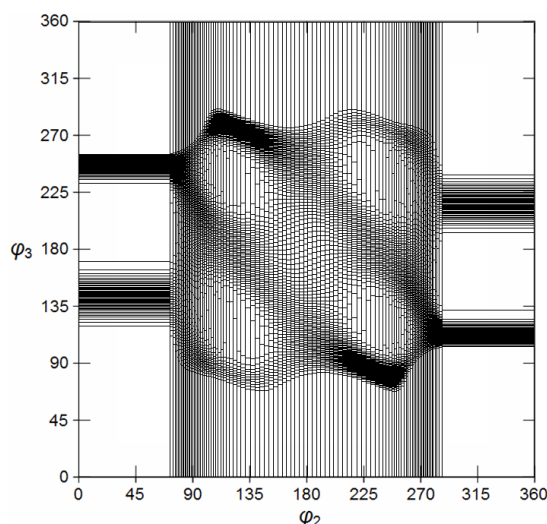


Рис. 6. Проекция границ параллелепипедов на плоскость (φ_2, φ_3) при $\varphi_1 = 180^\circ$ для фрагмента, изображенного на рис. 1

Легко видеть, что совокупность проекций границ параллелепипедов на рис. 6 воспроизводит очертания конформационной карты. Эти границы делят ось φ_2 фрагмента на полосы неравной ширины (это проекции "подслоев"), а каждую из полос по оси φ_3 — на прямоугольники (это проекции параллелепипедов) неравного размера. По количеству прямоугольников на единицу площади можно определить положения областей, в которых расположены минимумы энергии, — в них границы прямоугольников, действительно, сгущаются.

Необходимо также отметить следующее. Для расчета свойств цепных молекул часто применяется поворотнo-изомерная модель [13], в которой используются только дискретные состояния, отвечающие положениям минимумов энергии. Из рис. 3, 4 видно, что потенциальная поверхность данного молекулярного фрагмента сложна и имеет пологие, асимметричные минимумы. Такая ситуация характерна и для потенциальных поверхностей при других значениях угла φ_1 данного фрагмента, а также для потенциальных поверхностей других фрагментов молекул данного типа. Если ярко выраженных, симметричных минимумов энергии нет, то использование поворотнo-изомерного приближения затруднено, поскольку открытым остается весьма существенный вопрос о выборе значений параметров [13]. При использовании в настоящей работе непрерывного спектра конформаций учитывается асимметрия заселенности состояний. Это позволяет более адекватно моделировать макромолекулы заданной структуры, обеспечивает более широкие возмож-

ности в описании их характеристик в разных условиях, а также предсказании свойств в тех случаях, когда экспериментальные данные отсутствуют. Это создает основу для более корректного рассмотрения, например, эффектов, в которых важную роль играет "накопление" вдоль по цепи небольших отклонений углов внутреннего вращения от поворотнo-изомерных состояний. В этом состоят преимущества использования непрерывного спектра конформаций цепных молекул в теоретических моделях и, в частности, целесообразность разработки настоящего алгоритма. В последнем учитываются и другие важные факторы: воспроизводится точное химическое строение цепи и учитывается взаимозависимость каждого трех последовательных углов внутреннего вращения.

ГЕНЕРИРОВАНИЕ КОНФОРМАЦИЙ

При генерировании конформаций цепи с углами $\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_{N-1}$ составляли надлежащую последовательность фрагментов, воспроизводящих строение данной цепи, согласно принятому правилу: любые два соседних (смежных) фрагмента должны иметь два общих варьируемых торсионных угла из трех. Номерам каждой тройки углов во фрагментах присваивали последовательные номера $(\varphi_1, \varphi_2, \varphi_3)$, $(\varphi_2, \varphi_3, \varphi_4)$, $(\varphi_3, \varphi_4, \varphi_5)$ и т. д. вдоль по цепи. Для выбора значений первой тройки углов $(\varphi_1, \varphi_2, \varphi_3)$ в данной цепи методом МК выбирали один из 1000000 параллелепипедов из надлежащего концевой фрагмента, т. е. номера трех сторон параллелепипеда: номер $k^{(1)}$ слоя (от 1 до 100), номер $k^{(2)}$ подслоя (от 1 до 100) и номер $k^{(3)}$ состояния в подслое (от 1 до 100). Каждый номер выбирался с помощью генератора псевдослучайных чисел [27], равномерно распределенных в интервале $[0,1)$. Пусть $\eta^{(i)} \in [0,1)$ — такие числа ($i=1, 2, 3$), тогда $k^{(i)} = \text{INT}(100 \cdot \eta^{(i)}) + 1$, где INT — целая часть числа.

В пределах избранного параллелепипеда выбирали фиксированные значения углов φ_1, φ_2 и φ_3 , полагая, что ввиду большого количества (1000000) параллелепипедов, на которые разделено конфигурационное пространство каждого трех углов, изменениями плотности вероятности внутри каждого параллелепипеда можно пренебречь, т. е. распределение значений углов в нем считали равномерным (отметим, что при этом также вычисляли и реализуемую плотность вероятностей генерирования углов; см. ниже). При выборе этих значений использовали информацию о положе-

нии границ по осям φ_1 , φ_2 и φ_3 уже избранного параллелепипеда и три очередных псевдослучайных числа $\eta^{(i)} \in [0, 1]$; $i=4, 5, 6$. Например, если $\varphi_{1k^{(1)}}$, $\varphi_{1(k^{(1)}+1)}$ – значения границ $k^{(1)}$ -го слоя по углу φ_1 , то фиксированное значение $\varphi_{1k^{(1)}}^{(fix)}$ угла φ_1 внутри этого слоя вычисляли согласно $\varphi_{1k^{(1)}}^{(fix)} = \varphi_{1k^{(1)}} + (\varphi_{1(k^{(1)}+1)} - \varphi_{1k^{(1)}}) \cdot \eta^{(4)}$. Аналогично, для расчета фиксированного значения $\varphi_{2k^{(2)}}^{(fix)}$ угла φ_2 внутри $k^{(2)}$ -го слоя использовали положения соответствующих границ и число $\eta^{(5)}$, а для $\varphi_{3k^{(3)}}^{(fix)}$ угла φ_3 внутри $k^{(3)}$ -го слоя – положения границ и число $\eta^{(6)}$.

Очередной угол φ_4 выбирали из разбиения, отвечающего смежному фрагменту с тройкой углов $\varphi_2, \varphi_3, \varphi_4$. Но при этом рассматривали лишь те 100 параллелепипедов смежного фрагмента, в которые попали уже избранные на предыдущем шаге фиксированные значения двух предыдущих углов, φ_2 и φ_3 . Аналогично вышеприведенной процедуре выбирали некоторый номер $k^{(4)}$ одного из 100 этих параллелепипедов, для чего генерировали очередное (7-е) псевдослучайное число $\eta^{(7)} \in [0, 1]$ и вычисляли $k^{(4)} = \text{INT}(100 \cdot \eta^{(7)} + 1)$. Затем внутри избранного параллелепипеда выбирали фиксированное значение угла φ_4 . По-

следняя процедура, как и в приведенном выше примере, выполнялась с использованием известных значений границ по оси φ_4 данного параллелепипеда и еще одного генерированного псевдослучайного числа $\eta^{(8)} \in [0, 1]$.

Процедуру выбора фиксированных значений всех последующих углов φ_γ , где $\gamma = 5, 6, \dots, N-1$, проводили аналогично процедуре выбора значения угла φ_4 (выбирали номер одного параллелепипеда из 100 и затем – фиксированное значение угла), с использованием двух очередных псевдослучайных чисел, полученных с помощью датчика. Таким образом, при переходе к очередному смежному фрагменту выбирали только один (следующий) торсионный угол. Отметим, что в описанном алгоритме номера торсионных углов могут следовать не подряд вдоль по цепи, в общем случае используется последовательность только варьируемых углов.

Если бы при генерировании конформации молекулы в рамках описанной процедуры удалось воспроизвести требуемую плотность $(1/Z_{short}) \cdot \prod_{\gamma=1}^{N-3} \exp[-U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2})/k_B T]$ строго, то искомая оценка $\bar{H}_\omega$, полученная методом МК, имела бы вид, аналогичный выражению (4), т. е.

$$\bar{H}_\omega = \frac{\sum_{\nu=1}^{\omega} H(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu) \cdot \exp[-U_{long}(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)/k_B T]}{\sum_{\nu=1}^{\omega} \exp[-U_{long}(\varphi_1^\nu, \varphi_2^\nu, \dots, \varphi_{N-1}^\nu)/k_B T]}. \quad (10)$$

Но состояния в "кубах" вероятностей осуществления углов внутреннего вращения фрагментов задаются лишь с точностью до размеров параллелепипедов. Следовательно, необходимо применять выражение (9), или, в общем виде, выражения (6), (7), в которых следует вычислить реально использованную плотность вероятностей генерирования конформаций p . Это возможно (см. следующий раздел), поскольку размеры параллелепипедов, выбранных при генерировании очередной конформации цепной молекулы, известны. При этом соблюдается и общая рекомендация [12], согласно которой желательно выбирать плотность p по возможности пропорциональной подынтегральной функции, поэтому в итоге можно ожидать небольших дисперсий для оценок $\bar{H}_\omega$.

Плотность вероятностей

Размеры каждого параллелепипеда, который выбирается в процессе генерирования конформации цепи, можно идентифицировать

индексом фрагмента m_γ и тремя случайными номерами $\lambda_{1,\gamma}^\nu$, $\lambda_{2,\gamma}^\nu$, $\lambda_{3,\gamma}^\nu$ сторон 1, 2 и 3 конкретного параллелепипеда для каждого номера фрагмента γ (т. е. торсионного угла) в каждой конформации ν данной цепи, где $\lambda_{1,\gamma}^\nu = 1, 2, \dots, 100$; $\lambda_{2,\gamma}^\nu = 1, 2, \dots, 100$; $\lambda_{3,\gamma}^\nu = 1, 2, \dots, 100$. Пусть $(L_1)_{m_\gamma, \lambda_{1,\gamma}^\nu}$, $(L_2)_{m_\gamma, \lambda_{2,\gamma}^\nu}$, $(L_3)_{m_\gamma, \lambda_{3,\gamma}^\nu}$ – длины сторон, в угловых единицах, параллелепипеда, выбранного случайно в конформации номер ν во фрагменте номер γ типа m . Согласно предложенному алгоритму, фиксированные значения углов φ_1 , φ_2 , φ_3 в концовом (т. е. первом, $\gamma = 1$) фрагменте выбраны равновероятно в объеме $(L_1)_{m_1, \lambda_{1,1}^\nu} \cdot (L_2)_{m_1, \lambda_{2,1}^\nu} \cdot (L_3)_{m_1, \lambda_{3,1}^\nu}$, поэтому плотность вероятностей генерирования точки с фиксированными значениями углов $(\varphi_1, \varphi_2, \varphi_3)$ равна

$$100^{-3} \cdot [(L_1)_{m_1, \lambda_{1,1}^\nu} \cdot (L_2)_{m_1, \lambda_{2,1}^\nu} \cdot (L_3)_{m_1, \lambda_{3,1}^\nu}]^{-1}.$$

Множитель 100^{-3} обусловлен тем, что в данной процедуре выбирается также один из 100

номеров каждой из 3-х сторон параллелепипеда. Генерирование значения угла φ_4 производится по фрагменту $\gamma = 2$, в котором есть три торсионных угла: $\varphi_2, \varphi_3, \varphi_4$, но первые два уже избраны на предыдущем шаге. Плотность вероятностей выбора φ_4 есть $100^{-1} \cdot [(L_3)_{m_2, \lambda_{3,2}^\nu}]^{-1}$. Аналогичный сомножи-

тель появляется для всех последующих углов во всех фрагментах вдоль по цепи в конформации с номером ν , т. е. для очередного угла номер γ это $100^{-1} \cdot [(L_3)_{m_\gamma, \lambda_{3,\gamma}^\nu}]^{-1}$. Итоговая плотность вероятностей p , с которой реально осуществляется генерирование данной конформации, равна

$$p = 100^{-(N-1)} \cdot [(L_1)_{m_1, \lambda_{1,1}^\nu} \cdot (L_2)_{m_1, \lambda_{2,1}^\nu} \cdot (L_3)_{m_1, \lambda_{3,1}^\nu}]^{-1} \cdot \prod_{\gamma=2}^{N-3} [(L_3)_{m_\gamma, \lambda_{3,\gamma}^\nu}]^{-1}. \quad (11)$$

В общем случае, если варьируются не все углы, в этой формуле в произведении вместо $N - 3$ должно быть число $N_{angle} \leq N - 3$. Выражение p в итоге является простым по сути: в него входят лишь произведения размеров тех параллелепипедов, внутри которых были избраны значения углов. Очевидно, что p из (11) является плотностью, допустимой по отношению к любым функциям, поскольку, в силу конечности размеров всех параллелепипедов, $p > 0$.

Оценка средних

Для того чтобы осуществить варьирование не только торсионных углов, но и валентных углов и валентных связей (варьирование обобщенное и независимое для каждого валентного угла и связи), после генерирования совокупности торсионных углов проводили, как упоминалось выше, варьирование декартовых коор-

динат $x_1, x_2, \dots, x_n$ всех атомов молекулы. При решении ряда специальных задач эту процедуру не осуществляли. Оценка величины диапазона варьирования декартовых координат атомов проведена с использованием параметров поля CHARMM27 в варианте [21].

После завершения этапа варьирования декартовых координат всех атомов данной конформации ν и получения измененных наборов значений $\{l\}^\nu, \{\theta\}^\nu, \{\varphi\}^\nu$ проводили расчет полной энергии молекулы с использованием всех компонентов энергии поля CHARMM27 [21] и соответствующих Больцмановских множителей.

Формула (11) дает плотность вероятностей p генерирования конформации, полученной перед проведением этапа варьирования декартовых координат атомов. Именно это выражение следует подставить в (9) или (6), (7) вместо p для получения оценки $\bar{H}_\omega$. Например, используя выражение (7), получим

$$\bar{H}_\omega = \frac{\sum_{\nu=1}^{\omega} H^\nu \cdot \exp[-U^\nu/k_B T] \cdot [(L_1)_{m_1, \lambda_{1,1}^\nu} (L_2)_{m_1, \lambda_{2,1}^\nu} (L_3)_{m_1, \lambda_{3,1}^\nu}] \cdot \prod_{\gamma=2}^{N-3} (L_3)_{m_\gamma, \lambda_{3,\gamma}^\nu}}{\sum_{\nu=1}^{\omega} \exp[-U^\nu/k_B T] \cdot [(L_1)_{m_1, \lambda_{1,1}^\nu} (L_2)_{m_1, \lambda_{2,1}^\nu} (L_3)_{m_1, \lambda_{3,1}^\nu}] \cdot \prod_{\gamma=2}^{N-3} (L_3)_{m_\gamma, \lambda_{3,\gamma}^\nu}}. \quad (12)$$

Если варьирование декартовых координат атомов при генерировании конформаций не проводится, то в оценке (12) использовано точное выражение для плотности вероятностей p генерирования конформаций. Эта плотность, хотя и не равна строго требуемой плотности

$$(1/Z_{short}) \cdot \prod_{\gamma=1}^{N-3} \exp[-U_{m_\gamma}(\varphi_\gamma, \varphi_{\gamma+1}, \varphi_{\gamma+2})/k_B T],$$

но близка к ней. В итоге выражение (12) дает состоятельную оценку, $\bar{H}_\omega \xrightarrow{P} \langle H \rangle$ при $\omega \rightarrow \infty$.

Можно продемонстрировать степень эффективности (выгодности) разработанного алгоритма существенной выборки строго по торсионным углам по сравнению с алгоритмом простой выборки, осуществив 2 варианта ((i) и (ii)) компьютерного эксперимента, в ко-

торых варьирование декартовых координат атомов не проводится, – валентные связи и валентные углы фиксированы в равновесных положениях.

Ниже приведены оценки ряда средних характеристик для неразветвленной углеводородной цепи без двойных связей, содержащей 10 углеродных атомов (условное обозначение цепи 10:0). Расчеты проведены при температуре $T = 293$ К в так называемых Θ – условиях [13] молекулы. Доверительные интервалы во всех приведенных ниже результатах отвечают 95 % надежности согласно t -распределению Стьюдента.

Вариант (i). Цепь 10:0, простая выборка (без варьирования декартовых координат), $\omega = 9,6 \cdot 10^{10}$ конформаций. Оценки средних значений получены из выражения (9), в ко-

тором плотность p избрана постоянной, случайные значения торсионных углов распределены равномерно: расстояние между концевыми атомами углерода $9,345 \pm 0,003 \text{ \AA}$, квадрат расстояния между концевыми атомами углерода $88,48 \pm 0,06 \text{ \AA}^2$, радиус инерции $3,3434 \pm 0,0006 \text{ \AA}$.

Вариант (ii). Цепь 10:0, существенная выборка (без варьирования декартовых координат), $\omega = 7,2 \cdot 10^8$ конформаций. Оценки средних значений получены из выражения (9) с плотностью p распределения торсионных углов в генерируемой выборке согласно (11): расстояние между концевыми атомами углерода $9,344 \pm 0,001 \text{ \AA}$, квадрат расстояния между концевыми атомами углерода $88,45 \pm 0,02 \text{ \AA}^2$, радиус инерции $3,3433 \pm 0,0002 \text{ \AA}$.

Легко видеть, что оценки всех соответственных характеристик в (i) и (ii) в пределах доверительных интервалов совпадают, но в варианте (ii) использован примерно в 130 раз меньший размер выборки ω , чем в варианте (i), и при этом в 3 раза меньшим оказался доверительный интервал. Если учесть, что вероятная ошибка $\sim \omega^{-1/2}$ [3, 12], то для достижения одинаковых ошибок в (i) и (ii) можно было бы уменьшить ω в варианте (ii) еще примерно в 9 раз. В итоге ω для существенной выборки можно избрать примерно в 1000 раз меньше, чем для выборки простой.

Далее, расчеты показали, что суммарные времена t генерирования случайных точек и вычисления всех заданных случайных величин в обоих алгоритмах соизмеримы: на такой расчет для одной конформации цепи по алгоритму (ii) потребовалось всего в 1,25 раза больше времени, чем по алгоритму (i). В итоге произведение времени t на ω в варианте существенной выборки (ii) оказалось в несколько сотен раз меньше, чем в варианте простой выборки (i) и, таким образом, разработанный алгоритм существенной выборки является намного более эффективным, выгодным.

Варианты, в которых осуществляется варьирование декартовых координат всех атомов, требуют дополнительного анализа: такое варьирование приводит к изменению не только валентных связей и валентных углов (по отношению к своим равновесным значениям), но и к дополнительному изменению торсионных углов (по отношению к значениям, которые уже были получены в процедуре существенной выборки). Это изменение в формуле (11) не учтено, и для таких условий формула (11) дает лишь приблизительное выражение для плотности вероятностей p генерирования данной конформации.

В этой связи отметим, что проводимая в настоящей работе последовательность действий может быть рассмотрена и как приложение иных известных способов уменьшения дисперсии [12]. В частности, вычисление интегралов при фиксированных ("замороженных") значениях всех величин $\{l\}$ и $\{\theta\}$, осуществленное на 1-м этапе, фактически означает "интегрирование по части переменных (понижение порядка интеграла)" [12]. Доказано, что если аналитически взять интеграл по некоторым из переменных, а по остальным переменным использовать тот же метод МК, то дисперсия уменьшится. В данном случае этот этап проведен, и можно ожидать, что "размораживание" величин $\{l\}$ и $\{\theta\}$ на 2-м этапе приведет к некоторому увеличению дисперсии.

Далее, варьирование декартовых координат атомов (которое эквивалентно "размораживанию" $\{l\}$ и $\{\theta\}$), проводится лишь в малом диапазоне ($\pm 0,02 \text{ \AA}$) вокруг равновесных положений атомов; это отвечает малым колебаниям последних. Ограничение диапазона варьирования декартовых координат до конечных значений означает применение еще одного известного способа уменьшения дисперсии: "интегрирования по части области" [12]. Идея способа такова: если возможно (аналитически) вычислить интеграл по некоторой части области (допустим, что результат интегрирования равен W), то всегда выгодно вычислять простым методом МК только интеграл по оставшейся области и добавить W к итоговой оценке в качестве слагаемого. Этот способ для данного случая варьирования валентных углов и связей дает $W \rightarrow 0$, поскольку интегрирование проводится по той части области Ω , в которой энергия валентных углов и связей $\rightarrow \infty$, т. е. Больцмановский множитель в интеграле $\rightarrow 0$. Это есть вся область с большими отклонениями углов и связей. В итоге вычислению методом МК подлежит только интеграл по оставшейся области – вблизи положений равновесия атомов.

Для вычисления интеграла по этой оставшейся области можно рассчитать плотность p , – она будет иметь достаточно сложный вид. Более целесообразно, однако, использовать выражение (11) для генерирования торсионных углов, и рекомендуемый простой метод МК (равномерное распределение) для плотности декартовых координат атомов в узких "потенциальных ямах" ($\pm 0,02 \text{ \AA}$) вокруг равновесных положений. Типичные значения равновесных длин разных валентных связей в поле CHARMM27 [21] –

это $\sim 1,11 - 1,53 \text{ \AA}$ (что отвечает данным физического эксперимента), и амплитуда варьирования $\sim 0,02 \text{ \AA}$ соответствует лишь нескольким процентам от этих величин. Согласно исходным условиям, описанным выше, энергия таких колебаний должна быть сравнительно малой. Естественно ожидать, что влияние таких колебаний и на другие средние характеристики молекул также будет малым.

Методом МК получены оценки степени влияния малых колебаний на ряд характеристик. Для этого осуществлены компьютерные эксперименты (iii) и (iv) с варьированием декартовых координат всех атомов для молекулы 10:0 при $T = 293 \text{ K}$ в Θ – условиях [13].

Вариант (iii). Цепь 10:0, простая выборка (с варьированием декартовых координат), $\omega = 9,6 \cdot 10^{10}$ конформаций. Оценки средних значений получены из выражения (6) (или (7)), в котором положено $p = \text{Const}$, т. е. точки выбирались равномерно в пространстве торсионных углов и в малых областях вокруг равновесных значений атомов: расстояние между концевыми атомами углерода $9,347 \pm 0,007 \text{ \AA}$, квадрат расстояния между концевыми атомами углерода $88,51 \pm 0,10 \text{ \AA}^2$, радиус инерции $3,3441 \pm 0,0012 \text{ \AA}$. Эти оценки $\overline{H}_\omega \xrightarrow{P} \langle H \rangle$ при $\omega \rightarrow \infty$, поскольку все состояния выбирали равномерно, а дополнительное варьирование торсионных углов при этом не нарушает равномерности распределения по торсионным углам.

Вариант (iv). Цепь 10:0, существенная выборка (с варьированием декартовых координат), $\omega = 28,8 \cdot 10^8$ конформаций. Оценки средних значений получены из выражения (12), т. е. торсионные углы выбирали неравномерно, с плотностью p , но декартовые координаты в пределах областей $\pm 0,02 \text{ \AA}$ вокруг равновесных значений атомов – равномерно: расстояние между концевыми атомами углерода $9,348 \pm 0,001 \text{ \AA}$, квадрат расстояния между концевыми атомами углерода $88,54 \pm 0,02 \text{ \AA}^2$, радиус инерции $3,3443 \pm 0,0002 \text{ \AA}$. Размер выборки ω в варианте (iv) был увеличен в 4 раза по сравнению с таковым в варианте (ii) для того, чтобы добиться совпадения доверительных интервалов в (ii) и (iv).

Как и ожидалось, вследствие учета малых колебаний оценки геометрических характеристик в варианте (iii) увеличились по сравнению с (i) (и, соответственно, в варианте (iv) – увеличились по сравнению с (ii)). Вместе с тем относительный рост этих характеристик составил всего 0,02–0,04 %; т. е. влияние обсуждаемого варьирования декартовых координат

атомов данной цепной молекулы на ее геометрические характеристики, действительно, оказалось малым.

Отметим, что оценки всех соответствующих характеристик вариантов (iii) и (iv) в пределах доверительных интервалов совпадают (аналогично тому, как совпадают оценки характеристик в вариантах (i) и (ii)). Размер выборки ω в варианте (iii) примерно в 30 раз больше, чем в варианте (iv), а доверительные интервалы – больше примерно в 5–7 раз, т. е. для получения одинаковых интервалов в (iii) и (iv) следует увеличить ω в варианте (iii) еще примерно в 25–50 раз, и получим в ~ 750 –1500 раз больше, чем в варианте (iv). Поскольку времена расчета случайных величин соизмеримы, вариант существенной выборки с варьированием координат (iv) оказался также в сотни раз эффективнее, чем соответствующий вариант простой выборки (iii).

Разработанный алгоритм компьютерной имитации цепных молекул позволяет воспроизвести точное химическое строение цепи, учитывает взаимозависимость каждого трех последовательных углов внутреннего вращения и непрерывный спектр конформаций цепей, а при генерировании этих конформаций использует метод существенной выборки. В итоге он позволяет передать основные особенности молекул цепного строения и, обладая общностью, может быть адаптирован для конструирования и исследования свойств цепных молекул разных классов.

Работа выполнена при поддержке РФФИ (проект 10-03-00201а) и программы Президента РФ – Ведущие научные школы (НШ-1642.2012.4).

ЛИТЕРАТУРА

1. Биндер К. Методы Монте-Карло в статистической физике / Ред. К. Биндер. М.: Мир, 1982. С. 9–57.
2. Биндер К., Хеерман Д. В. Моделирование методом Монте-Карло в статистической физике: Введение. М.: Наука; Физматлит, 1995. 144 с.
3. Бусленко Н. П., Голенко Д. И., Соболев И. М., и др. Метод статистических испытаний (метод Монте-Карло). М.: ГИФМЛ, 1962. 332 с.
4. Гнеденко Б. В. Курс теории вероятностей. М.: Гос. изд. физ.-мат. лит., 1961. 408 с.
5. Гулд Х., Тобочник Я. Компьютерное моделирование в физике. М.: Мир, 1990. Ч. 2. 400 с.
6. Ермаков С. М. Метод Монте-Карло и смежные вопросы. М.: Наука; Гл. ред. физ.-мат. лит., 1971. 328 с.

7. Ермаков С. М., Михайлов Г. А. Статистическое моделирование. М.: Наука; Гл. ред. физ.-мат. лит., 1982. 296 с.
8. Зализняк В. Е. Основы научных вычислений. Введение в численные методы для физиков. М.: Едиториал УРПС, 2002. 296 с.
9. Ландау Л. Д., Лифшиц Е. М. Статистическая физика. Ч. 1. М.: Наука; Гл. ред. физ.-мат. лит., 1976. 584 с.
10. Методы компьютерного моделирования для исследования полимеров и биополимеров / Отв. ред. В. А. Иванов, А. Л. Рабинович, А. Р. Хохлов. М.: Книжный дом ЛИБРОКОМ, 2009. 696 с.
11. Рабинович А. Л., Иванов В. А. Методы компьютерного моделирования для исследования полимеров и биополимеров / Отв. ред. В. А. Иванов, А. Л. Рабинович, А. Р. Хохлов. М.: Книжный дом ЛИБРОКОМ, 2009. С. 63–119.
12. Соболев И. М. Численные методы Монте-Карло. М.: Наука; Гл. ред. физ.-мат. лит., 1973. 312 с.
13. Флори П. Статистическая механика цепных молекул. М.: Мир, 1971. 440 с.
14. Хеерман Д. В. Методы компьютерного эксперимента в теоретической физике. М.: Наука; Гл. ред. физ.-мат. лит., 1990. 176 с.
15. Allen M. P., Tildesley D. J. Computer Simulation of Liquids. Oxford: Clarendon Press, 1987. 385 p.
16. Feller S. E., MacKerell, Jr. A. D. An improved empirical potential energy function for molecular simulations of phospholipids // J. Phys. Chem. B. 2000. Vol. 104. P. 7510–7515.
17. Feller S. E., Yin D., Pastor R. W., MacKerell, Jr. A. D. Molecular dynamics simulation of unsaturated lipids at low hydration: parametrization and comparison with diffraction studies // Biophys. J. 1997. Vol. 73. P. 2269–2279.
18. Flory P. J. Foundations of rotational isomeric state theory and general methods for generating configurational averages // Macromolecules. 1974. Vol. 7, N 3. P. 381–392.
19. Frenkel D., Smit B. Understanding Molecular Simulation. From Algorithms to Applications. 2nd edition, San Diego: Academic Press, 2002. 638 p.
20. Gō N., Scheraga H. A. On the use of classical statistical mechanics in the treatment of polymer chain conformations // Macromolecules. 1976. Vol. 9, N 4. P. 535–542.
21. Högborg C.-J., Nikitin A. M., Lyubartsev A. P. Modification of the CHARMM force field for DMPC lipid bilayer // J. Comput. Chem. 2008. Vol. 29. P. 2359–2369.
22. Kahn H. Random sampling (Monte Carlo) techniques in neutron attenuation problems. 1 // Nucleonics. 1950. Vol. 6, N 5. P. 27–33.
23. Kahn H. Random sampling (Monte Carlo) techniques in neutron attenuation problems. 2 // Nucleonics. 1950. Vol. 6, N 6. P. 60–65.
24. Klauda J. B., Brooks B. R., MacKerell, Jr. A. D. et al. An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and DPPC bilayer // J. Phys. Chem. B. 2005. Vol. 109. P. 5300–5311.
25. Landau D. P., Binder K. A Guide to Monte Carlo Simulations in Statistical Physics. 2nd edition. Cambridge University Press, 2005. 432 с.
26. Simulation Methods for Polymers / Eds. Kotelyanskii M., Theodorou D.N. N.Y.: Marcel Dekker, Inc., 2004. 602 p.
27. Matsumoto M., Nishimura T. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator // ACM Transactions on Modeling and Computer Simulations: Special Issue on Uniform Random Number Generation. 1998. Vol. 8. P. 1–25.
28. Schlenkerich M., Brickmann J., MacKerell, Jr. A. D., Karplus M. In: Biological membranes: a molecular perspective from computation and experiment / K. M. Merz, B. Roux, Eds. Birkhauser, Boston. 1996. P. 31–81.

СВЕДЕНИЯ ОБ АВТОРАХ:

Рабинович Александр Львович

гл. науч. сотр., д. ф.-м. н.
Институт биологии Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: rabinov@krc.karelia.ru
тел.: (8142) 571879

Журкин Дмитрий Викторович

аспирант
Петрозаводский государственный университет,
физико-технический факультет
ул. Университетская, 10 А, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: zhurkin_dm@mail.ru
тел.: (8142) 719671

Rabinovich, Alexandr

Institute of Biology, Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: rabinov@krc.karelia.ru
tel.: (8142) 571879

Zhurkin, Dmitrii

Petrozavodsk State University, Faculty of Physical Engineering
10 A University St., 185910 Petrozavodsk, Karelia, Russia
e-mail: zhurkin_dm@mail.ru
tel.: (8142) 719671

УДК 519.2

О СЛУЧАЙНЫХ ПУАССОНОВСКИХ ЗАПОЛНЕНИЯХ ЯЧЕЕК

Е. В. Хворостянская

*Институт прикладных математических исследований
Карельского научного центра РАН
Петрозаводский государственный университет*

Рассматривается модель заполнения N различных ячеек частицами, в которой случайные величины $\eta_1, \dots, \eta_N$, равные числу частиц в ячейках, независимы и имеют распределение Пуассона. Для подмножества реализаций такой схемы, удовлетворяющих условию $\eta_1 + \dots + \eta_N \leq n$, при $N \rightarrow \infty$ получены предельные распределения числа ячеек μ_r , содержащих ровно r частиц.

Ключевые слова: размещение частиц по ячейкам, предельное распределение, число ячеек заданного объема.

Е. V. Khvorostyanskaya. ON RANDOM POISSON ALLOCATION TO CELLS

We study a model for particle allocation to N different cells in which the random variables $\eta_1, \dots, \eta_N$ equal to the number of particles in cells are independent and have a Poisson distribution. For a subset of realizations of such a scheme satisfying $\eta_1 + \dots + \eta_N \leq n$ limit distributions of cell number μ_r with exactly r particles are obtained as $N \rightarrow \infty$.

Key words: particle allocation to cells, limit distribution, number of cells with a given size.

Пусть n различных частиц независимо одна от другой размещаются по N ячейкам, пронумерованным числами от 1 до N , так, что каждая частица с вероятностью N^{-1} попадает в одну из ячеек. Такая схема размещения частиц по ячейкам носит название классической схемы размещения, и ее изучению посвящено множество работ (см., например, [2, 3]). Хорошо известно, что совместное распределение заполнений ячеек в классической схеме совпадает с условным распределением N независимых пуассоновских случайных величин при условии, что их сумма равна n . Используя это свойство, при $N \rightarrow \infty$ и всех возможных значениях n , r были, в частности,

получены предельные распределения числа ячеек, содержащих ровно r частиц.

В настоящей статье рассматривается следующая модель. Пусть случайные величины $\eta_1, \dots, \eta_N$, равные числу частиц в ячейках с номерами $1, \dots, N$ соответственно, независимы и имеют распределение Пуассона:

$$p_k = \mathbf{P}\{\eta_i = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

$i = 1, \dots, N$. Очевидно, что число всех возможных реализаций такой схемы бесконечно. Далее будем рассматривать подмножество реализаций, удовлетворяющих

условию $\eta_1 + \dots + \eta_N \leq n$. Ясно, что в этом случае случайные величины $\eta_1, \dots, \eta_N$ не являются независимыми. Пусть вероятностная мера на этом подмножестве индуцируется совместным распределением $\eta_1, \dots, \eta_N$. Тогда для целых неотрицательных чисел $k_1, \dots, k_N$ таких, что $k_1 + \dots + k_N \leq n$, справедливо равенство

$$\mathbf{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} = \frac{\lambda^{k_1 + \dots + k_N} e^{-\lambda N}}{k_1! \dots k_N!} \times \left(\sum_{i=0}^n \sum_{\substack{i_1, \dots, i_N \geq 0 \\ i_1 + \dots + i_N = i}} \frac{\lambda^i e^{-\lambda N}}{i_1! \dots i_N!} \right)^{-1}.$$

Введем независимые одинаково распределенные по закону Пуассона с параметром λ случайные величины $\xi_1, \dots, \xi_N$. Легко видеть, что

$$\mathbf{P}\{\xi_1 = k_1, \dots, \xi_N = k_N \mid \xi_1 + \dots + \xi_N \leq n\} = \frac{\lambda^{k_1 + \dots + k_N} e^{-\lambda N}}{k_1! \dots k_N!} \left(\sum_{i=0}^n \frac{(\lambda N)^i e^{-\lambda N}}{i!} \right)^{-1}.$$

Отсюда и из полиномиальной формулы следует, что выполнено соотношение

$$\mathbf{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} = \mathbf{P}\{\xi_1 = k_1, \dots, \xi_N = k_N \mid \xi_1 + \dots + \xi_N \leq n\}.$$

Эта модель является частным случаем схемы, рассмотренной в [4] и основанной на идеях обобщенной схемы размещения частиц по ячейкам [2]. Заметим, что в [4] нет конкретных примеров, удовлетворяющих условиям этой схемы.

Обозначим через μ_r случайную величину, равную числу ячеек, содержащих ровно r частиц, и пусть

$$y = \frac{n - \lambda N}{\sqrt{\lambda N}}, \quad f_r(\lambda) = \frac{(\lambda - r)^2 p_r}{\lambda}.$$

Справедливы следующие утверждения.

Теорема 1. Пусть $N \rightarrow \infty$, $y^2 \sqrt{p_r/N} + y^2 f_r(\lambda) \rightarrow 0$ при $y \rightarrow -\infty$ и выполнено одно из следующих условий:

- 1) $r \geq 0$, $\lambda \rightarrow \infty$;
- 2) $r \rightarrow \infty$, $0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty$;
- 3) $r \geq 2$, $\lambda \rightarrow 0$;
- 4) $r = 1$, $y \rightarrow \infty$, $\lambda \rightarrow 0$ так, что $\lambda^2 N \rightarrow \infty$.

Тогда

$$\mathbf{P}\{\mu_r = k\} = \frac{(N p_r)^k}{k!} e^{-N p_r} (1 + o(1))$$

равномерно относительно целых неотрицательных k , для которых $(k - N p_r) / \sqrt{N p_r}$ лежит в любом конечном фиксированном интервале.

Теорема 2. Пусть $N \rightarrow \infty$, $0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty$, $r \geq 0$ фиксировано и выполнено одно из следующих условий:

- 1) $y \rightarrow +\infty$;
- 2) $|y| \leq C < \infty$, $r - \lambda \rightarrow 0$;
- 3) $y \rightarrow -\infty$, $r - \lambda \rightarrow 0$ так, что $y(r - \lambda) \rightarrow 0$, $y^2 / \sqrt{N} \rightarrow 0$.

Тогда

$$\mathbf{P}\{\mu_r = k\} = \frac{1 + o(1)}{\sqrt{2\pi N p_r (1 - p_r)}} e^{-u^2/2}$$

равномерно относительно $u = (k - N p_r) / \sqrt{N p_r (1 - p_r)}$ в любом конечном фиксированном интервале. Утверждение теоремы остается справедливым, если выполнены условия теоремы 1 и $N p_r \rightarrow \infty$, а также в случае, когда $r = 0$, $y \rightarrow \infty$, $N \rightarrow \infty$, $\lambda \rightarrow 0$ так, что $\lambda^4 N \rightarrow \infty$.

Теорема 3. Пусть $r = 1$, $n \geq 1$ фиксировано, $\lambda \rightarrow 0$, $N \rightarrow \infty$ так, что $N \lambda \leq C < \infty$. Тогда для целых неотрицательных k

$$\mathbf{P}\{\mu_r = k\} = \frac{(N \lambda)^k e^{-N \lambda}}{k!} \times \left(\sum_{i=0}^n \frac{(N \lambda)^i e^{-N \lambda}}{i!} \right)^{-1} (1 + o(1))$$

при $0 \leq k \leq n$ и $\mathbf{P}\{\mu_r = k\} = 0$ при $k > n$.

Замечание 1. Нетрудно видеть, что для нашей модели теорема 1 имеет значительно более общий характер, чем теорема 9 работы [4], поскольку в ней условия 1), 2), 4) не рассматривались, а условие 3) требует дополнительных ограничений. Теорема 2 о сходимости к нормальному закону также охватывает существенно более широкую зону изменения параметров по сравнению с теоремами 4 и 5 [4].

Согласно теореме 3 [4], для $k = 0, 1, \dots, N$ справедливо равенство

$$\mathbf{P}\{\mu_r = k\} = \binom{N}{k} p_r^k (1 - p_r)^{N-k} \frac{\mathbf{P}\{\zeta_{N-k}^{(r)} \leq n - kr\}}{\mathbf{P}\{\zeta_N \leq n\}}, \quad (1)$$

где

$$\zeta_N = \xi_1 + \dots + \xi_N, \quad \zeta_{N-k}^{(r)} = \xi_1^{(r)} + \dots + \xi_{N-k}^{(r)},$$

независимые случайные величины $\xi_1^{(r)}, \dots, \xi_N^{(r)}$ имеют распределения

$$\mathbf{P} \left\{ \xi_i^{(r)} = k \right\} = \mathbf{P} \{ \xi_1 = k \mid \xi_1 \neq r \},$$

$i = 1, \dots, N$. Для доказательства теорем 1, 2 с помощью (1) нам потребуются вспомогательные утверждения, которые приводятся ниже в леммах 1, 2.

Лемма 1. Пусть $N \rightarrow \infty$ и $\lambda N \rightarrow \infty$. Тогда

$$\mathbf{P} \left\{ \frac{\zeta_N - \lambda N}{\sqrt{\lambda N}} \leq x \right\} = \frac{1 + o(1)}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz.$$

Доказательство. Обозначим через $\varphi(t)$, $\psi(t)$ характеристические функции случайных величин ξ_1 , $(\zeta_N - \lambda N) / \sqrt{\lambda N}$. Легко видеть, что

$$\varphi(t) = e^{\lambda(e^{it} - 1)},$$

$$\psi(t) = \exp \left\{ -it\sqrt{\lambda N} \right\} \varphi^N \left(t / \sqrt{\lambda N} \right).$$

Используя формулу Тейлора, при любом фиксированном t получаем, что

$$\begin{aligned} \ln \psi(t) &= -it\sqrt{\lambda N} + \lambda N \left(e^{it/\sqrt{\lambda N}} - 1 \right) \\ &= -t^2/2 + o(1). \end{aligned}$$

Отсюда и из теоремы непрерывности следует утверждение леммы. $\square$

Лемма 2. Пусть $s \rightarrow \infty$, $\lambda s \rightarrow \infty$ и при $\lambda \rightarrow 0$, $r = 1$ выполнено условие $\lambda^2 s \rightarrow \infty$. Тогда

$$\mathbf{P} \left\{ \frac{\zeta_s^{(r)} - m_r s}{\sigma_r \sqrt{s}} \leq x \right\} = \frac{1 + o(1)}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz,$$

где

$$m_r = \mathbf{E} \xi_1^{(r)} = \frac{\lambda - r p_r}{1 - p_r}, \quad (2)$$

$$\sigma_r^2 = \mathbf{D} \xi_1^{(r)} = \frac{\lambda(1 - p_r - f_r(\lambda))}{(1 - p_r)^2}.$$

Доказательство. При $\lambda \rightarrow 0$, $\lambda \rightarrow \infty$, $r \geq 2$ и $0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty$ слабая сходимость распределения $(\zeta_s^{(r)} - m_r s) / (\sigma_r \sqrt{s})$

к стандартному нормальному закону получена в [1], [3]. Учитывая явный вид характеристической функции случайной величины $(\zeta_s^{(r)} - m_r s) / \sigma_r \sqrt{s}$, можно показать, что утверждение леммы остается в силе и в оставшихся случаях. $\square$

Докажем теорему 1. Поскольку $p_r \rightarrow 0$, выполнено соотношение

$$\binom{N}{k} p_r^k (1 - p_r)^{N-k} = \frac{(N p_r)^k}{k!} e^{-N p_r} (1 + o(1)) \quad (3)$$

равномерно относительно k , для которых $u = (k - N p_r) / \sqrt{N p_r}$ лежит в любом конечном фиксированном интервале.

Пусть выполнено одно из условий теоремы и, кроме того, $\lambda N \rightarrow \infty$. С помощью лемм 1, 2 получаем, что

$$\frac{\mathbf{P} \left\{ \zeta_{N-k}^{(r)} \leq n - k r \right\}}{\mathbf{P} \{ \zeta_N \leq n \}} \quad (4)$$

$$\begin{aligned} &= \frac{\int_{-\infty}^{x_r} e^{-z^2/2} dz}{\int_{-\infty}^y e^{-z^2/2} dz} (1 + o(1)), \end{aligned}$$

где

$$x_r = \frac{n - k r - m_r(N - k)}{\sigma_r \sqrt{N - k}},$$

m_r , σ_r определены в (2). Для выбранных значений k несложно показать, что

$$x_r = \frac{n - \lambda N + u(\lambda - r)\sqrt{N p_r} / (1 - p_r)}{\sqrt{\lambda N} \left(1 - \frac{f_r(\lambda)}{1 - p_r} \right)^{1/2} \left(1 - u \frac{\sqrt{p_r/N}}{1 - p_r} \right)^{1/2}}.$$

При условиях 1)-3) теоремы $f_r(\lambda) \rightarrow 0$ и

$$x_r = y \left(1 + \frac{f_r(\lambda) + u\sqrt{p_r/N}}{2(1 - p_r)} + o(f_r(\lambda)) \right) \quad (5)$$

$$+ o \left(\sqrt{\frac{p_r}{N}} \right) + \frac{u\sqrt{f_r(\lambda)} \operatorname{sgn}(\lambda - r)}{1 - p_r} (1 + o(1)).$$

Если выполнено условие 4) теоремы, то $f_1(\lambda) \rightarrow 1$ и

$$x_1 = y \left(1 - \frac{f_1(\lambda)}{1 - p_1} \right)^{-1/2} \left(1 - u \frac{\sqrt{p_1/N}}{1 - p_1} \right)^{-1/2} \quad (6)$$

$$-u\sqrt{f_1(\lambda)}(1 + o(1)) = \frac{y}{\sqrt{2\lambda}}(1 + o(1)).$$

Поскольку (4) справедливо при любых сколь угодно больших значениях x_r и y , учитывая (5), (6), находим, что если $y \rightarrow \infty$, то $x_r \rightarrow \infty$ и

$$\frac{\mathbf{P}\left\{\zeta_{N-k}^{(r)} \leq n - kr\right\}}{\mathbf{P}\left\{\zeta_N \leq n\right\}} = 1 + o(1). \quad (7)$$

Если $|y| \leq C < \infty$, то из (4) получаем, что

$$\begin{aligned} & \frac{\mathbf{P}\left\{\zeta_{N-k}^{(r)} \leq n - kr\right\}}{\mathbf{P}\left\{\zeta_N \leq n\right\}} \\ &= \left(1 + \frac{\int_{-x_r}^{-y} e^{-z^2/2} dz}{\int_{-y}^{+\infty} e^{-z^2/2} dz}\right) (1 + o(1)), \end{aligned}$$

и, поскольку из (5) следует соотношение $x_r - y \rightarrow 0$, легко видеть, что выполнено (7).

Нетрудно проверить, что при любом $a > 0$

$$\begin{aligned} \int_a^\infty e^{-z^2/2} dz &\leq \frac{1}{a} \int_a^\infty z e^{-z^2/2} dz = \frac{e^{-a^2/2}}{a}, \\ \int_a^\infty e^{-z^2/2} dz &\geq a^2 \int_a^\infty \frac{e^{-z^2/2}}{z^2} dz = \frac{a}{1+a^2} e^{-a^2/2}. \end{aligned}$$

Используя эти неравенства и (5), получаем оценки

$$\begin{aligned} \frac{\int_{|x_r|}^\infty e^{-z^2/2} dz}{\int_{|y|}^\infty e^{-z^2/2} dz} &\leq \frac{1+y^2}{yx_r} e^{-(x_r^2-y^2)/2}, \\ \frac{\int_{|x_r|}^\infty e^{-z^2/2} dz}{\int_{|y|}^\infty e^{-z^2/2} dz} &\geq \frac{yx_r}{1+x_r^2} e^{-(x_r^2-y^2)/2}, \end{aligned}$$

где

$$\begin{aligned} \frac{1+y^2}{yx_r} &= (1+1/y^2) (1+o(1)), \\ \frac{yx_r}{1+x_r^2} &= \frac{1}{1+1/y^2} (1+o(1)), \end{aligned}$$

$$x_r^2 - y^2 = y^2 \left(f_r(\lambda) + u\sqrt{p_r/N} \right) (1+o(1)).$$

Отсюда и из (4) следует, что при $y \rightarrow -\infty$ и $y^2 f_r(\lambda) + y^2 \sqrt{p_r/N} \rightarrow 0$ справедливо (7).

Пусть теперь $\lambda N \rightarrow \alpha$, где α – некоторая неотрицательная постоянная. Очевидно,

что распределение суммы ζ_N сходится к закону Пуассона с параметром α . Для характеристической функции $\psi_r(t)$ случайной величины $\zeta_{N-k}^{(r)}$ выполнено соотношение

$$\begin{aligned} \ln \psi_r(t) &= (N-k) \ln \frac{e^{\lambda(e^{it}-1)} - e^{itr} p_r}{1-p_r} \\ &= N\lambda (e^{it} - 1) (1+o(1)), \end{aligned}$$

и по теореме непрерывности распределение $\zeta_{N-k}^{(r)}$ сходится к закону Пуассона с параметром α . Поскольку $Np_r \rightarrow 0$, получаем, что $k=0$ и справедливо равенство (7).

Утверждение теоремы 1 следует из (1), (3), (7).

Для доказательства теоремы 2 положим $k = Np_r + u\sqrt{Np_r(1-p_r)}$. Пусть $0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty$, $r \geq 0$ фиксировано. С помощью лемм 1, 2 находим, что выполнено равенство (4), где

$$\begin{aligned} x_r &= \left(y(1-p_r) + u\sqrt{f_r(\lambda)(1-p_r)} \operatorname{sgn}(\lambda-r) \right) \\ &\quad \times \frac{(1-p_r-f_r(\lambda))^{-1/2}}{\sqrt{1-p_r-u\sqrt{p_r(1-p_r)}/N}}. \end{aligned}$$

Аналогично доказательству теоремы 1 можно показать, что справедливо соотношение (7).

При $\lambda \rightarrow 0$, $r=0$ находим, что $\lambda^3(N-k) = \lambda^4 N(1+o(1)) \rightarrow \infty$, и из лемм 1, 2 следует (4), где $x_0 = \sqrt{2y/\lambda(1+o(1))}$. Очевидно, что если $y \rightarrow \infty$, то $x_0 \rightarrow \infty$ и выполнено (7).

При $Np_r(1-p_r) \rightarrow \infty$ имеет место равенство

$$\binom{N}{k} p_r^k (1-p_r)^{N-k} = \frac{1+o(1)}{\sqrt{2\pi Np_r(1-p_r)}} e^{-u^2/2}.$$

Отсюда и из (1), (7) получаем утверждение теоремы 2.

Доказательство теоремы 3 аналогично доказательству теоремы 8 [4].

Работа выполнена при поддержке Программы стратегического развития Петрозаводского государственного университета на 2012–2016 гг.

ЛИТЕРАТУРА

1. Колчин А. В. Предельные теоремы для обобщенной схемы размещения // Дискретная математика. 2003. Т. 15, № 4. С. 148–157.
2. Колчин В. Ф. Случайные графы. М.: Физматлит, 2000. 256 с.

3. Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. М.: Наука, 1976. 159 с.

4. Чупрунов А. Н., Фазекаш И. Аналог обобщенной схемы размещения. Предельные теоремы для числа ячеек заданного объема // Дискретная математика. 2012. Т. 24, № 1. С. 140–158.

СВЕДЕНИЯ ОБ АВТОРЕ:

Хворостянская Елена Владимировна

старший научный сотрудник, к. ф.-м. н.

Институт прикладных математических исследований
Карельского научного центра РАН

ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910

эл. почта: cher@krc.karelia.ru

тел.: (8142) 781218

Khvorostyanskaya, Elena

Institute of Applied Mathematical Research, Karelian
Research Centre, Russian Academy of Sciences

11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
Russia

e-mail: cher@krc.karelia.ru

tel.: (8142) 781218

УДК 512.2

СЛУЧАЙНЫЕ ЛЕСА: ОБЗОР

С. П. Чистяков

*Институт прикладных математических исследований
Карельского научного центра РАН*

В статье представлен обзор современного состояния исследований в области случайных лесов – статистического метода, предназначенного для решения задач классификации и регрессии. Приведен исторический экскурс развития деревьев решений и ансамблей классификаторов и описаны основные понятия (загрязненность, расщепление, баггинг, бустинг и др.), используемые при их построении. Рассматриваются вопросы, касающиеся состоятельности метода и сравнения его с другими методами классификации. Представлены возможности использования случайных лесов для задач нахождения наиболее информативных признаков, кластеризации, выделения аномальных наблюдений и определения прототипов классов. Кратко рассмотрены некоторые неклассические разновидности деревьев решений и случайных лесов, а именно: косоугольные деревья, случайные леса выживаемости, квантильные леса регрессий, логические случайные леса, вероятностные случайные леса и потоковые случайные леса. Также приведен обзор соответствующего программного обеспечения с акцентом на пакет R – свободно распространяемое программное обеспечение для статистических вычислений и графики, доступное на платформах Linux, Windows, Mackintosh.

Ключевые слова: деревья решений, ансамбли классификаторов, баггинг, случайные леса, классификация, регрессия, кластеризация, пакет R.

S. P. Chistiakov. RANDOM FORESTS: AN OVERVIEW

This paper presents an overview of the state-of-the-art in the studies of random forests – a statistical method designed to deal with problems of classification and regression. We tell about the history of decision trees and classifier ensembles and describe the corresponding basic ideas (impurity, split, bagging, boosting, etc.). Some issues of the consistency of the method are considered. Applicability of random forests to the problems of finding most informative features, clustering, finding outlier observations and class prototypes is surveyed. Several non-classical variants of decision trees and random forests is considered, namely: oblique trees, survival random forests, quantile regression forests, logical random forests, probabilistic random forests and streaming random forests. We also survey the corresponding software with the emphasis on R package – open source environment for statistical computing and graphics which is freely available for the computing platforms Linux, Windows, Mackintosh.

Key words: decision trees, classifier ensembles, bagging, random forests, classification, regression, clustering, R package.

ВВЕДЕНИЕ

Понятие *случайный лес* впервые было введено в научный обиход в работах [6, 7], см. также [8]. В этих статьях рассматривалось множество корневых лесов с помеченными вершинами, на котором задавалось равномерное распределение вероятностей. Позднее появилась монография [81], в которой изучались случайные леса с распределениями, отличными от равномерного. Таким образом, с точки зрения теории вероятностей, случайные леса являются частным случаем известного понятия *случайный элемент* (см. [5]). Однако в 2001 г. в статье [25] был предложен новый метод классификации и регрессии, также получивший название *случайный лес*. В этом смысле термин *случайный лес* широко используется в таких дисциплинах как машинное обучение (machine learning), распознавание образов, дисциплине известной как "Data Mining"¹ и, в меньшей степени, в прикладной статистике. Настоящий обзор посвящен этому методу.

Метод основан на построении большого числа (ансамбля) деревьев решений (это число является параметром метода), каждое из которых строится по выборке, получаемой из исходной обучающей выборки с помощью бутстрепа (т. е. выборки с возвращением). В отличие от классических алгоритмов построения деревьев решений [21, 91] в методе случайных лесов при построении каждого дерева на стадиях расщепления вершин используется только фиксированное число случайно отбираемых признаков обучающей выборки (второй параметр метода) и строится полное дерево (без усечения), т. е. каждый лист дерева содержит наблюдения только одного класса. Классификация осуществляется с помощью голосования классификаторов, определяемых отдельными деревьями, а оценка регрессии — усреднением оценок регрессии всех деревьев. Известно (см., например, [66]), что точность (вероятность корректной классификации) ансамблей классификаторов существенно зависит от разнообразия (diversity) классификаторов, составляющих ансамбль или, другими словами, от того, насколько коррелированы их решения. А именно, чем более разнообразны классификаторы ансамбля (меньше коррелированность их решений), тем выше вероятность корректной классификации. В случайных лесах решения составляющих их деревьев слабо кор-

релированы вследствие двойной "инъекции случайности" в алгоритм построения случайного леса — на стадии бутстрепа и на стадии случайного отбора признаков, используемых при расщеплении вершин деревьев.

Метод быстро получил признание как в статистическом сообществе, так и в среде исследователей, использующих методы распознавания образов в своей работе и в настоящее время является одним из наиболее популярных методов классификации и непараметрической регрессии. Причиной этого явилась не только высокая точность классификации (а по мнению автора и не столько), обеспечиваемая методом, но и другие его достоинства. Именно:

- метод гарантирует защиту от переобучения (overfitting)² даже в случае, когда количество признаков значительно превышает количество наблюдений. Это свойство выделяет метод "случайный лес" среди множества других методов классификации и является чрезвычайно ценным для решения многих прикладных задач;
- для построения случайного леса по обучающей выборке требуется задание всего двух параметров, которые требуют минимальной настройки (tuning);
- метод Out-Of-Bag (ООВ), предложенный Брейманом [23], обеспечивает получение естественной оценки вероятности ошибочной классификации случайных лесов на основе наблюдений, не входящих в обучающие бутстреп выборки, используемые для построения деревьев (эти наблюдения называются ООВ выборками);
- случайные леса могут использоваться не только для задач классификации и регрессии, но и для задач выявления наиболее информативных признаков, кластеризации, выделения аномальных наблюдений и определения прототипов классов;
- обучающая выборка для построения случайного леса может содержать признаки, измеренные в разных шкалах: числовой, порядковой и номинальной, что недопустимо для многих других классификаторов;

¹Общепринятого перевода этого термина на русский язык не существует. Иногда используемый дословный перевод "добыча данных" не выдерживает никакой критики.

²Под переобученкой понимается ситуация, при которой классификатор хорошо классифицирует наблюдения обучающей выборки, но непригоден для классификации наблюдений, не входящих в нее.

- метод допускает легкую параллелизацию (т. е. программную реализацию, пригодную для параллельных вычислений), что весьма существенно при больших объемах обучающей выборки.

Автору представляется, что развитие метода случайных лесов проходило в следующих направлениях:

- исследование свойств самого метода, т. е. аналитическая и экспериментальная работа по оценке точности, сравнению с другими ансамблевыми методами классификации и т. д.;
- развитие возможностей метода, ориентированных на решение задач, непосредственно не связанных с задачами классификации и регрессии (см. выше);
- разработка на основе метода других родственно связанных методов, таких как случайные леса выживаемости, квантильные регрессионные леса, логические случайные леса, вероятностные случайные леса и потоковые случайные леса;
- использование схемы построения случайных лесов для построения ансамблей классификаторов, не являющихся деревьями — ансамблей наивных байесовских классификаторов и мультиномиальных логистических моделей;
- разработка алгоритмов и программных средств, реализующих метод.

В соответствии с этой классификацией исследований и построен данный обзор.

Первый раздел представляет собой краткий экскурс в историю возникновения метода. Рассмотрены элементы метода — деревья решений, ансамбли классификаторов, баггинг (агрегированный бутстреп) и метод случайных подпространств. В этом разделе автор счел целесообразным также кратко описать основные понятия, используемые при построении деревьев решений — загрязненность (impurity), расщепление (split) и усечение (pruning) деревьев решений.

Во втором разделе приведен алгоритм построения случайного леса и его определение, принадлежащие Брейману. Приведены некоторые теоретические и экспериментальные результаты относительно состоятельности метода. Обсуждаются достоинства и недостатки метода.

В третьем разделе рассмотрены возможности использования случайных лесов для задач отбора наиболее информативных признаков, кластеризации, выявления аномальных наблюдений и определения прототипов классов.

В четвертом разделе рассмотрены разновидности случайных лесов — случайные леса выживаемости (random survival forests) [63], квантильные регрессионные леса (quantile regression forests) [73], логические регрессионные леса (logic forest) [117] и другие. В этом разделе также кратко рассмотрены ансамбли классификаторов, строящихся по той же схеме, что и случайные леса, а именно, ансамбли наивных байесовских классификаторов (naive bayes classifier) и ансамбли мультиномиальных логистических моделей [84].

Пятый раздел посвящен обзору программного обеспечения для построения деревьев решений и случайных лесов с акцентом на пакет R — свободно распространяемое программное обеспечение для статистических вычислений и графики, доступное на платформах Linux, Windows, Macintosh [85]. Пакет R не является пакетом в классическом понимании. Это среда для статистических вычислений и графики с интерпретируемым языком программирования (который также носит название R). За рубежом R получил широкое распространение как в среде профессиональных статистиков, так и в среде исследователей, регулярно применяющих статистические методы в своей работе. В настоящее время R содержит более 3000 пакетов практически по всем статистическим методам, методам распознавания образов, машинного обучения и Data Mining разработанные различными институтами, группами исследователей и отдельными исследователями по всему миру.

В заключении кратко обсуждаются возможные направления дальнейших исследований.

Обзор ни в коей мере не носит учебно-методический характер и предполагает, что читатель знаком с основными понятиями и методами распознавания образов и прикладной статистики (в частности, предполагается знакомство с деревьями решений). Разделы, посвященные деревьям решений, могут быть найдены практически в любом учебнике по машинному обучению и Data Mining. Обзор предназначен для читателей, чьи профессиональные интересы связаны с прикладной статистикой и распознаванием образов, однако автор надеется, что он также окажется полез-

ным и для исследователей в других областях науки.

ДЕРЕВЬЯ РЕШЕНИЙ

В этом разделе приведен краткий экскурс, посвященный истории возникновения и развития идей, синтез которых привел к появлению метода случайных лесов. Отправной точкой в этом изложении являются деревья решений. Идея, лежащая в основе деревьев решений, состоит в разбиении множества возможных значений вектора признаков (независимых переменных) на непересекающиеся множества и подгонке простой модели для каждого такого множества. Понятие дерева решений опирается на понятие дерева из теории графов и понятие обучающей выборки из распознавания образов. Ниже кратко описаны эти понятия.

Граф $G = (V, E)$ состоит из конечного непустого множества V , элементы которого называются вершинами и множества пар вершин E , называемых ребрами. Путем в графе называется последовательность ребер вида $(v_1, v_2), (v_2, v_3), \dots, (v_{m-1}, v_m)$. Если $v_1 = v_m$, то такой путь называется циклом. Если пара вершин v, w , образующая ребро (v, w) , является упорядоченной, то такое ребро называется ориентированным или дугой, ведущей из вершины v в вершину w . Если все ребра графа ориентированы, то такой граф называется ориентированным. Дерево представляет собой связный граф без циклов. Под корневым деревом понимается дерево, в котором одна вершина выделена и называется корнем. Далее рассматриваются только ориентированные корневые деревья, в которых дуги направлены по направлению от корня. Заметим, что такие деревья удовлетворяют следующим условиям:

- существует только одна вершина, называемая корнем, в которую не ведет ни одна дуга;
- в каждую вершину (исключая корень) ведет только одна дуга;
- существует единственный путь от корня к любой вершине.

Если (v, w) — некоторая дуга, то вершина v называется родителем w , а вершина w — потомком вершины v . Вершина, не имеющая потомков, называется терминальной вершиной или листом. Дерево называется бинарным, если каждая его вершина (за исключением терминальных вершин) имеет ровно двух потомков.

Понятие обучающей выборки является ключевым в распознавании образов. Под обучающей выборкой понимается независимая выборка $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^l$ из некоторого (неизвестного) распределения $P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$. Здесь $\mathbf{x}_i$, $i = 1, 2, \dots, l$ — векторы признаков (называемые прецедентами), координаты которых представляют значения n признаков (независимых переменных), измеренных на некотором объекте (образе). Множество всех возможных значений векторов признаков $\mathcal{X}$ называется пространством образов. Признаки могут быть измерены в различных шкалах — числовой, порядковой или номинальной. Соответствующие y_i представляют собой значения зависимой переменной. Если y_i может принимать только конечное число значений, т. е. $y_i \in \{\omega_1, \omega_2, \dots, \omega_c\}$, $c \geq 2$, то мы имеем задачу классификации. В этом случае y_i называется меткой класса и определяет принадлежность соответствующего объекта к одному из c классов, а сам признак называется классовым; если же y_i измерены в числовой шкале, то мы имеем задачу регрессии; в этом случае признак называется откликом;

Деревом решений называется дерево, с каждой вершиной t которого связаны [1]³:

1. Некоторое подмножество $\mathcal{X}_t \subset \mathcal{X}$; с корневой вершиной связывается все пространство образов $\mathcal{X}$;
2. Подвыборка $\mathcal{D}_t \subset \mathcal{D}$ обучающей выборки $\mathcal{D}$, такая, что $\mathcal{D}_t = \{(\mathbf{x}, y) \in \mathcal{D} : \mathbf{x} \in \mathcal{X}_t\}$; таким образом с корневой вершиной связывается вся выборка $\mathcal{D}$;
3. Некоторая функция (правило) $f_t : \mathcal{X} \rightarrow \{0, 1, \dots, k_t - 1\}$ (здесь $k_t \geq 2$ — количество потомков вершины t), определяющая разбиение множества $\mathcal{X}$ на k непересекающихся подмножеств. С терминальными вершинами не связывается никакая функция.

Обозначим $t_{i(t)}$, $i = 0, 1, \dots, k_t - 1$ вершину, являющуюся i -м потомком вершины t . Множество $\mathcal{X}_t$ и правило f_t определяют множества $\mathcal{X}_{t_{i(t)}}$ следующим образом:

$$\mathcal{X}_{t_{i(t)}} = \mathcal{X}_t \cap \{\mathbf{x} \in \mathcal{X} : f_t(\mathbf{x}) = i\}. \quad (1)$$

Цель построения дерева решений состоит либо в классификации векторов $\mathbf{x}$ из распределения $P(\mathbf{x})$, либо в оценке условного математического ожидания отклика при данном значении $\mathbf{x}$. Процесс принятия решений начинается с корневой вершины и состоит в последовательном применении правил, связанных с вершинами

³В этой работе использовался термин "древовидные классификаторы".

дерева. Результатом этого процесса является определение терминальной вершины t такой, что $\mathbf{x} \in \mathcal{X}_t$. В случае классификации вектор $\mathbf{x}$ относится к классу, являющемуся мажорантным (наиболее часто встречающимся) в подвыборке D_t , соответствующей данной терминальной вершине, а в случае регрессии оценка условного математического ожидания отклика представляет собой среднее значение отклика в этой подвыборке (в этом случае дерево решений часто называют деревом регрессий).

Существует несколько причин, обусловивших широкое применение деревьев решений в самых различных областях несмотря на их, вообще говоря, невысокую точность классификации. Основными из них, по мнению автора, являются следующие:

1. Обучающая выборка для построения деревьев решений может содержать признаки, измеренные как в числовой, так и в номинальной шкале, что недопустимо для многих других классификаторов, предполагающих существование некоторой объективной метрики, измеряющей близость наблюдений;
2. Иерархическая структура деревьев решений сводит принятие решения (классификацию) к последовательности более простых и интуитивно понятных решений;
3. Деревья решений служат удобной моделью представления знаний в экспертных системах;
4. Алгоритмы построения (индукции) деревьев решений инвариантны к масштабированию значений признаков.

Хотя деревья решений как инструмент классификации известны достаточно давно, однако их широкое применение началось с разработкой алгоритма CART (Classification And Regression Trees) и публикацией классической работы [21], идеи которой дали начало многочисленным исследованиям. Алгоритм CART является базовым (generic) алгоритмом, на основе которого может быть построено множество конкретных алгоритмов, приводящих к построению различных деревьев решений. Алгоритм основан на идее рекурсивного разбиения обучающей выборки на две более однородные подвыборки с помощью одного из признаков. Для реализации этой идеи необходимо определить понятие меры однородности.

Загрязненность

Обычно вместо меры однородности рассматривается противоположная по смыслу мера загрязненности (impurity). Пусть t — некоторая вершина дерева решений, $D(t)$ — подвыборка, связанная с этой вершиной и $i(t)$ — загрязненность вершины. Естественно потребовать, чтобы загрязненность вершины была равна 0, если $D(t)$ содержит прецеденты только одного класса и была бы максимальной в случае, если $D(t)$ содержит одинаковое число прецедентов каждого класса. Все приведенные ниже меры загрязненности удовлетворяют этому условию. Одной из наиболее используемых является мера загрязненности вершины, основанная на понятии энтропии (entropy impurity):

$$i(t) = - \sum_{j=1}^c P(\omega_j) \log_2 P(\omega_j), \quad (2)$$

где $P(\omega_j)$ есть доля примеров класса ω_j в подвыборке $D(t)$ и полагается $0 \log 0 = 0$. Другой популярной мерой является индекс Джини (Gini) [44], определяемый как

$$i(t) = 1 - \sum_{j=1}^c P^2(\omega_j). \quad (3)$$

Индекс Джини представляет собой частоту ошибочной классификации при случайном назначении меток классов наблюдениям подвыборки $D(t)$. Реже применяется мера загрязненности, основанная на частоте ошибочной классификации (misclassification impurity):

$$i(t) = 1 - \max_j P(\omega_j). \quad (4)$$

В этом случае $i(t)$ представляет собой частоту ошибочной классификации если все наблюдения выборки $D(t)$ относятся к мажорантному (наиболее часто встречающемуся) классу. Исследования показали [21], что выбор меры загрязненности не оказывает существенного влияния на точность классификации и более важным является выбор критериев остановки и усечения дерева решений, рассматриваемых ниже.

Расщепление вершин

Правило разбиения множества $\mathcal{X}$, связанное с каждой вершиной дерева решений, называется расщеплением (split). Количество подмножеств, на которые разбивается $\mathcal{X}$, в принципе может быть разным для разных вершин, однако, большинство алгоритмов основано на

построении бинарных деревьев, т. е. деревьев, в которых расщепление осуществляется на два подмножества. Это связано с тем, что для любого дерева решений можно построить эквивалентное ему (с точки зрения принимаемых решений) бинарное дерево и, кроме того (что весьма существенно), значительно облегчается программная реализация алгоритма. Бинарное расщепление вершины t можно рассматривать как функцию $f_t : \mathcal{X} \rightarrow \{0, 1\}$, $\mathbf{x} \in \mathcal{X}$, где в случае $f(\mathbf{x}) = 0$ вектор $\mathbf{x}$ относится к первому (левому) потомку, а в случае $f(\mathbf{x}) = 1$ – ко второму (правому). Обычно эта функция имеет простой вид и зависит от значений только одного признака. Именно, если некоторый признак x измерен в числовой шкале, то расщепление состоит в выборе x_s , минимизирующего используемую меру загрязненности и определении

$$f_t(\mathbf{x}) = \begin{cases} 0 & \text{при } x < x_s, \\ 1 & \text{при } x \geq x_s. \end{cases} \quad (5)$$

В этом случае, если признак x в выборке принимает m различных значений, то существует $m - 1$ возможное расщепление, сохраняющее порядок значений признака x в подвыборках. Легко видеть, что если все признаки обучающей выборки измерены в числовой шкале и используются только расщепления указанного вида, то области решений будут представлять собой многомерные параллелепипеды, а часть границы области решений, соответствующая данному расщеплению, будет представлять собой часть гиперплоскости, параллельной соответствующей координатной оси в пространстве $\mathbf{R}^n$. Аналогичные расщепления используются и для признаков, измеренных в порядковой шкале.

Если признак x измерен в номинальной шкале и варьирует на уровнях из множества C , то возможные расщепления имеют вид

$$f_t(\mathbf{x}) = \begin{cases} 0 & \text{при } x \in C_1 \\ 1 & \text{при } x \in C_2 \end{cases}, \quad (6)$$

где C_1 — произвольное непустое подмножество C и $C_2 = C - C_1$.

В работе [110] рассматривались расщепления, определяемые линейными комбинациями значений признаков:

$$f_t(\mathbf{x}) = \begin{cases} 0 & \text{при } \sum_{i=1}^n a_i x_i < x_s \\ 1 & \text{при } \sum_{i=1}^n a_i x_i \geq x_s \end{cases}, \quad (7)$$

что приводит к так называемым косоугольным деревьям решений (oblique trees). В этом случае минимизация загрязненности осуществляется по параметрам $a_1, a_2, \dots, a_n$ и x_s . Неко-

торые алгоритмы построения деревьев решений, использующих для расщепления значения нескольких признаков, представлены в [16, 30, 77, 78]. Пока эти алгоритмы не получили широкого распространения, что, по-видимому, связано со сложностью дальнейшей интерпретации получаемых деревьев. Одним из подходов к получению расщеплений (рекомендованный Брейманом), использующий значения нескольких признаков, является метод главных компонент, при котором из исходной обучающей выборки сначала выделяются главные компоненты, а на втором этапе по преобразованной выборке строится дерево решений.

Оптимальное расщепление

Как уже отмечалось, расщепление подвыборки естественно осуществлять таким образом, чтобы максимально уменьшить загрязненность. Уменьшение загрязненности вершины t для бинарных деревьев определяется как

$$\Delta i(t) = i(t) - P_L i(t_L) - P_R i(t_R), \quad (8)$$

где P_L и P_R есть доли примеров подвыборки $D(t)$, соответствующие левому и правому потомку (t_L и t_R). Наилучшим расщеплением вершины t естественно считать разбиение, которое максимизирует величину $\Delta i(t)$.

В общем случае уменьшение загрязненности определяется как

$$\Delta i(t) = i(t) - \sum_{k=1}^B P_k i(t_k), \quad (9)$$

где B — количество потомков вершины t , P_k — доля примеров подвыборки $D(t)$, соответствующая вершине t_k и $\sum_{k=1}^B P_k = 1$. Непосредственное использование этой формулы, однако, приводит к тому, что чаще для расщепления будут выбираться признаки, имеющие больше уровней по сравнению с остальными. Поэтому обычно используется следующая формула

$$\Delta i_B(t) = \frac{\Delta i(t)}{-\sum_{k=1}^B P_k \log_2 P_k}, \quad (10)$$

и выбирается расщепление, максимизирующее величину $\Delta i_B(t)$.

Критерии остановки расщепления

Процесс расщепления вершин имеет естественный предел, когда становится невозможно уменьшить загрязненность очередной вершины. Все такие вершины объявляются терминальными, а соответствующее дерево

полным. Обычно такая ситуация характеризуется тем, что каждая терминальная вершина содержит примеры только одного класса, а само дерево содержит большое количество вершин. Хорошо известно, что полное дерево, как правило, обладает низкой точностью классификации [21]. Это связано с так называемой проблемой переподргонки (overfitting), заключающейся в том, что статистическая модель фактически описывает только саму выборку и непригодна в качестве модели всей генеральной совокупности. В некоторой степени задачу предотвращения построения больших деревьев играют критерии останова расщепления. Существует несколько способов принятия решения об остановке расщепления. Наиболее простой способ состоит в задании минимального числа для количества наблюдений в подвыборках, соответствующих терминальным вершинам (или минимальной доли наблюдений обучающей выборки). Вторым способом заключается в установлении верхнего порога для загрязненности вершины, т. е. задается некоторое число a и расщепление вершины не происходит, если уменьшение загрязненности при расщеплении не превышает a . Третий способ заключается в применении кросс-проверки, а именно, сравнивается количество ошибочно классифицированных наблюдений до и после расщепления вершины и если сокращения ошибок не происходит, то вершина не расщепляется. В случае бинарных деревьев альтернативой приведенным способам является статистический подход [75], основанный на статистике

$$\chi^2 = \sum_{i=1}^2 \frac{(n_{iL} - n_{ie})^2}{n_{ie}}, \quad (11)$$

где n_{iL} – количество наблюдений класса ω_i , отнесенных в результате расщепления в левую подвыборку, и n_{ie} – ожидаемое количество наблюдений в левой подвыборке при случайном расщеплении вершины t . Если максимальное значение (по всем возможным расщеплениям) статистики χ^2 не превышает критического значения, соответствующего выбранному уровню значимости, то расщепления не происходит и вершина объявляется терминальной.

Усечение деревьев

Недостатком подхода, основанного на способах останова расщепления вершин для предотвращения построения больших деревьев, является то, что решение об остановке принимается без учета ситуации, которая могла бы сложиться при продолжении расщеп-

ления. Именно, не исключено, что дальнейшее расщепление потомков некоторой вершины могло бы существенно повысить точность классификации. В связи с этим ключевое значение приобрел альтернативный подход, основанный на построении и дальнейшем усечении (pruning) полных деревьев решений [49, 76, 79, 91]. Усечение означает процедуру замены в построенном полном дереве некоторой вершины и связанного с ней поддерева терминальной вершиной. Большинство методов усечения основано на оценке чувствительности поддеревьев по отношению к некоторой мере и удалению поддеревьев, которые оказывают минимальное влияние на эту меру. Эмпирическое сравнение различных методов усечения деревьев решений проведено в [45, 76]. Общим недостатком всех методов усечения относится то, что они не могут гарантировать нахождение оптимального решения, а также высокая вычислительная сложность.

Другим популярным алгоритмом построения деревьев решений является алгоритм C4.5 [92], предшественником которого является алгоритм ID3 [90]. Алгоритм ID3 предполагает, что все признаки обучающей выборки измерены в номинальной шкале. Если обучающая выборка содержит признаки, измеренные в числовой шкале, то они предварительно дискретизируются. Дискретизация означает разбиение области значений непрерывного признака на совокупность непересекающихся интервалов, каждый из которых затем трактуется как уровень номинального признака. Количество подвыборок, на которые расщепляется выборка, соответствующая некоторой вершине, равно количеству уровней признака, выбранного для расщепления в этой вершине. В алгоритме C4.5 номинальные признаки используются аналогично ID3, а непрерывные признаки аналогично CART.

Построение деревьев решений по обучающей выборке относится к методам рекурсивного разбиения. Подробное введение в эти методы содержится в работе [105]. В заключение заметим, что наряду с классическими деревьями решений разрабатывается также подход, основанный на теории нечетких множеств и приводящий к построению нечетких деревьев решений (fuzzy decision trees) [33, 112].

АНСАМБЛИ КЛАССИФИКАТОРОВ

Ансамбль классификаторов представляет собой множество классификаторов, чьи решения комбинируются некоторым образом для получения окончательной классификации наблюдений. Обычно синтез решений от-

дельных классификаторов, составляющих ансамбль, осуществляется путем их голосования (возможно, взвешенного). Основная причина, обусловившая интерес к данной тематике, состоит в том, что при определенных условиях ансамбли классификаторов обладают точностью, значительно превосходящей точность отдельных классификаторов и робастны (устойчивы) по отношению к "зашумлению" обучающей выборки. Необходимым и достаточным условием высокой точности ансамбля классификаторов является то, чтобы составляющие его классификаторы были сами достаточно точны и различны (diverse), т. е. совершали ошибки на различных прецедентах [54]. Существует несколько групп методов построения ансамблей классификаторов [42]:

1. Манипулирование примерами обучающей выборки;
2. Манипулирование признаками;
3. Инъекция случайности в индуктивный алгоритм;
4. Манипулирование метками классов;
5. Байесовское голосование.

Случайные леса синтезируют методы первых трех групп, поэтому мы остановимся только на них. Первая группа методов состоит либо в прогоне базового индуктивного алгоритма на различных подвыборках исходной обучающей выборки, либо в итеративном перевзвешивании наблюдений. Наиболее простой способ формирования подвыборок предложен Брейманом [22]. Метод основан на формировании обучающей выборки для каждого классификатора ансамбля с помощью бутстрепа (bootstrap), т. е. случайной выборки (того же объема, что и исходная обучающая выборка) с возвращением из исходной обучающей выборки и использовании метода голосования для агрегирования решений отдельных классификаторов. Метод получил название баггинг (bagging) или агрегированный бутстреп (bootstrap aggregating). Многочисленные экспериментальные исследования (см., например, [20, 22] показали, что использование баггинга существенно повышает точность классификации в случае неустойчивости базового классификатора, когда небольшие возмущения обучающей выборки приводят к существенным изменениям в классификации. Заметим, что деревья решений представляют собой пример неустойчивого классификатора. Теоретический анализ баггинга деревьев решений, проведенный в [15], показал, что баггинг

приводит к сокращению средней квадратичной ошибки классификации. Дальнейшие исследования показали, что это верно не для всех базовых моделей классификаторов. К методам этой группы относится также метод, основанный на формировании множества дизъюнктивных подвыборок обучающей выборки и использовании методики кросс-проверки [80] для формирования ансамбля классификаторов.

Метод, основанный на итеративном перевзвешивании наблюдений обучающей выборки, называемый бустингом (boosting), был предложен в работе [47]. Идея бустинга состоит в том, что классификаторы ансамбля строятся последовательно и на каждой итерации происходит коррекция (перевзвешивание) наблюдений обучающей выборки (на первой итерации веса всех наблюдений равны). Коррекция осуществляется таким образом, чтобы соответствующий классификатор делал меньше ошибок на тех наблюдениях, на которых часто делали ошибки классификаторы, построенные на предыдущих итерациях алгоритма. Кроме того, каждому классификатору приписывается некоторый вес исходя из количества допущенных им ошибок.

Отметим, что идея последовательной коррекции классификационных алгоритмов для компенсации предыдущих ошибок классификации развивалась и в рамках оптимизационных методов алгебраического подхода к распознаванию образов [2, 3, 9] (см., также обзор [4]).

Из методов второй группы отметим работу [32], в которой исходное множество признаков было разбито на несколько дизъюнктивных подмножеств и строился ансамбль нейронных сетей, каждая из которых включала признаки только из одного подмножества. К методам этой (и третьей) группы можно отнести метод случайных подпространств [56–58], рассматриваемый ниже.

Методы третьей группы основаны на том, что случайность вводится непосредственно в базовый алгоритм. Так, в работах [40, 41] алгоритм построения деревьев решений, составляющих ансамбль, был рандомизирован следующим образом. На каждом шаге расщепления вычислялось 20 наилучших расщеплений и затем осуществлялся случайный выбор одного из них. Во второй из этих работ было проведено сравнительное экспериментальное исследование трех методов построения ансамблей деревьев решений — бустинга, баггинга и рандомизации на 33 выборках. Результаты показали, что в случае незначительно "зашумленной" обучающей выборки рандомизация срав-

нима (и, возможно, несколько превосходит) с баггингом, но не так точно как бустинг. В случае сильно "за шумленными" выборки баггинг намного точнее бустинга и иногда превосходит рандомизацию.

Экспериментальное сравнение различных алгоритмов построения ансамблей деревьев решений на 57 реальных обучающих выборках было проведено в [14]. Точность решений соответствующих классификаторов оценивалась методом кросс-проверки с дальнейшей проверкой на статистическую значимость. По результатам кросс-проверки алгоритмы ранжировались в порядке уменьшения точности (для каждой выборки). Результаты этих исследований могут быть суммированы следующим образом:

- лучший алгоритм был статистически значимо лучше баггинга только на 8 из 57 выборок;
- проверка средних рангов алгоритмов по всем выборкам показала, что бустинг, случайные леса и рандомизированные деревья статистически значимо точнее баггинга.

В работе [100] было проведено сравнительное экспериментальное исследование двух методов построения ансамблей – баггинга и бустинга. В качестве базового классификатора использовался метод ближайшего среднего (сравнение с эталоном). Исследование показало, что баггинг "работает" намного лучше бустинга в случае, если количество признаков сравнимо с объемом обучающей выборки, а бустинг предпочтительнее баггинга если объем обучающей выборки намного превышает количество признаков. Причиной этого (как установлено в статье) является то, что в случае баггинга разнообразие классификаторов ансамбля уменьшается при увеличении объема обучающей выборки, а в случае бустинга увеличивается.

Случайные леса

Первые работы, связанные с построением ансамблей деревьев решений, были основаны на эвристических процедурах и относятся к началу 90-х гг. прошлого века [68, 97, 98, 115]. Применялся также подход, основанный на бустинге (boosting) деревьев решений [43].

Более строгий подход, названный методом случайных подпространств, был развит в работах Хо [56–58]. Суть подхода заключается в том, что для построения каждого дерева используется только фиксированная доля слу-

чайно отобранных признаков. Ансамбли построенных деревьев Хо назвал случайными лесами решений (Random Decision Forests).

В важной для формирования метода случайных лесов работе [12] для построения деревьев было предложено использовать для расщепления каждой вершины деревьев только фиксированную долю случайно отбираемых признаков, т. е. фактически использовать метод случайных подпространств на каждой итерации построения дерева.

И, наконец, Брейманом был предложен метод, известный как случайные леса. Фактически случайный лес Бреймана — это ансамбль деревьев решений, каждое из которых строится на основе бутстреп выборки из исходной обучающей выборки (баггинг) [22], причем для расщепления вершин аналогично [12] используется только доля случайно отбираемых признаков. Кроме того, строится полное дерево (без усечения). Классификация деревьев в ансамбле осуществляется большинством голосов.

Алгоритм индуктивного построения и использования случайного леса впервые был разработан Брейманом и Катлер [35] и реализован в ряде коммерческих пакетов. Вся информация, касающаяся деталей алгоритма, технические доклады, руководства, обзор и примеры применения метода случайных лесов может быть найдена по адресу:

URL: <http://www.stat.berkeley.edu/users/breiman/RandomForests>.

Алгоритм индукции случайного леса может быть представлен в следующем виде [99]:

1. Для $i = 1, 2, \dots, B$ (здесь B — количество деревьев в ансамбле) выполнить
 - Сформировать бутстреп выборку S размера l по исходной обучающей выборке $D = \{\mathbf{x}_i, y_i\}_{i=1}^l$;
 - По бутстреп выборке S индуцировать неусеченное дерево решений T_i с минимальным количеством наблюдений в терминальных вершинах равным n_{min} , рекурсивно следуя следующему подалгоритму:
 - (а) из исходного набора n признаков случайно выбрать p признаков;
 - (б) из p признаков выбрать признак, который обеспечивает наилучшее расщепление;
 - (в) расщепить выборку, соответствующую обрабатываемой вершине, на две подвыборки;

2. В результате выполнения шага 1 получаем ансамбль деревьев решений $\{T_i\}_{i=1}^B$;
3. Предсказание новых наблюдений осуществлять следующим образом:

(а) для регрессии:

$$\hat{f}_{rf}^B(\mathbf{x}) = \frac{1}{B} \sum_{i=1}^B T_i(\mathbf{x}) ;$$

(б) для классификации:

пусть $\hat{\omega}_i(\mathbf{x}) \in \{\omega_1, \omega_2, \dots, \omega_c\}$ – класс, предсказанный деревом решений T_i , т. е. $T_i(\mathbf{x}) = \hat{\omega}_i(\mathbf{x})$; тогда $\hat{\omega}_{rf}^B(\mathbf{x})$ – класс, наиболее часто встречающийся в множестве $\{\hat{\omega}_b(\mathbf{x})\}_{b=1}^B$.

Одно из достоинств случайных лесов состоит в том, что для оценки вероятности ошибочной классификации нет необходимости использовать кросс-проверку или тестовую выборку. Оценка вероятности ошибочной классификации случайного леса осуществляется методом "Out-Of-Bag" (OOB), состоящем в следующем [23]. Известно, что каждая бутстреп выборка не содержит примерно 37 % наблюдений исходной обучающей выборки (поскольку выборка с возвращением, то некоторые наблюдения в нее не попадают, а некоторые попадают несколько раз). Классифицируем некоторый вектор $\mathbf{x} \in \mathcal{D}$. Для классификации используются только те деревья случайного леса, которые строились по бутстреп выборкам, не содержащим $\mathbf{x}$, и как обычно используется метод голосования. Частота ошибочно классифицированных векторов обучающей выборки при таком способе классификации и представляет собой оценку вероятности ошибочной классификации случайного леса методом OOB. Практика применения оценки OOB показала, что в случае, если количество деревьев достаточно велико, эта оценка обладает высокой точностью. Если число деревьев мало, то оценка имеет положительное смещение [31].

Случайные леса обладают целым рядом привлекательных качеств, что обусловило их широкое применение, а именно:

1. Случайные леса обеспечивают существенное повышение точности, так как деревья в ансамбле слабо коррелированы вследствие двойной инъекции случайности в индуктивный алгоритм — посредством баггинга и использования метода случайных подпространств при расщеплении каждой вершины;

2. Методически и алгоритмически сложная задача усечения полного дерева решений снимается, поскольку деревья в случайном лесу не усекаются (это также приводит к высокой вычислительной эффективности);
3. Отсутствует проблема перепогонки (даже при количестве признаков, превышающем количество наблюдений обучающей выборки и большом количестве деревьев). Тем самым снимается сложная проблема отбора признаков, необходимая для других ансамблевых классификаторов;
4. Простота применения: единственными параметрами алгоритма являются количество деревьев в ансамбле и количество признаков, случайно отбираемых для расщепления в каждой вершине дерева. Подробные рекомендации по выбору этих параметров можно найти в [69];
5. Легкость организации параллельных вычислений.

Состоятельность случайных лесов

Брейман дает следующее определение случайного леса [25].

Определение 1. Случайным лесом называется классификатор, состоящий из набора деревьев $\{h(\mathbf{x}, \Theta_k) \mid k = 1, \dots, l\}$, где Θ_k — независимые одинаково распределенные случайные векторы и каждое дерево вносит один голос при определении класса $\mathbf{x}$.

Определение, данное Брейманом, является достаточно общим. В рамках этого определения возможны различные модели случайных лесов, в зависимости от того, каким образом "вводится случайность" в алгоритм индукции деревьев решений [12, 25, 41, 56]. Так, в соответствии с этим определением, случайным лесом является классификатор, состоящий из ансамбля деревьев решений, каждое из которых строится с использованием баггинга. В этом случае Θ_k представляют собой независимые одинаково распределенные l -мерные случайные векторы, координаты которых являются независимыми дискретными случайными величинами, принимающими значения из множества $\{1, 2, 3, \dots, l\}$ с равными вероятностями. Реализация случайного вектора Θ_k определяет номера прецедентов обучающей выборки, которые образуют бутстреп выборку, используемую при построении k -го дерева решений.

Исследование состоятельности случайных лесов, т. е. вопросов сходимости их решений при неограниченном возрастании объема обучающей выборки, было начато автором метода. В работе [24] исследовались общие вопросы сходимости решений ансамблей классификаторов. Для случайных лесов регрессий исследовалась среднеквадратичная сходимость. Именно, обозначим через $M(Y|\mathbf{X} = \mathbf{x})$ условное математическое ожидание непрерывного отклика при данном значении вектора признаков $\mathbf{X}$. Среднеквадратичная сходимость означает, что

$$\lim_{l \rightarrow \infty} |\hat{f}_{rf}(\mathbf{x}) - M(Y|\mathbf{X} = \mathbf{x})|^2 = 0. \quad (12)$$

В случае классификации состоятельность означает сходимость риска (вероятности ошибочной классификации) к байесовскому риску, т. е. риску оптимального байесовского классификатора.

Брейман отмечал, что хотя алгоритм индукции случайных лесов кажется простым, его теоретическая модель сложна для анализа. Поэтому в работе [27] рассматривалась упрощенная модель случайного леса, основанная на предположении, что шаг формирования бутстреп выборок отсутствует и каждая терминальная вершина содержит ровно одно наблюдение. В рамках этой модели Брейманом было показано, что случайный лес регрессий состоятелен и количество признаков, случайно выбираемых для расщепления, не зависит от количества наблюдений. Также для случая двух классов была показана состоятельность соответствующего классификатора. Анализ Бреймана существенно опирался на работу [70], в которой была установлена связь между случайными лесами и адаптивным методом ближайших соседей.

В работе [19] исследовался вопрос состоятельности модели случайного леса, близкой к модели, рассматривавшейся Брейманом [27]. Основной результат работы состоит в том, что модель состоятельна и скорость сходимости зависит только от количества "сильных" признаков и не зависит от количества иррелевантных ("шумовых") признаков.

В работах [17, 18] получен ряд результатов относительно состоятельности нескольких моделей случайных лесов — чисто случайного леса (purely random forest) и инвариантного к масштабу случайного леса (scale-invariant random forest). В основу работы положен результат, полученный авторами, о состоятельности

ансамблей классификаторов, основанных на процедуре усреднения рандомизированных базовых классификаторов. Суть результата состоит в том, что для состоятельности ансамбля классификаторов необходимо и достаточно, чтобы базовый классификатор был состоятелен.

СЛУЧАЙНЫЕ ЛЕСА КАК ИНСТРУМЕНТ СТАТИСТИЧЕСКОГО АНАЛИЗА

Основной недостаток случайных лесов по сравнению с деревьями решений состоит в отсутствии визуального представления процесса принятия решений и сложности интерпретации их решений. Однако предложенные Брейманом меры информативности признаков⁴ и возможность использовать случайные леса для построения матрицы близости (proximity) наблюдений более чем компенсируют этот недостаток. Меры информативности дают возможность выделения наиболее информативных признаков, что является одной из важнейших задач статистического анализа. Матрица близости дает возможность применения методов, непосредственно не связанных с классификацией и регрессией, таких как многомерное шкалирование, кластеризация, определение прототипов (т. е. наиболее типичных представителей) классов и выявление аномальных наблюдений. Отмеченные возможности во многом способствовали усилению популярности случайных лесов.

Наиболее используемой сферой применимости случайных лесов (помимо классификации и регрессии) является задача выделения наиболее информативных признаков. Брейманом было предложено 4 меры информативности признаков [25].

Меры информативности признаков, основанные на случайных лесах

Пусть x_i некоторый признак. Три первые меры информативности основаны на оценке влияния случайной перестановки значений этого признака в ООВ выборках на результаты классификации.

Первая мера вычисляется следующим образом:

1. Построить случайный лес и получить оценку вероятности ошибочной классификации e методом ООВ;
2. В ООВ выборках для каждого дерева из построенного случайного леса произвести

⁴Брейман использовал термин "меры важности" (importance measures); автор же придерживается терминологии, принятой в прикладной статистике.

случайную перестановку значений признака x_i ;

3. Получить оценку вероятности ошибочной классификации $\hat{e}_i$ по модифицированным ООВ выборкам;
4. Определить информативность признака x_i как $I_1(x_i) = \max_i(0, \hat{e}_i - e_i)$.

Вторая и третья меры используют понятие отступа (*margin*). Пусть $(\mathbf{x}, y)$ — элемент обучающей выборки. Отступ $marg(\mathbf{x}, y)$ определяется как разность между долей деревьев в лесу, правильно классифицирующих, $(\mathbf{x}, y)$ и максимумом из долей деревьев, относящих $(\mathbf{x}, y)$ в другие классы. В результате перестановки значений признака x_i отступ уменьшается. В качестве меры информативности берется среднее значение (по всем наблюдениям) уменьшения отступа, т. е.

$$I_2(x_i) = \frac{1}{l} \sum_{j=1}^l (marg(\mathbf{x}_j, y_j) - marg_i(\mathbf{x}_j, y_j)),$$

где $marg_i(\mathbf{x}, y)$ означает отступ, вычисленный по ООВ выборке со случайной перестановкой значений признака x_i .

Мера 3 равна разности между количеством отступов, которые уменьшились, и количеством отступов, которые увеличились в результате перестановки значений признака x_i :

$$I_3(x_i) = \frac{\#[marg(\mathbf{x}, y) > marg_i(\mathbf{x}, y)] - \#[marg(\mathbf{x}, y) < marg_i(\mathbf{x}, y)]}{\#[marg(\mathbf{x}, y) > marg_i(\mathbf{x}, y)] + \#[marg(\mathbf{x}, y) < marg_i(\mathbf{x}, y)]}. \quad (13)$$

Мера 4 определяется как среднее уменьшение загрязненности вершин, обусловленное данным признаком, именно

$$I_4(x_i) = \frac{1}{k} \sum_t \Delta i(t) I(i, t),$$

где суммирование осуществляется по всем вершинам деревьев случайного леса, $\Delta i(t)$ — уменьшение загрязненности в вершине t , и $I(i, t)$ — индикаторная функция, равная 1, если признак x_i был выбран для расщепления в вершине t .

Несмотря на многочисленные исследования как теоретического, так и экспериментального характера [51, 101–103], пока неизвестно, какая из предложенных мер предпочтительнее в той или иной ситуации. В работе Бреймана [26] обсуждаются некоторые вопросы, касающиеся различия между мерами информативности и приведен ряд примеров их применения. Подробное изложение вопросов, связанных с мерами информативности, содержится в работе [51].

В работе [101] показано, что описанные выше меры важности признаков не очень надежны в ситуациях, когда признаки измерены в различных шкалах или сильно варьируют по количеству уровней (для признаков, измеренных в номинальной или порядковой шкале). В этой работе предложено использовать другой алгоритм построения случайного леса, для которого соответствующие меры важности признаков "хорошо работают" в описанных выше ситуациях, а именно, использовать для формирования обучающих выборок используемых для построения деревьев решений не выборку с возвращением (баггинг), а выборку без возвращения объемом равным 0,62 объема исходной выборки. Такой метод получил название подбаггинг (*subbagging*).

График частной зависимости

Для случайных лесов регрессий возможно построение графиков частных зависимостей отклика от некоторых независимых переменных [55]. Пусть X_S — подвектор входных признаков $x_1, x_2, \dots, x_n$ с индексами из множества $S \subset \{1, 2, \dots, n\}$ и C — дополнительное ему множество $S \cup C = \{1, 2, \dots, n\}$. Оценка функции частной (маргинальной) зависимости отклика от значений подвектора входных признаков X_S определяется как

$$f_S(x_S) = \frac{1}{l} \sum_{i=1}^l f_S(x_S, x_{iC}).$$

Эта функция (и соответствующий график) позволяют выделить эффект воздействия подвектора входных признаков X_S на отклик после удаления эффекта воздействия X_C .

Построение матрицы близости

Индукция случайного леса может служить промежуточным этапом для построения матрицы близости (*proximity*) наблюдений [26]. Элемент (i, j) матрицы близости равен доле деревьев в лесу, таких, что элементы $\mathbf{x}_i$ и $\mathbf{x}_j$ классифицируются в одну терминальную вершину (лист). Ясно, что элементы, часто попадающие в одну терминальную вершину, "подобны" (близки) в некотором смысле. Матрица близости затем может быть использована для:

1. Кластеризации наблюдений;
2. Многомерного шкалирования;
3. Нахождения прототипов классов;
4. Выявления аномальных наблюдений.

Кластеризация наблюдений осуществляется с помощью генерации искусственных данных. Именно исходная выборка, подлежащая кластеризации, образует первый класс, а наблюдения второго класса генерируются искусственным образом. Генерация может осуществляться двумя способами:

- генерируются независимые бутстреп-выборки для каждого признака отдельно;
- значения каждого признака генерируются в соответствии с равномерным распределением на области его значений в исходной выборке.

После формирования такой двухклассовой выборки осуществляется построение случайного леса. Суть подхода состоит в том, что близкие точки исходной выборки будут часто попадать в одинаковые терминальные вершины и поэтому матрица близости может быть использована для кластеризации. Некоторые теоретические результаты такого подхода можно найти в [96], а практические примеры его использования в [26, 69].

Определение прототипов классов производится следующим образом [26]. Для каждого класса находится наблюдение, имеющее наибольшее количество наблюдений этого класса среди своих k ближайших соседей (определяемых по матрице близости). Затем для каждого признака находится медиана этих k ближайших соседей. Эти медианы и считаются прототипами рассматриваемого класса.

Мера аномальности наблюдения определяется как величина, обратная сумме квадратов близостей между этим наблюдением и другими наблюдениями того же класса.

Вопросы методического и практического применения рассмотренных выше методов можно найти в [26, 28].

РАЗНОВИДНОСТИ СЛУЧАЙНЫХ ЛЕСОВ

Случайные леса выживаемости

Обобщением случайных лесов на случай данных, цензурированных справа, важным примером которых являются данные о выживаемости, являются случайные леса выживаемости (Random Survival Forests). С их помощью возможно построение непараметрических регрессионных кривых выживаемости, не предполагающих пропорциональности риска. В R случайные леса выживаемости реализованы в пакете **randomSurvivalForest** [62, 63].

Квантильные леса регрессий

В случае непрерывного отклика случайные леса являются хорошим методом построения непараметрической регрессии, т. е. дают хорошее приближение к условному среднему отклика. В работе [73] предложено важное обобщение – квантильные леса регрессий (Quantile Regression Forests), которые дают существенно более полную картину условного распределения отклика, а именно, возможность непараметрической оценки квантилей условного распределения. Это позволяет, в частности, использовать их для построения доверительных интервалов и обнаружения аномальных значений отклика. В R квантильные леса регрессий реализованы в пакете **quantregForest** [74].

Логические леса

В работе [117] предложен метод построения ансамблей классификаторов, основанных на логических деревьях решений. Теоретической основой является логическая регрессия [88, 89], предполагающая, что все признаки, включая классовой, являются бинарными. Предсказание отклика в логической регрессии осуществляется с использованием логических комбинаций бинарных признаков, позволяя строить хорошо интерпретируемые модели сложных взаимодействий признаков. Однако при наличии шума возможности логической регрессии падают. Кроме классификации логические леса могут использоваться для выявления наиболее информативных признаков и их взаимодействий. В R логические леса реализованы в пакете **logicForest** [118].

Вероятностные случайные леса

Оценка вероятности ошибочной классификации случайных лесов, даваемая методом ООВ, является усредненной оценкой [23, 25]. В работе [29] предложен метод и алгоритм построения так называемых вероятностных случайных лесов (Probabilistic Random Forests), с помощью которых возможно получить оценку вероятности ошибочной классификации конкретного наблюдения в случае бинарной классификации. Matlab и C коды, реализующие алгоритм, могут быть загружены с сайта <http://ucsu.colorado.edu/breitenm/>.

Потоковые случайные леса

Многие приложения, такие как анализ телекоммуникационных данных, финансовые приложения, анализ интернет протоколов, электронной почты и т. д. основаны на потоках

данных, представляющих собой записи некоторых событий (причем о некоторых известна их принадлежность определенному классу, а о некоторых нет), изменяющихся во времени (и часто с высокой скоростью). В такой постановке классические алгоритмы распознавания образов не могут быть применены непосредственно и требуют определенной (не всегда очевидной) модификации. Основными требованиями к такой модификации являются высокая скорость обработки потока данных, инкрементальность (пошаговость) обучения (т. е. каждая запись должна обрабатываться не более одного раза) и адаптация к текущему распределению событий в потоке. Для случайных лесов одна из таких модификаций, ориентированная на задачи визуального слежения (visual tracking), была предложена в работе [93]. Программная реализация этого алгоритма в пакете "Online Multiclass LPBoost" может быть загружена с сайта <http://www.ymer.org/amir/software/online-random-forests/>.

Другая модификация, ориентированная на задачи Data Mining и получившая название потоковые случайные леса (Streaming Random Forests), была предложена и разрабатывалась в работах [10, 11].

Случайный наивный Байесовский классификатор и случайная мультиномиальная логит модель

Экспериментальные исследования и практика применения случайных лесов показали, что метод обладает рядом хороших качеств – высокой точностью, робастностью по отношению к зашумленности обучающей выборки и отсутствием переобучения. В связи с этим естественно предположить, что основные элементы алгоритма Бреймана – формирование подвыборок методом баггинга и случайный отбор признаков на стадии расщепления вершин деревьев решений могут быть успешно использованы и для некоторых других базовых классификаторов. Эта идея была реализована в работе [84]. В этой работе в качестве базовых классификаторов рассматривались наивный Байесовский классификатор и мультиномиальная логит модель. Авторами проведен ряд экспериментов по сравнению точности классификации соответствующих ансамблей с точностью случайных лесов и машин опорных векторов, которые показали, что случайный наивный Байесовский классификатор и случайная мультиномиальная логит модель в определенных ситуациях превосходят эти методы.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНДУКЦИИ ДЕРЕВЬЕВ РЕШЕНИЙ И СЛУЧАЙНЫХ ЛЕСОВ

Деревья решений и случайные леса реализованы во многих программных средствах как коммерческого назначения, так и свободно доступных. К свободно доступным относится программная система WEKA [53, 116], разрабатываемая с 1993 г. группой машинного обучения университета Вайкато (Новая Зеландия). Сайт проекта находится по адресу <http://www.cs.waikato.ac.nz/ml/weka> (дата обращения 15.04.2011), с которого можно загрузить систему, руководства по работе с ней и соответствующие публикации. Отметим, что WEKA содержит также большой выбор неркурсивных процедур распознавания образов и имеет интерфейс с пакетом R, называемый RWEKA [59]. Другой свободно доступной системой является система RapidMiner; сайт соответствующего проекта находится по адресу <http://www.rapidminer.com> (дата обращения 15.04.2011). В среде MATLAB имеется интерфейс к алгоритму, реализующему метод случайных лесов [113]. В работе [64] предложен свободно доступный C++ код, ориентированный на использование случайных лесов для задач кластеризации. Наиболее широкий спектр процедур, содержащий реализации не только классических алгоритмов CART для деревьев решений и алгоритма Бреймана для случайных лесов содержится в пакетах, разработанных для среды R [85] – свободно распространяемом программном обеспечении для статистических вычислений и графики, доступном на платформах Windows, Linux, Mackintosh. Пакет R может быть загружен с сайта <http://cran.r-project.org/>, где также имеется руководство по инсталляции и руководства по работе с пакетом. Полезные материалы по пакету R на русском языке могут быть найдены по адресу <http://herba.msu.ru/shipunov/software/r/r-ru.htm>. Ниже приводятся описания практически всех пакетов, реализующих описанные выше методы и имеющиеся на момент написания настоящего обзора. Эти пакеты также могут быть загружены с сайта <http://cran.r-project.org/>. Нет сомнений, что их количество будет увеличиваться. Все пакеты, разработанные для среды R, оформлены по единому стандарту – содержат описания всех содержащихся в них процедур и хорошо продуманные примеры их применения, которые могут служить в качестве шаблонов. Ряд пакетов, кроме того, описан в журнале R

Journal (до 2005 г. R News). В частности, один из номеров этого журнала содержит краткий обзор применения метода случайных лесов в среде R [99]. Ниже кратко описаны эти пакеты.

Пакеты **tree** [86] и **rpart** [106] содержат процедуры рекурсивного разбиения для решения задач классификации, регрессии и анализа выживаемости, основанных на идеях классической работы [21]. Адаптация процедур пакета **rpart** на случай многомерного отклика содержится в пакете **mvpart** [36, 37]. Имеется также вспомогательный пакет **maptree** [114], содержащий полезные процедуры для улучшенного графического представления и усечения дендрограмм, деревьев классификации и регрессий, построенных с использованием пакетов **cluster** [72] и **rpart**.

Пакет **rpartOrdinal** [13] содержит процедуры построения деревьев решений, ориентированные на случай, когда классовая переменная измерена в порядковой шкале.

Одним из наиболее мощных пакетов, реализующих алгоритмы рекурсивного разбиения для построения деревьев решений и регрессий, является пакет **party** [60]. Он включает возможность построения деревьев условного вывода (conditional inference trees), лесов условного вывода (conditional inference forests), основанных на теории условного вывода [60], а также процедуры рекурсивного разбиения для параметрических моделей (линейной и обобщенной линейной модели). Допускается использование признаков, измеренных в различных шкалах: числовой, номинальной и порядковой, а также цензурированные и многомерные отклики. Пакет содержит процедуры улучшенного представления бинарных деревьев и распределения отклика в вершинах деревьев. Ряд вопросов, касающихся применения пакета, рассмотрен в работе [104].

Процедуры построения косоугольных деревьев решений содержатся в пакете **oblique.tree** [109].

Пакет **Margin.tree** [108] реализует иерархическую версию метода опорных векторов для построения деревьев решений. Особенно полезен для обработки обучающих выборок с количеством классов большим 2 и когда количество признаков превышает количество наблюдений. Его применение описано в [107].

Пакет **ipred** [82] содержит процедуры построения деревьев решений для задач непрямой (т. е. использующей некоторые типы априорной информации) классификации, регрессии и анализа выживаемости. В пакете ре-

ализованы процедуры получения улучшенных оценок вероятности ошибочной классификации и предсказания. Примеры применения этого пакета для решения задач классификации в медицине содержатся в [83].

Основным пакетом, реализующим классический алгоритм случайных лесов Бреймана (и процедуры которого используются в некоторых других пакетах), является пакет **randomForest** [69]. Возможности пакета и вопросы его методического и практического применения рассмотрены в работах [69, 99].

Пакеты **varSelRF** [38, 39] и **Boruta**: [67] содержат процедуры отбора множества наиболее информативных признаков с использованием случайных лесов. В частности, пакет **varSelRF** включает процедуры, основанные на методе обратного исключения переменных и на спектре важности переменных, ориентирован на данные высокой размерности и предусматривает возможность организации параллельных вычислений. Пакет **Boruta** ориентирован на нахождение наиболее информативных признаков в информационных системах. Процедура поиска информативных признаков основана на оригинальном алгоритме, предусматривающем итеративное построение случайных лесов.

Пакет **quantregForest** [73] содержит процедуры построения квантильных лесов регрессий.

Случайные леса выживаемости реализованы в пакете **randomSurvivalForest** [62]. Вопросы применения пакета рассмотрены в [63].

Пакет **LogicForest** [118] содержит процедуры построения логического леса для выявления логических соотношений между откликом и независимыми переменными. Предполагается, что все признаки обучающей выборки являются бинарными.

Пакет **Gbevc** [94] ориентирован на построение ансамблей деревьев регрессий методом бутстинга для данных с ошибками измерений в независимых переменных.

Пакет **ModelMap** [46] содержит процедуры детального картирования для данных, представленных в виде изображений, основанные на методе случайных лесов и модели стохастического градиентного бутстинга.

ЗАКЛЮЧЕНИЕ

Говоря о периоде, прошедшем со времени возникновения метода случайных лесов, необходимо отметить, что он не только состоялся как метод классификации и регрессии и широко применяется в различных предметных областях, но и послужил основой для разработки

различных его модификаций, важных для приложений. Кроме того, он дал импульс для разработки методов, в основе которых лежит близкая схема, не использующая деревья решений в качестве базового классификатора. Нет сомнений, что в будущем это направление будет развиваться. Представляется, что потенциал метода не исчерпан и с точки зрения решения других статистических задач — нахождения наиболее информативных переменных, кластеризации и т. д. В связи с этим автор считает, что использование этого метода в научных учреждениях различного профиля нашей страны (в частности, в институтах Карельского научного центра РАН) для анализа статистических данных было бы весьма полезно.

Автор выражает признательность Ю. Л. Павлову за предложение написать данный обзор и полезные замечания.

ЛИТЕРАТУРА

1. Айвазян С. А., Бухштабер В. Н., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерностей. Справочное издание / Ред. Айвазян С. А. // Финансы и статистика. М., 1989. С. 607.
2. Воронцов К. В. О проблемно-ориентированной оптимизации базисов задач распознавания // ЖВМ и МФ. 1998. Т. 38, № 5. С. 870–880. URL: <http://www.ccas.ru/frc/papers/voron98jvm.pdf>. (дата обращения 15.04.2011).
3. Воронцов К. В. Оптимизационные методы линейной и монотонной коррекции в алгебраическом подходе к проблеме распознавания // ЖВМ и МФ. 2000. Т. 40, № 1. С. 166–176. URL: <http://www.ccas.ru/frc/papers/voron00jvm.pdf>. (дата обращения 15.04.2011).
4. Воронцов К. В. Обзор современных исследований по проблеме качества обучения алгоритмов // Таврический вестник информатики и математики. 2004. Р. 1–20. URL: www.ccas.ru/frc/papers/voron04twim.pdf (дата обращения 15.04.2011).
5. Колчин В. Ф. Случайные отображения. М.: Наука, 1984. 208 с.
6. Павлов Ю. Л. Предельные теоремы для числа деревьев заданного объема в случайном лесе // Математический сборник. 1977. Т. 103, № 3. С. 392–403.
7. Павлов Ю. Л. Асимптотическое распределение максимального объема дерева в случайном лесе // Теория вероятностей и ее применения. 1977. Т. 22, № 3. С. 523–533.
8. Павлов Ю. Л. Случайный лес // Вероятность и математическая статистика. М.: Изд.
9. Рудаков К. В., Воронцов К. В. О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания // Доклады РАН. 1999. Т. 367, № 3. С. 314–317. URL: <http://www.ccas.ru/frc/papers/rudvoron99dan.pdf> (дата обращения 15.04.2011).
10. Abdulsalam H., Skillicorn D. B., Martin P. Streaming Random Forests // Proceedings of the 11th International Database Engineering and Applications Symposium (IDEAS), (September 2007). 2007. P. 225–232.
11. Abdulsalam H., Skillicorn D. B., Martin P. Classification Using Streaming Random Forests // IEEE Transactions on Knowledge and Data Engineering. 2011. Vol. 23, N. 1. P. 22–36.
12. Amit H., Geman D. Shape quantization and recognition with randomized trees // Neural Computation. 1997. Vol. 9. P. 1545–1588.
13. Archer K. J. rpartOrdinal: An R Package for Deriving a Classification Tree for Predicting an Ordinal Response // Journal of Statistical Software. 2010. Vol. 34, N 7. P. 1–17. URL: <http://www.jstatsoft.org/v34/i07/> (дата обращения 15.04.2011).
14. Banfield R. E., Hall L. O., Bowyer K. W., Kegelmeyer W. P. A Comparison of Decision Tree Ensemble Creation Techniques // IEEE Trans. Pattern analysis and Machine Intelligence. 2007. Vol. 29, N 1. P. 173–180.
15. Behlmann P., Yu B. Analyzing bagging // Annals of Statistics. 2002. Vol. 30. P. 927–961.
16. Bennet K. P. Decision tree construction via linear programming // Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society Conference. (Utica, Illinois 1992). 1992. P. 97–101.
17. Biau G., Devroye L., Lugosi G. Consistency of random forests and other averaging classifiers // Journal of Machine Learning Research. 2008. Vol. 9. P. 2015–2033.
18. Biau G., Devroye L. On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification // Journal of Multivariate Analysis. 2010. Vol. 101. P. 2499–2518. URL: <http://www.lsta.upmc.fr/BIAU/bd4.pdf> (дата обращения 15.04.2011).
19. Biau G. Analysis of a Random Forests Model // Technical report, University Paris 6. 2010. P. 1–31. URL: <http://hal.archives-ouvertes.fr/docs/00/47/65/45/PDF/article2.pdf> (дата обращения 15.04.2011).

20. Borra S., Ciaccio A. Improving nonparametric regression methods by bagging and boosting // *Computational Statistics and Data Analysis*. 2002. Vol. 38, N 4. P. 407–420.
21. Breiman L., Friedman R., Olshen R., Stone C. *Classification and Regression Trees*. Belmont, California: Wadsworth International, 1984. 342 p.
22. Breiman L. Bagging predictors // *Machine Learning*. 1996. Vol. 24, N 2. P. 123–140.
23. Breiman L. Out-of-bag estimation // Technical report, Statistics Department University of California, Berkeley. 1996. P. 1–13. URL: <ftp://ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps.Z> (дата обращения 15.04.2011).
24. Breiman L. Some infinite theory for predictor ensembles // Technical Report 577, Statistics Department University of California, Berkeley, 2000. P. 1–30. URL: <http://www.stat.berkeley.edu/breiman> (дата обращения 15.04.2011).
25. Breiman L. Random forests // *Machine Learning*. 2001. Vol. 45, N 1. P. 5–32.
26. Breiman L. Manual on setting up, using, and understanding random forests v3.1. 2002. URL: http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf.
27. Breiman L. Consistency for a simple model of random forests // Technical Report 670, Statistics Department, UC Berkeley. 2004. P. 1–10. URL: <http://www.stat.berkeley.edu/breiman> (дата обращения 15.04.2011).
28. Breiman L., Cutler A. *Random Forests*. Berkeley. 2005. 56 p. URL: <http://www.stat.berkeley.edu/users/breiman/RandomForests>
29. Breitenbach M., Grudic G., Nielsen R. Probabilistic Random Forests: Predicting Data Point Specific Misclassification Probabilities // *Machine Learning*. 2009. Vol. 23, N 2. P. 48–73.
30. Brodley C. E., Utgoff P. E. Multivariate decision trees // *Machine Learning*. 1995. Vol. 19, N 1. P. 45–77.
31. Bylander T. Estimating generalization error on twoclass datasets using outofbag estimates // *Machine Learning*. 2002. Vol. 48. P. 287–297.
32. Cherkauer K. G. Human expert-level performance on a scientific image analysis task by a system using combined artificial neural networks / Eds. P. Chan // *Working Notes on the AAAI Workshop on Integrating Multiple Learned Models*. 2006. P. 15–21. URL: <http://www.cs.fit.edu/imlm> (дата обращения 15.04.2011).
33. Cios K. G., Sztandera L. M. Continuous ID3 algorithm with fuzzy entropy measures // *Proceedings of IEEE International Conference on Fussy Systems*. 1992. P. 469–476.
34. Cutler D. R., Edwards T. C. et al. Random forests for classification in ecology // *Ecology*. 2007. Vol. 88, N. 11. P. 2783–2792.
35. Cutler A., Breiman L. RAFT: RANdom Forest Tool. URL: <http://www.stat.berkeley.edu/users/breiman/RandomForests/> (дата обращения 15.04.2011).
36. De'ath G. Multivariate Regression Trees: A New Technique for Constrained Classification Analysis // *Ecology*. 2002. Vol. 83, N 4. P. 1103–1117.
37. De'ath G. mvpart: Multivariate partitioning // R package version 1.31. 2010. URL: <http://CRAN.Rproject.org/package=mvpart> (дата обращения 15.04.2011).
38. Diaz-Uriarte R. varSelRF: Variable selection using random forests // R package version 0.71. 2009. URL: <http://ligarto.org/rdiaz/Software/Software.html> (дата обращения 15.04.2011).
39. Diaz-Uriarte R., Andrr'es S. Gene Selection and classification of microarray data using random forest // *BMC Bioinformatics*. Vol. 7, N 3. P. 1–13.
40. Dietterich T. G., Kong E. B. Machine learning bias, statistical bias, and statistical variance of decision tree algorithms // Technical report, Department of computer science Oregon State University. 1995. P. 1–22. URL: <ftp://ftp.cs.orst.edu/pub/tgd/papers/tr-bias.ps.gz> (дата обращения 15.04.2011).
41. Dietterich T. G. An Experimental Comparison of Three Methods for Constructing Ansembles of Decision Trees: Bagging, Boosting, and Randomization // *Machine Learning*. 1999. P. 1–20.
42. Dietterich T. G. Ensemble methods in machine learning // *First International Workshop on Multiple Classifier Systems*. Lecture Notes in Computer Science. New York: Springer, 2000. P. 1–15.
43. Drucker H., Cortes C. Boosting decision trees / Eds. D. Touretsky, M. Mozer et al. // *Advances in Neural Information Processing Systems*. Cambridge: MA. MIT Press, 1996. Vol. 8. P. 479–485.
44. Duda R. O., Hart P. E., Stork D. G. *Pattern Classification*. NY.: John Wiley Sons, 2001. 639 p.
45. Esposito F., Malerba D., Semeraro G. A comparative analysis of methods for pruning decision trees // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997. Vol. 19, N 5. P. 476–491.
46. Freeman E., Frescino T. ModelMap: Modeling and Map production using Random

- Forest and Stochastic Gradient Boosting // USDA Forest Service, Rocky Mountain Research Station, 507 25th street, Ogden, UT, USA. 2009. P. 134–156.
47. *Freund Y., Shapire R. E.* Experiments with a new boosting algorithm // Proceedings of the 13rd International Conference on Machine Learning. P. Morgan Cauffman. 1996. P. 146–148.
48. *Freund Y., Shapire R. E.* Discussion of the paper “Arcing classifiers” by Leo Breiman // The Annals of Statistics. 1998. Vol. 26, N 3. P. 824–832.
49. *Furnkranz J.* Pruning algorithms for rule learning // Machine Learning. 1997. Vol. 27. P. 139–172.
50. *Gehrke J., Ramakrishnan R., Ganti V.* Rainforest a framework for fast decision tree construction of large datasets // Data Mining and Knowledge Discovery. 2000. Vol. 4, N 2/3. P. 127–162.
51. *Genuer R., Poggi J.-M., Tuleau C.* Random Forests: Some methodological insights // Research Report 6729, INRIA Saclay-Ile-de-France. 2008. P. 1–35. URL: <http://hal.inria.fr/inria-00340725/fr> (дата обращения 15.04.2011).
52. *Genuer R.* Risk bounds for purely uniformly random forests // Technical report 7318, INRIA. 2010. P. 1–22. URL: <http://www.math.u-psud.fr/~genuer/genuer.purf.pdf> (дата обращения 15.04.2011).
53. *Hall M., Frank E., Holmes G. et al.* The WEKA Data Mining Software: An Update // SIGKDD Explorations. 2009. Vol. 11, N 1. P. 1–9.
54. *Hansen L., Salamon P.* Neural Network Ensembles // IEEE Trans. Pattern analysis and Machine Intelligence. 1990. Vol. 12. P. 993–1001.
55. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. Springer, 2001. 533 p.
56. *Ho T. K.* Random Decision Forests // Proceedings of the 3rd International Conference on Document Analysis and Recognition, (Montreal, Canada 1995), 1995. P. 278–282.
57. *Ho T. K.* C4.5 Decision Forests // Proceedings of the 14th International Conference on Pattern Recognition. (Brisbane, Australia, 1998). 1998. P. 17–20.
58. *Ho T. K.* The Random Subspace Method for Constructing Decision Forests // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998. Vol. 20, N 8. P. 832–844.
59. *Hornik K., Buchta C., Zeileis A.* Open-Source Machine Learning: R Meets Weka // Computational Statistics. 2009. Vol. 24, N. 2. P. 225–232.
60. *Hothorn T., Hornik K., Zeileis A.* Unbiased Recursive Partitioning: A Conditional Inference Framework // Journal of Computational and Graphical Statistics. 2006. Vol. 15, N 3. P. 651–674. URL: <http://statmath.wu.wien.ac.at/~zeileis/papers/Hothorn+Hornik+Zeileis2006.pdf>. (дата обращения 15.04.2011).
61. *Hothorn T., Buehlmann P., Dudoit S. et al.* Survival Ensembles // Biostatistics. 2006. Vol. 7, N 3. P. 355–373.
62. *Ishwaran H., Kogalur U.* Random survival forests for R // R News. 2007. Vol. 7, N 2. P. 25–31.
63. *Ishwaran H., Kogalur U., Blackstone E., Lauer M.* Random survival forests // Ann. Appl. Statist. 2008. Vol. 2, N 3. P. 841–860.
64. *Karpievitch, Y. V., Leclerc, A. P., Hill, E. G., Almeida, J. S.* RF++: Improved Random Forest for Clustered Data Classification. URL: <http://www.ohloh.net/p/rfpp>. (дата обращения 15.04.2011).
65. *Kooperberg C., Ruczinski I., LeBlanc M., Hsu L.* Sequence Analysis using Logic Regression // Genetic Epidemiology. 2001. Vol. 21. P. 626–631. URL: <http://kooperberg.fhcrc.org/logic/documents/ingophdlogic.pdf> (дата обращения 15.04.2011).
66. *Kuncheva L. I.* Combining Pattern Classifiers: Methods and Algorithms / Hoboken, New Jersey: John Wiley Sons, 2004. 349 p.
67. *Kursa M. B., Rudnicki W. R.* Feature Selection with the Boruta Package // Journal of Statistical Software. 2010. Vol. 36, N 11. P. 1–13. URL: <http://www.jstatsoft.org/v36/i11/> (дата обращения 15.04.2011).
68. *Kwok S. W., Carter C.* Multiple decision trees // Uncertainty in Artificial Intelligence. 1990. Vol. 4. P. 327–335.
69. *Liaw A., Wiener M.* Classification and Regression by randomForest // R News. 2002. Vol. 2, N 3. P. 18–22. URL: <http://CRAN.R-project.org/doc/Rnews/>
70. *Lin Y., Jeon Y.* Random forests and Adaptive Nearest Neighbors // Technical report 1055. Dept. Statistics, Univ. Wisconsin, 2002. P. 1–31.
71. *Lin Y., Jeon Y.* Random forests and adaptive nearest neighbors // Journal of the American Statistical Association. 2006. Vol. 101. P. 578–590.
72. *Maechler M.* cluster: Cluster Analysis Extended Rousseeuw et al. // R package version 1.13.3. 2011. URL: <http://CRAN.Rproject.org/package=cluster> (дата обращения 15.04.2011).

73. *Meinshausen N.* Quantile Regression Forests // Journal of Machine Learning Research. 2006. Vol. 7. P. 983–999.
74. *Meinshausen N.* quantregForest: Quantile Regression Forests // R package version 0.22. 2007. URL: <http://www.stat.berkeley.edu/~nicolai> (дата обращения 15.04.2011).
75. *Mingers J.* Expert systems experiments with rule induction // Journal of the Operational Research Society. 1987. Vol. 38. P. 39–47.
76. *Mingers J.* An empirical comparison of pruning methods for decision tree induction // Machine Learning. 1989. Vol. 4. P. 227–243.
77. *Murthy S. K.* Automatic Construction of Decision Trees from Data: A MultiDisciplinary Survey // Data Mining and Knowledge Discovery. 1998. Vol. 2. P. 345–389.
78. *Murthy S., Kasif F., Salzberg S., Beigel R.* OCI: Randomized induction of oblique decision trees // Proceedings of the Eleventh National Conference on Artificial Intelligence, Mit Press, Boston. 1993. P. 322–327.
79. *Niblett T., Bratko I.* Learning decision rules in noisy domains / Eds. M. A. Bramer // Research and development in Expert Systems. Cambridge University Press, 1986. P. 25–34.
80. *Parmanto B., Munro P. W., Doyle H. R.* Improving committee diagnosis with resampling technique // Advances in Neural Information Processing Systems. 1996. Vol. 8. P. 882–888.
81. *Pavlov Yu. L.* Random Forests. Utrecht, Boston, Koln, Tokyo: VSP, 2000. 122 p.
82. *Peters A., Hothorn T.* ipred: Improved Predictors // R package version 0.8-11. 2011. URL: <http://CRAN.Rproject.org/package=ipred> (дата обращения 15.04.2011).
83. *Peters A., Hothorn T., Lausen B.* ipred: Improved Predictors // R News. 2002. Vol. 2, N 2. P. 33–36. URL: <http://CRAN.Rproject.org/doc/Rnews/> (дата обращения 15.04.2011).
84. *Prinzie A., Poel D.* Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB // Working paper, Department of Marketing, Ghent University, 2007. P. 1–12.
85. *R Development Core Team* R: A language and environment for statistical computing // R Foundation for Statistical Computing: Vienna, Austria, ISBN 3-900051-07-0, URL: <http://www.R-project.org> (дата обращения 15.04.2011).
86. *Ripley B. D.* tree: Classification and regression trees // R package version 1.026. 2007. URL: <http://CRAN.Rproject.org/package=tree> (дата обращения 15.04.2011).
87. *RobnikSikonja M.* Improving Random Forests / Eds. J. F. Boulicaut et al. ECML 2004, LNAI 3210, Berlin: Springer, 2004. P. 359–370. URL: <http://lkm.fri.unilj.si/rmarko/papers/> (дата обращения 15.04.2011).
88. *Ruczinski I., Kooperberg C., LeBlanc M.* Logic Regression methods and software / Eds: D. Denison et al. // Proceedings of the MSRI workshop on Nonlinear Estimation and Classification. New York: Springer, 2002. P. 333–344.
89. *Ruczinski I., Kooperberg C., LeBlanc M.* Logic Regression // Journal of Computational and Graphical Statistics. 2003. Vol. 12. P. 475–511.
90. *Quinlan J. R.* Learning efficient classification procedures and their application to chess end games / Eds. R. S. Michalski et al. // Machine Learning: An artificial intelligence approach. San Francisco: Morgan Kaufmann, 1983. P. 463–482.
91. *Quinlan J. R.* Simplifying decision trees // International Journal of ManMachine Studies. 1987. Vol. 27. P. 221–234.
92. *Quinlan J. R.* C4.5 Programs for Machine Learning. San Mateo, California: Morgan Kaufmann, 1993.
93. *Saffari A., Leistner C., Jakob Santner J. et al.* On-line Random Forests // 3rd IEEE ICCV Workshop on On-line Computer Vision, 2009. P. 112–127.
94. *Sexton J.* gbev: Gradient Boosted Regression Trees with Errors in Variables / R package version 0.1.1. 2009. URL: <http://CRAN.Rproject.org/package=gbev> (дата обращения 15.04.2011).
95. *Shapire R., Freund Y., Bartlett P., Lee W.* Boosting the margin: A new explanation for the effectiveness of voting methods // Annals of Statistics. 1998. Vol. 26, N 5. P. 1651–1686.
96. *Shi T., Horvath S.* Unsupervised Learning with Random Forest Predictors // Journal of Computational and Graphical Statistics. 2006. Vol. 15, N 1. P. 118–138.
97. *Shlien S.* Multiple Binary Decision Tree Classifiers // Pattern Recognition. 1990. Vol. 23, N 7. P. 757–763.
98. *Shlien S.* Nonparametric classification using matched binary decision trees // Pattern Recognition Letters. 1992. Vol. 13. P. 83–87.
99. *Siroky D.* Navigating Random Forests and related advances in algorithmic modeling // Statistics Surveys. 2009. Vol. 3. P. 147–163.
100. *Skurichina M., Kuncheva L., Duin R.* Bagging and boosting for the nearest mean classifier: Effects of sample size on diversity and accuracy / Eds. J. K. F. Roli // Multiple Classifier Systems, Proc. Third International

- Workshop MCS. (Cagliari, Italy 2002). Berlin: Springer, 2002. Vol. 2364. P. 62–71. <http://citeseer.nj.nec.com/539135.html>
101. *Strobl C., Boulesteix A. L., Augustin T., Zeileis A.* Bias in RandomForest Variable Importance Measures: Illustrations, Sources and a Solution // BMC Bioinformatics. 2007. Vol. 8, N 25. URL: <http://www.biomedcentral.com/14712105/8/25/abstract>. (дата обращения 15.04.2011).
 102. *Strobl C., Boulesteix A. L., Augustin T., Zeileis A.* Conditional variable importance for Random Forests // BMC Bioinformatics. Vol. 9, N 307. P. 67–78. <http://www.stat.uni-muenchen.de/carolin/research.html> (дата обращения 15.04.2011).
 103. *Strobl C., Zeileis A.* Danger: High Power! Exploring the Statistical Properties of a Test for Random Forest Variable Importance // Technical Report 017. Department of Statistics, University of Munich, 2008. P. 1–9. URL: <http://www.stat.uni-muenchen.de> (дата обращения 15.04.2011).
 104. *Strobl C., Hothorn T., Zeileis A.* Party On! A New, Conditional Variable-Importance Measure for Random Forests Available in the party Package // R Journal. 2009. Vol. 1, N 2. P. 14–17.
 105. *Strobl C., Malley J., Tutz G.* An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests // Technical Report 55. Department of Statistics University of Munich, 2009. P. 1–50. URL: <http://www.stat.uni-muenchen.de> (дата обращения 15.04.2011).
 106. *Therneau T., Atkinson B.* rpart: Recursive Partitioning // R package version 3.145. 2009. URL: <http://CRAN.Rproject.org/package=rpart> (дата обращения 15.04.2011).
 107. *Tibshirani R., Hastie T.* Margin trees for highdimensional classification // Technical report. Stanford University, 2006. P. 1–21. URL: <http://www.stanford.edu/hastie/pub.htm> (дата обращения 15.04.2011).
 108. *Tibshirani R.* marginTree: margin trees for highdimensional classification // R package version 1.01. 2010. URL: <http://CRAN.Rproject.org/package=marginTree> (дата обращения 15.04.2011).
 109. *Truong A.* oblique.tree: Oblique Trees for Classification Data // R package version 1.0. 2009. URL: <http://CRAN.Rproject.org/package=oblique.tree> (дата обращения 15.04.2011).
 110. *Truong A.* Fast Growing and Interpretable Oblique Trees via Probabilistic Model // Univ. of Oxford, A thesis submitted for the degree of Doctor of Philosophy, Trinity term. 2009. P. 1–119.
 111. *Wang X., Chen B., Qian G., Ye F.* On the optimization of fuzzy decision trees // Fussy Sets and Systems. 2000. Vol. 11, N 2. P. 117–125.
 112. *Wang L. X., Mendel J. M.* Generating fuzzy rules by learning from examples // IEEE Transaction on Systems, Man and Cybernetics. 1992. Vol. 22. P. 1414–1427.
 113. *Wang T.* MATLAB R13. // URL: <http://lib.stat.cmu.edu/matlab> (дата обращения 15.04.2011).
 114. *White D.* maptree: Mapping, pruning, and graphing tree models // R package version 1.4-6. 2010. <http://CRAN.R-project.org/package=maptree>
 115. *Williams G. J.* Combining decision trees: Initial results from the MIL algorithm / Eds. J. S. Gero, R. B. Stanton // Artificial Intelligence Developments and Applications. NorthHolland: Elsevier Science Publishers, 1988. P. 273–289.
 116. *Witten I., Frank E.* Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition. San Francisco: Morgan Kaufmann, 2005. 567 p.
 117. *Wolf B. J., Slate E. H., Hill E. G.* Logic Forest: An ensemble classifier for discovering logical combinations of binary markers // Bioinformatics. 2010. Vol. 26, N 17. P. 2183–2189.
 118. *Wolf B. J.* LogicForest: Logic Forest // R package version 1.0. 2010. URL: <http://CRAN.Rproject.org/package=LogicForest> (дата обращения 15.04.2011).
 119. *Zeileis A., Hothorn T.* ModelBased Recursive Partitioning // Journal of Computational and Graphical Statistics. 2008. Vol. 17, N 2. P. 492–514.

СВЕДЕНИЯ ОБ АВТОРЕ:

Чистяков Сергей Павлович

младший научный сотрудник
Институт прикладных математических исследований
Карельского научного центра РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: chistiakov@krc.karelia.ru
тел.: (8142) 763370

Chistiakov, Sergey

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: chistiakov@krc.karelia.ru
tel.: (8142) 763370

ПРАВИЛА ДЛЯ АВТОРОВ

Серия «Математическое моделирование и информационные технологии»

(требования к работам, представляемым к публикации
в «Трудах Карельского научного центра Российской академии наук»)

«Труды Карельского научного центра Российской академии наук» (далее – Труды КарНЦ РАН) публикуют результаты завершенных оригинальных исследований в различных областях современной науки: теоретические и обзорные статьи, сообщения, материалы о научных мероприятиях (симпозиумах, конференциях и др.), персоналии (юбилеи и даты, потери науки), статьи по истории науки. Представляемые работы должны содержать новые, ранее не публиковавшиеся данные.

Статьи проходят обязательное рецензирование. Решение о публикации принимается редакционной коллегией серии или тематического выпуска Трудов КарНЦ РАН после рецензирования, с учетом научной значимости и актуальности представленных материалов. Редколлегия серий и отдельных выпусков Трудов КарНЦ РАН оставляет за собой право возвращать без регистрации рукописи, не отвечающие настоящим правилам.

При получении редакцией рукопись регистрируется (в случае выполнения авторами основных правил ее оформления) и направляется на отзыв рецензентам. Отзыв состоит из ответов на типовые вопросы «Анкеты» и может содержать дополнительные расширенные комментарии. Кроме того, рецензент может вносить замечания и правки в текст рукописи. Авторам высылаются электронная версия «Анкеты» и комментарии рецензентов. Доработанный экземпляр автор должен вернуть в редакцию вместе с первоначальным экземпляром и ответом на все вопросы рецензента не позднее, чем через месяц после получения рецензии.

Почтовый адрес редакции: 185910, г. Петрозаводск, ул. Пушкинская, 11, КарНЦ РАН, редакция Трудов КарНЦ РАН. Телефон: (8142) 780109.

Содержание номеров Трудов КарНЦ РАН, аннотации и полнотекстовые электронные варианты статей, а также другая полезная информация, включая настоящие Правила, доступна на сайте – <http://transactions.krc.karelia.ru>.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСИ

Статьи публикуются на русском или английском языке. Рукописи должны быть тщательно выверены и отредактированы авторами.

Статьи должны быть подписаны всеми авторами.

Объем рукописи (включая таблицы, список литературы, подписи к рисункам, рисунки) не должен превышать: для обзорных статей – 30 страниц, для оригинальных – 25, для сообщений – 15, для хроники и рецензий – 5–6. Объем рисунков не должен превышать 1/4 объема статьи. Рукописи большего объема (в исключительных случаях) принимаются при достаточном обосновании по согласованию с ответственным редактором.

Рукописи присылаются в электронном виде, а также в двух экземплярах, напечатанных на одной стороне листа формата А4. Все страницы, включая список литературы и подписи к рисункам, должны иметь сплошную нумерацию в нижнем правом углу. Страницы с рисунками не нумеруются.

ОБЩИЙ ПОРЯДОК РАСПОЛОЖЕНИЯ ЧАСТЕЙ СТАТЬИ

Элементы статьи должны располагаться в следующем порядке: *УДК* курсивом на первой странице, в левом верхнем углу; заглавие статьи на русском языке заглавными буквами полужирным шрифтом; инициалы, фамилии всех авторов на русском языке полужирным шрифтом; полное название организации – место работы каждого автора в именительном падеже на русском языке курсивом (если авторов несколько и работают они в разных учреждениях, то следует отметить арабскими цифрами соответствие фамилий авторов учреждениям, в которых они работают; если все авторы статьи работают в одном учреждении, можно не указывать место работы каждого автора отдельно); аннотация на русском языке; ключевые слова на русском языке; инициалы, фамилии всех авторов на английском языке полужирным шрифтом; название статьи на английском языке заглавными буквами полужирным шрифтом; аннотация на английском языке; ключевые слова на английском языке; текст статьи (статьи экспериментального характера, как правило, должны иметь разделы: ВВЕДЕНИЕ. МАТЕРИАЛЫ И МЕТОДЫ. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. ВЫВОДЫ. ЛИТЕРАТУРА); благодарности; литература (с новой страницы).

Дополнительные сведения об авторах: фамилия, имя, отчество всех авторов полностью на русском и английском языке; полный почтовый адрес каждой организации (страна, город) на русском и англ-

лийском языке; должности авторов; адрес электронной почты для каждого автора; телефон для контактов с авторами статьи (можно один на всех авторов).

ЗАГЛАВИЕ СТАТЬИ должно точно отражать содержание статьи и содержать не более 8–10 значащих слов.

АННОТАЦИЯ должна быть лишена вводных фраз, содержать только главную информацию статьи, не превышать объем – 15 строк.

Отдельной строкой приводится перечень **КЛЮЧЕВЫХ СЛОВ**. Ключевые слова или словосочетания отделяются друг от друга запятой, в конце фразы ставится точка.

МАТЕРИАЛЫ И МЕТОДЫ должны содержать сведения об объекте исследования с обязательным указанием латинских названий и сводок, по которым они приводятся, авторов классификаций и пр. Транскрипция географических названий должна соответствовать атласу последнего года издания. Единицы физических величин приводятся по Международной системе СИ. Желательна статистическая обработка всех количественных данных. Необходимо возможно точнее обозначать местонахождения (в идеале – с точным указанием географических координат).

ИЗЛОЖЕНИЕ РЕЗУЛЬТАТОВ должно заключаться не в пересказе содержания таблиц и графиков, а в выявлении следующих из них закономерностей. Автор должен сравнить полученную им информацию с имеющейся в литературе и показать, в чем заключается ее новизна. Следует ссылаться на табличный и иллюстративный материал так: на рисунки, фотографии и таблицы в тексте (рис. 1, рис. 2, табл. 1, табл. 2 и т. д.), фотографии, помещаемые на вклейках (рис. I, рис. II). Обсуждение завершается формулировкой основного вывода, которая должна содержать конкретный ответ на вопрос, поставленный во Введении. Ссылки на литературу в тексте даются номерами в квадратных скобках.

ТАБЛИЦЫ нумеруются в порядке упоминания их в тексте, каждая таблица имеет свой заголовок. Диаграммы и графики не должны дублировать таблицы. Материал таблиц должен быть понятен без дополнительного обращения к тексту. Все сокращения, использованные в таблице, должны быть пояснены в Примечании, расположенном под ней. При повторении цифр в столбцах нужно их повторять, при повторении слов – в столбцах ставить кавычки. Таблицы могут быть книжной или альбомной ориентации.

ПОДПИСИ К РИСУНКАМ должны содержать достаточно полную информацию, для того чтобы приводимые данные могли быть понятны без обращения к тексту (если эта информация уже не дана в другой иллюстрации). Аббревиации расшифровываются в подрисуночных подписях.

СОКРАЩЕНИЯ. Разрешаются лишь общепринятые сокращения – названия мер, физических, химических и математических величин и терминов и т. п. Все сокращения должны быть расшифрованы, за исключением небольшого числа общеупотребительных.

БЛАГОДАРНОСТИ. В этой рубрике выражается признательность частным лицам, сотрудникам учреждений и фондам, оказавшим содействие в проведении исследований и подготовке статьи, а также указываются источники финансирования работы.

ЛИТЕРАТУРА. Пристатейные ссылки и/или списки пристатейной литературы следует оформлять по ГОСТ Р 7.0.5-2008. Библиографическая ссылка. Общие требования и правила составления (http://www.bookchamber.ru/GOST_P_7.0.5.-2008). Список работ представляется в алфавитном порядке. Все ссылки даются на языке оригинала (названия на японском, китайском и других языках, использующих нелатинский шрифт, пишутся в русской транскрипции). Сначала приводится список работ на русском языке и на языках с близким алфавитом (украинский, болгарский и др.), а затем – работы на языках с латинским алфавитом. В списке литературы между инициалами ставится пробел.

Электронные версии статей выпусков серии «Математическое моделирование и информационные технологии» принимаются в формате .tex (LaTeX 2ε) с использованием стилевого файла, который находится по адресу <http://transactions.krc.karelia.ru/section.php?id=755>.

TABLE OF CONTENTS

Preface	3
B. Z. Belashev. ANALISYS OF THE DISTRIBUTION OF SEISMIC EVENTS FROM THE NEIC CATALOG	4
G. A. Borisov, T. P. Tikhomirova. METHODS TO MINIMIZE ENERGY AND POWER LOSSES IN THE REGIONAL FUEL-AND-ENERGY SECTOR	12
V. N. Gridin, A. Yu. Golubin, M. N. Petrova. OPTIMAL PER-CLAIM INSURANCE AND OPTIMAL REINSURANCE OF SUMMARY RISK IN A STATIC SETTING	17
A. L. Zabrovskiy. THE CRITERION OF QUALITY ASSESSMENT OF EDUCATIONAL MULTIMEDIA STREAMS TRANSMITTED IN REAL TIME MODE	26
Yu. V. Zaika. MODEL OF HYDROGEN THERMODESORPTION FROM METAL-ORGANIC FRAMEWORKS	33
K. A. Zamuraev, A. V. Prasolov. MODEL OF REGULATING THE ST. PETERSBURG ECONOMY BY THE INVESTMENT FLOW	38
N. V. Kablukova, A. V. Sokolov. MATHEMATICAL ANALYSIS OF A WAY OF PRESENTING TWO FIFO QUEUES IN SHARED MEMORY	46
V. A. Kashtanov, O. B. Zaytseva. ON MINIMAX APPROACHES TO PROBLEMS OF SAFETY	55
A. N. Kirillov. FINITE TIME STABILIZATION OF CONTROLLABLE DYNAMICAL SYSTEMS	68
A. O. Oleynikov. NUMERICAL OPTIMIZATION OF PARALLEL PROCESSING IN A STATIONARY ENVIRONMENT	73
E. S. Petrova. ON THE LIMIT BEHAVIOUR OF NUMBER OF TREES OF A GIVEN SIZE IN A RANDOM UNLABELLED UNROOTED FOREST	79
A. A. Pechnikov. APPLICATION OF WEBOMETRIC TECHNIQUES FOR INVESTIGATION OF THE WEB SPACE OF SCIENTIFIC ORGANIZATION (A CASE STUDY OF KARELIAN RESEARCH CENTRE OF RAS)	86
A. L. Rabinovich, D. V. Zhurkin. AN IMPORTANCE SAMPLING FOR MONTE CARLO SIMULATION OF A CONTINUOUS SPECTRUM OF CONFORMATIONS OF MACROMOLECULES	96
E. V. Khvorostyanskaya. ON RANDOM POISSON ALLOCATION TO CELLS	112
S. P. Chistiakov. RANDOM FORESTS: AN OVERVIEW	117
INSTRUCTIONS FOR AUTHORS	137

Научное издание

**Труды Карельского научного центра
Российской академии наук**

№ 1, 2013

Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Вып. 4

*Печатается по решению
Президиума Карельского научного центра РАН*

Свидетельство о регистрации СМИ ПИ № ФС77-48848 от 02.03.2012 г.
выдано Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций

Редактор М. А. Радостина
Оригинал-макет Е. Н. Спектор
Стилевой файл А. С. Румянцев

Подписано в печать 04.02.2013. Формат 60x84¹/₈.
Гарнитура CMR. Печать офсетная. Уч.-изд. л. 16,8. Усл. печ. л. 16,28.
Тираж 500 экз. Изд. № 349. Заказ 112

Карельский научный центр РАН
Редакционно-издательский отдел
185003, г. Петрозаводск, пр. А. Невского, 50