



Институт прикладных
математических исследований
Карельского научного центра
Российской академии наук

**IX Международная
Петрозаводская конференция**

**Вероятностные методы
в дискретной математике**

30 мая — 3 июня 2016 г.
Петрозаводск, Россия

РАСШИРЕННЫЕ ТЕЗИСЫ

Петрозаводск 2016

Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences

**The IX International
Petrozavodsk Conference**

**Probabilistic Methods
in Discrete Mathematics**

May 30 – June 3, 2016,
Petrozavodsk, Russia

EXTENDED ABSTRACTS

Petrozavodsk 2016

УДК 519.1(063)
ББК 22.176
В35

Научный редактор *В. В. Мазалов*
Editor *V. V. Mazalov*

Ответственный редактор *Е. Н. Спектор*
Responsible editor *E. N. Spector*

Вероятностные методы в дискретной математике,
В35 IX Международная Петрозаводская конференция (30 мая — 3 июня 2016 г., Петрозаводск, Россия) : расширенные тезисы / Ин-т прикладных мат. исследований Карел. науч. центра Рос. акад. наук ; [науч. ред. В. В. Мазалов ; отв. ред. Е. Н. Спектор]. — Петрозаводск : [Издательство ПетрГУ], 2016. — 106, [2] с. : ил.
ISBN 978-5-8021-2950-0
Сборник содержит расширенные тезисы докладов участников IX Международной Петрозаводской конференции «Вероятностные методы в дискретной математике» (30 мая — 3 июня 2016 г.).

Probabilistic Methods in Discrete Mathematics, the IX International Petrozavodsk Conference (May 30 — June 3, 2016, Petrozavodsk, Russia) : Extended Abstracts / Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences ; [ed. V. V. Mazalov ; resp. ed. E. N. Spector]. — Petrozavodsk : [PetrSU Press], 2016. — 106, [2] c. : il.

ISBN 978-5-8021-2950-0

The present volume contains extended abstracts accepted for the IX International Petrozavodsk Conference "Probabilistic Methods in Discrete Mathematics" (May 30 — June 3, 2016, Petrozavodsk, Russia).

УДК 519.1
ББК 22.176

ISBN 978-5-8021-2950-0

© Институт прикладных
математических исследований
КарНЦ РАН, 2016
© Издательство ПетрГУ, 2016

Main organizers

INSTITUTE OF APPLIED MATHEMATICAL RESEARCH
OF THE KARELIAN RESEARCH CENTRE RAS,
PETROZAVODSK STATE UNIVERSITY
RUSSIAN FEDERATION ACADEMY OF CRYPTOGRAPHY
STEKLOV MATHEMATICAL INSTITUTE OF RAS
"TRANSACTIONS OF KARELIAN RESEARCH CENTRE OF RAS"
JOURNAL EDITORIAL BOARD
"MATEMATICHESKAYA TEORIYA IGR I PRILOZHENIA" (IN RUSSIAN)
JOURNAL EDITORIAL BOARD

Sponsors

Russian Foundation for Basic Research (grant 16-01-20205)

Program committee

V.F. Kolchin (co-chair)
A.M. Raigorodskii (co-chair)
N. Alon
G.I. Ivchenko
Yu.S. Kharin
A.P. Kovalenko
T. Matsuhisa
Yu.I. Medvedev
V.G. Mikhailov
L.A. Petrosian
V.N. Sachkov
J. Spencer
K. Szajowski
P.V. Tetali
A.N. Tikhomirov
V.A. Vatutin
A.B. Zhizhchenko
A.M. Zubkov

Organizing committee

chair: Vladimir Mazalov

co-chair Yury Pavlov

secretary: Marina Leri

V.I. Khokhlov,

A.V. Voronin,

A.A. Rogov,

E.E. Ivashko,

N.N. Nikitina,

E.N. Spector,

A.V. Chirkov,

Yu.V. Chirkova

Scientific program

Probability and statistical problems of discrete mathematics Theory
of random graphs and hypergraphs Combinatorial methods in the Internet
data analysis Mathematical problems of information security Applied statistics
Statistical modeling and simulation Game theory and stochastic optimization
problems

Invited speakers

Ivchenko Grigori Ivanovich, Medvedev Yuri Ivanovich

(Moscow State Institute of Electronics and Mathematics HSE, Russia)

Kharin Yuri Semenovich

(Belarusian State University, Belarus)

Raigorodskii Andrei Mikhailovich

(Moscow State University, Russia)

Shabanov Dmitry Aleksandrovich

(Moscow State University, MIPT, Russia)

Tikhomirov Aleksandr Nikolaevich

(Department of Mathematics Komi SC UrD RAS, Russia)

Zhukovskii Maksim Evgenievich

(MIPT, Russia)

Contents

Барковский Е. А., Соколов А. В.

Вероятностная модель для задачи оптимального управления work-stealing деками при различных стратегиях перехвата работы..... 11

Бородина А. В.

Оценивание эффективной пропускной способности узла тандемной сети на основе различных типов регенерации..... 14

Волгин А. В.

О законе квадратного корня в задаче выявления вкраплений в цепях Маркова с неизвестной матрицей переходных вероятностей..... 17

Воронов Р. В.

Обобщение задачи определения метрической размерности графа..... 20

Гиппиус М. Б.

Методы распознавания строк и алгоритмы классификации символов в стенограммах..... 23

Gusev A. S.

Clique Numbers of Random Subgraphs of Distance Graphs..... 25

Zhukovskii M. E.

First-order Logic of Random Graphs..... 27

Зубков А. М., Серов А. А.

Итерации случайных отображений конечных множеств..... 28

Калинина К. А.

Влияние случайного суммирования на регенеративную оценку эффективной пропускной способности узла сети..... 29

Колмогоров А. В.

Об оптимальных стратегиях двухальтернативной обработки данных на начальном этапе..... 32

<i>Kolchin A. V.</i>	
Urn Schemes and Partitions of a Positive Integer	34
<i>Кондратьев А. Ю.</i>	
Правила коллективного выбора на основе минимальных выигрывающих коалиций кандидатов	37
<i>Крижановский А. А., Кириллов А. Н.</i>	
Модель синсета: геометрия и статистика	38
<i>Лазутченко А. Н.</i>	
Оптимизация параллельного управления в случайной среде с нормально распределенными доходами и различными дисперсиями	41
<i>Лёвин В. В., Гуськов С. Ю.</i>	
Локальная предельная теорема для разделимых статистик в полиномиальной схеме	43
<i>Лукашенко О. В., Пагано М.</i>	
Об оценке вероятности редких событий в гауссовских очередях	45
<i>Малодушев С. В., Воронов Р. В.</i>	
Математическая модель эвакуации людей из здания	47
<i>Matushkin A.D.</i>	
Zero-One Law for Random Uniform Hypergraphs	50
<i>Минаков А. А.</i>	
Пуассоновская аппроксимация числа монотонных цепочек с началом в локальном минимуме в цепи Маркова	51
<i>Митрофанова О. А., Буре В. М.</i>	
Две задачи статистического анализа экологических данных	54
<i>Михайлов В. Г., Шойтов А. М., Волгин А. В.</i>	
О структурной эквивалентности s -цепочек в цепях Маркова	57

<i>Морозов Е. В., Румянцев А. С., Пешкова И. В.</i>	
Метод регенеративных огибающих для оценивания вы-	
сокопроизводительных многосерверных систем	63
<i>Morozov E. V., Potakhina L. V.</i>	
Asymptotic Analysis of a System with Random Volume	
Customers.....	65
<i>Никитина Н. Н., Ивашко Е. Е.</i>	
Цена анархии в игре заполнения для поиска лекарств	67
<i>Павлов Ю. Л.</i>	
О случайных конфигурационных графах в случайной	
среде.....	70
<i>Павлов Ю. Л.</i>	
Асимптотика числа вершин заданной степени конфи-	
гурационного графа вблизи критических точек	72
<i>Переходцева Э. В.</i>	
Статистические модели распознавания и прогнозиров-	
ание метеоситуаций, приводящих к возникновению тор-	
надо и сильных шквалов умеренных широт	74
<i>Потехина Е. А.</i>	
Применение произведения Адамара в одной из комби-	
наторно-вероятностных задач	78
<i>Rettieva A. N.</i>	
Equilibria in Dynamic Multicriteria Games with Incom-	
plete Information.....	81
<i>Румянцев А. С.</i>	
Ускоренная оценка стационарности вероятностной мо-	
дели процесса нагрузки вычислительного кластера ...	84
<i>Самуйлов К. Е., Сопин Э. С., Вихрова О. Г.</i>	
К анализу стационарных характеристик системы мас-	
сового обслуживания со случайными требованиями ..	87
<i>Семенов А. С.</i>	
Число независимости случайного разреженного	
k-однородного гиперграфа.....	90

<i>Староверова К. Ю., Буре В. М.</i>	
Кластеризация районов Санкт-Петербурга по показателю заболеваемости	92
<i>Стафеев С. В.</i>	
Идентифицируемость модели факторного анализа с негауссовскими общими факторами	95
<i>Хонов С. А.</i>	
Современные статистические выводы применительно к одному конфликту при сортировке алмазов	97
<i>Чеплюкова И. А.</i>	
О распределении числа вершин заданной степени конфигурационного графа	100
<i>Чиркова Ю. В.</i>	
Цена анархии в задаче максимизации минимальной поддержки машин в системе обслуживания	103
<i>Шабанов Д. А.</i>	
Вероятностные методы в задачах о раскрасках гиперграфов и теории Рамсея	106

Вероятностная модель для задачи оптимального управления work-stealing деками при различных стратегиях перехвата работы

Е. А. Барковский, А. В. Соколов

Петрозаводский государственный университет,
Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: barkevgen@gmail.com, avs@krc.karelia.ru

В методе динамической балансировки параллельных вычислений на многоядерной архитектуре применяют механизм "work-stealing" [1, 2]. Процессоры, ставшие пустыми, пытаются перехватить часть работы у других процессоров. Здесь каждый процессор (поток) имеет пул задач, информация о которых хранится в деке этого процессора. То есть мы работаем (в терминологии Д. Кнута) с несколькими деками с ограниченным входом [3]. В [1] было предложено перехватывать один элемент, в [4] – половину элементов дека.

В этой работе мы предлагаем математическую модель последовательного циклического представления work-stealing деков (как FIFO-очереди) в общей памяти, где каждый дек расположен в отдельном участке памяти [5]. На каждом шаге дискретного времени могут произойти операции с заданными вероятностями. Ранее, такие модели были построены нашим коллективом для представления некоторых динамических структур данных ([6, 7, 8] и другие).

В докладе будет рассматриваться работа двух деков. Этот конкретный случай важен не только как первый шаг к построению модели – у него есть и свое независимое значение. Так, среди многоядерных архитектур есть такие, в которых отсутствует кэш памяти. Например, в архитектуре AsAP-II каждое ядро имеет два FIFO буфера, а в архитектуре SEAForth – два стека. Work-stealing дека могут быть реализованы аппаратно, по принципу вышеупомянутых архитектур. В этом случае важной задачей является исследование оптимального управления двумя деками.

Пусть в памяти размера m мы работаем с двумя последовательными циклическими деками, где все элементы имеют фиксированный размер в одну условную единицу. Для последовательного представления деков выделим каждому некоторое количество единиц памяти из общего объема m . Обозначим s количество единиц памяти, выделенное первому деку, $(m - s)$ – количество памяти, выделенное второму деку, x и y текущие длины первого и второго дека соответственно. Известны вероятностные характеристики операций, производимых над деками. Исключение элемента из пустого дека не является аварийной ситуацией – пустой дек начинает перехватывать элементы у другого дека (при условии, что эти элементы имеются). В качестве математической модели мы рассматриваем случайное блуждание по двумерной целочисленной решетке в области $-1 \leq x \leq s + 1$, $-1 \leq y \leq m - s + 1$. Блуждание начинается в начале координат. Прямые $x = s + 1$ и $y = m - s + 1$ образуют два поглощающих экрана – когда процесс попадает на эти экраны память переполняется и программа аварийно прекращается. Прямые $x = -1$, $y = -1$ образуют два отражающий экрана, которые определены для случая исключения элемента из пустого дека. Введение данных экранов позволяет установить процесс перехвата элемента. Формально деки попадают на экраны, но фактически в область $1 \leq x < s$, $1 \leq y < m - s$. Конкретная точка перехода зависит от стратегии перехвата.

В задаче оптимального разделения памяти, нужно найти такое s , при котором система работает дольше.

Для оптимального перераспределения памяти ситуация иная. Весь процесс работы системы можно разбить на два этапа: до перераспределения и после. Так, до перераспределения, прямые $x = s + 1$ и $y = m - s + 1$ (в области блуждания) образуют отражающий экран – попадая на эти прямые, процесс отражается внутрь области, т.е. система переходит в некоторую точку перераспределения z . На втором этапе прямые $x = s + 1$ и $y = m - s + 1$ становятся поглощающими экранами, и процесс начинает свое блуждания из точки z .

Пусть $1 \leq o \leq s$ – некоторое количество элементов. В этой задаче требуется найти такое значение o , при котором система, начинающая свою работу из точки перераспределения z с координатами $(s - o + 1, y + o)$ или $(x + o, m - s - o + 1)$, в среднем работает дольше. В докладе мы рассматриваем случай $s = m/2$ (пусть m – четное), т.е. каждому деку выделена половина доступной памяти.

Разработаны алгоритмы и программы нахождения оптимального разбиения памяти между деками, в зависимости от вероятностных характеристик очередей (эта задача была апробирована в [9]) и нахождения оптимального количества элементов для перераспреде-

ния. Для решения поставленных задач использовался аппарат случайных блужданий, поглощающих цепей Маркова, система LAPACK. Вычисления производились с помощью кластера КарНЦ РАН.

В докладе рассматривается математическая модель этих процессов и приведены некоторые результаты вычислений (указанные результаты были подтверждены имитационными экспериментами).

Работа выполнена при финансовой поддержке РФФИ (грант № 15-01-03404-а).

Список литературы

- [1] Blumofe R.D., Leiserson C.E. Scheduling multithreaded computations by work stealing // Journal of the ACM. 1999. N. 46. P. 720–748.
- [2] Herlihy M., Shavit N. The Art of Multiprocessor Programming. Elsevier, 2008.
- [3] Knuth D. The Art of Computer Programming. Vol. 1. Addison-Wesley, 2001.
- [4] Hendler D., Shavit N. Non-blocking Steal-half Work Queues // The Twenty-first Annual ACM Symposium on Principles of Distributed Computing. 2002. P. 280–289.
- [5] Chase D., Lev Y. Dynamic Circular Work-Stealing Deque // 17th ACM Symposium on Parallelism in Algorithms and Architectures. 2005. P. 21–28.
- [6] Aksenova E.A., Lazutina A.A., Sokolov A.V. Study of a Non-Markovian Stack Management Model in a Two-Level Memory // Programming and Computer Software. 2004. Vol. 30, No. 1. P. 25–33.
- [7] 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.
- [8] Соколов А.В., Драц А.В. Моделирование некоторых методов представления n FIFO очередей в памяти одного уровня // Эвристические алгоритмы и распределенные вычисления. 2014. Т. 1, № 1. С. 40–52.
- [9] Соколов А.В., Барковский Е.А. Вероятностная модель для задачи оптимального управления Work-stealing деками // Обзорение прикладной и промышленной математики. 2015. Т. 22, № 4, С. 503.

Оценивание эффективной пропускной способности узла тандемной сети на основе различных типов регенерации

А. В. Бородина

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводский государственный университет,
Петрозаводск, Россия

E-mail: borodina@krc.karelia.ru

Проблема оценивания основных показателей качества обслуживания (QoS) в современных коммуникационных системах достаточно известна. В частности, одним из важных показателей надежности является вероятность потери (или перегрузки системы) $P_b := P(W > b)$, а именно, вероятность превышения стационарным процессом нагрузки W некоторого высокого уровня b . Оценивание вероятности перегрузки системы тесно связано с определением другого параметра качества обслуживания: *эффективной пропускной способности (ЭПС)*.

Задача выбора ЭПС состоит в определении такого значения C мощности обслуживающего устройства, которое гарантирует, что вероятность P_b не превысит заданной (малой) величины Γ , т. е.

$$P_b := P(W > b) \leq \Gamma. \quad (1)$$

Если входящий процесс нагрузки, поступающий на обслуживание, обладает регенеративной структурой, то возможно применения регенеративной оценки для ЭПС (см. [1, 2]). Таким образом, для оценивания ЭПС последнего узла тандемной сети вида $GI/G/1 \rightarrow \dots \rightarrow \cdot/G/1$, состоящей из $m + 1$ узлов, необходимо реализовать алгоритмическое определение моментов регенерации. Пусть на вход в первый узел подается процесс восстановления, а узел с номером $m + 1$ представляет систему с постоянной скоростью обслуживания C . Подчеркнем, что при этом выходной поток узла m (который является входным потоком в узел $m + 1$) является регенеративным.

Пусть W_n есть незавершенная нагрузка в системе, которую встречает заявка с номером n . Тогда процесс незавершенной нагрузки в узле $m + 1$ моделируется с использованием рекурсии Линдли

$$W_{n+1} = [W_n + v_n - C\tau_n]^+, n \geq 0; W_0 = 0,$$

построенной по моментам $\{t_n\}$ прихода заявок, т. е. $W_n = W(t_n^-)$, $\tau_n = t_{n+1} - t_n$, v_n — есть работа, приносимая заявкой с номером n , а C — искомая ЭПС.

Для построения оценки ЭПС C последнего узла сети с помощью регенеративного метода можно использовать следующую аппроксимацию (см. [3, 4]):

$$C = \frac{\ln E e^{\theta^* \hat{X}}}{\theta^* ED}, \quad (2)$$

где $D =_{st} \sum_{k=0}^{\alpha-1} \tau_k = T_1$, $\frac{1}{k} \sum_{i=1}^k D_i \rightarrow ED$, $k \rightarrow \infty$, а $D_i = T_{i+1} - T_i$ есть длина i -го цикла регенерации в непрерывном времени. Таким образом, для регенеративного входного потока в узел $m + 1$ необходимо определить моменты начала циклов регенерации T_i , длины циклов D_i в непрерывном времени и длины циклов α_i в дискретном времени (количество моментов прихода на цикле). Значение параметра θ^* определяется согласно принципу больших уклонений (см. [5, 6]) для стационарного процесса незавершенной нагрузки W следующим образом:

$$\theta^* = -\ln \Gamma/b$$

и связано с условием (1).

В работе представлены результаты сравнительного анализа регенеративных оценок ЭПС, построенных на основе различных типов регенерации (классическая, квази-регенерация) для узла тандемной сети с использованием аппроксимации (2).

Список литературы

- [1] А. В. Бородина, Е. В. Морозов. Сравнение двух оценок эффективной пропускной способности системы обслуживания // Труды Карельского научного центра РАН, №6, 2012, с. 8-17.
- [2] А. В. Бородина, Е.В. Морозов. Об оценивании эффективной пропускной способности в системе с регенеративным входным процессом // Информатика и ее применения, 2013. Т. 7. Вып. 2. С. 13-20.

- [3] А. В. Бородина, Е.В. Морозов Оценивание эффективной пропускной способности узла в инфокоммуникационной тандемной сети // Системы и средства информатики т. 24 N. 2, 2014, pp. 37-54
- [4] Alexandra V. Borodina, Ksenia A. Kalinina, Evsey V. Morozov. On the accuracy of the effective bandwidth regenerative estimation, Proceedings of ICUMT 2014: The 6th International Congress on Ultra Modern Telecommunications and Control systems and Workshops, Saint-Petersburg, pp. 652-656, 2014,IEEE, Springer.
- [5] A. Ganesh, N. O'Connell and D. Wischik. Big Queues. Springer-Verlag, Berlin, 2004.
- [6] P. W. Glynn, W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. JAP, 31, 1994, 131-156.

О законе квадратного корня в задаче выявления вкраплений в цепях Маркова с неизвестной матрицей переходных вероятностей

А. В. Волгин

Лаборатория ТВП, Москва, Россия

E-mail: artem.volgin@bk.ru

Пусть $X = \{X_1, X_2, \dots\}$ — простая, конечная, неразложимая и ацикличная цепь Маркова с множеством состояний $\mathcal{A} = \{0, 1, \dots, N-1\}$ и фиксированной матрицей переходных вероятностей $\Pi = (\pi_{ab})_{N \times N}$, $a, b \in \mathcal{A}$. Стационарное распределение цепи обозначим через $\pi = (\pi_1, \dots, \pi_N)$, $\pi_0, \dots, \pi_N \in (0, 1)$. Рассмотрим процедуру внесения вкраплений в цепь Маркова X , в которой используются две последовательности: $Z = \{Z_1, Z_2, \dots\}$ и $\Delta = \{\delta_1, \delta_2, \dots\}$, $Z_i \in \mathcal{A}$, $\delta_i \in \{0, 1\}$, $i \geq 0$. При этом Z и δ являются реализациями независимых испытаний и $\mathbf{P}\{Z_i = a\} = p_a$, $a \in \mathcal{A}$, и $\mathbf{P}\{\delta_i = 0\} = \tau = \tau(n)$, $i \geq 0$, $p_a, \tau \in [0, 1]$.

В результате внесения вкраплений образуется последовательность $Y = \{Y_1, Y_2, \dots\}$:

$$Y_i = X_i \mathbf{I}\{\delta_i = 1\} + Z_i \mathbf{I}\{\delta_i = 0\}, \quad i \geq 0, \quad Y_i \in \mathcal{A}, \quad (1)$$

где через $\mathbf{I}\{A\}$ обозначается индикатор события A . При этом предполагается, что последовательности X , Z и Δ являются независимыми.

Постановка задачи заключается в следующем. Наблюдается отрезок последовательности $(Y_0, \dots, Y_{n-1})$ длины $n \in \mathbb{N}$. Относительно способа образования данного отрезка выдвигаются две сложные гипотезы $H_0 : \tau = 0$ и $H_1 : \tau > 0$ об отсутствии и наличии вкраплений в цепь Маркова X соответственно. Причем при обеих гипотезах будем предполагать, что матрица Π , а также величины p_a , $a \in \mathcal{A}$ и τ — неизвестны. Задача заключается в построении критерия различения гипотез: H_0 и H_1 .

В [3] приведены асимптотические (при $n \rightarrow \infty$) условия, при выполнении которых существует процедура различения гипотез H_0 и H_1 , при этом предполагается, что матрица переходных вероятностей

П цепи Маркова известна. В докладе рассматривается статистический критерий различия указанных гипотез в случае неизвестной матрицы Π .

Рассмотрим вероятностные пространства $(\Omega, \mathfrak{A}, \mathbf{P}_{H_i})$, $i \in \{0, 1\}$, где $\Omega = \{w_0, w_1, \dots, w_{n-1} : w_j \in \{0, 1\}, j = 0, \dots, n-1\}$, $n \in \mathbb{N}$, $\mathfrak{A}$ – совокупность всех подмножеств множества Ω , $\{\mathbf{P}_{H_i}(A) : A \in \mathfrak{A}\}$, $i \in \{0, 1\}$, – распределения вероятностей событий при гипотезах H_0 и H_1 соответственно. Множества $\{X_0, \dots, X_{n-1}\}$, $\{Y_0, \dots, Y_{n-1}\}$ и $\{Z_0, \dots, Z_{n-1}\}$, $n \in \mathbb{N}$, будем рассматривать как наборы случайных величин, заданных на измеримом пространстве $(\Omega, \mathfrak{A})$.

В [2] анонсирован подход к различению гипотез H_0 и H_1 на основе статистики следующего вида:

$$S = \sum_{a,b,c \in \mathcal{A}} \frac{(\nu_{abc} - \nu_{ab}\nu_{bc}/\nu_b)^2}{\nu_{ab}\nu_{bc}/\nu_b}, \quad (2)$$

$$\nu_{abc} = \sum_{i=0}^{n-3} \mathbf{I}\{Y_i = a, Y_{i+1} = b, Y_{i+2} = c\},$$

$$\nu_{ab} = \sum_{i=0}^{n-2} \mathbf{I}\{Y_i = a, Y_{i+1} = b\},$$

$$\nu_a = \sum_{i=0}^{n-1} \mathbf{I}\{Y_i = a\}.$$

Известно ([1]), что при гипотезе H_0 статистика (2) асимптотически при $n \rightarrow \infty$ сходится к распределению хи-квадрат с двумя степенями свободы. Обозначим через $t_{\chi^2_2, 1-\alpha}$ – квантиль уровня $1 - \alpha$ распределения хи-квадрат с двумя степенями свободы. Рассмотрим критерий согласия с гипотезой H_0 , основанный на статистике (2) :

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если } S < t_{\chi^2_2, 1-\alpha}, \\ H_1, & \text{если } S \geq t_{\chi^2_2, 1-\alpha}, \end{cases} \quad (3)$$

где $\alpha = \mathbf{P}\{S \geq t_{\chi^2_2, 1-\alpha} | H_0\}$ – вероятность ошибки первого рода.

Теорема 1. *Если существуют числа $a, b, c \in \mathcal{A}$, для которых выполняется условие:*

$$\sqrt{n}\tau(1-\tau)^2(C_1(1-\tau) + C_2\tau) \rightarrow \infty, \quad n \rightarrow \infty, \quad (4)$$

где

$$C_1 = p_b\pi_b(\pi_{a*c} - \pi_a\pi_{bc}) + \pi_a\pi_b p_b(\pi_{bc} - \pi_c), \quad C_2 = p_b(\pi_{a*c} - \pi_a\pi_c),$$

$$\pi_{a*c} = \mathbf{P}\{Y_0 = a, Y_2 = b\},$$

то критерий (3) является состоятельным.

Замечание 1. Если $C_1, C_2 > 0$, то критерий, предложенный в теореме 1, является состоятельным при выполнении двух условий:

$$\sqrt{n}\tau \rightarrow \infty, \quad n \rightarrow \infty, \quad (5)$$

и

$$\sqrt{n}(1 - \tau)^2 \rightarrow \infty, \quad n \rightarrow \infty. \quad (6)$$

Условие (5) означает, что число вкраплений должно превосходить квадратный корень из длины наблюдаемого отрезка. Условие (6) является в некотором смысле противоположным (5) и означает, что вкраплений не должно быть «слишком много».

Далее рассмотрим случай, когда цепь Маркова $X = X_1, X_2, \dots$ является неразложимой и циклической с периодом $d \in \mathbb{N}$. Это означает, что множество состояний разбивается на d циклических подклассов. Процедуру внесения вкраплений в циклические цепи Маркова будем осуществлять аналогичным (1) образом. Далее рассмотрим метод различения гипотез H_0 и H_1 об отсутствии и наличии вкраплений соответственно.

Зафиксируем $m \in \mathbb{N}$ и будем полагать, что $m \geq d$. Рассмотрим множество регулярных выборок из отрезка последовательности Y :

$$M = \left\{ Y^{(t)} = \left(Y_0, Y_t, Y_{2t}, \dots, Y_{t \lfloor \frac{n-1}{t} \rfloor} \right) : t = 2, \dots, m \right\}. \quad (7)$$

Для каждой регулярной выборки $Y^{(t)}$ обозначим через $\mu_0^{(t)}$ статистику числа «пустых ящиков» при размещении $t \lfloor \frac{n-1}{t} \rfloor$ дробин по N ячейкам. Рассмотрим критерий различия гипотез H_0 и H_1 :

$$\text{принимается гипотеза} \begin{cases} H_0, \text{ если существует } t = 2, \dots, m, \\ \quad \text{при котором } \mu_0^{(t)} > 0, \\ H_1, \text{ если } \mu_0^{(2)} = \dots = \mu_0^{(m)} = 0. \end{cases} \quad (8)$$

Теорема 2. Критерий (8) различия гипотез H_0 и H_1 является состоятельным.

Замечание 2. При рассмотрении критерия различия гипотез H_0 и H_1 для циклических цепей Маркова не требуется выполнение «закона квадратного корня».

Список литературы

- [1] Г.И. Ивченко, Ю.И. Медведев. Введение в математическую статистику. М.: ЛКИ, 2010. 600 с.
- [2] А.М. Шойтов. О выявлении факта зашумления конечной цепи Маркова с неизвестной матрицей переходных вероятностей / Тезисы докладов конференции SIBECRYPT'10, 2010. С. 44-45.
- [3] T. Filler, A.D. Ker, J. Fridrich. The square root law of steganographic capacity for Markov covers / Proc. SPIE. 2009. V.7254. P.31-47.

Обобщение задачи определения метрической размерности графа

Р. В. Воронов

Петрозаводский государственный университет,
г. Петрозаводск, Россия

E-mail: rvoronov@sampo.ru

В настоящее время широкое распространение получают системы определения местоположения мобильных объектов внутри помещений. Такие системы состоят из нескольких стационарных точек доступа и мобильного устройства, местоположение которого необходимо определить. Местоположение определяется на основе мощности принимаемого сигнала или времени его передачи [1]. При этом точность зависит от многих факторов, один из которых – число и расположение стационарных точек доступа [2].

В некоторых моделях систем позиционирования помещение представляется в виде неориентированного графа. Вершинам графа соответствуют места возможного расположения объекта и точек доступа, ребрам – возможные пути перемещения объекта. Ряд исследователей связывают задачу расстановки точек доступа с задачей определения метрической размерности графа. Рассмотрим более подробно этот подход.

Пусть задан взвешенный неориентированный связный граф $G = (V, E, \omega)$, где V – множество вершин, E – множество ребер, ω – весовая функция:

$$\omega: E \rightarrow \mathbb{N}.$$

Расстояние $d(u, v)$ между двумя вершинами u и v графа – это минимальная длина связывающей их цепи. Подмножество вершин $W \subset V$ называется разрешающим, если для каждой пары вершин $u, v \in V$ найдется вершина $w \in W$, для которой $d(w, u) \neq d(w, v)$. Метрическая размерность $\beta(G)$ графа G – это минимальное число вершин в разрешающем множестве [4]. Разрешающее множество минимальной мощности называется базисом метрик графа.

Множеству точек доступа можно сопоставить разрешающее множество вершин графа. Минимально необходимое число точек доступа соответствует метрической размерности графа.

Известно, что задача определения метрической размерности графа в общем случае является NP-полной. Для некоторых частных случаев получены ее решения в явном виде [3]. Рассмотрим один из частных случаев – граф ходов шахматного короля $G_{m,n}$, изображающий все возможные ходы короля на шахматной доске размера m на n . Вес каждого ребра принимается равным единице.

Можно доказать, что метрическая размерность графа $G_{m,n}$ такова:

$$\beta(G_{m,n}) = \begin{cases} 1, & \text{если } m = 1 \text{ или } n = 1, \\ 3, & \text{если } m = n \text{ и } n > 1, \\ \left\lceil \frac{m-1}{n-1} \right\rceil + 1, & \text{если } m > n \text{ и } n > 1, \\ \left\lceil \frac{n-1}{m-1} \right\rceil + 1, & \text{если } n > m \text{ и } m > 1, \end{cases}$$

где $\lceil \cdot \rceil$ – округление вверх.

Описанный способ решения задачи расстановки точек доступа не учитывает возможные ошибки в определении расстояния между точками доступа и мобильным объектом, а также случайные изменения уровня сигналов в зонах помещения. Рассмотрим математическую модель, учитывающую случайные ошибки при определении расстояний.

Пусть $n = |V|$ и зафиксирован индекс $i \in \{1, \dots, n\}$. Рассмотрим двумерную случайную величину (ξ, η_i) , где ξ принимает значения из множества $\{1, \dots, n\}$ и ее значение соответствует номеру вершины, η_i принимает целые значения, которые будем интерпретировать как результат определения расстояния между вершиной v_i , соответствующей точке доступа, и вершиной v_ξ , соответствующей местоположению мобильного объекта.

Пусть заданы вероятности

$$\begin{aligned} P(\xi = j) &= p_j, & j &= 1, \dots, n, \\ P(\eta_i = s \mid \xi = j) &= q_{ijs}, & s &\in \mathbb{Z}, \\ \sum_{s \in \mathbb{Z}} q_{ijs} &= 1, & i &= 1, \dots, n, \quad j = 1, \dots, n. \end{aligned}$$

Пусть $D \subset \{1, \dots, n\}$ – множество, состоящее из номеров вершин, соответствующих потенциально возможным точкам доступа. Обозначим $E(d(v_\xi, u) \mid \bar{\eta}_D)$ – условное математическое ожидание расстояния между вершиной v_ξ графа и вершиной u относительно случайного вектора $\bar{\eta}_D$, $\bar{\eta}_D = \{\eta_i, i \in D\}$.

Пусть система локации после регистрации вектора измерений расстояний $\bar{s}_D \in \mathbb{Z}^{|D|}$ в качестве искомой возвращает вершину u с минимальным значением $E(d(v_\xi, u) \mid \bar{s}_D)$. Обозначим

$$\Delta(\bar{\eta}_D) = \min_{u \in V} \{E(d(v_\xi, u) \mid \bar{\eta}_D)\}.$$

Математическое ожидание погрешности локации мобильного объекта равно:

$$E(\Delta(\bar{\eta}_D)) = \sum_{\bar{s}_D \in \mathbb{Z}^{|D|}} \Delta(\bar{s}_D) P(\bar{\eta}_D = \bar{s}_D).$$

Обозначим

$$\beta'(\varepsilon) = \min\{|D| : D \subset \{1, \dots, n\}, E(\Delta(\bar{\eta}_D)) \leq \varepsilon\}.$$

Значение $\beta'(\varepsilon)$ существует при $\varepsilon \geq \varepsilon^* = E(\Delta(\bar{\eta}_V))$.

В результате можно поставить две оптимизационные задачи.

Задача 1. При заданном значении $\varepsilon \geq \varepsilon^*$ определить $\beta'(\varepsilon)$ и соответствующее множество вершин D .

Задача 2. При заданном m ($1 \leq m \leq n$) найти множество вершин D , для которого $|D| = m$ и $E(\Delta(\bar{\eta}_D))$ примет минимальное значение.

Решение поставленных задач позволяет найти оптимальную расстановку точек доступа системы локации с учетом погрешности определения местоположения объекта.

Исследования выполнялись в рамках реализации программы стратегического развития ПетрГУ на 2012 – 2016 годы.

Список литературы

- [1] Малодушев С.В., Рогов А.А. Определение локации в корпоративных Wi-Fi сетях // Вестник южно-уральского государственного университета. Серия: Программирование. 2016. Т. 9, № 1. С. 92–104.
- [2] R. Battiti, M. Brunato and A. Delai, “Optimal wireless access point placement for location-dependent services”, Technical Report DIT-03?052, 2003.
- [3] Klein D. J., Yi E. A comparison on metric dimension of graphs, line graphs, and line graphs of the subdivision graphs // European Journal of Pure and Applied Mathematics. 2012. V. 5. N 3. pp. 302–316.
- [4] Slater P. J. Leaves of trees // Proc. 6th Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1975), Congressus Numerantium 14, Winnipeg: Utilitas Math., pp. 549–559.

Методы распознавания строк и алгоритмы классификации символов в стенограммах

М. Б. Гиппиев

Петрозаводский государственный университет,
Петрозаводск, Россия

E-mail: gippiev@gmail.com

Преобразование изображений рукописного, машинописного или печатного текста в текстовые данные сегодня является обыденным явлением. Разработано большое количество различных программных средств, осуществляющих оптическое распознавание символов (optical character recognition, OCR). К системам, поддерживающим русский язык, можно отнести «ABBYY FineReader», «CuneiForm», «Google Tesseract» и другие. Данные системы успешно справляются с распознаванием печатных текстов, даже при наличии в них искажений, характерных для сканированных печатных документов. Однако проблемы возникают при попытке распознавания рукописных текстов, к которым относятся стенографические документы, то есть документы, записанные посредством особых знаков и целого ряда сокращений.

Одной из таких проблем является сегментация строк. Обусловлено это тем, что любой рукописный документ несет в себе индивидуальные особенности, которые связаны с привычками автора, скоростью письма, аккуратностью и некоторыми другими факторами. На определение строк влияют: наклон текста в ту или иную сторону, заваливание, исправление и зачеркивание текста. Поэтому методы выделения строк в печатных текстах, используемые в системах OCR, окажутся бесполезными в случае стенографических документов.

Основным методом, используемым при распознавании печатных строк системами OCR, можно назвать метод определения строк при помощи построения проекции символов на ось перпендикулярную строкам текста. Пики в построенной проекции будут соответствовать строкам. Данный метод не подходит для выделения строк в стенограммах,

так как, он подразумевает, что строки расположены строго горизонтально, а в случае стенограмм данное условие выполняется необязательно, и строки могут быть выпуклыми, вогнутыми или быть расположены под некоторым углом.

В связи с существующими проблемами для распознавания строк в стенографических документах были разработаны метод построения графа связей и метод построения графа связей с последующим объединением строк. Главная идея метода построения графа связей заключается в нахождении расстояний между символами в стенограмме и последовательном объединении символов, начиная с символов, расстояние между которыми минимально. Метод построения графа связей с последующим объединением строк отличается от метода построения графа связей наличием постобработки строк, заключающейся в объединении непересекающихся по горизонтали строк, расстояния между которыми не превышают максимально заданного.

При распознавании строк возникает задача определения к какому типу (основной, надстрочный, подстрочный или комплексный) относится каждый символ в строке. Для решения данной задачи были разработаны алгоритмы классификации символов методами одинарной аппроксимации, двойной аппроксимации, одинарной аппроксимации с фиксированной высотой строк, а также основанные на них нейросетевые алгоритмы классификации символов. В случае алгоритма классификации символов методом одинарной аппроксимации для каждой строки строится линия аппроксимации по центрам ее символов и в зависимости от расположения каждого символа относительно линии аппроксимации определяется его тип. В случае алгоритмов классификации символов методами двойной аппроксимации и одинарной аппроксимации с фиксированной высотой строк определяются границы каждой строки и в зависимости от расположения каждого символа относительно границ строки определяется его тип. При использовании метода двойной аппроксимации верхней границей строки является линия аппроксимации, построенная по точкам, являющимся серединами верхних сторон прямоугольников, описывающих символы строки, а нижней границей строки является линия аппроксимации, построенная по точкам, являющимся серединами нижних сторон таких прямоугольников. При использовании метода одинарной аппроксимации с фиксированной высотой строк расстояния от линии аппроксимации, построенной по центрам символов строки, до границ строки являются постоянными для всего документа и определяются статистическим способом.

В таблице 1 представлены оценки качества работы алгоритмов классификации символов методом одинарной аппроксимации и мето-

дом двойной аппроксимации для двух стенограмм. К представленным оценкам относятся точность, полнота, F-мера для каждого типа символов, а также средняя F-мера.

Таблица 1: Оценки алгоритмов классификации символов

Оценка	Алгоритм классификации символов методом одинарной аппроксимации			Алгоритм классификации символов методом двойной аппроксимации		
Тип символов	Осн.	Надстр.	Подстр.	Осн.	Надстр.	Подстр.
Стенограмма 1						
Полнота	0,7838	0,5333	0,4854	0,952	0,2667	0,4078
Точность	0,8651	0,1143	0,5495	0,8617	0,2	0,84
F-мера	0,8225	0,1882	0,5155	0,9046	0,2286	0,549
Ср. F-мера	0,5087			0,5607		
Стенограмма 2						
Полнота	0,9422	0,7333	0,3784	0,9558	0,6667	0,4595
Точность	0,9142	0,55	0,6087	0,9183	0,6667	0,68
F-мера	0,928	0,6286	0,4667	0,9367	0,6667	0,5484
Ср. F-мера	0,6744			0,7172		

Работа выполнена при поддержке Программы стратегического развития (ПСР) ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности на 2012–2016 гг.

Список литературы

- [1] *Likforman-Sulem L., Zahour A., Taconet B.* Text line segmentation of historical documents: a survey // International Journal of Document Analysis and Recognition (IJ DAR). 2007. Vol. 9 (2-4). P. 123–138.
- [2] *Гунпиев М. Б., Жуков А. В., Rogov A. A., Скабин А. В.* Распознавание строк в стенографических документах // Современные проблемы науки и образования, 2013. № 4 [Электронный ресурс]. Режим доступа: www.science-education.ru/110-9725
- [3] *Гунпиев М. Б., Rogov A. A.* Экспериментальное сравнение алгоритмов классификации символов в стенографических документах // Ученые записки Петрозаводского государственного университета. Сер. «Естественные и технические науки», 2015. № 4 (149). С. 115–121.

Clique Numbers of Random Subgraphs of Distance Graphs

A. S. Gusev

MSU, Moscow, Russia

E-mail: aretalogus@inbox.ru

In our talk, we are concerned with the classical Nelson–Hadwiger problem on finding the chromatic numbers of distance graphs in $\mathbb{R}^n$. We study a class of graphs $G(n, r, s) = (V(n, r), E(n, r, s))$ defined as follows:

$$V(n, r) = \{x = (x_1, x_2, \dots, x_n) : x_i \in \{0, 1\}, x_1 + x_2 + \dots + x_n = r\},$$

$$E(n, r, s) = \{\{x, y\} : (x, y) = s\},$$

where (x, y) is the Euclidean scalar product.

We also study the random graphs $\mathcal{G}(G(n, r, s), p)$ whose edges are chosen independently from the set $E(n, r, s)$, each with probability p . If r and s are constant, we obtain sharp asymptotic bounds for the clique numbers of such random graphs. For $r = r(n)$ and $s = s(n)$, we also get a series of asymptotic results. In our talk, we shall present these results.

First-order Logic of Random Graphs

M. E. Zhukovskii

MIPT, Moscow, Russia

E-mail: zhukmax@gmail.com

The random graph $G(n, p)$ *obeys Zero-One Law* if for each first-order property its probability tends to 0 or tends to 1. The random graph $G(n, p)$ *obeys Limit Law* if for each first-order property its probability converges. In 1969 Y.V. Glebskii, D.I. Kogan, M.I. Liagonkii, V.A. Talanov (and independently in 1976 R. Fagin) proved that $G(n, p)$ obeys Zero-One Law, if p does not depend on n . This result can be extended on p such that $\min\{p, 1 - p\}n^\alpha \rightarrow \infty$ for all $\alpha > 0$. In 1988 S. Shelah and J. Spencer showed that if α is an *irrational* positive number and $p(n) = n^{-\alpha+o(1)}$, then $G(n, n^{-\alpha})$ obeys Zero-One Law. If α is positive and rational, then $G(n, n^{-\alpha})$ does not obey Zero-One Law for $\alpha \leq 1$ and $\alpha = 1 + \frac{1}{m}$ for some $m \in \mathbb{N}$. For all other rational α Zero-One Law holds. Moreover, S. Shelah and J. Spencer proved that $G(n, n^{-\alpha})$ does not obey Limit Law for rational $\alpha \in (0, 1)$. However, in 1992 J. Lynch proved that $G(n, n^{-\alpha})$ obeys Limit Law for $\alpha = 1$ and $\alpha = 1 + \frac{1}{m}$.

The random graph $G(n, p)$ *obeys Zero-One k -Law* if for each first-order property which is expressed by a first-order formula with quantifier depth at most k its probability tends to 0 or tends to 1. The random graph $G(n, p)$ *obeys Limit k -Law* if for each first-order property which is expressed by a first-order formula with quantifier depth at most k its probability converges.

In the talk, we present the results on Zero-One k -Law and Limit k -Law for $G(n, n^{-\alpha})$, $\alpha > 0$, and the distributions of positive rational numbers $0 < \alpha < 1$ such that Zero-One k -Law for $G(n, n^{-\alpha})$ does not hold (in this case, we say that α is in *k -spectrum*). In particular, we found minimal and maximal points of k -spectrum and bounds on minimal and maximal limit points of k -spectrum. Moreover, we prove that the minimal k such that k -spectrum is infinite is either 4 or 5.

Итерации случайных отображений конечных множеств

А. М. Зубков, А. А. Серов

Математический институт им. В.А. Стеклова РАН, Москва, Россия

E-mail: zubkov@mi.ras.ru, serov@mi.ras.ru

В докладе планируется сделать обзор задач, связанных со случайными отображениями конечных множеств, результатов, полученных при их исследовании, а также областей применений этих результатов.

Пусть $\mathcal{N}$ — множество из N элементов и $F_1, F_2, \dots$ — последовательность случайных независимых отображений $\mathcal{N} \rightarrow \mathcal{N}$. Для подмножества $S_0 \subseteq \mathcal{N}$ рассматриваются различные характеристики последовательности его образов $S_k = F_k(\dots F_2(F_1(S_0))\dots)$, $k = 1, 2, \dots$. Примерами таких характеристик являются числа $|S_k|$ элементов множеств S_k , времена достижения последовательностью $|S_k|$ заданного уровня (в частности, уровня 1), числа элементов в объединениях $\Psi_k = S_1 \cup \dots \cup S_k$, $k = 1, 2, \dots$

Влияние случайного суммирования на регенеративную оценку эффективной пропускной способности узла сети

К. А. Калинина

Институт прикладных математических исследований,
Карельский научный центр РАН, Петрозаводск, Россия

E-mail: smesharikk@mail.ru

Введение

В статье рассматривается проблема переоценивания эффективной пропускной способности (ЭПС) узла сети регенеративным методом, ранее исследованная в [1, 4, 2, 5]. Напомним, что под ЭПС понимается такое значение мощности (пропускной способности) обслуживающего прибора C , для которого выполняется следующее требование качества обслуживания: вероятность превышения стационарным процессом нагрузки W некоторого (заданного) высокого уровня b не превышает (заданного) малого значения

$$P_b := \mathbb{P}(W > b) \leq \Gamma. \quad (1)$$

Требование (1) является важным в системах высокой надежности, его нарушение крайне нежелательно. Это привело к тому, что самой предпочтительной из возможных стала регенеративная оценка ЭПС. (Сравнение разных оценок ЭПС на основе результатов моделирования представлено, например, в [2].) Ранее было установлено наличие переоценивания Δ (избыточной пропускной способности) в случае использования этого метода в сетях разных типов [1, 4] ($10\% \leq \Delta \leq 17\%$). Это означает, что регенеративная оценка гарантирует выполнение требования (1) с небольшим запасом.

Оценивание ЭПС на основе случайного суммирования

На основе исследований свойств регенеративной оценки и сравнения ее с оценками, построенными другими способами [2, 4, 5], можно предположить, что возможной и наиболее вероятной причиной появления избыточной пропускной способности является случайное суммирование данных, принадлежащий одному циклу регенерации. Открытым остается вопрос о зависимости полученного переоценивания от способа реализации случайного суммирования. Для изучения данного вопроса построим оценку ЭПС, основанную на искусственных случайных циклах (вместо регенеративных циклов, которые получаются естественным путем). Для изучения влияния случайного суммирования будем менять распределение длин циклов.

Мы моделируем тандемную сеть из двух узлов (выходной процесс первого узла создает регенеративный входной процесс во второй узел, для которого и необходимо рассчитать ЭПС). В таблице 1 представлены результаты моделирования для длин искусственных циклов с распределением Вейбулла с функцией $F(x) = 1 - e^{-0.2x^2}$. Таблица 2 содержит результаты для случая, когда длины циклов имеют распределение Парето с функцией $F(x) = 1 - (\frac{1}{x})^{1.7}$ (для $x \geq 1$).

Таблица 1: Искусственные циклы, Weibull β

Γ	$\hat{C}_{rnd}$	$\hat{\Gamma}_{rnd}$	Δ_{rnd}	$\hat{\Gamma}_{\Delta_{rnd}}$
10^{-4}	5.70	$1.39 \cdot 10^{-5}$	0.14	$2.58 \cdot 10^{-4}$
10^{-5}	6.16	$4.78 \cdot 10^{-7}$	0.16	$4.22 \cdot 10^{-5}$
10^{-6}	6.42	$2.08 \cdot 10^{-8}$	0.18	$1.05 \cdot 10^{-6}$

Таблица 2: Искусственные циклы, Pareto β

Γ	$\hat{C}_{rnd}$	$\hat{\Gamma}_{rnd}$	Δ_{rnd}	$\hat{\Gamma}_{\Delta_{rnd}}$
10^{-4}	6.18	$1.64 \cdot 10^{-6}$	0.19	$5.37 \cdot 10^{-4}$
10^{-5}	7.03	$5.02 \cdot 10^{-8}$	0.23	$7.23 \cdot 10^{-5}$
10^{-6}	6.82	$3.68 \cdot 10^{-9}$	0.28	$8.26 \cdot 10^{-6}$

Результаты моделирования показали, что переоценивание в случае искусственных циклов сильно зависит от типа распределения длин циклов и в целом сравнимо с переоцениванием в случае использования регенерации: $0.14 \leq \Delta_{rnd} \leq 0.18$, хотя и достигает достаточно высоких значений 25%- 28% в случае распределения Парето (см. таб.

лицу 2). В связи с этим регенеративная оценка все же остается более предпочтительной.

Выводы

Изучение свойств регенеративной оценки и результаты моделирования показали, что наиболее вероятной причиной установленного переоценивания является случайное суммирование данных на циклах регенерации. Однако оценка, построенная на основе искусственных случайных циклов, сильно зависит от распределения количества слагаемых, что приводит к нежелательному увеличению переоценивания в некоторых случаях. Этот факт позволяет считать регенеративную оценку ЭПС более предпочтительной с точки зрения гарантии выполнения требования качества обслуживания (1) и относительно умеренной величины переоценивания.

Список литературы

- [1] Borodina A., Kalinina K., Morozov E. *On the accuracy of the effective bandwidth regenerative estimation*, ICUMT, Saint-Petersburg: 2014. P. 652-656.
- [2] Kalinina K., Morozov E., Rykov V. *Effective bandwidth estimation in highly reliable regenerative networks* // The Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management, Israel, 2016: Proceedings, P. 323-327.
- [3] Kelly F. *Notes on effective bandwidths* // Stochastic Networks: Theory and Applications / Ed. F. P. Kelly, S. Zachary, I. B. Ziedins. Royal Statistical Society Lecture Notes Series, 4: Oxford University Press, 1996. P. 141–168.
- [4] Morozov E., Kalinina K. *On the effective bandwidth estimation in communication network* // 29th European Conference on Modelling and Simulation ECMS 2015: Proceedings, P.423-429.
- [5] Калинина К. *Об эффективной пропускной способности узлов коммуникационной сети*, Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем: материалы Всероссийской конференции с международным участием. М.: РУДН, 2015, 24-26.

Об оптимальных стратегиях двухальтернативной обработки данных на начальном этапе

А. В. Колногоров

Новгородский государственный университет имени Ярослава
Мудрого, Великий Новгород, Россия

E-mail: Alexander.Kolnogorov@novsu.ru

Рассматривается минимаксная постановка для задачи управления обработкой больших объемов данных, если для обработки имеются два альтернативных метода с различными фиксированными, но априори неизвестными эффективностями [1]. Требуется в процессе обработки определить более эффективный метод и обеспечить его преимущественное применение.

Для решения задачи может быть использована основная теорема теории игр, согласно которой минимаксный риск совпадает с байесовским, вычисленным относительно наихудшего априорного распределения, а минимаксная стратегия совпадает с соответствующей байесовской. В случае близких эффективностей используемых методов данные можно обрабатывать группами, в том числе параллельно. Если число групп достаточно велико, то такая обработка выполняется практически без потери качества управления, т.е. без увеличения минимаксного риска. Например, 50000 данных можно обработать за 50 этапов партиями по 1000 данных приблизительно с такими же максимальными потерями, как если бы данные обрабатывались оптимально по одному.

Если гарантировать близость эффективностей используемых методов нельзя, то полные потери могут существенно расти за счет начального этапа управления, когда к группам данных применяются по очереди оба метода. Чтобы этого не произошло, управление на начальном этапе следует модифицировать.

Первый подход состоит в уменьшении объемов партий данных, обрабатываемых на начальном этапе. Второй подход состоит в использовании стратегии, которая на начальном этапе производит последовательное сравнение методов и отбрасывает худший метод, если раз-

ница между текущими оценками эффективностей превысит некоторый порог. Показано, что такие модификации стратегии управления практически не увеличивают значение минимаксного риска.

Работа выполнена при финансовой поддержке проектной части государственного задания в сфере научной активности Министерства образования и науки Российской Федерации, проект № 1.949.2014/К.

Список литературы

- [1] Колногоров А.В. Робастное параллельное управление в случайной среде (задаче о двуруком бандите) // Автоматика и телемеханика. 2012, № 4, С. 114–130.

Urn Schemes and Partitions of a Positive Integer

A. V. Kolchin

Moscow Automobile and Road Construction
State Technical University — MADI,
Moscow, Russia

E-mail: andrei.kolchin@gmail.com

Any representation of a positive integer in the form of a sum of positive integers (parts) is referred to as a partition of the number, which is primarily of combinatorial and number theoretical nature. Questions concerning the partitions have played an important part in mathematics.

Gian-Carlo Rota wrote in his preface to the monograph [1] that the theory of partitions is one of the very few branches of mathematics that can be appreciated by anyone who is endowed with little more than a lively interest in the subject. Its applications are found wherever discrete objects are to be counted or classified, whether in the molecular and the atomic studies of matter, in the theory of numbers, or in combinatorial problems from all sources.

Partitions are investigated in combinatorics and in the theory of numbers; classical combinatorial problems concern counting partitions, and in the theory of numbers, problems on additive representations of numbers are being solved under arithmetical constraints imposed on the parts. Serious difficulties may arise while solving problems on partitions, though; so a great body of special methods in the theory of partitions have been elaborated (see, e.g., [1]).

If one considers the applications of partitions in various branches of mathematics, one is struck by the interplay of combinatorial and asymptotic methods.

We consider the problem on the number of partitions of a positive integer n into s positive integer parts which do not exceed a given integer r (those partitions which differ in the order of parts only are counted as one). Let $C_{n,s,r}$ stand for the number of these partitions.

It turns out that there is an easy way to arrive at a compact asymptotic expression for the number of these partitions with the use of probabilistic

reasoning (see, e.g., [2, 3]). It is not difficult to see, indeed, that the partition of a number is described by the classical scheme of equiprobable allocation of n particles into s cells under the condition that each cell contains not more than r particles.

Let $\xi_1, \xi_2, \dots, \xi_s$ be independent random variables which take values $1, 2, \dots, r$ with equal probabilities, which can be considered as the contents of the corresponding cells in the above allocation scheme. It is easily seen that

$$\mathbf{P}\{\xi_1 + \dots + \xi_s = n\} = \sum_{\substack{k_1 + \dots + k_s = n \\ k_1, \dots, k_s \leq r}} \mathbf{P}\{\xi_1 = k_1, \dots, \xi_s = k_s\} = C_{n,s,r} \frac{1}{r^s}. \quad (1)$$

In addition, $\mathbf{E}\xi_1 = (r+1)/2 = m$, $\mathbf{Var}\xi_1 = (r^2 - 1)/12 = \sigma^2$. So, in order to investigate the asymptotic behaviour of $C_{n,s,r}$ one is able to utilise the well-developed apparatus of local limit theorems of probability theory (see, e.g., [4, 5, 6]).

First, let us analyse the behaviour of the number of partitions in the so-called ‘central’ domain of variation of the parameters.

It is obvious that

$$\mathbf{P}\{\xi_1 + \dots + \xi_s = n\} = \mathbf{P}\left\{\frac{\xi_1 + \dots + \xi_s - sm}{\sigma} = \frac{n - sm}{\sigma}\right\}.$$

Let $x = (n - sm)/\sigma$. Using the above relations for the expectation m and variance σ^2 , we find that $x = (2n - s(r+1))/\sqrt{(r^2 - 1)/3}$.

For the sake of simplicity, let r be fixed. From relation (1), with the use of the local convergence to the standard normal law we see that for fixed r , while the parameters n and s tend to infinity in such a way that the ratio of n to s lies in some finite interval, the relation

$$\frac{1}{r^s} C_{n,s,r} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (1 + o(1))$$

holds true uniformly with respect to all x in an arbitrary fixed finite interval. Thus, the following assertion is valid.

Theorem 1. *Let r be a fixed integer. If the parameters n and s tend to infinity in such a manner that the ratio of n to s remains inside some finite interval, then the number $C_{n,s,r}$ of partitions of a positive integer n into s positive integer parts not exceeding r obeys the equality*

$$C_{n,s,r} = \frac{r^s}{\sqrt{2\pi}} e^{-x^2/2} (1 + o(1)),$$

which holds true uniformly in all $x = (2n - s(r+1))/\sqrt{(r^2 - 1)/3}$ which fall into an arbitrary fixed finite interval.

The author believes that the analysis of behaviour of this number of partitions in other domains of variation of the parameters n , s , r certainly deserves a separate presentation.

In conclusion, it is worth noticing that the reverse approach which consists of studying asymptotic properties of characteristics of a classical allocation scheme with the use of apparatus of the theory of partitions of a number seems to be very fruitful as well. For example, in [7, 8, 9], with the use of very careful estimates at the roots of unity, an asymptotic formula was obtained for the number $A(n, s, r)$ of partitions of n into at most s parts each not larger than r in the domain where n is close to $\frac{1}{2}sr$. The relation $C_{n,s,r} = A(n, s, r) - A(n, s - 1, r)$ of course holds.

Список литературы

- [1] G. E. Andrews, *The Theory of Partitions*. Addison–Wesley, Reading, Mass., 1976.
- [2] J. Spencer, P. Erdős, *Probabilistic Methods in Combinatorics*. Akadémiai Kiadó, Budapest, 1974.
- [3] N. Alon, J. Spencer, *The Probabilistic Method*. Wiley, New York, 1992.
- [4] B. V. Gnedenko, A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Addison–Wesley, Cambridge, 1954.
- [5] V. V. Petrov, *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Clarendon Press, Oxford, 1995.
- [6] V. F. Kolchin, *Asymptotic Methods in Probability Theory*. MIEE, Moscow, 1988 (in Russian).
- [7] G. Almkvist, *How to compute the number of restricted partitions*. Preprint, Dept. Math., Lunds Tekniska Högskola, 1988.
- [8] G. Almkvist, *A Hardy–Ramanujan formula for restricted partitions*. Preprint, Dept. Math., Lunds Tekniska Högskola, 1988.
- [9] G. Almkvist, G. E. Andrews, A Hardy–Ramanujan formula for restricted partitions. *J. Number Theory* (1991) **38**, 135–144.

Правила коллективного выбора на основе минимальных выигрывающих коалиций кандидатов

А. Ю. Кондратьев

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: kondratev@krc.karelia.ru

Рассматривается задача выбора победителя среди кандидатов по результатам голосования избирателей. Каждый избиратель заполняет бюллетень, в котором указывает строгий линейный порядок своих личных предпочтений на множестве кандидатов. Правило коллективного выбора для каждого возможного профиля предпочтений избирателей однозначно назначает одного или нескольких победителей. Здесь существует два основных подхода.

В подходе на основе принципа большинства победителем объявляется кандидат, набравший больше половины голосов избирателей на первом месте в бюллетене. Если такого кандидата нет, то определенным образом из профиля последовательно исключаются один или несколько кандидатов, пока такой кандидат не появится. На выборах, например, президента России по результатам голосования в первом туре исключаются все кандидаты кроме двух, набравших больше всех голосов.

В подходе на основе принципа Кондорсе победителем объявляется кандидат, который побеждает любого другого кандидата при парном сравнении. Парадоксом Кондорсе называется ситуация, когда победителя по Кондорсе не существует.

Мы предлагаем рассматривать всевозможные минимальные выигрывающие коалиции кандидатов, которые могут выставить единого кандидата, который побеждает по принципу большинства (или по принципу Кондорсе). Выигрыш коалиции это количество голосов, набранных единым кандидатом. Затем для определения сильнейшего кандидата мы применяем методы кооперативной теории игр. Такой способ определения победителя обладает рядом положительных свойств. В частности, не нужно последовательно исключать кандидатов.

Модель синсета: геометрия и статистика

А. А. Крижановский,
А. Н. Кириллов

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: andew.krizhanovsky@gmail.com, kirillov@krc.karelia.ru

Трудности в определении понятия синонима, влекущие к неоднозначности в понимании этого термина, приводят к необходимости введения некоторой формализации, которая позволила бы дать количественные характеристики для описания соотношений между словами. В докладе предложен подход к математическому моделированию понятия синсета (набора синонимов).

Для разработки модели синсета слова представляются в виде векторов. Рассмотрим некоторый словарь и пронумеруем некоторым образом все слова, входящие в него. Пусть $|D|$ — количество слов в словаре, i — номер слова.

Определение 1. Векторным словарем назовем множество $D = \{w_i \in \mathbb{R}^{|D|}\}$, где i -ая компонента вектора w_i равна 1, а остальные компоненты — нули.

Задача векторного представления слов состоит в построении линейного отображения $L : D \rightarrow \mathbb{R}^N$, где $N \ll |D|$, а вектор $v = L(w)$, $w \in D$, v имеет компоненты $v_j \in \mathbb{R}$. Полагая что линейное отображение L реализуется с помощью матрицы W , получаем $v = Ww$, причем для нахождения матрицы W используют различные методы, в частности, основанные на нейронных сетях. Наибольшую популярность в самое последнее время приобрели CBOW и Skip-gram методы, предложенные в [1] и являющиеся, по сути, модификацией метода максимального правдоподобия. При этом в методе Skip-gram матрица W максимизирует функцию $F(W)$ вида

$$F(W) = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \ln p(w_{t+j} | w_t)$$

$$p(w_{t+j}|w_t) = \frac{\exp u_{t+j}}{\sum_{i=1}^{|D|} \exp u_i}, \quad u_i = (Ww_i, Ww_t)$$

где $(\cdot, \cdot)$ — символ скалярного произведения, T — объем обучающего контекста. Здесь по слову w_t находится содержащий его контекст, составляющий «окно» размера $2c$. В методе CBOW (continuous bag of words), наоборот, по контексту находится слово, входящее в него. Для максимизации $F(W)$ используется метод стохастического градиентного спуска.

Введем обозначения для нормированных сумм векторов: $M((a_i), n) = \frac{\sum_{i=1}^n a_i}{\|\sum_{i=1}^n a_i\|}$, $M((a_i, v), n+1) = \frac{\sum_{i=1}^n a_i + v}{\|\sum_{i=1}^n a_i + v\|}$. Рассмотрим синсет $S = \{v_k, k = 1, \dots, |S|\}$.

Определение 2. *Внутренностью $IntS$ синсета S называется множество всех векторов $v_l \in S$, удовлетворяющих условию*

$$IntS = \{v_l \in S : (M((v_{i_s}), q) - M((V_{i_s}, v_l), n+1), M((v_{j_p}), r)) < 0\}$$

для любых дизъюнктивных разбиений $S \setminus \{v_l\} = \{v_{i_s}\} \sqcup \{v_{j_p}\}$, $s = 1, \dots, q$, $p = 1, \dots, r$, $q + r = |S| - 1$, $i_s \neq j_p$.

Смысл определения состоит в том, что добавление вектора $v_l \in S$ в любое из двух подмножеств множества $S \setminus \{v_l\}$, образующих его дизъюнктивное разбиение, уменьшает расстояние между этими подмножествами.

Расстояние между множествами $\{a_1, \dots, a_n\}$ и $\{b_1, \dots, b_m\}$ определим значением скалярного произведения нормированных средних векторов этих множеств $(M((a_i), n), M((b_j), m))$. Введём понятие **ранга синонима** $v \in S$. Дизъюнктивное разбиение на два множества будем называть разбиением. Пусть $P_v = \{p_i, i = 1, \dots, 2^{n-2} - 1\}$ — множество всех пронумерованных каким-либо образом разбиений $(n-1)$ -элементного множества $S \setminus v$.

Пусть sim_i — расстояние между элементами разбиения p_i ; $sim_i(v, 1)$, $sim_i(v, 2)$ — расстояния между одним из элементов разбиения и множеством, являющимся объединением другого элемента и слова v .

Введем функцию $r : P_v \rightarrow \{-1, 0, 1\}$. Пусть $r(p_i) = r_i$, где $r_i = -1, 1$ или 0 , если добавление слова v к каждому из элементов разбиения p_i увеличивает, уменьшает расстояние sim_i или добавление к одному элементу увеличивает, а к другому — уменьшает расстояние sim_i , соответственно.

Определение 3. *Рангом синонима $v \in S$ называется целое число вида*

$$rank v = \sum_{i=1}^{|S \setminus v|} r_i$$

Легко видеть, что если $v \in \text{Int}S$, то $\text{rank } v = d = 2^{|S|-2} - 1$ — это число всех непустых дизъюнктивных разбиений $(|S| - 1)$ —элементного множества $S \setminus v$, т.е. $\text{rank } v$ максимален и совпадает с числом Стирлинга второго рода: $\{n\}_k = \{n\}_2$, где n — мощность разбиваемого множества, а k — число подмножеств, здесь два.

Определение 4. *Границей синсета S называется множество $\text{Bd}S = S \setminus \text{Int}S$ (synset's edge).*

Поясним, что граница синсета включает такие синонимы из S , добавление которых в любые непустые разбиения синсета не сближает средние векторы этих разбиений.

Пример 1. Дан синсет (баюкать, убаюкивать, укачивать, усыплять). Значение ранга для слова "усыплять" в этом синсете равно -1. В $\text{Int}S$ вошли векторы, соответствующие словам "убаюкивать" и "баюкать". Очевидно, что эта пара наиболее близка по смыслу среди четвёрки слов: "баюкать" "убаюкивать" "укачивать" "усыплять". Значения ранга и степени центральности для слов в синсете представлены в Табл. 1.

synonym	усыплять	укачивать	убаюкивать	баюкать
rank	-1	1	3	3
IntS	—	—	+	+

Таблица 1: Ранг (rank) и включение в $\text{Int}S$ для слов синсета

Проведены эксперименты для апробации предложенной модели синсета. Были использованы две матрицы W , построенные по корпусу НКРЯ и по Веб корпусу.

Для нескольких тысяч синсетов, извлечённых из Русского Викисловаря, вычислен ранг и определена внутренность синсета $\text{Int}S$. Эксперименты показали, что для редких в корпусе слов $\text{Int}S$ может оказаться пустым. Интересно исследовать зависимость качества разрешения лексической многозначности от $\text{Int}S$.

Работа поддержана грантом РГНФ (проект № 15-04-12006).

Список литературы

- [1] Mikolov, Tomas and Chen, Kai and Corrado, Greg and Dean, Jeffrey, *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781, 2013.

Оптимизация параллельного управления в случайной среде с нормально распределенными доходами и различными дисперсиями

А. Н. Лазутченко

Новгородский государственный университет имени Ярослава
Мудрого, Великий Новгород, Россия

E-mail: aleskey@hotmail.ru

Рассматривается задача об оптимальном управлении в случайной среде с нормально распределенными доходами в минимаксной постановке, если имеются два альтернативных действия, которым соответствуют доходы с известными дисперсиями и неизвестными математическими ожиданиями. Такие среды имеет смысл рассматривать в силу того, что использование нормального распределения тесно связано с параллельной обработкой данных. Предположим, что необходимо обработать объем данных T , разделенный на $M \cdot K$ данных, результат обработки которых имеет бинарное распределение с вероятностями $0 < p_1, p_2 < 1$, где M — количество групп, K — количество данных в одной группе. Для всех K данных в одной группе мы можем применить параллельно один и тот же способ обработки, сформировав новый процесс. В силу центральной предельной теоремы распределение этого процесса будет близким к нормальному, имеющему плотность

$$f_{D_\ell}(x|m_\ell) = \frac{1}{\sqrt{2\pi D_\ell}} e^{-\frac{(x-m_\ell)^2}{2D_\ell}},$$

где m_ℓ — математические ожидания, D_ℓ — дисперсии, $\ell = 1, 2$ — действия среды.

Эта задача для случая единичных дисперсий $D_1 = D_2 = 1$ решена в [2]. Были получены рекуррентные уравнения для вычисления байесовского риска и потерь методами динамического программирования.

В некоторых задачах целью оптимизации выступают не данные, а например, время обработки данных. В таких случаях дисперсии могут быть совершенно произвольными, их значения оцениваются на начальных этапах управления. В работе [1] рассматриваемая задача обобщается на случай равных произвольных дисперсий на действиях. Здесь мы развиваем эти результаты для случайных сред с различными известными дисперсиями, выбираемыми из условия $D_0 \leq D_\ell \leq 1$, $D_0 > 0$. Для этого случая также найдены уравнения для вычисления байесовского риска и потерь, а также стратегии, обеспечивающие оптимальное управление.

Работа выполнена при финансовой поддержке проектной части государственного задания в сфере научной активности Министерства образования и науки Российской Федерации, проект № 1.949.2014/К.

Список литературы

- [1] Лазутченко А. Н. О робастном управлении в случайной среде, характеризующейся доходами с различными дисперсиями // КарНЦ. Сер.: Математическое моделирование и информационные технологии. 2015. № 10. С. 107–113
- [2] Колногоров А. В. Нахождение минимаксных стратегии и риска в случайной среде (задача о двуруком бандите) // АиТ. 2011. № 5. С. 127–138

Локальная предельная теорема для разделимых статистик в полиномиальной схеме

В. В. Лёвин, С. Ю. Гуськов

МГТУ им Баумана, Москва, Россия

E-mail: vladimir.levin.51@mail.ru, ser0432@yandex.ru

Пусть задан случайный вектор $v_N = (v_{1N}, \dots, v_{NN})$, имеющий полиномиальное распределение $M(n, p_{1N}, \dots, p_{NN})$. Изучаются условия применимости локальной предельной теоремы (л.п.т.), при $N \rightarrow \infty$, к распределению разделимой статистики (р.с.):

$$L_N = \sum_{m=1}^N f_{mN}(v_{mN})$$

Здесь $\{f_{mN}(x), \quad m = 1, \dots, N, \quad N \geq 1\}$ - заданное множество функций, определенных для целых значений аргумента x и принимающих целые значения. Понятие разделимых статистик было введено в работах Медведева Ю.И. ([1,2]), изучению асимптотического поведения р.с. посвящено большое число работ (см. обзор [3]). Условия асимптотической нормальности р.с. L_N изучались, в частности, в [4]. Как следует из результатов работы [4], при выполнении следующих условий (1), (2), (3) распределение р.с. L_N является асимптотически нормальным при $n, N \rightarrow \infty$: для некоторых констант $A, B, C > 0$

$$n/N = \alpha + o(1), \quad 0 < \alpha < \infty, \quad (1)$$

$$Np_{mN} \leq A, \quad m = 1, \dots, N, \quad N \geq 1, \quad (2)$$

$$|f_{mN}(x)| < Be^{Cx}. \quad (3)$$

Определение. Множество $\mathcal{A}$, $\text{card}(\mathcal{A}) \geq 2$, будем называть решетчатым с шагом $h \geq 1$, если н.о.д. $\{m - n > 0 | m, n \in \mathcal{A}\} = h$.

Обозначим D_{mN} - множество значений $f_{mN}(x)$, $m = 1, \dots, N$, $N \geq 1$. Сформулируем условие (4), которое совместно с условиями (1), (2), (3), обеспечивает справедливость л.п.т. для р.с. L_N .

Существует δ , $0 < \delta < 1$, и конечное множество целых чисел D , являющееся решетчатым с шагом $h = 1$, такие, что

$$D \subseteq D_{mN}, \quad m = 1, \dots, N, \quad N \geq 1,$$

и

$$\inf\{P(v_{mN} = k), \quad k \in D, \quad m = 1, \dots, N, \quad N \geq 1\} \geq \delta. \quad (4)$$

Список литературы

1. Медведев Ю.И. Разделимые статистики в полиномиальной схеме I // Теор.вероятн. и ее примен.. 1977. Т. 22, № 1. С. 3 – 17;
2. Медведев Ю.И., Разделимые статистики в полиномиальной схеме II // Теор.вероятн. и ее примен. 1977. Т. 22. № 3. С. 623 – 631;
3. В.А.Иванов, Г.И.Ивченко, Ю.И.Медведев, Дискретные задачи в теории вероятностей. Итоги науки и техники. Сер. теор. вероятн. мат. стат. теор. кибернет. 1984. Т. 22. С. 3 – 60.
4. Ивченко Г.И., Левин В.В., Асимптотическая нормальность одного класса статистик в полиномиальной схеме // Теор.вероятн. и ее примен. 1976. Т. 21. № 1. С. 190 – 195;

Об оценке вероятности редких событий в гауссовских очередях

О. В. Лукашенко¹, М. Пагано²

¹Институт прикладных математических исследований
Карельского научного центра РАН, Петрозаводск, Россия

²Университет г. Пизы, Италия

E-mail: lukashenko@krc.karelia.ru, m.pagano@iet.unipi.it

Рассматривается задача оценки следующей вероятности

$$\pi(\mathbb{T}) := \mathbb{P}(\forall t \in \mathbb{T} : X_t > t), \quad (1)$$

где X – гауссовский процесс со стационарными приращениями, с ковариационной функцией $\Gamma_{s,t}$, $\mathbb{T} = [0, T] \subseteq \mathbb{R}_+$. Данную вероятность можно интерпретировать как вероятность занятости системы обслуживания с входным потоком X и постоянной скоростью обслуживания. К сожалению, получить выражение в явном виде для данной характеристики затруднительно, поэтому для ее оценки используется метод Монте-Карло. При больших значениях T искомая вероятность мала, а значит возникает проблема оценки редких событий. Прямой метод Монте-Карло в данной ситуации неприменим ввиду неприемлемо большого времени моделирования, необходимого для вычисления оценки с заданной точностью. Поэтому для эффективного вычисления оценок вероятностей редких событий необходимо использовать специальные методы, понижающие дисперсию оценки.

Для вычисления требуемой вероятности предлагается использовать метод условных математических ожиданий [1], который гарантированно дает уменьшение дисперсии оценки. В данном случае метод основан на представлении вероятности $\pi(\mathbb{T})$ в виде математического ожидания от некоторой функции так называемого моста Y гауссовского процесса X [2, 3]. Зафиксируем $\tau \in \mathbb{T}$ и рассмотрим процесс

$$Y_t = X_t - \psi_t X_\tau, \quad (2)$$

где функция ψ_t определяется следующим образом

$$\psi_t := \frac{\Gamma_{t,\tau}}{\Gamma_{\tau,\tau}}.$$

Обозначим

$$\bar{Y} := \sup_{t \in \mathbb{T}} \frac{t - Y_t}{\psi_t}. \quad (3)$$

Можно показать, что искомая вероятность (1) может быть записана в следующем виде

$$\pi(\mathbb{T}) = \mathbb{E} \left[\Phi \left(\frac{\bar{Y}}{\sqrt{\Gamma_{\tau, \tau}}} \right) \right],$$

где Φ – хвост распределения стандартной нормальной случайной величины. На основе сгенерированной последовательности независимых реализаций $\{\bar{Y}^{(i)}, i = 1, \dots, N\}$ случайной величины $\bar{Y}$, строится оценка вероятности $\pi(\mathbb{T})$:

$$\hat{\pi}_N = \frac{1}{N} \sum_{i=1}^N \Phi \left(\frac{\bar{Y}^{(i)}}{\sqrt{\Gamma_{\tau, \tau}}} \right). \quad (4)$$

Список литературы

- [1] Ross, S.M.: Simulation. Elsevier (2006)
- [2] Giordano, S., Gubinelli, M., Pagano, M.: Bridge Monte-Carlo: a novel approach to rare events of Gaussian processes. In: Proc. of the 5th St.Petersburg Workshop on Simulation, pp. 281–286, St. Petersburg, Russia, (2005)
- [3] Giordano, S., Gubinelli, M., Pagano, M.: Rare events of Gaussian processes: a performance comparison between Bridge Monte-Carlo and Importance Sampling. In: Next Generation Teletraffic and Wired/Wireless Advanced Networking, pp. 269–280, St. Petersburg, Russia, (2007)

Математическая модель эвакуации людей из здания

С. В. Малодушев, Р. В. Воронов

Петрозаводский государственный университет, Петрозаводск, Россия

E-mail: svm@petrsu.ru, rvoronov76@gmail.com

В задачах эвакуации часто применяют следующие типы моделей – клеточные автоматы и потоковые, рассмотренные в работах [1, 2]. Данная работа является продолжением исследования [3], направленного на определение путей эвакуации с помощью потоковой модели.

В качестве модели здания рассмотрим связный граф $G = (V, E)$, где V – множество вершин, E – множество ребер, $E \subset V \times V$. Ребрам графа соответствуют части коридоров здания, вершинам графа – места разветвления коридоров или аудиторий, а также места скопления людей на момент начала эвакуации.

Пусть каждая вершина $v \in V$ имеет неотрицательный вес w_{v0} , равный статистической оценке числа людей, определенной при помощи алгоритмов локации, рассмотренных в работе [4]. Процесс эвакуации людей из здания будем рассматривать как поток на графе. Поток определяет количество перемещаемых по ребрам графа частиц (людей), из вершин графа $v \in V$ в вершины множества $U \subseteq V$ (формальное определение потока дано ниже). Данный поток можно интерпретировать как определение направлений движения эвакуируемых людей из аудиторий по коридорам здания к его выходам.

Рассмотрим случай, когда при эвакуации запрещены разделения потоков людей. Тогда ребра графа с положительной величиной потока должны образовывать лес корневых деревьев с корнями из множества U . Все пути в этом лесу идут в сторону корней деревьев.

Если имеется ребро $\{u, v\}$, инцидентное вершинам u и v , то множеству E принадлежат обе пары (u, v) и (v, u) . Порядок вершин в записи ребра однозначно определяет направление движения потока по ребру.

Назовем подграф $H = (V, A)$ графа G покрывающим корневым лесом (A – множество дуг), если:

© Малодушев С. В., Воронов Р. В., 2016

Работа выполнена при финансовой поддержке программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научноисследовательской деятельности.

1. Граф H является ориентированным;
2. Полустепени исхода всех вершин $u \in U$ в графе H равны нулю;
3. Все вершины $v \in V \setminus U$ имеют в графе H полустепень исхода равную одному;
4. Для каждой вершины $v \in V \setminus U$ найдется такая вершина $u \in U$, что в графе H существует путь из v в u .

Вершины $u \in U$ являются корнями деревьев графа H . Каждая из вершин $v \in V$ принадлежит какому-нибудь дереву графа H .

Обозначим Ω – множество покрывающих корневых лесов графа G . Обозначим $\Gamma(v)$ – множество вершин, из которых существуют дуги в вершину v в графе H :

$$\Gamma(v) = \{u \in V \mid (u, v) \in A\}, \quad v \in V. \quad (1)$$

Пусть $a(v)$ – вершина, в которую ведет дуга из вершины $v \in V \setminus U$ в графе H .

Для каждого ребра $(u, v) \in E$ задана Q_{uv} – пропускная способность, т.е. число людей, которые могут пройти по соответствующему ребру коридора за единицу времени. Положим, что $Q_{uv} = Q_{vu}$ для всех $u, v \in E$. Для каждой вершины обозначим w_{vt} – число частиц в вершине v в момент времени t , W_v – ограничение сверху на значения w_{vt} , $t = 1 \dots T$ – множество моментов времени. Пусть T – максимальное время на эвакуацию, T' – фактическое время эвакуации. Пусть $(Z, +)$ – множество целых положительных чисел.

Определение. *Поток* – это функция $x : A \times t \rightarrow (Z, +)$, которая каждой дуге в каждый момент времени ставит в соответствие целое положительное число, обладающее следующими свойствами:

- 1) Ограничение на пропускную способность:

$$0 \leq x_{uv} \leq Q_{uv} \text{ для всех } u, v \in A. \quad (2)$$

- 2) Уравнение баланса:

$$w_{vt} = w_{v,t-1} - x_{va(v)t} + \sum_{u \in \Gamma(v)} x_{uv,t-1} \text{ для всех } v \in V \setminus U, t = 1, \dots, T. \quad (3)$$

$$w_t = w_{v,t-1} + \sum_{u \in \Gamma(v)} x_{uv,t-1} \text{ для всех } v \in U, t = 1, \dots, T. \quad (4)$$

- 3) Ограничение на число частиц в вершинах:

$$w_t \leq W_v \text{ для всех } v \in V, t = 1, \dots, T. \quad (5)$$

4) Величина потока:

$$\sum_{u \in U} w_{uT'} = \sum_{v \in V \setminus U} w_{v0}, \quad T' \leq T. \quad (6)$$

Здесь x_{uv} обозначает значение функции x на паре (u, v) .

Задача состоит в поиске корневого покрывающего леса $H \in \Omega$ графа G с корнями, принадлежащими множеству U . В качестве критериев оптимальности работы модели примем наименьшее время, требуемое на перемещение всех частиц в $u \in U$, а также наибольшее количество эвакуировавшихся людей, т.к. не все люди могут быть эвакуированы из здания за отведенное на эвакуацию время:

$$\max(T - T'), \quad (7)$$

$$\min\left(\sum_{u \in U} w_{uT'} - \sum_{v \in V \setminus U} w_{v0}\right), \quad (8)$$

В качестве решения задачи, будет выбран такой покрывающий лес H , при движении частиц к корням деревьев по которому потребуются наименьшее время и будет эвакуировано наибольшее количество людей. Алгоритм построения покрывающего дерева описан в работе [5].

Список литературы

- [1] Егоров А.А. Математические модели и алгоритмы эвакуации людей в аварийных ситуациях в учебных заведениях: дис. канд. техн. наук. В. 2008. С. 11-83.
- [2] Самошин Д.А. Современные программные комплексы для моделирования процесса эвакуации людей // Пожарная безопасность в строительстве. 2011. №1. С. 62-65.
- [3] Малодушев С.В., Воронов Р.В. Определение оптимальных путей эвакуации из здания // Фундаментальные исследования. 2015. №10. С. 495-502.
- [4] Малодушев С.В., Рогов А.А. Определение локации в корпоративных Wi-Fi сетях // Вестник ЮУрГУ. Серия: Программирование. 2016. Т.9, №1. С. 92-104.
- [5] Воронов Р.В., Воронова А.М., Пискунов М.А. Задача покрытия гиперсети взвешенным корневым деревом и ее приложение для оптимального проектирования схем волоков на лесосеках // Информатика и системы управления. 2012. №1. С. 56-64.

Zero-One Law for Random Uniform Hypergraphs

A. D. Matushkin

Moscow Institute of Physics and Technology, Dolgoprudny, Russia

E-mail: 1alexmatushkin1@gmail.com

The random s -uniform hypergraph $G^s(n, p(n))$ obeys Zero-One Law if for each first-order property its probability tends to 0 or tends to 1. In [1] S. Shelah and J. Spencer showed that if α is an *irrational* positive number and $p = n^{-\alpha}$, then the random graph $G(n, p)$ obeys Zero-One Law. In the same paper it was proved that if $\alpha \in (0, 1]$ is a *rational* number, then $G(n, n^{-\alpha})$ does not obey Zero-One Law. We extended these results to the case of the uniform hypergraphs.

Theorem 1. *If α is a positive irrational number then the random s -hypergraph $G^s(n, n^{-\alpha})$ obeys Zero-One Law.*

We say that an s -uniform hypergraph G is *strictly balanced* if its density $e(G)/v(G)$ is greater than a density of any its proper sub-hypergraph. The next theorem is a generalization of the result of A.Ruciński and A.Vince (see [2]).

Theorem 2. *Let $s \in \mathbb{N}$, $s \geq 2$, $\rho \in \mathbb{Q}$. Then there exists a strictly balanced s -hypergraph with a density ρ if and only if $\rho \geq \frac{1}{s-1}$ or $\rho = \frac{k}{1+k(s-1)}$ for some $k \in \mathbb{N}$.*

It is easy to show that the asymptotical distribution of the number of copies of a strictly balanced s -uniform hypergraph G in $G^s(n, n^{-\alpha})$ is Poisson when $\alpha = \frac{v(G)}{e(G)}$. Hence, as the property of containing a particular sub-hypergraph G could be expressed by a first-order formula, $G^s(n, n^{-\alpha})$ does not obey Zero-One Law for $\alpha \in [0, s-1] \cap \mathbb{Q}$ and $\alpha = s-1 + 1/k$, $k \in \mathbb{N}$.

References

- [1] S. Shelah, J.H. Spencer, *Zero-one laws for sparse random graphs*, J. Amer. Math. Soc., **1**: 97–115, 1988.
- [2] A.Ruciński, A.Vince, *Strongly balanced graphs and random graphs*, J. Graph Theory, **10**: 251–264, 1986.

Пуассоновская аппроксимация числа монотонных цепочек с началом в локальном минимуме в цепи Маркова

А. А. Минаков

Московский государственный институт радиотехники, электроники
и автоматики, Москва, Россия

E-mail: minak-ski@yandex.ru

Пусть $X_1, X_2, \dots, X_n$ — отрезок последовательности случайных величин, образующих стационарную, неприводимую и апериодическую цепь Маркова $\{X_a\}_{a \in \mathbb{Z}}$ со значениями из множества $\mathcal{A} = \{1, \dots, N\}$, $N \geq 1$, стационарным распределением $\vec{\pi} = (\pi_i)_{i \in \mathcal{A}}$ и матрицей переходных вероятностей $\mathbb{P} = (p_{ij})_{(i,j) \in \mathcal{A} \times \mathcal{A}}$.

Монотонной цепочкой длины $s \geq 2$ с началом в t в локальном минимуме назовем событие $E_{t,s} = \{X_t > X_{t+1} \leq \dots \leq X_{t+s}\}$.

Обозначим через $\mathbf{I}\{B\}$ индикатор события B . Рассмотрим случайную величину

$$\xi_n(s) = \sum_{t=1}^n \mathbf{I}\{E_{t,s}\}, \quad (1)$$

равную числу монотонных цепочек длины s с началом в локальном минимуме, которые начинаются на отрезке $X_1, X_2, \dots, X_n$.

Результаты исследований монотонных серий в конечных бесповторных последовательностях можно найти в статьях [2, 8]. В статье [7] доказана предельная теорема для распределения числа непересекающихся монотонных серий заданной длины в стационарной цепи Маркова. Многомерную нормальную теорему для числа монотонных серий заданной длины в случайной последовательности можно найти в статье [3]. В статье [4] исследовано распределение числа монотонных цепочек заданной длины в случайной последовательности.

Через $d(\Phi, \Psi)$ обозначим расстояние по вариации между распределениями Φ и Ψ . Распределение случайной величины ζ будем обозначать $\mathcal{L}(\zeta)$.

Определим матрицу $\mathbb{M} = (m_{ij})_{(i,j) \in \mathcal{A} \times \mathcal{A}}$, у которой, элементы стоящие ниже главной диагонали, равны 0, а остальные совпадают с соответствующими элементами матрицы $\mathbb{P}$. Матрица $\mathbb{M}$ определяет вероятность перехода состояния $i \in \mathcal{A}$ в состояние $j \in \mathcal{A}$ при условии $i \leq j$.

По теореме о жордановой канонической форме (см. [5]) для верхнетреугольной матрицы $\mathbb{M}$ существует невырожденная трансформирующая матрица $\mathbb{T} = (t_{ij})_{(i,j) \in \mathcal{A} \times \mathcal{A}}$ и обратная ей $\mathbb{T}^{-1} = (t'_{ij})_{(i,j) \in \mathcal{A} \times \mathcal{A}}$ для преобразования подобия $\mathbb{M} = \mathbb{T} \cdot \mathbb{J} \cdot \mathbb{T}^{-1}$, где $\mathbb{J}$ — жорданова форма матрицы $\mathbb{M}$.

Далее положим

$$\lambda_1 = n \cdot \sum_{l=1}^N \sum_{k=1}^{N-1} w_{k,l} \sum_{t=k+1}^N \pi_t p_{t,k}, \quad (2)$$

где $w_{k,l} = \sum_{i=1}^N t'_{i,l} \sum_{j=1}^i t_{k,j} p^{s-i+j-1} \binom{s-1}{s-i+j-1}$, если $k \leq l$; и $w_{k,l} = 0$, если $k > l$.

Определим константу

$$Q' = \sum_{k=1}^{N-1} \sum_{t=2}^N \pi_t p_{t,k}. \quad (3)$$

Для любого $i \in \{1, \dots, N\}$ определим константу

$$C_i = \left(t_{1,i} \sum_{l=2}^N \pi_l p_{l,1} + t_{2,i} \sum_{l=3}^N \pi_l p_{l,2} + \dots + t_{i,i} \sum_{l=i}^N \pi_l p_{l,i} \right) \cdot \sum_{k=i}^N t'_{i,k}. \quad (4)$$

Теперь положим

$$\lambda_2 = np_{1,1}^{s-1} C_1 + \dots + np_{N,N}^{s-1} C_N. \quad (5)$$

На основе метода Стейна и результатов работы [6] автором доказаны следующие две теоремы.

Теорема 1. Пусть на главной диагонали матрицы $\mathbb{P}$ все вероятности равны $p \in (0, 1]$, тогда

$$d\left(\mathcal{L}(\xi_n(s)), \text{Pois}(\lambda_1)\right) \leq \frac{(1 - e^{-\lambda_1})(2s + 3)\lambda_1}{n},$$

где λ_1 определена в (2).

Теорема 2. Пусть вероятности на главной диагонали матрицы $\mathbb{P}$ различны, тогда

$$d\left(\mathcal{L}(\xi_n(s)), \text{Pois}(\lambda_2)\right) \leq \frac{(1 - e^{-\lambda_2})(2s + 3)\lambda_2}{n},$$

где λ_2 определена в (5).

Используя теоремы 1 и 2, доказаны следующие теоремы для случайной величины $\xi_n(s)$.

Теорема 3. Пусть на главной диагонали матрицы $\mathbb{P}$ все вероятности равны $p \in (0, 1]$ и выполнено условие $n, s \rightarrow \infty$ так, что $sn^{-1} \rightarrow 0$ и $np^{s-1}(s-1)^N \rightarrow Q \in (0, \infty)$, то

$$\mathcal{L}(\xi_n(s)) \rightarrow \text{Pois}(\lambda'),$$

где $\lambda' = Q' \cdot Q \in (0, \infty)$, а константа Q' определена выражением (3).

Теорема 4. Пусть среди вероятностей на главной диагонали матрицы $\mathbb{P}$ существует единственная максимальная вероятность $p_{m,m}$, и выполняется условие $n, s \rightarrow \infty$ так, что $np_{m,m}^{s-1} \rightarrow C \in (0, \infty)$ и $sn^{-1} \rightarrow 0$, то

$$\mathcal{L}(\xi_n(s)) \rightarrow \text{Pois}(\lambda^{(m)}),$$

где $\lambda^{(m)} = C \cdot C_1 \in (0, \infty)$, а константа C_1 определена выражением (4).

Список литературы

- [1] Гантмахер Ф.Р. Теория матриц. М., 1966 г., 576 с.
- [2] Гончаров В.Л. Из области комбинаторики. Изв. АН СССР. Сер. матем., 1944, том 8, выпуск 1. С.3-48.
- [3] Меженная Н.М. Многомерная нормальная теорема для числа монотонных серий заданной длины в равновероятной случайной последовательности. Обзорение прикладной и промышленной математики. 2007. Т.14, вып.3, с.503-505.
- [4] Минаков А.А. Аппроксимация распределения числа монотонных цепочек заданной длины в случайной последовательности сложным распределением Пуассона. Прикладная дискретная математика № 2 (28), 2015 г., с. 21-29.
- [5] Хорн Р.А., Джонсон Ч.Р. Матричный анализ. Мир, 1989 г., 656 с.
- [6] Barbour A.D., Holst L., Janson S. Poisson approximation, volume 2 of Oxford Studies in Probability. The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publication.
- [7] Chrysaphinou O., Papastavridis S., Vaggelatos E. Poisson Approximation for the Non-Overlapping Appearances of Several Words in Markov Chains. Combinatorics, Probability and Computing. 2001, v.10, № 4, p.293–308.
- [8] Wolfowitz J. Asymptotics distribution of runs up and down. Ann. Math. Statist. 1944. v.15, p.163-172.

Две задачи статистического анализа экологических данных

О. А. Митрофанова¹, В. М. Буре^{1,2}

¹Агрофизический научно-исследовательский институт

²Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия

E-mail: omitrofa@gmail.com, vlb310154@gmail.com

Построение калибровочных кривых. В настоящее время при исследовании многих экологических проблем важное значение имеют различные аспекты статистического анализа экологических данных, а также методы анализа цифровых изображений. Рассмотрим одну из таких задач, которая заключается в оценке обеспеченности растений азотом по данным аэрофотосъемки.

Решение этой задачи основано на анализе цветовых характеристик растений по аэрофотоснимкам. Для интерпретации цветовых параметров применяется модель цветового пространства CIELAB. Аэрофотоснимки получены с помощью беспилотных летательных аппаратов (БЛА) самолетного и вертолетного типа. Каждое цифровое изображение поля содержит тестовые площадки - небольшие участки поля, на которых нам известны экологические параметры (в данном случае была внесена определенная доза азота). Таким образом, средние значения цветовых характеристик растений на каждой из тестовых площадок являются эталонами, на основе которых можно оценить содержание азота на остальных участках поля. Данную оценку можно провести двумя путями: напрямую сравнивать цветовые характеристики с эталонными либо построить калибровочную кривую также на основе характеристик эталонов.

Для получения корректных результатов целесообразно построить калибровочную кривую, описывающую зависимость дозы азота на тестовых площадках от характеристики цвета растений. А для этого необходимо представить цветовые характеристики в виде единой величины. В связи с этим предложено свернуть вектор параметров цвета в одно значение с использованием выпуклой линейной комбинации. Данное значение условно названо обобщенной характеристикой цвета:

$$C_{Lab} = \alpha L^* + \beta_1 a^* + \beta_2 b^*,$$

где $\alpha \geq 0$, $\beta_1 \geq 0$ и $\beta_2 \geq 0$ – эмпирически подбираемые коэффициенты для каждого цифрового изображения посева, причем

$$\alpha + \beta_1 + \beta_2 = 1.$$

Подбор коэффициентов α , β_1 , β_2 осуществляется путем обеспечения максимально выраженной линейной связи между дозой азота и характеристикой цвета. Для этого строится линейная регрессионная модель. При построении предполагается, что линейная модель наилучшим образом характеризует зависимость между содержанием азота N и характеристикой цвета C_{Lab} :

$$N = \beta_0 + \beta_1 C_{Lab} + \varepsilon,$$

в которой β_0 и β_1 – параметры модели, а ε – случайная величина (возмущение), характеризующая влияние неучтенных факторов. Уравнение регрессии при этом примет вид:

$$N^* = b_0 + b_1 C_{Lab}.$$

Коэффициенты b_0 и b_1 вычисляются методом наименьших квадратов.

После этого проводится проверка адекватности модели. Соответственно рассчитываются полная сумма квадратов вариации отклика, а также две ее составляющие: сумма квадратов, объясненная регрессией и остаточная сумма квадратов. Для оценки меры выраженности линейной связи используется коэффициент детерминации. Для того чтобы иметь общее суждение о качестве модели, по каждому наблюдению из относительных отклонений определяется средняя ошибка аппроксимации. Оценка значимости уравнения регрессии в целом проводится на основе F -критерия Фишера, строится F -статистика.

Получив оптимальный набор коэффициентов, строится калибровочная кривая, с помощью которой можно определить содержание азота на каждом элементарном участке поля.

Применение кригинга и бинарной регрессии для анализа экологических данных. В качестве примера рассмотрим задачу, схожую с предыдущей, которая заключается в оценке обеспеченности растений азотом. Предположим, что имеется набор экологических или агрохимических данных, измеренных контактным способом (в данном случае измерения содержания азота в растениях на сельскохозяйственном поле), а также аэрофотоснимок рассматриваемого объекта. Предварительно с помощью метода классификации определим однородные зоны поля на снимке. Предполагается, что в каждой выделенной зоне имеется набор экологических данных $Z(x_i)$. Далее будем рассматривать каждую зону отдельно.

Вначале построим набор оценок содержания азота с помощью метода обычного кригинга:

$$\hat{Z}(x) = \sum_{i=1}^n \lambda_i Z(x_i), \sum_{i=1}^n \lambda_i = 1.$$

Затем зададим пороговое значение содержания азота d для данной зоны и введем фиктивную переменную:

$$y(x) = \begin{cases} 1, & Z(x) > d, \\ 0, & Z(x) \leq d. \end{cases}$$

Таким образом, получаем основу для логистической регрессии:

$$P(y(x_i) = 1 \mid \Phi_i) = p_i = \frac{1}{1 + \exp(-\Phi_i^T \beta)},$$

где в факторы Φ_i входит набор оценок, спрогнозированных методом кригинга $\hat{Z}(x_i)$. Кроме того, в эти факторы могут входить цветковые характеристики с аэрофотоснимка.

В результате сможем для каждой точки поля вычислять вероятность $P(y(x) = 1)$, в случае, если она окажется близка к 1, есть основания полагать, что в этой точке содержание азота превышает пороговый уровень d , а если вероятность близка к 0, есть основания считать, что содержание азота ниже порогового уровня.

Список литературы

1. Митрофанова О. А., Буре В. М., Канаиш Е. В. Математический подход к автоматическому построению калибровочных кривых для оценки обеспеченности растений азотом по данным дистанционного зондирования // Применение средств дистанционного зондирования земли в сельском хозяйстве. 2015. С. 134-137.
2. Митрофанова О. А., Буре В. М., Канаиш Е. В. Математический модуль для автоматизации колориметрического метода оценки обеспеченности растений азотом // Вестн. С.-Петерб. ун-та. Сер. 10 Прикладная математика. Информатика. Процессы управления. 2016. Вып. 1. С. 87-93.
3. Fernandes G. B., Artes R. Spatial dependence in credit risk and its improvement in credit scoring // European Journal of Operational Research. 2016. Vol. 249. P. 517-524.

О структурной эквивалентности s-цепочек в цепях Маркова

В. Г. Михайлов, А. М. Шойтов, А. В. Волгин

Лаборатория ТВП, Москва, Россия

E-mail: artem.volgin@bk.ru

Пусть $(x_1, \dots, x_s)$ и $(y_1, \dots, y_s)$ — s -цепочки (цепочки) букв алфавита $\{1, \dots, N\}$. Они называются эквивалентными, если существует такая перестановка $\sigma = (\sigma(1) \dots, \sigma(N))$ знаков алфавита $\{1, \dots, N\}$, при которой $(y_1, \dots, y_s) = (\sigma(x_1), \dots, \sigma(x_s))$. Эквивалентные цепочки всегда имеют одинаковую структуру, а именно, если $x_i = x_j$, то и $y_i = y_j$, и наоборот, если $x_i \neq x_j$, то и $y_i \neq y_j$. Свойство эквивалентности цепочек $(x_1, \dots, x_s)$ и $(y_1, \dots, y_s)$ запишем $(x_1, \dots, x_s) \sim (y_1, \dots, y_s)$.

Существует понятие, относительно которого совпадение цепочек и совпадение лишь их структур являются частными случаями.

Пусть H — группа перестановок на множестве $\{1, \dots, N\}$. Цепочки $(a_1, \dots, a_s)$, $(b_1, \dots, b_s)$, составленные из элементов множества $\{1, \dots, N\}$, называют H -эквивалентными, если найдется перестановка $h \in H$, при которой

$$b_i = h(a_i), \quad i = 1, \dots, s.$$

Для H -эквивалентных цепочек $a = (a_1, \dots, a_s)$, $b = (b_1, \dots, b_s)$ будем использовать запись aHb . Если надо подчеркнуть, что цепочки a и b не являются H -эквивалентными, то используем запись $\bar{H}$.

Если H состоит из одной тождественной перестановки e : $H = \{e\}$, то соотношение $a\{e\}b$ эквивалентно равенству $a = b$.

В качестве H можно рассмотреть и произвольное множество перестановок знаков алфавита $\{1, \dots, N\}$. В этом случае вместо термина «эквивалентные цепочки» принято использовать термин «связанные цепочки». Если $H_1 \subseteq H_2 \subseteq S_N$ (S_N — множество перестановок на $\{1, \dots, N\}$), то H_1 -эквивалентные (H_1 -связанные) цепочки являются также H_2 -эквивалентными (H_2 -связанными) цепочками. Отсюда следует, что все эти цепочки обладают одинаковой структурой.

Наличие в последовательности знаков алфавита большого числа цепочек с одинаковой структурой снижает ее сложность. Повышенное

или пониженное число H -эквивалентных цепочек в последовательности может свидетельствовать о способе ее построения.

Задача о количестве эквивалентных цепочках в случайных дискретных последовательностях впервые была рассмотрена в работе [1], где были указаны достаточные условия выполнения предельной теоремы Пуассона для числа пар эквивалентных цепочек в последовательности независимых случайных величин, распределенных равномерно на множестве $\{1, \dots, N\}$.

В [2] были получены достаточные условия выполнения предельной теоремы Пуассона для числа таких пар цепочек в последовательности независимых случайных величин, распределенных равномерно на множестве $\{1, \dots, N\}$, в которых вторая цепочка получается из первой применением перестановки σ из множества перестановок, образующих «латинский прямоугольник». С этой работы началось изучение числа появлений в случайной последовательности H -эквивалентных цепочек.

Дальнейшее развитие этого направления отражено в обзорной работе [3], где описаны результаты работ, появившихся до 2003 года, а также анонсирован ряд результатов, опубликованных немного позже в [4] и [6]. Большое количество работ, появившихся с 2003 по 2013 год (см. [5], [7] – [16]) содержит много новых постановок и результатов, касающихся свойств H -эквивалентных цепочек в последовательности полиномиальных испытаний. Эти результаты должны еще найти свое описание в новом обзоре, который, как надеются авторы, не заставит себя долго ждать.

Среди этих работ следует отдельно упомянуть [5], где описывается явление, которое было названо «редукцией группы перестановок». Оно связано с поведением распределения числа $\xi_r(H)$ наборов по r H -эквивалентных цепочек в неравновероятной полиномиальной схеме. Группе H и набору $P = (p_1, \dots, p_N)$ вероятностей полиномиальной схемы ставится в соответствие подгруппа

$$H_P = \{h \in H : p_i = p_{h(i)} \forall i \in \{1, \dots, N\}\}.$$

Оказалось, что в случае сходимости к распределениям пуассоновского типа предельные распределения величин $\xi_r(H)$ и $\xi_r(H_P)$ совпадают. Использование этого свойства заметно упростило исследования предельного поведения случайной величины $\xi_r(H)$.

В последнее время проявился интерес к изучению аналогичных задач для цепочек в однородной неразложимой цепи Маркова с конечным числом состояний. В работах [17] – [19], описываются достаточные условия выполнения предельных теорем пуассоновского типа для числа наборов одинаковых цепочек в такой цепи Маркова. Используя методы этих работ, удалось доказать следующие результаты.

Пусть $\mathbf{X} = \{X_1, \dots, X_n, \dots\}$ — простая однородная цепь Маркова с исходами $1, \dots, N$, матрицей переходных вероятностей $\mathbb{P} = \|p_{a,b}\|$, $p_{a,b} > 0$, $a, b = 1, \dots, N$ (для случая одинаковых цепочек ($H = \{e\}$) требование $p_{a,b} > 0$, $a, b = 1, \dots, N$, можно ослабить до требования эргодичности цепи $\mathbf{X}$), стационарным распределением $\pi = (\pi_1, \dots, \pi_N)$, где все $\pi_k > 0$, и произвольным начальным распределением. Положим

$$p = \max_{a,b} p_{a,b}, \quad \Pi_2 = \sum_{k=1}^N \pi_k^2.$$

Нас интересуют события

$$\left\{ Y_{i_1-1} \overline{H} Y_{i_2-1}, Y_{i_1}(s) H Y_{i_2}(s) \right\}, \text{ где } Y_i(s) = (X_i, \dots, X_{i+s-1}),$$

закрывающиеся в том, что в моменты i_1 и i_2 начинается серия H -повторений s -цепочек. Каждая такая серия имеет вообще говоря свою (случайную) длину.

Рассмотрим число серий H -повторений s -цепочек, начинающихся до момента n включительно:

$$\xi_2(n, s, H) = \sum_{2 \leq i_1 < i_2 \leq n} I \left\{ Y_{i_1-1} \overline{H} Y_{i_2-1}, Y_{i_1}(s) H Y_{i_2}(s) \right\}.$$

По аналогии с [5] введем понятие «редуцированной группы»:

$$H\mathbb{P} = \{h \in H : p_{k,l} = p_{h(k),h(l)} \quad \forall k, l = 1, \dots, N\}.$$

Пусть матрица $\mathbb{P}_2(h) = \|p_{a,b} p_{h(a),h(b)}\|$ (она неразложима и примитивна), $\rho(h) > 0$ — ее максимальное характеристическое число матрицы. Для всех $h \in H\mathbb{P}$ выполнено $\rho(h) = \rho(e) = \rho$. Здесь и далее в случае, когда h совпадает с единичной перестановкой e , упоминание о h опускаем, например, пишем $\mathbb{P}_2$ вместо $\mathbb{P}_2(e)$. Известно ([20]), что

$$\mathbb{P}_2(h) = Z(h) \rho(h) \mathbb{P}'_2(h) Z(h)^{-1},$$

где $Z(h)$ — диагональная матрица, на главной диагонали которой стоит собственный вектор $(z_1(h), \dots, z_N(h))$, все $z_k(h) > 0$, соответствующий характеристическому числу $\rho(h)$, а $\mathbb{P}'_2(h)$ — стохастическая матрица, которая определяет неразложимую и непериодическую цепь Маркова с некоторым стационарным распределением $\pi'(h) = (\pi'_1(h), \dots, \pi'_N(h))$, все $\pi'_k(h) > 0$, и для некоторой константы $c_h > 0$

$$|\mathbb{P}'_2(h)^s - \Pi'(h)| = O(e^{-c_h s}), \quad s \rightarrow \infty.$$

Матрица $\mathbf{\Pi}'(h)$ состоит из одинаковых строк $\pi'(h)$, а норма матрицы A выражается формулой

$$|A| = \max_{1 \leq i \leq N} \sum_{j=1}^N |a_{i,j}|.$$

Введем обозначения $(\pi', z^{-1}) = \sum_{k=1}^N \pi'_k z_k^{-1}$, $(\pi^2, z) = \sum_{l=1}^N \pi_l^2 z_l$,

$$R^2 = \rho^{s-1} (1 - \rho) (\pi', z^{-1}) (\pi^2, z) |H_{\mathbb{P}}|.$$

Теорема 1. Пусть $p^2 < \rho$ и при $n, s \rightarrow \infty$ выполнено условие

$$n^2 R/2 \rightarrow \lambda \in (0, \infty).$$

Тогда распределение случайной величины $\xi_2(n, s, H)$ сходится к распределению Пуассона с параметром λ .

Пусть $\tau(s, H) = \min\{n : \xi_2(n, s, H) > 0\}$ — момент первого появления H -эквивалентных цепочек.

Теорема 2. Пусть $p^2 < \rho$ и $n \rightarrow \infty$. Тогда для любого $x > 0$

$$\mathbf{P}\{\tau(s, H)R \leq x\} \rightarrow 1 - e^{-x^2/2}.$$

Пусть $\mu(n, H) = \max\{s : \xi_2(n, s, H) > 0\}$ — максимальная длина имеющихся H -эквивалентных цепочек.

Теорема 3. Пусть $p^2 < \rho$, и $n \rightarrow \infty$. Тогда для любого $k \in \mathbb{Z}$

$$\begin{aligned} & \mathbf{P}\left\{\mu(n, H) - \left\lceil \frac{2 \ln n}{|\ln \rho|} \right\rceil < k\right\} = \\ & = \exp\left(-\frac{1-\rho}{2} (\pi', z^{-1}) (\pi^2, z) |H_{\mathbb{P}}| \rho^{k-1 - \left\{\frac{2 \ln n}{|\ln \rho|}\right\}}\right) + o(1), \end{aligned}$$

где $[u]$ и $\{u\}$ — целая и дробная части числа u .

СПИСОК ЛИТЕРАТУРЫ

1. Буравлев С. М. Повторения с точностью до перестановок в последовательности независимых испытаний // Дискретная математика, 1999, т. 11, вып. 2, с. 53-75.
2. Буравлев С.М. «Повторения с точностью до перестановок, образующих латинский прямоугольник» // Дискретная математика. 2000. Т. 12, вып. 1. С. 24-46.
3. Михайлов В. Г., Шойтов А. М. Структурная эквивалентность s -цепочек в случайных дискретных последовательностях // Дискретная математика, 2003, т. 15, вып. 4, с. 7-34.
4. Михайлов В. Г., Предельные теоремы пуассоновского типа для числа пар H -связанных цепочек // Труды по дискретной математике. Том 7, с. 138-155. М.: ФИЗМАТЛИТ, 2003.
5. Шойтов А. М. Об одной особенности асимптотического поведения числа наборов H -эквивалентных n -цепочек в неравновероятной полиномиальной схеме // Труды по дискретной математике. Том 7. с. 227-238. М.: ФИЗМАТЛИТ, 2003.
6. Михайлов В. Г. Об особенностях асимптотического поведения числа пар структурно близких цепочек // Труды по дискретной математике. Том 8. с. 176-186. М.: ФИЗМАТЛИТ, 2004.
7. Шойтов А. М. Предельные распределения случайных величин, характеризующих связь цепочек полиномиальной схемы структурной эквивалентностью // Труды по дискретной математике. Том 8. с. 312-326. М.: ФИЗМАТЛИТ, 2004.
8. Шойтов А. М. Пуассоновское приближение для числа повторений значений дискретной функции от цепочек // Дискретная математика, 2005, т. 17, вып. 2, с. 59-69.
9. Шойтов А. М. Повторения неперекрывающихся цепочек полиномиальной схемы с точностью до перестановочной эквивалентности // Труды по дискретной математике. Том 9. с. 401-414. М.: Гелиос АРВ, 2006
10. Шойтов А. М. Сложное распределение Пуассона для числа повторений значений дискретной функции от цепочек. — Дискретная математика, 2007, т. 19, вып. 2, с. 6-26.
- 11 Шойтов А. М. Нормальное приближение в задаче об эквивалентных цепочках // Труды по дискретной математике. Том 10. с. 326-349. М.: ФИЗМАТЛИТ, 2007
12. Шойтов А. М. Дискретные предельные распределения для числа цепочек с одинаковой структурой в последовательности равновероятных испытаний // Обзорение прикл. промышл. математики, 2007, т. 14, вып. 3, с. 474-484

13. Михайлов В. Г. Об асимптотическом поведении вероятности наличия в последовательности эквивалентных цепочек с нетривиальной структурой // Дискретная математика, 2008, т. 20, вып. 4, с. 113-119.

14. Шойтов А. М. Асимптотические свойства распределения числа пар метрически близких и функционально связанных цепочек полиномиальной схемы // Труды по дискретной математике. Том 11. с. 151-165. М.: ФИЗМАТЛИТ, 2008.

15. Шойтов А. М. Структурно эквивалентные цепочки в равновероятной полиномиальной схеме // Математические вопросы криптографии, 2012, т. 3, вып.3, с. 129-151.

16. Михайлов В. Г. , Шойтов А. М. О числах множеств эквивалентных цепочек в последовательности независимых случайных величин // Математические вопросы криптографии, 2013, т. 4, вып.1, с. 77-86.

17. Михайлов В. Г., Шойтов А. М. О длинных повторениях цепочек в цепи Маркова // Дискретная математика, 2014, т. 26, вып. 3, с. 79-89.

18. Михайлов В. Г., Шойтов А. М. Многократные повторения длинных цепочек в цепи Маркова // Математические вопросы криптографии, 2015, т. 6, вып.3, с. 117-134.

19. Михайлов В. Г. Оценки точности пуассоновской аппроксимации для распределения числа серии повторений длинных цепочек в цепи Маркова // Дискретная математика, 2015, т. 27, вып. 4, с. 67-78.

20. Гантмахер Ф.Р. Теория матриц. 5-е изд. М.: ФИЗМАТЛИТ, 2004. 560 с.

Метод регенеративных огибающих для оценивания высокопроизводительных многосерверных систем

Е. В. Морозов, А. С. Румянцев, И. В. Пешкова

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводский государственный университет, Петрозаводск, Россия

E-mail: emorozov@karelia.ru, ar0@krc.karelia.ru,
iaminova@petrsu.ru

Рассматривается новый подход к доверительному оцениванию стационарных характеристик высокопроизводительных многосерверных систем. Метод основан на построении двух новых процессов, траектории которых являются, соответственно, верхней и нижней границами для траектории исследуемого процесса исходной системы. Преимущество построенных *огибающих* процессов состоит в том, что для них существуют моменты регенерации. В случае, когда в исходной системе моменты регенерации отсутствуют, либо крайне редки, данный подход позволяет получать верхнюю и нижнюю границы исследуемых стационарных характеристик с заданной надежностью на основе регенеративного моделирования.

Суть подхода состоит в построении для исходной m -серверной системы Σ с входным процессом восстановления и неограниченным буфером *мажорантной* системы $\bar{\Sigma}$ и *минорантной* системы $\underline{\Sigma}$ с использованием *метода каплинга*. Затем применяется свойство монотонности процессов, описывающих число заявок в системе или в очереди. Обозначим через t_n – момент прихода заявки с номером n (ввиду каплинга, эти моменты одинаковы для всех трех систем). В момент t_n^- в исходной системе обозначим через ν_n число заявок в системе, через Q_n – число заявок, ожидающих в очереди, через $S_i(n)$ – остаточное время обслуживания заявки с номером i , а через $\mathcal{M}_n$ – множество номеров обслуживаемых заявок. (Для построенных ниже систем $\bar{\Sigma}$ и $\underline{\Sigma}$

используются, соответственно, обозначения $\bar{\nu}_n, \bar{Q}_n$ и $\underline{\nu}_n, \underline{Q}_n$.) Зафиксируем целое число $\sigma \geq 0$, константы $0 \leq a \leq b < \infty$ и рассмотрим момент

$$\beta_1 = \inf \left\{ k : \nu_k = \sigma, S_i(k) \in [a, b], i \in \mathcal{M}_k \right\}.$$

В этот момент все ненулевые остаточные времена обслуживания заменяются на *верхнюю границу* b , что увеличивает соответствующие времена обслуживания. Полученную в результате такого преобразования мажорантную систему обозначим $\bar{\Sigma}$. Для построения минорантной системы $\underline{\Sigma}$ в момент β_1 все остаточные времена обслуживания заменяются на *нижнюю границу* a . Таким образом, марковский процесс $\{\nu_k = \sigma, S_i(k) \in [a, b], i \in \mathcal{M}_k\}$ в момент β_1 стартует в системе $\bar{\Sigma}$ в состоянии $(\sigma, \mathbf{b})$, а в системе $\underline{\Sigma}$ - в состоянии $(\sigma, \mathbf{a})$, где m -мерный vector $\mathbf{b}$ ($\mathbf{a}$) содержит $\min(\sigma, m)$ компонент равных b (a) и $(m - \sigma)^+$ нулей. Далее, по аналогии, в системе $\bar{\Sigma}$ строятся моменты $\bar{\beta}_n, n > 1$, когда величина очереди снова равна σ , а остаточные времена обслуживания принадлежащие интервалу $[a, b]$ заменяются на величину b и, кроме того, все заявки, обслуживаемые в момент времени $\bar{\beta}_n$, должны покинуть систему до момента $\bar{\beta}_{n+1}$. Подобным же образом строятся моменты $\{\beta_n\}$ в минорантной системе $\underline{\Sigma}$. При этом в исходной системе Σ времена обслуживания $\{S_n\}$ остаются неизменными. Так как во всех системах входные потоки совпадают, а времена обслуживания (в очевидном обозначении) связаны неравенством: $\bar{S}_n \geq S_n \geq \underline{S}_n, n \geq 1$, то верны такие неравенства:

$$\bar{\nu}_n \geq \nu_n \geq \underline{\nu}_n, \bar{Q}_n \geq Q_n \geq \underline{Q}_n, n \geq 1. \quad (1)$$

Показано, что $\{\beta_n\}$ и $\{\bar{\beta}_n\}$ являются моментами регенерации системы $\underline{\Sigma}$ и $\bar{\Sigma}$, соответственно. На основе метода регенеративного моделирования это позволяет строить с заданной точностью интервальные оценки стационарной длины очереди в системах $\underline{\Sigma}, \bar{\Sigma}$, а в силу неравенств (1) и для исходной системы.

Важно подчеркнуть, что при построении систем $\underline{\Sigma}$ и $\bar{\Sigma}$ исходный процесс обслуживания подвергается изменению. Однако численные эксперименты на модели высокопроизводительного кластера показали, что различие в средней длине очереди во всех трех системах совершенно незначительно, и это позволяет говорить о применимости предложенного метода для надежного оценивания загрузки высокопроизводительных вычислительных систем.

Asymptotic Analysis of a System with Random Volume Customers

E. V. Morozov, L. V. Potakhina

Institute of Applied Mathematical Research, Karelian Research Centre,
Petrozavodsk State University, Petrozavodsk, Russia

E-mail: emorozov@karelia.ru, lpotakhina@gmail.com

We consider a FIFO multiserver $GI/G/m$ -type queueing system with renewal input with generic interarrival time τ and rate $\lambda := 1/E\tau \in (0, \infty)$, in which customer n has two parameters: service time S_n and volume v_n , $n \geq 0$. Such systems have been used to model various practical problems occurring in the design of computer and communication systems [4, 5]. We assume the pairs (S_n, v_n) to be independent identically distributed (iid), while, for given n , a dependence between S_n and v_n may exist. Then the volumes $\{v_i\}$ are iid with generic element v , and the service times $\{S_n\}$ are iid as well, with generic element S . We assume both unlimited space for the buffered customers and the limited space case. (The latter case has been considered in [3] as well.) At the arrival instant of customer n we consider its waiting time in queue W_n , the number of present customers ν_n (queue size) and the total accumulated volume in the system V_n , $n \geq 0$. We establish the following *asymptotic solidarity* property:

$$V_n \Rightarrow \infty \text{ iff } \nu_n \Rightarrow \infty \text{ iff } W_n \Rightarrow \infty, \quad (1)$$

where $\Rightarrow$ stands for the convergence in probability.

Furthermore, under stability conditions

$$\rho := \lambda ES < m, \quad P(\tau > S) > 0, \quad (2)$$

we show that the processes $\{V_n\}$, $\{\nu_n\}$, $\{W_n\}$ are tight.

Let V, ν be the stationary summary volume and the stationary queue size, respectively, at arbitrary time instant. Also let $W_n \Rightarrow W$. Using regenerative approach [2], we prove the following expression,

$$EV = \lambda[EW E v + E(S v)], \quad (3)$$

which becomes the celebrated classical Little's formula when $v_n \equiv 1$ [1]:

$$E\nu = \lambda(EW + ES). \quad (4)$$

Finally, for the single-server system with Poisson input and limited space, we suggest an approximation of the well-known Pollaczek-Khintchine formula [1],

$$EW = \frac{\lambda ES^2}{2(1 - \lambda ES)}, \quad (5)$$

and verify its accuracy by simulation.

Acknowledgements. This work is supported by Russian Foundation for Basic Research, projects 15-07-02341, 15-07-02354, 15-07-02360 and the Program of Strategic development of Petrozavodsk State University.

References

- [1] Asmussen, S.: Applied probability and queues. Springer-Verlag, New York (2003).
- [2] Morozov, E., Delgado, R.: Stability analysis of regenerative queueing systems. Automation and Remote control, 70 (12), 1977-1991 (2009).
- [3] Morozov, E., Nekrasova, R., Potakhina, L., Tikhonenko, O.: Asymptotic Analysis of Queueing Systems with Finite Buffer Space. Communications in Computer and Information Science. 431, 223-232 (2014).
- [4] Tikhonenko, O. M.: The problem of determination of the summarized messages volume in queueing systems and its applications. Journal of Information Processing and Cybernetics. 32 (7), 339-352 (1987)
- [5] Tikhonenko, O.: Computer Systems Probability Analysis. Akademicka Oficyna Wydawnicza EXIT, Warsaw (2006) (in Polish).

Цена анархии в игре заполнения для поиска лекарств

Н. Н. Никитина, Е. Е. Ивашко

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: nikitina@krc.karelia.ru, ivashko@krc.karelia.ru

Разработка нового лекарства — весьма ресурсоемкий и времяемкий процесс. На первом его этапе химическое пространство мощностью, в общем случае, порядка 10^{60} элементов, сокращается до $10-10^4$ потенциальных лекарств, которые будут тестироваться в лаборатории. Благодаря развитию высокопроизводительных вычислительных технологий, отбор множества потенциальных лекарств в большинстве случаев производится автоматизированно, в ходе процесса виртуального скрининга. Виртуальный скрининг представляет собой вычислительную процедуру, производимую на большом числе независимых заданий, и выполняется на множестве вычислительных узлов, которые можно рассматривать как независимых агентов, каждый из которых стремится выполнить как можно больше полезной работы.

В данной работе мы моделируем процесс виртуального скрининга в виде игры заполнения между вычислительными узлами, которые совместно используют совокупность подмножеств вычислительных заданий. Выигрыш каждого игрока зависит не только от ценности выбранного ресурса, но и от числа игроков, выбравших данный ресурс, то есть от уровня заполнения. Определим социальную полезность как сумму выигрышей всех игроков, и получим точные нижние границы цены анархии в игре заполнения с двумя ресурсами.

Рассмотрим вычислительную систему с M идентичными вычислительными узлами ($M \geq 2$), или игроками, $C = C_1, \dots, C_M$, каждый из которых имеет производительность, т.е. число операций в секунду, *ops*. Рассмотрим также множество вычислительных заданий T . Разобьем множество T на N непересекающихся подмножеств заданий $T = T_1, \dots, T_N$ таким образом, что ожидаемая доля полезных результатов в подмножестве T_j будет составлять p_j . Определим приоритет подмножества T_j следующим образом:

$$\sigma_j = \frac{p_j}{p_1 + \dots + p_M}, 0 \leq \sigma_j \leq 1. \quad (1)$$

Подмножество с более высоким приоритетом будет выбираться первым. Предположим, что все вычислительные задания имеют одну и ту же сложность, то есть число операций, θ .

После того как вычислительный узел обработал порцию заданий, он отправляет результаты центральному узлу и запрашивает следующую порцию. Пусть выигрыш узла C_i на интервале времени τ выражает количество полезной работы, выполненной им на данном интервале времени. Это количество зависит от числа выполненных заданий из выбранного подмножества, приоритета подмножества, а также от числа других узлов, выбравших данное подмножество. Чем меньше узлов выполняют подмножество заданий T_j одновременно, тем более ценна работа каждого из них. Это условие позволяет обеспечить разнообразие промежуточных полезных результатов. Обозначим как n_j число узлов, выбравших подмножество T_j на рассматриваемом интервале времени, а $\delta(n_j)$ — коэффициент заполнения подмножества:

$$\delta(n_j) = \frac{M + 1 - n_j}{M}. \quad (2)$$

Тогда выигрыш узла C_i , выбравшего подмножество T_j , составит

$$U_{ij} = \sigma_j \delta(n_j) \frac{ops}{\theta} \tau. \quad (3)$$

Таким образом, на каждом интервале времени имеется игра заполнения $\Gamma = \langle C, T, U \rangle$, где C — множество игроков (вычислительных узлов), T — множество стратегий (подмножеств заданий, из которых каждый узел выбирает ровно одно), U — множество функций выигрыша. Профиль стратегий в данной игре — вектор-расписание $\vec{s} = (s_1, \dots, s_M)$, где компонента $s_i = j$ равна номеру подмножества T_j , выбранного игроком C_i .

Доказано, что игра Γ всегда имеет равновесие по Нэшу в чистых стратегиях [1]. Ситуация равновесия означает, что ни один вычислительный узел не может увеличить количество своей полезной работы, в одностороннем порядке отклонившись от равновесного расписания. Известно также, что динамики лучших и наилучших ответов сходятся к равновесию за полиномиальное время [2].

Рассмотрим социальную полезность как сумму выигрышей игроков,

$$SOC(\vec{s}) = \sum_{i=1}^M U_i(\vec{s}) = \frac{ops \times \tau}{\theta \times M} \sum_{i=1}^N \sigma_i n_i (M + 1 - n_i). \quad (4)$$

Тогда цена анархии в игре Γ может быть определена следующим образом:

$$PoA(\Gamma) = \min_{\vec{s}-NE} \frac{SOC(\vec{s})}{OPT(\Gamma)}, \quad (5)$$

где $OPT(\Gamma) = \max_{\vec{s}} SOC(\vec{s})$ — максимальная социальная полезность среди всех возможных расписаний.

В работе [1] было показано, что игра Γ имеет потенциал, который, в наших предположениях, имеет вид

$$P(\vec{s}) = \frac{ops \times \tau}{2\theta \times M} \sum_{i=1}^N \sigma_i n_i (2M + 1 - n_i). \quad (6)$$

Рассмотрим случай с двумя подмножествами, $N = 2$. Без потери общности будем полагать $\sigma_1 \geq \sigma_2$. В сделанных предположениях имеет место следующая теорема.

Теорема. Цена анархии в игре Γ имеет нижнюю оценку $PoA > PoA_e$, где

$$PoA_e = \begin{cases} \frac{8}{3(M+2)}, & \text{если } M \text{ — четное и } \sigma_1 > \frac{M}{M+1}, \\ \frac{9}{4\left(M+1-\frac{1}{M+1}\right)}, & \text{если } M \text{ — нечетное и } \sigma_1 > \frac{M}{M+1}, \\ \frac{4\lfloor\sigma_1(M+1)\rfloor(M-\lfloor\sigma_1(M+1)\rfloor)}{M(M+2)}, & \text{если } M \text{ — четное и } \sigma_1 \leq \frac{M}{M+1}, \\ \frac{4\lfloor\sigma_1(M+1)\rfloor(M+1-\lfloor\sigma_1(M+1)\rfloor)}{(M+1)^2}, & \text{если } M \text{ — нечетное и } \sigma_1 \leq \frac{M}{M+1}. \end{cases}$$

Все приведенные оценки точные, т.е. для любого заданного числа игроков M возможно построить игру с ценой анархии, сколь угодно близкой к приведенным нижним оценкам.

Список литературы

- [1] R. Rosenthal. A Class of Games Possessing Pure-strategy Nash Equilibria. International Journal of Game Theory, Vol. 2, I. 1, pp. 65–67, 1973.
- [2] S. Ieong, R. McGrew, E. Nudelman, Y. Shoham, Q. Sun. Fast and Compact: A Simple Class Of Congestion Games. AIII Proceedings, pp. 1–6, 2005.

О случайных конфигурационных графах в случайной среде

Ю. Л. Павлов

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: pavlov@krc.karelia.ru

При моделировании современных сложных сетей коммуникаций, таких как Интернет и системы мобильной связи, широко используются методы теории случайных графов. Известно [1, 2], что одними из наиболее подходящих для такого моделирования видов графов являются конфигурационные графы [3] со случайными независимыми одинаково распределенными степенями вершин. Рассмотрим конфигурационный граф с N вершинами, среди которых нет изолированных. Степень каждой вершины определяет число выходящих из нее различных полуребер, т. е. ребер, для которых смежные вершины еще не заданы. Для обеспечения четности суммы степеней в случае необходимости в граф вводится вспомогательная вершина единичной степени. Ребра графа образуются путем попарного равновероятного соединения полуребер.

Для исследования структуры и динамики развития сетей коммуникаций при $N \rightarrow \infty$ в [2] предложено использовать распределение:

$$p_k = \mathbf{P}\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \quad k = 1, 2, \dots; \quad \tau > 0, \quad (1)$$

где ξ – случайная величина, равная степени любой вершины графа. В последнее время стали появляться работы, в которых отмечается, что, по мере развития сетей, распределения степеней вершин могут меняться и даже быть случайными. Предположим, что параметр τ распределения (1) является случайной величиной, равномерно распределенной на отрезке $[a, b]$, $0 < a < b < \infty$. Поскольку в этом случае распределение степеней является случайным, будем говорить, что граф развивается в случайной среде. Заметим еще, что, как показывают наблюдения [1], для большинства реальных сетей $\tau \in (1, 2)$. Из (1) следует, что

$$p_1 = \mathbf{P}\{\xi = 1\} = 1 - \frac{1}{(b-a) \ln 2} \left(\frac{1}{2^a} - \frac{1}{2^b} \right),$$

$$p_k = \mathbf{P}\{\xi = k\} = \frac{1}{(b-a)\ln k} \left(\frac{1}{k^a} - \frac{1}{k^b} \right) - \frac{1}{(b-a)\ln(k+1)} \left(\frac{1}{(k+1)^a} - \frac{1}{(k+1)^b} \right), \quad k \geq 2.$$

Далее рассматриваются условные случайные графы при условии, что сумма степеней вершин известна и равна n . Введем вспомогательную случайную величину η такую, что

$$p_k(\lambda) = \mathbf{P}\{\eta = k\} = \lambda^k p_k / B(\lambda), \quad (2)$$

где $k = 1, 2, \dots$; $0 < \lambda < 1$, $B(\lambda) = \sum_{k=1}^{\infty} \lambda^k p_k$. Обозначим m математическое ожидание распределения (2) и пусть параметр λ этого распределения является решением уравнения $m = n/N$. При различном характере стремления N и n к бесконечности найдены предельные распределения случайной величины $\xi_{(N)}$, равной максимальной степени вершины рассматриваемого случайного графа. В частности, справедливо следующее утверждение.

Теорема 1. Пусть $N, n \rightarrow \infty$ так, что $1 < C_1 \leq n/N \leq C_2 < \infty$, а $r = r(N, n)$ выбраны так, что

$$\frac{aN\lambda^{r+1}}{(b-a)B(\lambda)r^{a+1}\ln r} \rightarrow \gamma,$$

где γ – некоторая положительная постоянная. Тогда для любого фиксированного $k = 0, \pm 1, \pm 2, \dots$

$$\mathbf{P}\{\xi_{(N)} \leq r + k\} = \exp\{-\gamma\lambda^k(1-\lambda)^{-1}\}(1 + o(1)).$$

Работа выполнена при поддержке РФФИ, грант 16-01-00005.

Список литературы

- [1] C. Faloutsos, P. Faloutsos, M. Faloutsos. On power-law relationship of the Internet topology. *Computer communication rev.*, **29**, 1999, 251-262.
- [2] I. Norros, H. Reittu. On the power-law random graph model of massive data networks. *Performance Evaluation*, **55**, 2004, 3-23.
- [3] B. Bollobas. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Comb.*, **1**, 1980, 311-316.

Асимптотика числа вершин заданной степени конфигурационного графа вблизи критических точек

Ю. Л. Павлов

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: pavlov@krc.karelia.ru

Изучается случайный конфигурационный граф с N вершинами, степени которых являются независимыми одинаково распределенными случайными величинами, подчиняющимися дискретному степенному закону. Это распределение определяет число полуребер, выходящих из каждой вершины. Обозначим ξ случайную величину, равную степени вершины, и пусть

$$p_k = \mathbf{P}\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \quad k = 1, 2, \dots; \quad \tau > 0.$$

Граф образуется путем попарного равновероятного соединения полуребер для формирования ребер. В случае необходимости в граф вводится вспомогательная вершина единичной степени, обеспечивающая четность суммы степеней всех вершин. Рассмотрим подмножество графов, в которых сумма степеней основных вершин равна n и обозначим μ_r число вершин, степени которых равны r . В статье [1] получены предельные теоремы для μ_r при $N, n \rightarrow \infty$ и фиксированных значениях τ . Известно, что значения $\tau = 1, 2, 3$ являются критическими, поскольку при переходе τ через них свойства графа резко изменяются. Очевидно, что если τ не фиксировано, но не приближается к критическим значениям, то результаты работы [1] остаются в силе. В докладе рассматривается предельное поведение μ_r при $\tau \rightarrow 1, 2, 3$.

Введем независимые одинаково распределенные случайные величины $\eta_1, \dots, \eta_N$, для которых

$$p_k(\lambda) = \mathbf{P}\{\eta_1 = k\} = \lambda^k p_k(1 - (1 - \lambda)\Phi(\lambda, \tau, 1))^{-1}, \quad k = 1, 2, \dots; \quad \tau > 0,$$

где $0 < \lambda < 1$, а $\Phi(z, s, a)$ – трансцендентная функция Лерча:

$$\Phi(z, s, a) = \sum_{j=0}^{\infty} \frac{z^j}{(j+a)^s}.$$

Обозначим $m = \mathbf{E} \eta_1, \sigma^2 = \mathbf{D} \eta_1$. В качестве параметра λ будем использовать единственное решение уравнения $m = n/N$. Известно [1], что если $\tau > 1$ и $n/N \sim \zeta(\tau)$ при $N, n \rightarrow \infty$, где $\zeta(\tau)$ – значение дзета-функции Римана в точке τ , то $\lambda \rightarrow 1$. Это же верно и в случае $N, n \rightarrow \infty$ так, что $n/N \rightarrow \infty$ и $\tau \leq 1$. В этих условиях для фиксированных r доказана следующая теорема, где символом c обозначены некоторые положительные постоянные, не обязательно одинаковые.

Теорема 1. Пусть $\sigma_{rr}^2 = p_r(\lambda)(1 - p_r(\lambda) - (m - r)^2 p_r(\lambda)/\sigma^2)$ и выполнено одно из следующих условий:

1. $\tau \downarrow 3, \quad (1 - \lambda)^{\tau-3} \leq c < 1, \quad \sqrt{N}(\tau - 3) \rightarrow \infty;$
2. $\tau \rightarrow 3, \quad (1 - \lambda)^{\tau-3} \rightarrow 1, \quad \sqrt{N}/|\ln(1 - \lambda)| \rightarrow \infty;$
3. $\tau \uparrow 3, \quad (1 - \lambda)^{\tau-3} \geq c > 1, \quad \sqrt{N}(3 - \tau)(1 - \lambda)^{3-\tau} \rightarrow \infty;$
4. $\tau \downarrow 2, \quad (1 - \lambda)^{\tau-2} \leq c < 1, \quad \sqrt{N}(1 - \lambda)^{3-\tau}/(\tau - 2) \rightarrow \infty;$
5. $\tau \rightarrow 2, \quad (1 - \lambda)^{\tau-2} \rightarrow 1, \quad \sqrt{N}(1 - \lambda)|\ln(1 - \lambda)| \rightarrow \infty;$
6. $\tau \uparrow 2, \quad (1 - \lambda)^{\tau-2} \geq c > 1, \quad \sqrt{N}(1 - \lambda)/(2 - \tau) \rightarrow \infty;$
7. $\tau \downarrow 1, \quad (1 - \lambda)^{\tau-1} \leq c < 1, \quad N(1 - \lambda)^2/(\tau - 1) \rightarrow \infty;$
8. $\tau \rightarrow 1, \quad (1 - \lambda)^{\tau-1} \rightarrow 1, \quad N(1 - \lambda)^2|\ln(1 - \lambda)| \rightarrow \infty;$
9. $\tau \uparrow 1, \quad (1 - \lambda)^{\tau-1} \geq c > 1, \quad N(1 - \lambda)^{\tau+1}/(1 - \tau) \rightarrow \infty.$

Тогда для целых неотрицательных k равномерно относительно $u_r = (k - Np_r(\lambda))/(\sigma_{rr}\sqrt{N})$ в любом фиксированном конечном интервале

$$\mathbf{P}\{\mu_r = k\} = (\sigma_{rr}\sqrt{2\pi N})^{-1} e^{-u_r^2/2} (1 + o(1)).$$

Работа выполнена при поддержке РФФИ, грант 16-01-00005.

Список литературы

- [1] Ю.Л. Павлов, И.А. Чеплюкова. Случайные графы Интернет-типа и обобщенная схема размещения // *Дискретная математика*, **20**, вып. 3, 2008, 3-18.

Статистические модели распознавания и прогнозирование метеоситуаций, приводящих к возникновению торнадо и сильных шквалов умеренных широт

Э. В. Переходцева

ФГБУ «Гидрометцентр России», Москва, Россия

E-mail: perekhod@mecom.ru

Возникновение летнего штормового ветра со скоростью $V > 24\text{ м/с}$, включая торнадо и сильные летние шквалы умеренных широт непосредственно связано с активной конвекцией, развитием мощной кучево-дождевой облачности и со скоростью нисходящих потоков порядка $30 - 50\text{ см/с}$. Эта облачность образуется там, где атмосфера у Земли стратифицирована неустойчиво, велика влажность воздуха в тропосфере, у земли существует значительная конвергенция воздушных потоков. Для возникновения таких явлений в атмосфере должны сложиться определенные синоптические и термодинамические условия, характеризующиеся значениями целого ряда параметров (предикторов). В синоптической практике прогноз таких явлений практически не осуществляется. Иногда синоптиками за 3ч дается штормовое предупреждение о сильном ветре скоростью 25 м/с и более, но обычно при приближении холодного атмосферного фронта летом прогнозируются порывы ветра при грозе скоростью $15 - 18\text{ м/с}$.

К сожалению, гидродинамический прогноз таких явлений, основанный на численном решении уравнений движения, притока тепла и состояния атмосферы пока еще не является успешным, предупрежденность сильных шквалов и тем более торнадо на территории России не превышает 10%. Несмотря на то, что эти торнадо чаще всего слабее, чем в США, они зачастую приносят очень большой экономический ущерб, а порой и человеческие жертвы.

Первые достаточно успешные результаты объективного прогноза этих явлений на текущий день (по независимым испытаниям) с заблаговременностью 12 – 18ч были получены по нашей первой статистической модели в 1984 г., а затем уже успешно прошли оперативные испытания. В данной модели метеорологическая ситуация, способствующая возникновению указанных явлений, представляется как многомерный вектор $\mathbf{X}(A) = (x_1(A), x_2(A), \dots, x_n(A))$, где n – число параметров атмосферы (предикторов). При разработке модели использовались значения двадцати шести потенциальных предикторов, отражающих особенности метеорологических ситуаций, способствующих возникновению смерчей и штормового ветра. Были составлены выборки векторов наличия явлений A – архив $\mathbf{X}_1(A)$ и их отсутствия в условиях неустойчиво стратифицированной атмосферы – архив $\mathbf{X}_2(A)$. Отбор наиболее информативных и слабо зависимых предикторов проводился нашим новым методом диагонализации средней матрицы корреляции $\mathbf{R}$, основанным на теории графов. В качестве критериев информативности использовались критерий Махаланобиса и критерий минимальной энтропии Вапника-Червоненкиса. Модель распознавания и прогноза таких ситуаций была построена с использованием байесовского подхода, а прогноз торнадо и шквалов осуществлялся по значениям линейной дискриминантной функции $U(\mathbf{X})$ если $U(\mathbf{X}) > 0$. В результате процесса выбора были отобраны из 26-ти потенциальных предикторов шесть наиболее информативных мало зависимых предикторов ($\mathbf{V}_{700}$, $\mathbf{H}_0$, $(\mathbf{T}' - \mathbf{T})_{500}$, $d\mathbf{t}/d\mathbf{n}$, $\mathbf{T}_{max}$, $\mathbf{T}_{dmax}$). Значение прогностической дискриминантной линейной функции $U(\mathbf{X})$ было рассчитано для случая сильнейшего торнадо в г. Иваново 9 июня 1984 года, $U(\mathbf{X}) = 3,6$. Для 3-х торнадо, наблюдавшихся в 1986 году в г. Пенза (21.08.1986), в г. Рязань (26.08.1986), в г. Москва (31.08.1986) значение прогностической $U(\mathbf{X})$ составило $U(\mathbf{X}) = 7,3 - 7,9$. Поэтому при значениях $U(\mathbf{X}) > 3$ прогнозируется с большой вероятностью в данном пункте или его окрестности торнадо или очень сильные шквалы с $V \geq 25\text{м/с}$.

Затем в 1993 году (для $V > 19\text{м/с}$) и в 2000 году (для $V > 24\text{м/с}$) была разработана модель и получены результаты независимых испытаний полностью автоматизированного прогноза с заблаговременностью 12 – 24ч с использованием новой статистической модели прогноза с использованием и выходной продукции первой оперативной полусферной модели Гидрометцентра России (автор – Беркович Л.В.). В 2006 г. были проведены испытания прогнозов по этой гидродинамико-статистической модели с заблаговременностью 36ч по нескольким УГМС Европейской территории России (ЕТР). Оценки T критерия Пирси-Обухова составили $T = 0,6 - 0,9$. В новой стати-

стической модели потенциальными предикторами являлись 38 параметров из первой гидродинамической полусферной модели. Она была создана на основе того же байесовского подхода. После проведения отбора наиболее информативного вектора-предсказателя была получена новая дискриминантная функция для прогноза торнадо, сильных шквалов и штормового ветра со скоростью $V \geq 25\text{ м/с}$ $F_1(\mathbf{X})$, значения которой и значения вероятностей прогноза торнадо и шквалов $\mathbf{P}(\mathbf{X})$, $\mathbf{P}(\mathbf{X}) = 100\% / (1 + \exp(-F_1(\mathbf{X})))$ рассчитывались в узлах сетки $150 \times 150\text{ км}$. По пороговым значениям вероятностей в узлах модельной сетки $\mathbf{P}(\mathbf{X}) > 60\%$ для заблаговременности 12–24 ч на карте ЕТР выделялись области этих прогнозируемых явлений. В докладе приводятся прогнозы торнадо по этой модели в Москве в 1998 году, в 2001 году, в Дубне в 2006 году и другие случаи.

В течение 2007–2008 гг проводилась адаптация данной статистической модели к выходной продукции новой оперативной региональной модели Гидрометцентра России. Для случая сильного торнадо в Московской области в г. Краснозаводск вечером 3 июня 2009 года в докладе приводятся карты прогноза по этой гидродинамико-статистической модели с заблаговременностью 24 ч и 48 ч по сроку прогноза 00ч ВСВ. При этом на севере области выделяется прогнозируемая область с высокими вероятностями $\mathbf{P}(\mathbf{X}) > 85\%$. Изолинии этой вероятности выделяют область, охватывающую данный район.

В настоящее время гидродинамико-статистическая модель прогноза сильных шквалов и торнадо со скоростью $V \geq 25\text{ м/с}$ ($V \geq 90\text{ км/ч}$) адаптирована к гидродинамическим прогнозам метеозадающих элементов из новой региональной модели Гидрометцентра России с горизонтальным разрешением $75 \times 75\text{ км}$, в узлах которой, покрывающих территорию ЕТР рассчитываются по той же выше приведенной формуле вероятности прогноза очень сильных шквалов и смерчей. По независимой оценке в 2015 году предупрежденность летних ветров с $V \geq 25\text{ м/с}$ с допуском до 22 м/с составила $W_{pr} = 75–77\%$ для заблаговременности 12–36 ч, а предупрежденность отсутствия этих явлений – $W_{ab} = 85\%$ и выше. При наличии в области прогноза вероятностей $\mathbf{P} = 75\%$ и выше ($\mathbf{P} > 75\%$) дается вероятностный прогноз смерчопасной ситуации.

По данной гидродинамико-статистической модели были успешно даны прогнозы торнадо (смерчей) в г. Краснозаводск (Подмосковье) в июне 2009 года, в Санкт-Петербурге в августе 2010 года, в Одессе в мае 2013 г, в Сочи в сентябре 2013 года и др. Особенно сильный разрушительный смерч (торнадо) третьей категории был отмечен в Янаульском районе Башкортостана (пос. Дюртюли и др.) 29 августа 2014 года. При этом на 6 станциях были отмечены (предсказанные нами) шквалы скоростью 22 м/с и выше. На 26 станциях, вошедших

в область, ограниченную изолинией $P = 65\%$, были отмечены ветры скоростью 19 м/с. Примеры прогноза торнадо и сильных шквалов последних лет будут представлены в докладе.

Следует отметить, что результаты прогноза достаточно близки при расчете прогноза по сроку 00ч ВСВ как на текущий день, так и на вторые сутки с заблаговременностью 36ч (в докладе приводятся соответствующие примеры), что позволяет синоптикам заблаговременно принять решение о необходимости штормового предупреждения, а администрации принимать предохранительные меры, т.к. цена “пропуска цели” существенно выше цены “ложной тревоги”.

В настоящее время наиболее корректной формой выдачи прогноза этих явлений является вероятностный комплексный прогноз, при этом используются результаты прогноза по обоим статистическим моделям с эмпирически найденными весами.

Применение произведения Адамара в одной из комбинаторно-вероятностных задач

Е. А. Потехина

Череповецкий государственный университет, Череповец, Россия

E-mail: peajk7@mail.ru

Произведение Адамара находит применение в ряде задач перечислительной комбинаторики и дискретной теории вероятностей.

Произведением Адамара формальных степенных рядов $G(x) = \sum_{k=0}^{\infty} g_k x^k$ и $H(x) = \sum_{k=0}^{\infty} h_k x^k$ называется степенной ряд $G(x) * H(x) = \sum_{k=0}^{\infty} g_k h_k x^k$.

Рассмотрим задачу замощения прямоугольника плитками в динамике. Будем укладывать бесконечную полосу ширины k плитками размеров $1 \times 1, 1 \times 2, 1 \times 3, \dots$, имеющими соответственно вероятности выпадения для первого ряда — $q_{11}, q_{12}, q_{13}, \dots$, для второго ряда — $q_{21}, q_{22}, q_{23}, \dots$, соответственно, и т. д., для k -го ряда — $q_{k1}, q_{k2}, q_{k3}, \dots$, соответственно. Укладка плиток начинается с первого ряда и выполняется по следующему алгоритму. Каждая новая плитка укладывается в тот ряд, который в данный момент имеет наименьшую длину относительно начала отсчета. Началом отсчета считается момент времени, в который ни один из заполненных плитками рядов не являлся выступающим относительно других. Если несколько рядов плиток имеют одинаковую длину и она является наименьшей, то укладка выполняется в ряд с наименьшим номером. Таким образом, чередуя укладку рядов, заполняем полосу.

Пусть $(X_{1j})_{j=1}^{\infty}, (X_{2j})_{j=1}^{\infty}, \dots, (X_{kj})_{j=1}^{\infty}$ — не зависящие друг от друга последовательности независимых одинаково распределённых случайных величин, принимающих неотрицательные целые значения,

$$g_i(x) = \sum_{j=0}^{\infty} \mathbf{P}(X_{ij} = j) x^j = \sum_{j=0}^{\infty} q_{ij} x^j \quad (i = 1, \dots, k),$$

причем q_{ij} — вероятность того, что плитка i -го ряда будет иметь длину j .

Рассмотрим производящую функцию

$$G_k(x, t) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} p_{k,n,m} x^n t^m,$$

где $p_{k,n,m}$ — вероятность того, что в случайном процессе на каком-то шаге появится прямоугольник размера $k \times n$, состоящий из m плиток.

В [1] доказана следующая теорема.

Теорема 1. *В принятых выше обозначениях справедливо равенство:*

$$G_k(x, t) = \frac{1}{1 - g_1(x)t} * \frac{1}{1 - g_2(x)t} * \dots * \frac{1}{1 - g_k(x)t}.$$

Если случайные величины в рассмотренной выше задаче имеют равномерное распределение, то для вычисления произведений Адамара может быть применена следующая теорема.

Теорема 2. *Для любых целых m, n, k ($m \geq 2, n \geq 2, k < m$) справедлива следующая формула:*

$$\frac{1}{1 - d_1 x - \dots - d_n x^n} * \frac{x^k}{1 - b_1 x - \dots - b_m x^m} = \frac{\det \mathbf{G}}{\det \mathbf{H}},$$

где $\mathbf{H}$ — матрица порядка n , имеющая вид

$$\begin{pmatrix} D_1 & L_{12} & L_{13} & \dots & L_{1,n} \\ D_2 & 1 + L_{22} & L_{23} & \dots & L_{2,n} \\ D_3 & -b_1 + L_{32} & 1 + L_{33} & \dots & L_{3,n} \\ \dots & \dots & \dots & \dots & \dots \\ D_n & -b_{n-2} + L_{n,2} & -b_{n-3} + L_{n,3} & \dots & 1 + L_{n,n} \end{pmatrix},$$

$$L_{ij} = - \sum_{s=i}^m b_s x^{s+j-i} \sum_{t=j}^{s+j-i} d_t f_{s-t+j-i},$$

$$D_i = - \sum_{s=0}^{m-i+1} b_{s+i-1} f_s x^s,$$

$$i = 1, \dots, n; j = 2, \dots, n; b_0 = -1,$$

$$f_s = f_s(d_1, \dots, d_n) = \begin{cases} d_1 f_{s-1} + \dots + d_n f_{s-n} & \text{при } s > 0, \\ 1 & \text{при } s = 0, \\ 0 & \text{при } s < 0, \end{cases}$$

а матрица $\mathbf{G}$ получается из $\mathbf{H}$ заменой первой строки строкой

$$(C_1 \ C_2 \ C_3 \ \dots \ C_n),$$

$$\text{где } C_i = x^{k+i-1} \sum_{s=i}^n d_s f_{k-s+i-1}.$$

Теорема 3. Пусть $(X_j)_{j=1}^{\infty}, (Y_j)_{j=1}^{\infty}$ — не зависящие друг от друга последовательности независимых случайных величин, принимающих неотрицательные целые значения, имеющих равномерное распределение и соответствующие производящие функции $g_1(x) = t^{-1}(1+x+\dots+x^{m-1})$, $g_2(x) = n^{-1}(1+x+\dots+x^{n-1})$, где $n=4$. Тогда в принятых выше обозначениях справедливо равенство

$$G_2(x, t) = 4m(4-t)^{m-4} R(x, t) / S(x, t),$$

$$\begin{aligned} R(x, t) = & (m-t)^2(4-t)^3 - t(m-t)^2(4-t)^2 x - mt(m-t)(4-t)^2 x^2 - \\ & - m^2 t(4-t)^2 x^3 + t^2(m-t)(4-t)^2 (f_{m-2} + 2f_{m-3}) x^m + \\ & + t^2(4-t)^2 (mf_{m-2} + (m-t)(f_{m-1} + f_{m-2})) x^{m+1} - \\ & - t^3(4-t)(m-t)(f_{m-2} + 2f_{m-3}) x^{m+1} + \\ & + mt^2(4-t)(t(f_{m-4} - f_{m-2}) + (4-t)f_{m-1}) x^{m+2} + \\ & + t^4(4-t)(f_{m-3}^2 - f_{m-4}f_{m-2}) x^{2m} + \\ & + t^4(t(f_{m-4}f_{m-2} - f_{m-3}^2) + (4-t)(f_{m-3}f_{m-2} - f_{m-4}f_{m-1})) x^{2m+1}, \\ S(x, t) = & (m-t)^3(4-t)^m - mt(m-t)^2(4-t)^{m-1} x - \\ & - m^2 t(m-t)(4-t)^{m-1} x^2 - m^3 t(4-t)^{m-1} x^3 + \\ & + t(m-t)^2(4-t)^{m-1}((4-t)f_m + t(f_{m-2} + 2f_{m-3})) x^m - \\ & - mt^3(m-t)(4-t)^{m-2}(f_{m-2} + 2f_{m-3}) x^{m+1} + \\ & + mt^2(m-t)(4-t)^{m-1}(2f_{m-1} + 3f_{m-2}) x^{m+1} - \\ & - m^2 t^2(4-t)^{m-2}(t(f_{m-2} - f_{m-4}) - 2(4-t)f_{m-1}) x^{m+2} + \\ & + t^3(m-t)(4-t)^{m-1}(f_{m-2}f_m - f_{m-1}^2) x^{2m} + \\ & + 2t^3(m-t)(4-t)^{m-1}(f_{m-3}f_m - f_{m-2}f_{m-1}) x^{2m} + \\ & + t^4(m-t)(4-t)^{m-2}(f_{m-3}^2 - f_{m-4}f_{m-2}) x^{2m} + \\ & + mt^3(4-t)^{m-1}(f_{m-2}f_m - f_{m-1}^2) x^{2m+1} + \\ & + mt^4(4-t)^{m-3}(f_{m-4}f_{m-2} - f_{m-3}^2) x^{2m+1} - t^{m+3} x^{3m}, \\ f_m = & \sum_{i=1}^3 f_{m-i} t(4-t)^{-1} \text{ при } m > 0, f_m = 1 \text{ при } m = 0, f_m = 0 \text{ при } m < 0. \end{aligned}$$

Теорема 3 позволяет рассчитать вероятности $p_{k,n,m}$ для случая, когда распределение вероятностей того, что плитка, укладываемая в конкретный ряд, будет иметь определенную длину, равномерно.

Список литературы

1. Потехина Е. А. Приложение произведения Адамара к некоторым комбинаторным и вероятностным задачам // Дискретная математика. 2016. Т. 28, вып. 1. С. 101–112.

Equilibria in Dynamic Multicriteria Games with Incomplete Information

A. N. Rettieva

Institute of Applied Mathematical Research
Karelian Research Centre RAS, Petrozavodsk, Russia

E-mail: annaret@krc.karelia.ru

Mathematical models involving more than one objective seem more adherent to the real problems. Often players have more than one goal and they can be not comparable. These situations are typical for game-theoretic models in economy and ecology. Hence, multicriteria game approach helps to make decisions in multi-objective problems.

Traditionally, equilibrium analysis in multicriteria problems bases on the static or steady-state variant. For dynamic multicriteria games proposed equilibrium concepts do not assist in evaluating players' behavior. Furthermore, the situation with incomplete information, namely then the players' planning horizons are different and random, is considered. The model with random planning horizons in the bioresource exploitation process is most adequate to the reality: external random factors can cause the game breach and the participants know nothing about them a priori. As a matter of fact, equilibrium behavior of participants has not been examined in this case to date.

According to the aforesaid, equilibrium behavior design in dynamic models with many objectives calls for elaborating new methods. Thus and so, this work dedicated to linking multicriteria games with dynamic games.

We consider a bicriteria dynamic game with two participants in discrete time. Players exploit the common resource and both wish to optimize two different criteria. The state dynamics is in the form

$$x_{t+1} = f(x_t, u_{1t}, u_{2t}), \quad x_0 = x, \quad (1)$$

where $x_t \geq 0$ is the resource size at step t , $u_{it} \in U_i$ indicates the strategy of player i , $i = 1, 2$.

The payoff functions of the players over infinite time horizon are defined by

$$J_1 = \begin{pmatrix} J_1^1 = \sum_{t=0}^{\infty} \delta^t g_1^1(u_{1t}, u_{2t}) \\ J_1^2 = \sum_{t=0}^{\infty} \delta^t g_1^2(u_{1t}, u_{2t}) \end{pmatrix}, J_2 = \begin{pmatrix} J_2^1 = \sum_{t=0}^{\infty} \delta^t g_2^1(u_{1t}, u_{2t}) \\ J_2^2 = \sum_{t=0}^{\infty} \delta^t g_2^2(u_{1t}, u_{2t}) \end{pmatrix}, \quad (2)$$

where $g_i^j(u_{1t}, u_{2t}) \geq 0$ gives the instantaneous utility, $i, j = 1, 2$, $\delta \in (0, 1)$ means the common discount factor.

In the present work we design the equilibrium in multicriteria game using the Nash bargaining solution. Therefore, we begin with construction of guaranteed payoffs which play the role of the status quo points.

There are three possible concepts to determine the guaranteed payoffs $G_1^1, G_1^2, G_2^1, G_2^2$.

In the first one four guaranteed payoff points are obtained as the solutions of zero-sum games. In particular, the first guaranteed payoff point is a solution of zero-sum game where player 1 wishes to maximize her first criterion and player 2 wants to minimize it. Other points are obtained by analogy.

The second approach can be applied when the players' objectives are comparable. Consequently, the guaranteed payoff points for player 1 are obtained as the solution of zero-sum game where she wants to maximize the sum of her criteria and player 2 wishes to minimize it. And, by analogy, for player 2.

In the third approach the guaranteed payoff points are constructed as the Nash equilibrium with the first and the second criteria of both players, respectively.

To construct multicriteria payoff functions we adopt the Nash products. The role of the status quo points belongs to the guaranteed payoffs of the players:

$$H_1(u_{1t}, u_{2t}) = (J_1^1(u_{1t}, u_{2t}) - G_1^1)(J_1^2(u_{1t}, u_{2t}) - G_1^2), \quad (3)$$

$$H_2(u_{1t}, u_{2t}) = (J_2^1(u_{1t}, u_{2t}) - G_2^1)(J_2^2(u_{1t}, u_{2t}) - G_2^2). \quad (4)$$

Next definition presents the suggested solution concept.

Definition. *Strategy profile (u_{1t}^*, u_{2t}^*) is called multicriteria Nash equilibrium of the problem (1)–(2) if*

$$H_1(u_{1t}^*, u_{2t}^*) \geq H_1(u_{1t}, u_{2t}^*) \quad \forall u_{1t} \in U_1, \quad (5)$$

$$H_2(u_{1t}^*, u_{2t}^*) \geq H_2(u_{1t}^*, u_{2t}) \quad \forall u_{2t} \in U_2. \quad (6)$$

Just like in classical Nash equilibrium approach it is not profitable for both players to deviate from equilibrium strategies. But under presented equilibrium concept players maximize the product of the differences between optimal and guaranteed payoffs (3)–(4).

A dynamic multicriteria model related with the bioresource management problem (fish catching) is investigated to show how suggested concept works. Furthermore we explore the model, where players possess heterogeneous planning horizons.

References

1. Patrone F., Pusillo L. and Tijs S.H. Multicriteria games and potentials // *Top*. 2007. V. 15. P. 138–145.
2. Pusillo L., Tijs S. E-equilibria for multicriteria games // *Annals of ISDG*. 2013. V. 12. P. 217–228.
3. Rettieva A.N. A discrete-time bioresource management problem with asymmetric players // *Automation and Remote Control*. 2014. V. 75(9). P. 1665–1676.
4. Shapley L.S. Equilibrium points in games with vector payoffs // *Naval Research Logistic Quarterly*. 1959. V. 6. P. 57–61.
5. Voorneveld M., Grahn S. and Dufwenberg M. Ideal equilibria in non-cooperative multicriteria games // *Mathematical Methods of Operations Research*. 2000. V. 52. P. 65–77.

Ускоренная оценка стационарности вероятностной модели процесса нагрузки вычислительного кластера

А. С. Румянцев

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: ar0@krc.karelia.ru

Высокопроизводительный вычислительный кластер (ВК) представляет собой особый вид многосерверной системы массового обслуживания, в котором заявке может требоваться одновременно более одного сервера. Заявка занимает и освобождает выделенные ей серверы одновременно. Поток заявок на такую систему формируется множеством пользователей, которые получают доступ к ресурсам в удаленном режиме. Разделение ресурсов ВК между заявками осуществляет менеджер очереди. В простейшем случае заявки обслуживаются в порядке поступления (дисциплина FCFS). Для настройки параметров менеджера очереди, а также оценки характеристик ВК, целесообразно применение стохастических моделей [2].

При разработке новых и эксплуатации существующих ВК, важной задачей является оценка стационарности системы. Стационарность ВК при заданных характеристиках потока заявок означает, что очередь системы не растет неограниченно. Критерий стационарности вероятностной модели процесса нагрузки ВК получен в работе [4].

Предположим, что в s -серверный ВК поступает простейший поток заявок интенсивности λ . Каждой заявке i требуется одновременно N_i серверов на случайное время, экспоненциально распределенное с параметром μ . Если число свободных серверов меньше N_i , либо в системе есть ожидающие заявки, то заявка i поступает в очередь, ожидая освобождения требующихся серверов. Случайные величины $\{N_i\}$ — н.о.р., имеют дискретное распределение $\mathbf{p} = (p_1, \dots, p_s)$,

$$p_k := P\{N_i = k\}, \quad k = 1, \dots, s.$$

Назовем *фазой системы* вектор $m = (m_1, \dots, m_s) \in \mathcal{M} := \{1, \dots, s\}^s$, где m_i есть количество серверов, требуемых i -той старшей заявкой

среди находящихся в системе. Для данной фазы m число заявок, находящихся на обслуживании, есть $\sigma(m) := \max\{i : \sum_{j=1}^i m_j \leq s\} \leq s$. Система стационарна тогда и только тогда, когда (см. [4])

$$\rho := \frac{\lambda}{\mu} C(s) < 1, \quad (1)$$

где

$$C(s) := \sum_{m \in \mathcal{M}} \frac{1}{\sigma(m)} \prod_{i=1}^s p_{m_i}. \quad (2)$$

Таким образом, для оценки стационарности необходимо вычислять значение $C(s)$ при заданном распределении $\mathbf{p}$. Отметим, что мощность множества $\mathcal{M}$ есть s^s , что не позволяет применять прямой метод вычисления (2) даже для сравнительно небольших систем (например при $s = 100$). В этой связи в работе [4] предложен метод ускоренного вычисления величины $C(s)$. Метод основан на разбиении в (2) исходной суммы по множеству $\mathcal{M}$ следующим образом:

$$C(s) = \sum_{k=1}^s \sum_{n \in \mathcal{N}_k} \frac{(\sum_{i=1}^s n_i - 1)!}{\prod_{i=1}^s n_i!} \prod_{i=1}^s p_i^{n_i} \sum_{j=s-k+1}^s p_j, \quad (3)$$

где $n = (n_1, \dots, n_s) \in \mathcal{N}_k$ есть векторная форма записи целочисленного разбиения числа занятых серверов $k = 1, \dots, s$, т.е.,

$$k = n_1 + 2n_2 + \dots + sn_s.$$

Вычисление величины $C(s)$ в соответствии с формулой (3) требует порядка $se^{\sqrt{s}}$ операций, что существенно быстрее, чем прямой метод для больших s . Отметим, что порядок суммирования в (3) предполагает перебор сначала по числу занятых серверов k , а затем по возможным разбиениям этого числа между заявками, находящимися на обслуживании. При перестановке порядка суммирования в (3), внешняя сумма предполагает перебор по числу заявок, находящихся на обслуживании, а внутренняя — по характеристикам этих заявок. Именно,

$$C(s) = \sum_{k=1}^s \frac{1}{k} \sum_{m \in \mathcal{M}^{(k)}} \prod_{i=1}^s p_{m_i}, \quad (4)$$

где $\mathcal{M}^{(k)} = \{m \in \mathcal{M} : \sigma(m) = k\}$ есть множество фаз, в которых на обслуживании находится ровно k заявок. В работе [3] доказано, что

$$C(s) = \sum_{k=1}^s \frac{1}{k} \sum_{j=k}^s p_j^{k*} \sum_{t=s+1-j}^s p_t, \quad (5)$$

где p^{k*} есть k -кратная дискретная свертка вектора $\mathbf{p}$. При этом суммирование в (5) производится сначала по числу заявок, находящихся на обслуживании, а затем по числу занятых процессоров. Вычисление величины $C(s)$ из (5) требует порядка $s^3 \log s$ операций. Кроме того, дополнительного ускорения можно добиться за счет использования быстрого преобразования Фурье для вычисления дискретной свертки. Кроме того, вычисление может вестись в параллельном режиме, в связи с отсутствием зависимости между слагаемыми. Наконец, отметим, что, обрывая вычисление внешней суммы при $k = q < s$, можно получить оценку для $C(s)$, которая сходится к реальному значению с ростом q . Отметим, что функция для вычисления нормирующей константы $C(s)$ доступна в программном пакете `hpcwld` для среды вычислений R [1].

Список литературы

- [1] CRAN - Package `hpcwld`, <http://cran.r-project.org/web/packages/hpcwld/index.html>
- [2] Feitelson, Dror G. Workload modeling for computer systems performance evaluation. Cambridge University Press, 2015
- [3] Rumyantsev A., Morozov E.: Accelerated Verification of Stability of Simultaneous Service Multiserver Systems. In: 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, pp. 239–242. IEEE (2015)
- [4] Rumyantsev A., Morozov E.: Stability criterion of a multiserver model with simultaneous service. *Ann Oper Res.* 1–11 (2015)

К анализу стационарных характеристик системы массового обслуживания со случайными требованиями

К. Е. Самуйлов¹, Э. С. Сопин^{1,2}, О. Г. Вихрова¹

¹ Российский университет дружбы народов,

² Институт проблем информатики ФИЦ ИУ РАН,
Москва, Россия

E-mail: ksam@sci.pfu.edu.ru, esopin@sci.pfu.edu.ru,
o.vikhrova@gmail.com

В современных сотовых сетях связи каждому из пользователей необходимо некоторое количество ресурсов системы для установления соединения, в зависимости от используемого планировщиками такими ресурсами могут быть диапазон частот, излучаемая мощность или доля времени [1,2]. Такого рода сеть может быть смоделирована при помощи системы массового обслуживания, изображенной на рисунке 1, в которой каждый пользователь занимает случайный объем ресурса различных типов [3]. Если для обслуживания некоторого пользователя система не может выделить достаточный объем ресурсов, то запрос на обслуживание будет отклонен. При завершении обслуживания пользователя, система должна освободить столько же ресурса, сколько было выделено данному пользователю в начале его обслуживания.

В [4] был описан общий подход к анализу вышеописанной системы, а также предложено её упрощение. Упрощенная система отличается от исходной тем, что объем ресурсов, который освобождается по завершении обслуживания пользователя, отличается от того, который был ему изначально выделен, и также является случайной величиной. Зная общий объем занятых ресурсов и количество пользователей в системе, можно получить функцию распределения объема ресурсов, который освободится по завершении обслуживания пользователя, так как этот объем не зависит от состояния системы в прошлом. Случайный процесс, который описывает поведение упрощенной СМО, более

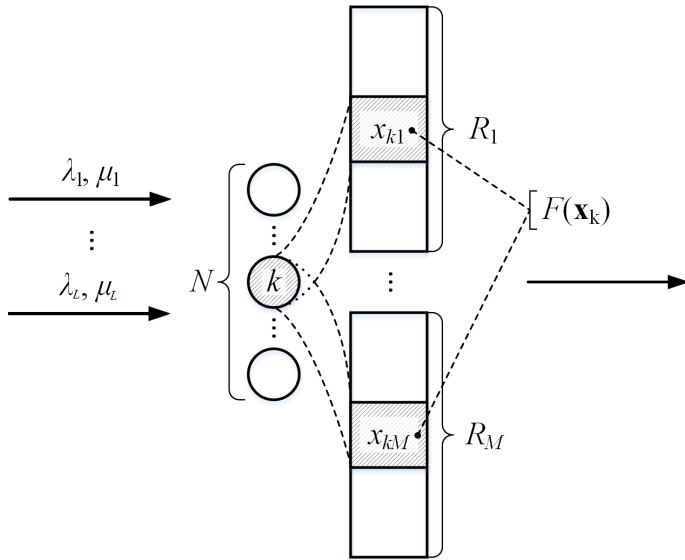


Рис. 1: Многолинейная СМО со случайными требованиями и множественными ресурсами

удобен в анализе, поскольку он не подразумевает запоминания объема ресурса, который был выделен каждому пользователю, а требует только знания, какой объем ресурсов занят пользователями в целом. При этом имитационное моделирование в [4] показало, что стационарные характеристики исходной и упрощенной модели практически совпадают для различных типов входящих потоков, широкого набора распределений времени обслуживания заявки и распределений требований к ресурсам.

В случае дискретных случайных величин требований заявок к ресурсам получен ряд результатов, позволяющих вычислять стационарное распределение вероятностей с использованием матричных методов, которые, однако обладают довольно высокой вычислительной сложностью [5].

В данной работе проведен анализ СМО с несколькими входящими потоками. Для данной системы была доказана мультипликативность распределения стационарных вероятностей. Также затрагивается проблема вычислительной сложности алгоритмов расчета вероятности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 15-07-03051, 15-07-03608, 16-37-60103, 16-07-00766

блокировки запросов, поступающих в систему и среднего объема занятых ресурсов, разработан рекуррентный алгоритм расчета данных характеристик для случая дискретных требований к ресурсам. Кроме того, показана приложениз данной модели к анализу современных телекоммуникационных систем четвертого и пятого поколений.

Список литературы

1. Kim H. and de Veciana G. Leveraging dynamic spare capacity in wireless systems to conserve mobile terminals' energy. // IEEE/ACM Transactions on Networking, 18(3), 2010, pp. 802–815.
2. Elshaer H., Boccardi F., Dohler M., Irmer R.. Downlink and uplink decoupling: a disruptive architectural design for 5G networks // Global Communications Conference (GLOBECOM), 2014 IEEE, pp. 1798-1803.
3. Наумов В.А., Самуйлов К.Е. О моделировании системы массового обслуживания с множественными ресурсами. // Вестник РУДН. Серия Математика. Информатика. Физика, №3, 2014, с. 60–64.
4. Naumov V., Samouylov K., Sopin E., Andreev S. Two approaches to analyzing dynamic cellular networks with limited resources. // 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2014, pp. 485–488.
5. Вихрова О.Г., Самуйлов К.Е., Сопин Э.С., Шоргин С.Я. К анализу качества обслуживания в современных беспроводных сетях // Информатика и ее применения, №4, 2015, с. 48 - 55.

Число независимости случайного разреженного k -однородного гиперграфа

А. С. Семенов

Московский государственный университет имени М.В. Ломоносова,
ФИБТ МФТИ (Государственный университет),
Москва, Россия

E-mail: alexsemenov1992@mail.ru

Гиперграфом называется пара множеств $H = (V, E)$, где $V = V(H)$ – некоторое множество, называемое множеством вершин гиперграфа, а $E = E(H)$ – произвольная совокупность подмножеств множества V , называемых ребрами гиперграфа. Гиперграф называется *k -однородным*, если каждое его ребро содержит ровно k вершин. Через $K_n(k)$ будем обозначать полный k -однородный гиперграф на n вершинах.

Подмножество вершин гиперграфа $\mathcal{I} \subset V$ называется *независимым*, если никакие вершины из этого подмножества не образуют ребра

$$\forall e \in E(H) : |\mathcal{I} \cap e| < |e|.$$

Числом независимости $\alpha(H)$ гиперграфа $H(V, E)$ называется размер максимального независимого подмножества его вершин.

В данной работе рассматривается задача о поиске асимптотики числа независимости случайных k -однородных гиперграфов. В качестве случайного гиперграфа рассматривается *биномиальная модель* случайных подмножеств. Пусть $p \in [0, 1]$, а $\mathcal{H}(k)$ – множество всех остовных подгиперграфов $K_n(k)$. Будем называть случайным гиперграфом $H(n, k, p)$ случайный элемент, принимающий значения в $\mathcal{H}(k)$ и имеющий на нем следующее распределение:

$$P(H(n, k, p) = F) = p^{|E(F)|} (1 - p)^{C_n^k - |E(F)|}, \forall F \in \mathcal{H}(k).$$

В разреженном случае (малые значения p) гиперграф будет представлять объединение гипердеревьев, гиперграфов без циклов в которых ребра имеют максимум одну общую вершину. С помощью метода интерполяции [1] показывается истинность следующей теоремы.

Теорема 1. Для случайного гиперграфа $H(n, k, p)$, где $p = c/C_{n-1}^{k-1}$ и $c > 0$, существует постоянная $\alpha(k, c)$, такая что

$$\lim_{n \rightarrow \infty} \frac{E[\alpha(H(n, k, p))]}{n} = \alpha(k, c).$$

Кроме того, из данной теоремы можно получить следующее полезное следствие.

Следствие. Для случайного гиперграфа $H(n, k, p)$, где $p = c/C_{n-1}^{k-1}$ и $c > 0$, существует постоянная $\alpha(k, c)$, такая что при $n \rightarrow \infty$

$$\frac{\alpha(H(n, k, p))}{n} \xrightarrow{P} \alpha(k, c).$$

Далее при помощи алгоритмического подхода [2] в случае малых значений константы c можно найти значение предела из предыдущей теоремы.

Теорема 2. Для случайного гиперграфа $H(n, k, p)$, $p = c/C_{n-1}^{k-1}$ и $0 < c \leq \frac{1}{k-1}$ выполнено

$$\lim_{n \rightarrow \infty} \frac{E[\alpha(H(n, k, p))]}{n} = r + \frac{c(k-1)}{k} r^k,$$

где r - это решение уравнения $r = e^{-cr^{k-1}}$.

Работа выполнена при финансовой поддержке гранта Президента РФ № МД-5650.2016.1

Список литературы

[1] M. Bayati, D. Gamarnik, and P. Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. The Annals of Probability. Vol 41, №. 6, 2013, pp. 4080-4115.

[2] R. Karp and M. Sipser, Maximum matchings in sparse random graphs, 22nd Annual Symposium on Foundation of Computer Science, 1981, pp. 364-375.

Кластеризация районов Санкт-Петербурга по показателю заболеваемости

К. Ю. Староверова, В. М. Буре

Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия

E-mail: ksenygnirps@gmail.com

Цель исследования — произвести кластеризацию 18 районов Санкт-Петербурга по показателю первичной общей заболеваемости на 1000 человек населения с 1999 по 2014 годы.

Метод Фреше [1]. Пусть M — множество всевозможных последовательностей m пар, которые сохраняют хронологический порядок: $r = ((X_{a_1}, Y_{b_1}), \dots, (X_{a_m}, Y_{b_m}))$, где $r \in M, a_i, b_j \in \{1, \dots, T\}$ и $\forall i, j \in \{1, \dots, m-1\}$ выполнены условия $a_1 = b_1 = 1; a_m = b_m = T; a_{i+1} = a_i$ или $a_{i+1} = a_i + 1; b_{i+1} = b_i$ или $b_{i+1} = b_i + 1$. Тогда мера схожести Фреше для временных рядов X, Y , состоящих из T наблюдений определяется как

$$d_F(X, Y) = \min_{r \in M} \left(\max_{i \in \{1, \dots, m\}} |X_{a_i} - Y_{b_i}| \right).$$

Этот метод хорош тем, что он прост для вычислений, может работать с временными рядами разной длины и учитывает зависимость наблюдений от времени (рис. 1).

Регулируемая мера сходства [2]. Введем аналог коэффициента корреляции, который будет равен 1, если на всем временном интервале ряды S_1 и S_2 убывают или возрастают с одинаковыми коэффициентами, и равен -1 , если на всем временном интервале ряды S_1 и S_2 имеют строго противоположную динамику. Определим функцию схожести параметра $k \geq 0$, которая будет учитывать и динамику, и расстояние между наблюдениями как

$$D_k(S_1, S_2) = \frac{2}{1 + e^{kCorT(S_1, S_2)}} \delta(S_1, S_2),$$

здесь $\delta(S_1, S_2)$ — это любая классическая мера схожести (Евклида, Фреше и др.), а $CorT(S_1, S_2)$ — аналог коэффициента корреляции (рис. 2).

быть ограничена сверху и снизу [4]. Пусть MBP имеет n размерностей. Представим временные ряды S_1 и S_2 в виде матриц $\widehat{S}_1 = [s_1^1, \dots, s_1^n]$, $\widehat{S}_2 = [s_2^1, \dots, s_2^n]$. Если $V_1 = [v_1^1, \dots, v_1^n]$, $V_2 = [v_2^1, \dots, v_2^n]$ — это матрицы правых сингулярных векторов, полученные SVD разложением матриц ковариаций, w — это вектор весовых коэффициентов (сингулярных чисел), тогда расстояние между рядами по методу Eros вычисляется по формуле

$$D_{Eros}(S_1, S_2, w) = \sqrt{2 - 2 \sum_{i=1}^n w^i |\langle s_1^i, s_2^i \rangle|}.$$

Метод реализован на языке R, результаты многомерной кластеризации часто сложно интерпретировать, одним из показателей качества кластеризации может служить коэффициент вариации (таблица 1).

Таблица 1: Коэффициент вариации в кластерах.

	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Дети	11,80	16,60	20,27	23,06
Подростки	25,69	18,42	28,09	34,81
Взрослые	12,86	23,30	35,82	25,93

Список литературы

- [1] Montero P., Vilar J. TSclust: An R package for time series clustering // Journal of Statistical Software. 2015. No 62.1. P. 1–43.
- [2] Douzal Chouakria A., Nagabhushan P.N. Adaptive dissimilarity index for measuring time series proximity // Advances in Data Analysis and Classification. March 2007. Vol. 1. Iss. 1. P. 1–43.
- [3] Li S. J. , Zhu Y. L., Zhang X. H. and Wan D. BORDA counting method based similarity analysis of multivariate hydrological time series // Journal of Hydraulic Engineering. 2009. Vol. 40, No 3. P. 378–384.
- [4] Yang K., Shahabi C. A PCA-based similarity measure for multivariate time series // MMDB '04 Proceedings of the 2nd ACM international workshop on Multimedia databases. 2004. P. 65–74.

Идентифицируемость модели факторного анализа с негауссовскими общими факторами

С. В. Стафеев

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: stafeev@krc.karelia.ru

Рассмотрим модель факторного анализа:

$$\mathbf{X} = \mathbf{A}\mathbf{H} + \mathbf{Y}, \quad (1)$$

где $\mathbf{X} = \{X_1, \dots, X_n\}^t$ вектор с матрицей ковариаций $\Sigma = (\sigma_{ij})$ и множеством третьих моментов $\Gamma_X = \{\mathbf{M}(X_i X_j X_l), i, j, l = 1, \dots, n\}$; $\mathbf{H} = \{H_1, \dots, H_k\}^t$ – множество независимых латентных (скрытых) случайных величин с $\mathbf{M}(H_j) = 0$, $\mathbf{M}(H_j)^2 = 1$ и $\mathbf{M}(H_j)^3 = d_j$, $j = 1, \dots, k$; $\mathbf{A} = (a_{ij})$ – матрица факторных нагрузок. Вектор остатков $\mathbf{Y} = \{Y_1, \dots, Y_n\}^t$ имеет распределение с нулевым вектором математических ожиданий, матрицей ковариаций $\Theta_G = (\theta_{ij})$ и множеством третьих моментов $\Gamma_Y = \{\mathbf{M}(Y_i Y_j Y_l), i, j, l = 1, \dots, n\}$. Векторы $\mathbf{Y}$ и $\mathbf{H}$ являются независимыми. Взаимосвязь между компонентами вектора $\mathbf{Y}$ представляется в виде графовой модели со структурой $G = (V, E)$, где $V = \{1, \dots, n\}$ и $(i, j) \notin E$, если Y_i и Y_j являются независимыми.

Параметры модели (1) состоят из матрицы факторных нагрузок $\mathbf{A}$, вектора третьих моментов (коэффициентов асимметрии) факторов $\mathbf{D} = \{d_1, \dots, d_k\}$ и ненулевых элементов матрицы Θ_G и множества Γ_Y .

Модель (1) называется глобально идентифицируемой, если по матрице Σ и множеству Γ_X матрица $\mathbf{A}$ определяется с точностью до перестановки и знаков столбцов, элементы вектора $\mathbf{D}$ с точностью до знака, а элементы матрицы Θ_G и множества Γ_Y определяются однозначно.

Пусть $\overline{G} = (V, \overline{E})$ – дополнительный к G граф. Образует граф $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, где $\mathbf{V} = \{\mathbf{i} = (i_1, \dots, i_k), 1 \leq i_1 < i_2 < \dots < i_k \leq k\}$, а $\mathbf{E} = \{(\mathbf{i}, \mathbf{j}) | \mathbf{i} = (i_1, \dots, i_k), \mathbf{j} = (j_1, \dots, j_k), (i_s, j_l) \in \overline{E}, s, l = 1, \dots, k\}$. Каждому ребру графа $\mathcal{G}$ сопоставим в соответствие детерминант $|\Sigma_{\mathbf{i}, \mathbf{j}}|$

матрицы $\Sigma_{\mathbf{i}, \mathbf{j}} = (\sigma_{i_m, j_l})_{m, l=1}^k$. Данный $k \times k$ минор матрицы Σ будем называть весом ребра $(\mathbf{i}, \mathbf{j})$.

Теорема 1. Пусть все элементы вектора факторов $\mathbf{D}$ имеют конечные и ненулевые коэффициенты асимметрии и пусть вектор остатков $\mathbf{Y}$ обладает конечными третьими моментами. Модель (1) является глобально идентифицируемой, если граф, полученный из графа $\mathcal{G}$ с помощью удаления ребер, соответствующих нулевым весам, содержит компоненту связности $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ и $\cup_{\mathbf{i} \in \mathbf{V}'} \mathbf{i} = V$.

СПИСОК ЛИТЕРАТУРЫ

1. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. М.: ФИЗМАТЛИТ, 1972, 656 с.
2. Bonhomme S., Robin J.M. Consistent Noisy Independent Component Analysis. – Journal of Econometrics, 2009, v. 149(1), p. 12-25.
3. Стафеев С. В. Об условиях глобальной идентифицируемости для моделей факторного анализа // Труды Карельского научного центра Российской академии наук, 2011, в. 5, с. 111-114.

Современные статистические выводы применительно к одному конфликту при сортировке алмазов

С. А. Хонов

E-mail: sergey.khonov@yandex.ru

Газета "Коммерсант" в номере от 28.08.2015 опубликовала статью Е.Киселёвой 'Ущерб может вырасти многократно', в которой описала спорную ситуацию, возникшую после сортировки и оценки Гохраном партии алмазов ОАО "Севералмаз". Суть разногласий сторон относилась к тому, что часть сырья якобы была подменена на худшее по стоимостным характеристикам. Или в нашей терминологии приёщиком были изменены размеры классов, составляющих партию.

Аттестация и сортировка партий алмазов производится в соответствии со стандартами ГОСТ Р 51519.1-99 и Р 51519.2-99. Алмазы классифицируются по утверждённому перечню градаций размерности, формы, дефектности и цвета.

По каждому признаку определено порядка 10 градаций. Причём, по итогам сортировки многие 'мусорные' подклассы объединяются. В итоге мы имеем не $10 \times 10 \times 10 \times 10$ разных градаций, а намного меньше. Подробнее с литературным описанием конфликтной ситуации с претензиями поставщика можно ознакомиться по адресу: <http://www.kommersant.ru/doc/2796814>

Поскольку каждый ювелирный алмаз имеет свой персональный сертификат, наш интерес к этой газетной публикации вызван лишь в части, относящейся к выборочному оцениванию составляющих партию классов относительно дешёвых промышленных алмазов, аттестуемых на основании проб (выборок). Причём, такая аттестация вполне удовлетворяет условиям статистической модели, рассматриваемой автором. А именно, (*структура совокупности*) рассматривается расслоенная конечная совокупность

$$U = \bigcup_{i=1}^k U_i; U_i \cap U_j = \emptyset, i \neq j; i, j = 1, \dots, k,$$

для которой число классов k фиксировано и известно, а

(*объект оценивания*) оцениваемые параметры - размеры классов (слов) $N_1, \dots, N_k$ удовлетворяют соотношениям:

$$N_j = \beta_j N, \delta_0 \leq \beta_j \leq 1 - \delta_0, \delta_0 > 0, j = 1, \dots, k, \quad (1)$$

$$0 < \alpha_1 \leq \alpha = n/N \leq \alpha_2 < \infty \quad (2)$$

так, что N и все $N_j, j = 1, \dots, k$, имеют те же порядки, что и общее число наблюдений n .

Для получения статистической информации используется (*выборочный план*) обобщённая схема выбора (ОСВ), состоящая в последовательном извлечении s бесповторных выборок фиксированного размера $m_i, i = 1, \dots, s$ из всей, полной совокупности U , число исходов в ОСВ равно

$$\Psi_s(N) = \prod_{i=1}^s \binom{N}{m_i};$$

и пусть

$$G(\alpha) = \prod_{i=1}^s \left(1 - \alpha \frac{m_i}{n}\right) / \alpha \quad (3)$$

При этом, число выборок s фиксировано, размеры выборок m_i возрастают вместе с $n = m_1 + \dots + m_s$ так, что

$$\frac{m_i}{n} = v_i, v_i \in [c, 1 - c], \exists c > 0, i = 1, \dots, s.$$

В статье [2] было детально исследовано асимптотическое поведение статистики $\bar{\eta} = (\eta_1, \dots, \eta_k)$, где η_j - наблюдаемое число различных элементов j -го класса за n наблюдений, $\eta = \eta_1 + \dots + \eta_k$, и построены состоятельные асимптотически нормальные оценки $\hat{\beta}_n = (\hat{\beta}_{1n}, \dots, \hat{\beta}_{kn}) = \bar{\eta}/\eta$ для вектора долей каждого класса $\beta = (\beta_1, \dots, \beta_k)$. А также - для параметра α , когда общий размер совокупности N неизвестен.

Результаты работы [2] позволяют рассчитывать соответствующие доверительные интервалы и области для $\hat{\alpha}$ и $\hat{\beta}$, а также строить соответствующие критерии для проверок гипотез о составе расслоенной совокупности.

Поскольку в силу многопараметричности рассматриваемой статистической модели, дальнейшие построения могут утратить связь с задачей сортировки алмазов, сформулируем два базовых результата, на основании которых можно уже в продолжение исследований рассматривать самые разные постановки задач.

Итак, в теореме 6 из [2] доказано, что при выполнении условий (1) и (2) статистики $\hat{\alpha}_n = G^{-1}(\eta/n)$ и $\hat{\beta}_n = \bar{\eta}/\eta$ являются состоятельными асимптотически нормальными оценками α и $\bar{\beta}$ соответственно. Найдены конкретные параметры асимптотических нормальных распределений. Причём, что важно, в приведенных выражениях для вторых моментов указанных нормальных распределений неизвестные параметры α и $\bar{\beta}$ могут быть заменены их оценками $\hat{\alpha}_n$ и $\hat{\beta}_n$.

Отметим, что помимо состоятельных асимптотически нормальных оценок параметров модели α и $\bar{\beta}$ доказанная теорема позволяет строить как состоятельные асимптотически нормальные оценки максимального правдоподобия $\tau(\hat{\alpha}_n, \hat{\beta}_n)$ для произвольных гладких функций $\tau(\alpha, \bar{\beta})$, так и асимптотические доверительные интервалы для них.

Следуя [2], получим, что асимптотическое γ -доверительное множество для вектора $\bar{\beta}$ имеет вид

$$\mathcal{G}_{\gamma_n} = \left\{ \bar{\beta} : \sum_{i=1}^k \frac{(\hat{\beta}_{jn} - \beta_j)^2}{\hat{\beta}_{jn}} < \frac{\chi_{\gamma, k-1}^2 \psi(\hat{\alpha}_n)}{n}, \sum_{j=1}^k \beta_j = 1 \right\}, \quad (4)$$

где $\chi_{\gamma, k-1}^2$ есть γ -квантиль распределения χ^2 с $k-1$ степенями свободы, $\psi(\alpha) = (1 - \alpha G(\alpha)) / G(\alpha)$, определение функции $G(\alpha)$ см. (3).

С учётом изложенного, только и если в результате испытаний мы получим значения $\hat{\beta}$, лежащие вне указанной γ -доверительной области, то только тогда мы сможем говорить, что значения $\hat{\beta}$ статистически значимо выходят за границы допустимых отклонений, и на этих основаниях предъявлять математически обоснованные претензии.

Все другие значения $\hat{\beta}$, попавшие внутрь γ -доверительной области, не дают статистически обоснованных оснований для того, чтобы считать их значимо отличными от гипотетических.

Что не исключает взаимных претензий сторон на основаниях, уже не связанных с математическими статистическими выводами.

Список литературы

1. Киселёва Е. Ущерб может вырасти многократно. Газета Коммерсант от 28.08.2015, <http://www.kommersant.ru/doc/2796814>
2. Ивченко Г. И., Хонов С. А. Об асимптотическом оценивании для расслоенных конечных совокупностей // Дискретная математика (1989) 1, №3, 87-95.
3. Ивченко Г. И., Хонов С. А. Статистическое оценивание состава конечной совокупности // Дискретная математика (1996) 8, №1, 1-40.

О распределении числа вершин заданной степени конфигурационного графа

И. А. Чеплюкова

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: chia@krc.karelia.ru

В последнее время уделяется большое внимание изучению структуры и свойств случайных графов, предназначенных для моделирования сложных сетей коммуникаций (см., например, [1, 3]). Одна из наиболее известных моделей – конфигурационная модель с независимыми одинаково распределенными степенями вершин. Построение этой модели состоит из двух этапов. На первом этапе построения конфигурационного графа, состоящего из N основных и одной фиктивной вершины, для каждой из N основных вершин определяется ее степень в соответствии с некоторым распределением вероятностей. Для удобства изложения процесса построения такой модели часто используется понятие полуребра, введенное в [4]. Число полуребер выходящих из вершины равно степени данной вершины. Предполагается, что все вершины и полуребра различны. На втором этапе построения происходит последовательное образование ребер: на каждом шаге два полуребра выбираются равновероятно и, соединившись, образуют ребро. Дополнительная вершина носит вспомогательный характер, ее степень равна 1, если сумма всех полуребер основных вершин является нечетным числом, в противном случае степень равна 0. Одним из основных свойств большого числа реальных сетей является то, что число вершин со степенью k пропорционально $k^{-\tau}$ при $k \rightarrow \infty$, где $\tau > 0$ (см., например, [2]).

В данной работе рассматривается конфигурационная модель, в которой степени вершин с номерами от 1 до N являются независимыми одинаково распределенными случайными величинами $\eta_1, \dots, \eta_N$ с распределением

$$p_k = \mathbf{P}\{\eta_i = k\} = \frac{h(k)}{k^\tau \Sigma(1, \tau)}, \quad (1)$$

$$i = 1, \dots, N, \quad k = 1, 2, \dots,$$

где $h(x)$ интегрируемая на любом конечном интервале медленно меняющаяся функция, $\tau > 1$ и

$$\Sigma(x, y) = \sum_{k=1}^{\infty} x^k \frac{h(k)}{k^y}. \quad (2)$$

Далее мы будем рассматривать множество конфигурационных графов при условии, что сумма степеней вершин $\eta_1 + \dots + \eta_N = n$. Для таких графов получены предельные распределения числа вершин заданной степени при $n, N \rightarrow \infty$ так, что $n/N \nearrow \Sigma(1, \tau - 1)/\Sigma(1, \tau)$.

Введем независимые одинаково распределенные случайные величины $\xi_1, \dots, \xi_N$, распределение которых имеет вид

$$p_r(\lambda) = \mathbf{P}\{\xi_i = k\} = \frac{\lambda^k p_k \Sigma(1, \tau)}{\Sigma(\lambda, \tau)}, \quad (3)$$

$$i = 1, \dots, N, \quad k = 1, 2, \dots,$$

где λ , $0 < \lambda < 1$, - параметр распределения. Из (1)-(3) несложно получить, что

$$m = \mathbf{E}\xi_1 = \frac{\Sigma(\lambda, \tau - 1)}{\Sigma(\lambda, \tau)}, \quad (4)$$

$$\sigma^2 = \mathbf{D}\xi_1 = \frac{\Sigma(\lambda, \tau - 2)}{\Sigma(\lambda, \tau)} - m^2.$$

Пусть параметр распределения (3) выбран так, что выполнено равенство

$$\frac{\Sigma(\lambda, \tau - 1)}{\Sigma(\lambda, \tau)} = n/N. \quad (5)$$

Кроме того введем следующие условия: пусть существует $\epsilon > 0$ такое что,

$$(A1) \quad \tau > 4;$$

$$(A2) \quad 3 < \tau \leq 4 \quad (1 - \lambda)^{\tau-4-\epsilon}/\sqrt{N} \rightarrow 0;$$

$$(A3) \quad 5/2 < \tau \leq 3 \quad N(1 - \lambda)^{11-3\tau+\epsilon} \geq C_1 > 0;$$

$$(A4) \quad \tau = 5/2, \quad N(-\ln(1 - \lambda))^2(1 - \lambda)^{7/2+\epsilon} \geq C_2 > 0;$$

$$(A5) \quad 1 < \tau < 5/2, \quad N(1 - \lambda)^{6-\tau+\epsilon} \geq C_3 > 0.$$

Обозначим μ_r число вершин степени r . Справедливо следующее утверждение.

Теорема 1. Пусть $n, N \rightarrow \infty$ так, что $n/N \nearrow \Sigma(1, \tau-1)/\Sigma(1, \tau)$, r фиксированное натуральное число и для параметров n, N, τ выполнены условия (A1)-(A5). Тогда равномерно относительно k таких, что $u_r = (k - Np_r(\lambda))/(\sigma_{rr}\sqrt{N})$ лежит в любом фиксированном конечном интервале

$$\mathbf{P}\{\mu_r = k\} = \frac{1}{\sigma_{rr}\sqrt{2\pi N}} \exp\left\{-\frac{u_r^2}{2}\right\} (1 + o(1)),$$

где

$$\sigma_{rr}^2 = p_r(\lambda) \left(1 - p_r(\lambda) - \frac{(m-r)^2}{\sigma^2} p_r(\lambda)\right).$$

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант 16-01-00005.

Список литературы

- [1] *Faloutsos M., Faloutsos P., Faloutsos Ch.* On power-law relationships of the internet topology // Computer Communications. 1999. Rev. 29. P. 251–262.
- [2] *Hofstad R.* Random graphs and complex networks. 2011. 386 p.
- [3] *Newman M. E. Y., Strogatz S. H., Watts D. Y.* Random graphs with arbitrary degree distribution and their applications // Physical Review E. 2001. 64. 026118. doi:10.1103/PhysRevE.64.026118.
- [4] *Reittu H., Norros I.* On the power-law random graph model of massive data networks // Performance Evaluation. 2004. Vol. 55. P. 3–23. doi:10.1016/S0166-53/6(3)00097-x.

Цена анархии в задаче максимизации минимальной задержки машин в системе обслуживания

Ю. В. Чиркова

ИПМИ КарНЦ РАН, Петрозаводск, Россия

E-mail: julia@krc.karelia.ru

В работе исследуется игра максимизации минимальной задержки системы обслуживания. Игроки распределяют свои задачи различного объема между машинами, различающимися скоростями обслуживания. Каждый игрок стремится минимизировать время обслуживания своей задачи на выбранной им машине. Выигрышем системы является минимальная среди всех машин задержка. Оптимальным для системы распределением задач по машинам является такое, где максимизируется наименьшая среди всех машин задержка.

Рассмотрим систему $S = S(N, v)$, состоящую из N обслуживающих машин со скоростями $v_1 = 1 \leq \dots \leq v_N = s$. Система используется множеством игроков $U = U(n, w)$, где каждый из n игроков выбирает машину для обслуживания своей задачи. Объем задачи игрока j равен w_j , $j = 1, \dots, n$. Суммарный объем всех задач обозначим как $W = \sum_{j=1}^n w_j$. Время выполнения задачи объемом w на свободной машине i со скоростью v_i равно w/v_i .

Рассмотрим следующую игру $\Gamma = \langle S(N, v), U(n, w), \lambda \rangle$ в чистых стратегиях. Полагаем, что каждый игрок может выбирать любую из машин. Стратегией игрока j является номер машины l_j , которую он выбирает для выполнения своей задачи. Тогда профиль стратегий в игре Γ – это вектор $L = (l_1, \dots, l_n)$. Загрузку машины i , то есть суммарный объем задач на ней, обозначим как $\delta_i(L) = \sum_{j=1, \dots, n: l_j=i} w_j$.

Задержка машины i обозначается как

$$\lambda_i(L) = \sum_{j=1, \dots, n: l_j=i} w_j / v_i = \frac{\delta_i(L)}{v_i},$$

заметим, что она одинакова для всех игроков, выбравших данную машину.

Предполагаем, что целью системы является минимизация простоя наименее занятой машины, то есть максимизация времени ее работы или задержки на ней. Выигрыш системы определяется как минимальная среди всех машин задержка

$$SC(L) = \min_{i=1,\dots,N} \lambda_i(L).$$

Обозначим

$$OPT = OPT(S, U) = \max_{L \text{ профиль в } \Gamma(S, U, \lambda)} SC(L)$$

оптимальный выигрыш, или выигрыш системы в оптимальном случае, где максимум находится среди всех возможных профилей стратегий в игре $\Gamma(S, U, \lambda)$.

Профиль стратегий L , где ни одному игроку не выгодно единолично менять выбранную в L машину на другую для выполнения своей задачи, называется чистым равновесием по Нэшу. Для того, чтобы дать формальное определение, обозначим $L(j \rightarrow i) = (l_1, \dots, l_{j-1}, i, l_{j+1}, \dots, l_n)$ профиль, получаемый из профиля L , если игрок j меняет выбранную им в L машину l_j на некоторую машину i , а все остальные игроки сохраняют свои стратегии неизменными.

Профиль стратегий L называется чистым равновесием по Нэшу тогда и только тогда, если каждый игрок выбрал машину с минимальной задержкой, то есть для каждого игрока $j = 1, \dots, n$ выполняется $\lambda_{l_j}(L) \leq \lambda_i(L(j \rightarrow i))$ для всех машин $i = 1, \dots, N$. Ценой анархии для системы S называется максимум отношения оптимального выигрыша к выигрышу системы в наихудшем равновесии:

$$PoA(S) = \max_U \frac{OPT(S, U)}{\min_{L - \text{равновесие по Нэшу в } \Gamma(S, U, \lambda)} SC(L)}.$$

В данной работе получена нижняя оценка цены анархии для системы с числом машин от трех, также найдено точное значение оценки цены анархии для трех машин со скоростями $1 \leq r \leq s < 2$:

$$\begin{cases} \frac{2+s}{(1+r)^{(2-s)}} & \text{для } rs \leq 2, \\ \frac{2}{r(2-s)} & \text{для } rs > 2. \end{cases}$$

Также для трех машин доказано, что при добавлении в систему новой машины цена анархии не уменьшается, либо растет. Для случая N

машин предлагается метод нахождения точного значения цены анархии путем решения ряда задач линейного программирования. Данный метод описан на примере трех машин и реализован в виде программы, позволяющей получать графики цены анархии как функции от скорости самой быстрой из трех машин и визуально сравнивать их с графиками оценок.

Работа поддержана грантом РГНФ №15-02-00352_а.

Вероятностные методы в задачах о раскрасках гиперграфов и теории Рамсея

Д. А. Шабанов

Московский государственный университет имени М. В. Ломоносова
Москва, Россия

Московский физико-технический институт, Москва, Россия

Email: shabanov@mech.math.msu.su, dmitry.shabanov@phystech.edu

Аннотация

В 1961 П. Эрдемем и А. Хайналом была поставлена задача о нахождении минимального числа ребер в однородном гиперграфе с хроматическим числом больше двух. Несмотря на простоту постановки, поставленный ими вопрос оказался весьма трудным. На сегодняшний день мы далеки даже от асимптотического ответа. Однако из проблемы Эрдеша–Хайнала выросло целое направление комбинаторики, связанное с раскрасками гиперграфов, в рамках которого были разработаны сильные вероятностные инструменты, ряд из которых стали широко применяться в остальной дискретной математике.

В докладе будет дан обзор последних достижений в экстремальных задачах типа Эрдеша–Хайнала о раскрасках гиперграфов, а также о близких к ним классических проблемах теории Рамсея. Мы расскажем и о недавних продвижениях в развитии вероятностных методов, с помощью которых были получены указанные результаты.

Работа выполнена при частичной финансовой поддержке гранта Президента РФ № МД-5650.2016.1.

Н а у ч н о е и з д а н и е

IX МЕЖДУНАРОДНАЯ
ПЕТРОЗАВОДСКАЯ КОНФЕРЕНЦИЯ
ВЕРОЯТНОСТНЫЕ МЕТОДЫ
В ДИСКРЕТНОЙ МАТЕМАТИКЕ
30 мая — 3 июня 2016 г.
Петрозаводск, Россия

РАСШИРЕННЫЕ ТЕЗИСЫ

*Печатается по решению Ученого совета
Института прикладных математических исследований
Карельского научного центра РАН*

Издается в авторской редакции
Ответственный за выпуск А. Б. Соболева
Оригинал-макет Е. Н. Спектор

Подписано в печать 16.05.2016.
Формат 60х84 1/16. Бумага офсетная.
Уч.-изд. л. 5. Тираж 70 экз. Изд. № 84

Институт прикладных математических исследований
Карельского научного центра Российской академии наук

Отпечатано в типографии Издательства ПетрГУ
185910, г. Петрозаводск, пр. Ленина, 33

ISBN: 978-5-8021-2950-0



9 785802 129500