

ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2016

№ 4 (37)

Зарегистрирован в Федеральной службе по надзору в сфере
массовых коммуникаций, связи и охраны культурного наследия
(свидетельство о регистрации ПИ № ФС 77-29497 от 27 сентября 2007 г.)

Подписной индекс в объединённом каталоге «Пресса России» 44031

Журнал включен в «Перечень рецензируемых научных изданий, не входящих
в международные реферативные базы данных и системы цитирования,
в которых должны быть опубликованы основные научные результаты диссертаций
на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук»,
Высшей аттестационной комиссии

Founder – Tomsk State University

EDITORIAL BOARD

Alexander Gortsev – Editor-in-Chief, Doctor of Sc., Prof., Head of the Operation Research Department, Dean of the Faculty of Applied Mathematics and Cybernetics Tomsk State University. Tel: +72822529599

Valery Smagin – Deputy Editor-in-Chief, Doctor of Sc., Prof. of the Applied Mathematics Department Tomsk State University. Tel: +72822529599

Lyudmila Nezhelskaya – Executive Editor, Cand. of Sc., Docent of the Operation Research Department.
E-mail: vestnik_uvti@mail.tsu.ru

Sergey Vorobeychikov – Doctor of Sc., Prof. of the Higher Mathematics and Mathematical Modeling Department Tomsk State University

Vladimir Vishnevsky – Doctor of Sc., Prof. Head of the laboratory Institute of Control Sciences of Russian Academy of Sciences (Moscow, Russia).

Gennady Koshkin – Doctor of Sc., Prof. of the Theoretical Cybernetics Department Tomsk State University

Yury Kostyuk – Doctor of Sc., Prof. of the Theoretical Informatics Department Tomsk State University

Anjela Matrosova – Doctor of Sc., Prof., Head of the Programming Department Tomsk State University

Anatoly Nazarov – Doctor of Sc., Prof., Head of the Probability Theory and Mathematical Statistics Department Tomsk State University

Konstantin Samouylov – Doctor of Sc., Prof., Head of the Applied Probability and Informatics Department RUDN University (Moscow, Russia)

Eugene Semekin – Doctor of Sc., Prof. System Analysis and Operations Research Department Reshetnev Siberian State Aerospace University (Krasnoyarsk, Russia)

Sergey Sushchenko – Doctor of Sc., Prof., Head of the Applied of Information Department, Dean of the Faculty of Informatics Tomsk State University

Mais Farkhadov – Doctor of Sc., Head of the laboratory Institute of Control Sciences of Russian Academy of Sciences (Moscow, Russia).

Gurami Tsitsiashvili – Doctor of Sc., Prof., Chief researcher Institute for Applied Mathematics Far Eastern Branch of RAS, Prof. Far Eastern Federal University (Vladivostok, Russia)

Editorial address:

Faculty of applied mathematics and cybernetics
National Research Tomsk State University
36 Lenina Avenue, Tomsk, 634050
Telephone / fax: +72822529599
E-mail: vestnik_uvti@mail.tsu.ru

EDITORIAL COUNCIL

Ana Rosa Cavalli
PhD, Prof.

University VII
Paris, France

Vladimir Dombrovskii
Doctor of Sc., Prof.
Tomsk State University
Russia

Alexander Dudin
Doctor of Sc., Prof.
Belarusian State University
Minsk, Republic Belorussia

Enco Orzinger
PhD, Prof.
University of Rome
Italy

Paolo Prinetto
Prof.
Politechnic Institute
Torino, Italy

Gilbert Saporta
PhD, Prof.

Pierre and Marie Curie
University, Paris, France

Raimund Ubar
Doctor of Sc., Prof.
University of Technology
Tallinn, Estonia

Reindert Nobel
Doctor of Sc., Associate Prof.
Vrije University of Amsterdam Netherlands

Nina Yevtushenko
Doctor of Sc., Prof.
Tomsk State University
Russia

Yervant Zorian
PhD, Fellow & Chief Architect, Synopsys, Mountain View, CA, USA

JOURNAL INFO

Tomsk State University Journal of Control and Computer Science is an independent peer-reviewed research journal that welcomes submissions from across the world.

Tomsk State University Journal of Control and Computer Science is issued four times per year, and can be subscribed to in the Russian Press Joint Catalogue (Subscription Index 44031)

The publication in the journal is free of charge and may be in Russian or in English.

The topics of the journal are the following:

- control of dynamical systems,
- mathematical modeling,
- data processing,
- informatics and programming,
- discrete function and automation,
- designing and diagnostics of computer systems.

Rules of registration articles are given in a site:

<http://journals.tsu.ru/informatics/>

ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

Учредитель – Томский государственный университет

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Горцев Александр Михайлович – гл. редактор, проф., д-р техн. наук, зав. кафедрой исследования операций, декан факультета прикладной математики и кибернетики ТГУ. Тел. +72822529599

Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, проф. кафедры прикладной математики ТГУ. Тел. +72822529599

Нежельская Людмила Алексеевна – ответственный секретарь, доц., канд. техн. наук, доц. кафедры исследования операций ТГУ.
E-mail: vestnik_uvti@mail.tsu.ru

Воробейчиков Сергей Эрикович – проф., д-р физ.-мат. наук, проф. кафедры высшей математики и математического моделирования ТГУ

Вишневский Владимир Миронович – проф., д-р техн. наук, зав. лабораторией Института проблем управления РАН (г. Москва).

Кошкин Геннадий Михайлович – проф., д-р физ.-мат. наук, проф. кафедры теоретической кибернетики ТГУ

Костюк Юрий Леонидович – проф., д-р техн. наук, проф. кафедры теоретической информатики ТГУ

Матросова Анжела Юрьевна – проф., д-р техн. наук, зав. кафедрой программирования ТГУ

Назаров Анатолий Андреевич – проф., д-р техн. наук, зав. кафедрой теории вероятностей и математической статистики ТГУ

Самуйлов Константин Евгеньевич – проф., д-р техн. наук, зав. кафедрой прикладной информатики и теории вероятностей РУДН (г. Москва)

Семенкин Евгений Станиславович – проф., д.т.н., проф. каф. системного анализа и исследования операций, СГАУ им. акад. М.Ф. Решетнева (г. Красноярск)

Сущенко Сергей Петрович – проф., д-р техн. наук, зав. кафедрой прикладной информатики, декан факультета информатики ТГУ

Фархадов Маис Паша Оглы – д-р техн. наук, зав. лаборатории Института проблем управления РАН (г. Москва)

Цициашвили Гурами Шалвович – проф., д-р физ.-мат. наук, гл. науч. сотр. Института прикладной математики ДВО РАН, проф. ДВФУ (г. Владивосток)

Адрес редакции и издателя: 634050, Томск, пр. Ленина, 36

Национальный исследовательский Томский государственный университет, факультет прикладной математики и кибернетики

Телефон / факс: +72822529599

E-mail: vestnik_uvti@mail.tsu.ru

РЕДАКЦИОННЫЙ СОВЕТ

Ана Роза Ковали
д-р философии, проф.

Университет VII,
Париж, Франция

Владимир Домбровский
д-р техн. наук, проф.
ТГУ, Томск, Россия

Александр Дудин
д-р физ.-мат. наук, проф.
БГУ, Минск,
Республика Беларусь

Енцо Орзингер
д-р философии, проф.
Римский университет,
Италия

Паоло Принетто
проф.
Политехнический институт,
Турин, Италия

Жильберт Сапорта

д-р философии, проф.
Университет им. Пьера и Марии,
Кюри, Париж, Франция

Раймонд Убар
проф.
Технологический университет,
Таллинн, Эстония

Рейндерт Нобель
д-р, доцент
Свободный университет,
Амстердам, Нидерланды

Нина Евтушенко

д-р техн. наук, проф.
ТГУ, Томск, Россия

Ервант Зориан
д-р философии, гл. науч. сотр.
фирмы «Синописис», США

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке

Статьи публикуются на русском и английском языках.

Тематика публикаций журнала:

- управление динамическими системами,
- математическое моделирование,
- обработка информации,
- информатика и программирование,
- дискретные функции и автоматы,
- проектирование и диагностика вычислительных систем.

Журнал входит в систему Российского Индекса Научного Цитирования (РИНЦ).

Правила оформления статей приведены на сайте:

<http://journals.tsu.ru/informatics/>

ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

СОДЕРЖАНИЕ

CONTENTS

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Богданов А.Л., Богданова Ю.В. Адаптивный детектор моментов мгновенной смены сцен в видеопотоке и метод его обучения на основе признаков содержимого видеопотока: тёмный / светлый, спокойный / динамичный	4
Ефанов Д.В. К вопросу синтеза генераторов модифицированных кодов с суммированием взвешенных информационных разрядов с последовательностью весовых коэффициентов, образующей натуральный ряд чисел	13
Игнатьев Н.А. Индексирование объектов по индивидуальным наборам информативных признаков	27
Лохвицкий В.А., Уланов А.В. Численный анализ системы массового обслуживания с гиперэкспоненциальным «охлаждением»	36
Медведев Г.А., Павлив Д.А. О квадратичных моделях доходности в риск-нейтральной среде	44
Назаров А.А., Фёдорова Е.А. Модификация метода асимптотического анализа в условии большой загрузки на примере RQ-системы M M 1	57
Ниссенбаум О.В., Харченко А.М. Адаптивная оценка весового коэффициента в алгоритме динамической кластеризации данных	65
Шуленин В.П. Асимптотические свойства модифицированных средних разностей Джини	73
Цициашвили Г.Ш. Пуассоновские потоки в системе с повторным обслуживанием	83

Bogdanov A.L., Bogdanova Yu.V. An adaptive video-stream's scene boundaries detector and method of its learning based on the video stream's content characteristics: dark/light, slow/quick	4
Efanov D.V. On the problem of modified codes with summation of weighted data bits with natural numbers sequence of weight indexes generators synthesis	13
Ignat'ev N.A. Indexing of objects on the individual sets of informing features	27
Lokhvitsky V.A., Ulanov A.V. The numerical analyses of queuing system with hyperexponential distribution of cooling time	36
Medvedev G.A., Pavliv D.A. On quadratic models of yield in risk-neutral world	44
Nazarov A.A., Fedorova E.A. Modification of the asymptotic analysis method under heavy load condition on the example of the study of the retrieval queueing system M M 1	57
Nissenbaum O.V., Kharchenko A.M. Adaptive Estimation of Weight Coefficient in a Time-weighted Incremental EM-algorithm for Data Streams	65
Shulenin V.P. The asymptotic properties of the modified estimators of Gini's mean differences	73
Tsitsiashvili G.Sh. Poisson flows in systems with retrieval queues	83

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

DISCRETE FUNCTION AND AUTOMATONS

Бурдонов И.Б., Косачев А.С. Обобщенная модель системы автоматов	89
Сведения об авторах	98

Burdonov I.B., Kossatchev A.C. Generalized model of automata system	89
Brief information about the authors	98

ОБРАБОТКА ИНФОРМАЦИИ

УДК 004.021, 004.023, 004.932.2

DOI: 10.17223/19988605/37/1

А.Л. Богданов, Ю.В. Богданова

АДАПТИВНЫЙ ДЕТЕКТОР МОМЕНТОВ МГНОВЕННОЙ СМЕНЫ СЦЕН
В ВИДЕОПОТОКЕ И МЕТОД ЕГО ОБУЧЕНИЯ НА ОСНОВЕ ПРИЗНАКОВ
СОДЕРЖИМОГО ВИДЕОПОТОКА: ТЁМНЫЙ / СВЕТЛЫЙ,
СПОКОЙНЫЙ / ДИНАМИЧНЫЙ

Предложен детектор (алгоритм и метод обучения) моментов мгновенной смены сцен в видеопотоке на основе признаков содержимого: тёмный / светлый и спокойный / динамичный. Приведены результаты эксперимента на реальных данных, подтверждающие работоспособность предложенного подхода.

Ключевые слова: видеосегментация; обнаружение границ видеосцен; обнаружение момента смены сцены.

Достижения в области сжатия потокового видео и рост пропускной способности каналов сети Интернет привели к широкому распространению интернет-сервисов, связанных с той или иной формой обработки видеоданных. Рост объёмов накапливаемых массивов видеоданных требует эффективных инструментов их индексации для последующей разработки различных систем контекстного поиска и других сервисов. Разбиение видеопотока на *сцены* (*shots*, *scenes*) – неразрывные последовательности кадров, снятые одной камерой и поддающиеся интерпретации, – является одной из первых процедур его обработки. В дальнейшем извлечённые сцены могут использоваться для индексации, кластеризации, распознавания объектов и людей и т.д.

Существует несколько типов процесса смены сцены: *резкая*, или *мгновенная*, смена сцены (*abrupt*, или *cut*) является результатом последовательного соединения двух различных сцен; *плавная* смена сцены (*gradual transition*) продолжается несколько кадров и является результатом применения специальных видеоэффектов, таких как затемнение (*fade-in*), просветление (*fade-out*), растворение (*dissolve*) и т.д. В данной работе рассматривается задача обнаружения моментов мгновенной смены сцен, которая формулируется следующим образом: требуется построить детектор (классификатор), который каждый кадр видео-потока относит либо к классу 1 (первый кадр сцены), либо к классу 0 (внутренний кадр сцены).

В работе [1] сделан большой обзор подходов к решению этой задачи. В основе большинства подходов [2–7] лежит применение специальных функций, которые принято называть *метриками* и которые количественно оценивают схожесть кадров. Если у близлежащих кадров значение метрики мало, т.е. кадры похожи, то, скорее всего, они относятся к одной сцене; если же значение метрики велико, то кадры существенно отличаются друг от друга и, возможно, принадлежат разным сценам. Превышение значения метрики заданного порогового значения T_m интерпретируется как сигнал о начале новой сцены. Таким образом, задача построения детектора сводится к выбору метрики и порогового значения.

Естественно, в работе детектора возможны ошибки: момент смены сцены может детектироваться тогда, когда его нет (*ложное срабатывание*, *false positive*); момент смены сцены не детектируется (*пропуск*, *false negative*). Пусть по результатам обработки видеопотока детектором стали известны следующие величины: TN – количество верно классифицированных кадров класса 0, FP – количество ложных срабатываний, FN – количество пропусков и TP – количество верно классифицированных кадров класса 1 и пусть P и N – количество кадров видеопотока, принадлежащих классу 1 и классу 0 соответственно, тогда для оценки качества работы детектора можно использовать следующие показатели:

$$Se = \frac{TP}{P}, \quad Sp = \frac{TN}{N}, \quad Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}, \quad (1)$$

которые называются соответственно *чувствительность*, *специфичность*, *точность*, *полнота* и *F-мера*. Значения всех показателей лежат в интервале [0; 1]. Чем точнее работает детектор, тем ближе к единице значения этих показателей.

В работе [8] предложен подход, позволяющий повысить качество работы детектора путём разбиения кадров видеопотока на непересекающиеся прямоугольные блоки. Суть подхода состоит в следующем. Для каждой пары одинаково расположенных блоков близлежащих кадров вычисляется значение метрики. Если её значение превышает пороговое значение T_m , то считается, что блок сигнализирует о наличии смены сцены; если доля таких блоков превышает заданное пороговое значение T_b , то детектор принимает решение об обнаружении момента смены сцены.

На качество работы детектора влияют выбор метрики и количество блоков, на которые разбивается кадр, выбор пороговых значений T_b и T_m и характер содержимого видео-потока. Причём последнее обстоятельство существенно ограничивает возможности построения полностью автоматизированных систем разбиения видеопотоков на сцены. Устранению этого ограничения посвящена данная работа.

1. Метрики и эвристики, используемые при построении детектора

Прежде чем описывать методологию построения детектора, необходимо отметить, что её разработка велась в компании TomskLabs (<http://www.tomsklabs.com>) в рамках работ по созданию системы видеоклассификации. На начальном этапе были изучены и опробованы некоторые известные методы решения данной задачи; подход, предложенный в работе [8], был признан удачным и лёг в основу разрабатываемого детектора; также были сделаны следующие выводы:

1. Увеличение количества блоков, на которые следует разбивать кадры, с 4 (2×2) до 100 (10×10) приводит к заметному повышению качества работы детектора, дальнейшее же увеличение количества блоков неэффективно. В данной работе используется разбиение на 100 (10×10) блоков.

2. Хорошие результаты показали следующие метрики:

$$\mu_{ij}^E(n) = \frac{1}{255} |E_{ij}(n) - E_{ij}(n-1)|, \quad (2)$$

$$\mu_{ij}^S(n) = \frac{1}{255} |S_{ij}(n) - S_{ij}(n-1)|, \quad (3)$$

где $E_{ij}(n)$ и $S_{ij}(n)$ – соответственно среднее значение и среднеквадратичное отклонение яркостей пикселей (i, j) -го блока кадра с номером n (предполагается, что кадры видеопотока представлены в градациях серого цвета, преобразование выполняется простым усреднением RGB-компонент, максимально возможное значение яркости одного пикселя равно 255). Решающее правило детектора при использовании этих метрик имеет вид: если значение одной из метрик (1) или (2) превысит заданное пороговое значение T_m , то принимается решение, что (i, j) -й блок кадра с номером n сигнализирует о смене сцены.

3. Пороговые значения T_b и T_m следует выбирать из интервалов [0,2; 0,8] и [0,03; 0,07] соответственно.

4. При выборе пороговых значений T_b и T_m следует обращать внимание на такие признаки видеопотока, как *тёмный / светлый* и *спокойный / динамичный*. Если фрагмент видеопотока состоит из тёмных кадров и события в нём разворачиваются медленно, то пороговые значения T_b и T_m следует выбирать ближе к левому краю указанных интервалов, а если кадры светлые и события происходят быстро, то, наоборот, – ближе к правому краю. Во всех остальных случаях значения порогов следует выбирать из середины интервалов.

Из последнего вывода следует, что должны иметь место зависимости $T_m = f(T_b)$ и $T_b = g(L, D)$, где L и D – некоторые величины, количественно описывающие такие характеристики фрагмента видеопотока, как *тёмный / светлый* и *спокойный / динамичный* соответственно.

Для оценки величины L необходимо уметь оценивать яркость одного кадра. Для этой цели естественно использовать нормированное среднее значение яркостей пикселей кадра

$$E(n) = \frac{1}{255} \frac{1}{wh} \sum_{i=1}^w \sum_{j=1}^h p_{ij}(n),$$

где $p_{ij}(n)$ – значение яркости (i, j) -го пикселя кадра с номером n ; w и h – соответственно ширина и высота кадра в пикселях.

Для оценки величины D необходимо уметь оценивать степень произошедших изменений в двух последовательных кадрах, для этого была разработана метрика

$$\Delta(n) = \begin{cases} 0, & n \leq 0, \\ 1, & n = 1, \\ \frac{1}{2} (\Delta_h(n) + \Delta_v(n)), & n > 1, \end{cases}$$

где

$$\Delta_h(n) = \frac{1}{510} \frac{1}{n_h(n_v-1)} \sum_{i=1}^{n_h} \sum_{j=1}^{n_v-1} |(E_{i,j}(n) - E_{i,j+1}(n)) - (E_{i,j}(n-1) - E_{i,j+1}(n-1))|,$$

$$\Delta_v(n) = \frac{1}{510} \frac{1}{(n_h-1)n_v} \sum_{i=1}^{n_h-1} \sum_{j=1}^{n_v} |(E_{i,j}(n) - E_{i+1,j}(n)) - (E_{i,j}(n-1) - E_{i+1,j}(n-1))|,$$

$E_{ij}(n)$ – среднее значение яркости пикселей его (i, j) -го блока кадра с номером n ; n_h и n_v – количество блоков по горизонтали и по вертикали, на которые разбивается кадр.

Так как видеопоток, как правило, не является *однородным* (некоторые его фрагменты могут быть, например, светлыми и динамичными, другие – тёмными и спокойными), то оценку величин L и D следует производить отдельно для каждого кадра, опираясь на данные близлежащих кадров. В работе для этой цели использовалось простое усреднение метрик $E(n)$ и $\Delta(n)$ по k предыдущим и k последующим кадрам:

$$L(n, k) = \frac{1}{2k+1} \sum_{i=-k}^k E(n+i), \quad (4)$$

$$D(n, k) = \frac{1}{2k+1} \sum_{i=-k}^k \Delta(n+i). \quad (5)$$

Экспериментально установлено, что при использовании метрик (2), (3) и указанном методе оценивания величин L и D упомянутые выше зависимости f и g хорошо описываются линейной и правой полулогарифмической моделями соответственно, т.е.

$$f(T_b) = a_0 + a_1 T_b, \quad (6)$$

$$g(L, D) = b_0 + b_1 \ln(L) + b_2 \ln(D), \quad (7)$$

где a_0, a_1, b_0, b_1, b_2 – параметры, значения которых необходимо найти.

При использовании формул (6)–(9) необходимо учитывать следующее:

1. При вычислении величин $L(n, k)$ и $D(n, k)$ для первых и последних кадров потока в формулах (4) и (5) необходимо изменить пределы суммирования и соответствующим образом нормирующий множитель, чтобы избежать выхода за границы видеопотока.

2. Так как экспериментально установлено, что пороговые значения T_b и T_m должны находиться внутри интервалов $[0, 2; 0, 8]$ и $[0, 03; 0, 07]$ соответственно, то вычисляемые по формулам (6) и (7) значения необходимо скорректировать следующим образом:

$$T_b = \max(0, 2; \min(0, 8; g(L, D))), \quad (8)$$

$$T_m = \max(0, 03; \min(0, 07; f(T_b))). \quad (9)$$

3. Величины $L(n)$ и $D(n)$ могут принимать нулевые значения, если, например, в видеопотоке имеются полностью чёрные кадры или последовательность полностью идентичных кадров, тогда воспользоваться формулой (7) для расчёта порогового значения T_b нельзя. В этом случае пороговое значение T_b следует установить равным 1.

2. Алгоритм работы адаптивного детектора моментов мгновенной смены сцены

Алгоритм работы предлагаемого детектора для каждого обрабатываемого кадра видеопотока производит подбор оптимальных пороговых значений T_b и T_m , опираясь на признаки тёмный / светлый и спокойный / динамичный близлежащих кадров. Псевдокод алгоритма имеет следующий вид:

Вход: последовательность кадров видеопотока, параметр k .

Выход: размеченная последовательность кадров.

1. Объявить первый кадр видеопотока начальным кадром новой сцены.

2. Для каждого кадра с номером $n > 1$ делать:

2.1. Вычислить значения величин $L(n, k)$ и $D(n, k)$ по формулам (4), (5).

2.2. Вычислить пороговые значения T_b и T_m по формулам (6), (7).

2.3. Скорректировать пороговые значения T_b и T_m по формулам (8), (9).

2.4. Для каждого блока кадра делать:

2.4.1. Вычислить значения метрик $\mu^E(n)$ и $\mu^S(n)$ по формулам (2) и (3).

2.4.2. Если $\mu^E(n) > T_m$ или $\mu^S(n) > T_m$, то объявить блок сигнализирующим о смене сцены.

3. Если доля блоков кадра, сигнализирующих о смене сцены, превышает пороговое значение T_b , то объявить текущий кадр начальным кадром новой сцены, иначе объявить текущий кадр внутренним кадром сцены.

3. Обучение детектора

Для работы адаптивного детектора требуются значения параметров a_0, a_1, b_0, b_1, b_2 , входящих в формулы (6), (7). Для их нахождения предлагается процедура *обучения* детектора, в основе которой лежат три ключевых момента:

1. Для решения задачи требуется обучающая выборка *однородных* видеопотоков.

2. Для каждого видеопотока обучающей выборки необходимо определить степень выраженности признаков тёмный / светлый и спокойный / динамичный.

3. Для каждого видеопотока обучающей выборки необходимо определить оптимальные пороговые значения T_b и T_m .

Однородность видеопотоков выборки будем понимать в том смысле, что каждый видеопоток может быть отнесён лицом, осуществляющим обучение детектора, к одному из четырёх классов тёмный-спокойный, тёмный-динамичный, светлый-спокойный, светлый-динамичный. Добиться такого свойства от выборки можно путём «нарезки» предназначенных для включения в выборку изначально неоднородных видеопотоков на короткие однородные фрагменты.

Пусть $X_L = \{x_i | i = 1, \dots, N_L\}$ – репрезентативная выборка однородных видеопотоков. Степень выраженности признаков тёмный / светлый и спокойный / динамичный видеопотока x_i определим путём усреднения по всем кадрам видеопотока значения метрик $E(n)$ и $\Delta(n)$ соответственно, т.е.

$$L_i = \frac{1}{n_i} \sum_{n=1}^{n_i} E(n) \text{ и } D_i = \frac{1}{n_i} \sum_{n=1}^{n_i} \Delta(n), \quad (12)$$

где n_i – количество кадров в видеопотоке x_i .

Оптимальность пороговых значений T_{bi} и T_{mi} видеопотока x_i будем понимать в смысле минимума суммарной ошибки ($FN + FP$) *детектора с фиксированными порогоми*, псевдокод алгоритма работы которого имеет следующий вид:

Вход: последовательность кадров видеопотока, параметры T_b и T_m .

Выход: размеченная последовательность кадров.

1. Объявить первый кадр видеопотока начальным кадром новой сцены.

2. Для каждого кадра с номером $n > 1$ делать:

2.1. Для каждого блока кадра делать:

2.1.1. Вычислить значения метрик μ^E и μ^S по формулам (2) и (3).

2.1.2. Если $\mu^E > T_m$ или $\mu^S > T_m$, то объявить блок сигнализирующим о смене сцены.

3. Если доля блоков кадра, сигнализирующих о смене сцены, превышает пороговое значение T_b , то объявить текущий кадр начальным кадром новой сцены, иначе объявить текущий кадр внутренним кадром сцены.

Алгоритм работы детектора с фиксированными порогами отличается от алгоритма адаптивного детектора отсутствием шагов 2.1–2.3, отвечающих за динамическое обновление пороговых значений T_b и T_m , которые в данном случае являются задаваемыми извне параметрами.

Поиск оптимальных пороговых значений T_{bi} и T_{mi} для видеопотока x_i будем производить путём перебора значений в диапазонах $[0,2; 0,8]$ и $[0,03; 0,07]$ с шагами 0,1 и 0,01 соответственно и вычисления для каждой пары пороговых значений суммарной ошибки ($FN + FP$) детектора с фиксированными порогами. Выполнив перебор, получим множество, в котором каждый элемент состоит из пары (T_{bi}, T_{mi}) и соответствующего значения суммарной ошибки. Извлечём из этого множества подмножество «топ-5», состоящее из пар (T_{bi}, T_{mi}) , соответствующих пяти наименьшим значениям суммарной ошибки (таких пар может быть несколько), и выберем в качестве оптимальных пороговых значений T_{bi} и T_{mi} центральную точку этого подмножества (рис. 1).

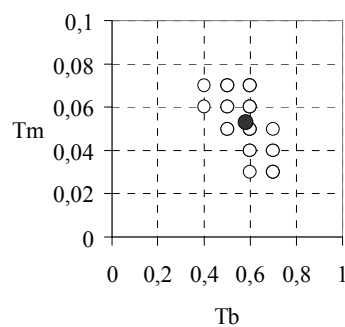


Рис. 1. Множество «топ-5» (белые точки) и его центральная точка (чёрная точка) для одного из видеопотоков обучающей выборки

В результате выполнения описанной процедуры получим набор $\{L_i, D_i, T_{bi}, T_{mi} \mid i = 1, \dots, N_L\}$, используя который найдём методом наименьших квадратов оценки параметров a_0, a_1, b_0, b_1, b_2 .

4. Эксперимент на реальных данных

Для проверки работоспособности предложенного метода построения адаптивного детектора мгновенной смены сцен была подготовлена выборка, состоящая из фрагментов популярных художественных фильмов. Источником первичных данных являлся сайт youtube.com. С него было загружено 20 фрагментов, из которых было «нарезано» 27 видеопотоков, содержащих только мгновенные смены сцен. Длина каждого видеопотока составляла 1 499 кадров, что соответствует примерно одной минуте видео (скорость воспроизведения в разных видеопотоках варьировалась от 24 до 30 кадров в секунду), общее количество сцен 1 098.

Для обучения и оценки качества работы адаптивного детектора исходная выборка «нарезанных» видеопотоков разделялась случайным образом на три равных части: обучающую, валидационную и тестовую. Обучающая выборка использовалась для оценки значений параметров a_0, a_1, b_0, b_1, b_2 . Валидационная выборка использовалась для определения оптимального в смысле минимума суммарной ошибки адаптивного детектора на валидационной выборке значения параметра k (поиск проводился в интервале $[1; 30]$ с шагом 1). Тестовая выборка, хотя и состоящая из однородных видеопотоков, но по сути представляющая собой один большой неоднородный видеопоток, использовалась для объективной оценки качества работы детектора. На данных тестовой выборки рассчитывались показатели качества $Se, Sp, Precision$ и $F1$ (1). Для полноты картины на тех же тестовых данных производилась оценка качества работы детектора с фиксированными порогами при $T_b = 0,6$ и $T_m = 0,06$ и гипотетического (наилучшего) детектора, алгоритм работы которого отличался от алгоритма детектора с фиксирован-

ными порогами тем, что для каждого видеопотока, входящего в тестовую выборку, использовались соответствующие оптимальные пороговые значения T_b и T_m . Было проведено 1 000 экспериментов. Их результаты представлены на рис. 2, 3 и в табл. 1, 2.

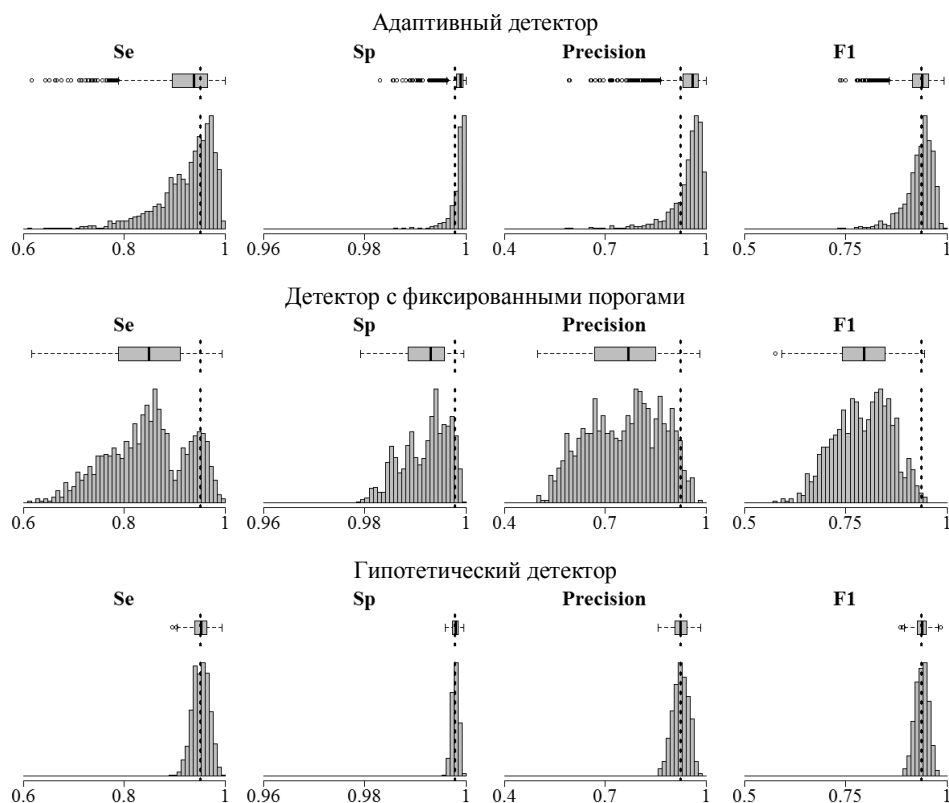


Рис. 2. Гистограммы распределения показателей качества работы различных детекторов

На рис. 2 приведены гистограммы распределения значений показателей качества работы трёх детекторов. К каждой гистограмме добавлены стандартная диаграмма размаха соответствующего показателя и вертикальная пунктирная линия, соответствующая среднему значению показателя гипотетического (наилучшего) детектора. В табл. 1 представлены основные статистические характеристики показателей качества работы детекторов: средние значения и среднеквадратичные отклонения.

Таблица 1

Основные статистические характеристики показателей качества работы детекторов

Тип детектора	Характеристика	Se	Sp	Pr	$F1$
Адаптивный	Среднее	0,9213	0,9983	0,9428	0,9295
	с.к.о.	0,0595	0,0019	0,055	0,0374
С фиксированными порогами	Среднее	0,8452	0,9918	0,7563	0,7918
	с.к.о.	0,079	0,0046	0,1096	0,0693
Гипотетический (наилучший)	Среднее	0,9512	0,9978	0,9234	0,9369
	с.к.о.	0,0166	0,0007	0,0239	0,0165

Интерес также представляет то, как соотносятся между собой показатели качества работы детекторов в каждом отдельно взятом эксперименте. На рис. 3 приведены гистограммы разностей значений соответствующих показателей качества адаптивного детектора и детектора с фиксированными порогами (вверху) и адаптивного детектора и гипотетического детектора (внизу), вычисленные по результатам каждого эксперимента. К каждой гистограмме добавлены стандартная диаграмма размаха разности показателей и вертикальная пунктирная линия, соответствующая нулевому значению; столбцы гистограмм, расположенные правее этой линии, свидетельствуют в пользу адаптивного детектора. В табл. 2

представлены основные статистические характеристики разностей значений показателей качества – их средние значения и среднеквадратичные отклонения, а также указаны доли от общего количества экспериментов, в которых эти разности положительны и превышают значение $-0,05$.

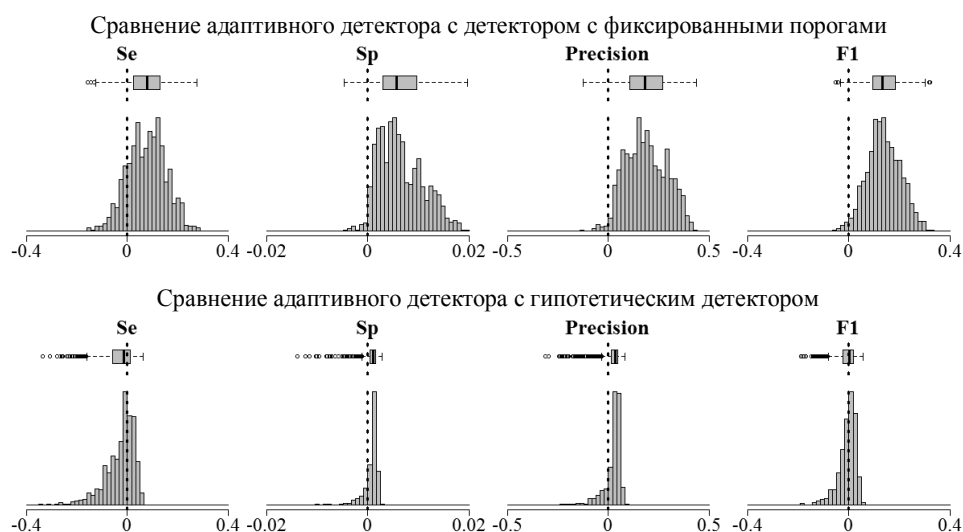


Рис. 3. Гистограммы распределения разности значений показателей качества работы различных детекторов

Таблица 2

Сравнение качества (случай адаптивного детектора с оптимальным значением параметра k)

Вариант сравнения	Характеристика	ΔSe	ΔSp	ΔPr	$\Delta F1$
Адаптивный детектор против детектора с фиксированными порогами	Среднее	0,0762	0,0065	0,1865	0,1376
	с.к.о.	0,0752	0,0044	0,1012	0,0654
	Доля > 0	83,5%	98,1%	98,5%	98,4%
	Доля $> -0,05$	95%	100%	99,6%	99,9%
Адаптивный детектор против гипотетического детектора	Среднее	-0,0299	0,0005	0,0194	-0,0075
	с.к.о.	0,0599	0,0018	0,0485	0,036
	Доля > 0	35,5%	80,4%	81,3%	50,9%
	Доля $> -0,05$	70,7%	100%	91,7%	88,9%

Проанализировав графики на рис. 2, 3 и данные табл. 1, 2, можно сделать вывод, что адаптивный детектор демонстрирует качество работы, близкое к качеству работы гипотетического (наилучшего) детектора – среднее значение его чувствительности на 3% меньше среднего значения чувствительности гипотетического детектора при сравнимой точности классификации, в то время как у детектора с фиксированными порогами при сопоставимой чувствительности точность составляет около 50%; в 35,5% от общего числа экспериментов адаптивный детектор продемонстрировал лучшую чувствительность, чем у гипотетического (наилучшего) детектора и в 70,7% уступил ему не более чем на 5% при сопоставимых значениях других показателей качества.

В описанных выше экспериментах оптимальное значение параметра k определялось на основе данных валидационной выборки. На рис. 4 представлены гистограмма распределения и диаграмма размаха оптимальных значений параметра, вертикальные пунктирные линии соответствуют среднему значению 11,85. Представляет интерес, как изменились бы показатели качества работы адаптивного детектора при фиксированном значении параметра k . Для ответа на данный вопрос были проведены две дополнительные серии экспериментов с тем же самым разбиением на обучающую, валидационную (не использовалась) и тестовую выборки при фиксированных значениях $k = 6$ и $k = 12$. В табл. 3 представлены основные статистические характеристики разностей значений показателей качества адаптивного детектора и гипотетического (наилучшего) детектора – их средние значения, среднеквадратичные отклонения и доли от общего количества экспериментов, в которых указанные разности положительны и

превышают значение $(-0,05)$. Сравнив данные табл. 2 и 3, можно сделать вывод, что и при фиксированном правильно выбранном значении параметра k качество работы адаптивного детектора близко к качеству гипотетического (наилучшего) детектора.

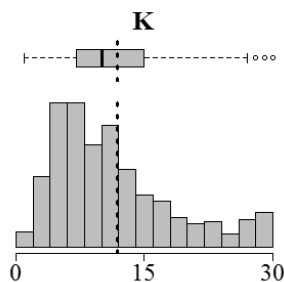


Рис. 4. Гистограмма распределения оптимального значения параметра k

Таблица 3

Сравнение качества (случай адаптивного детектора при фиксированном значении параметра k)

Вариант сравнения	Характеристика	ΔSe	ΔSp	ΔPr	$\Delta F1$
Адаптивный детектор с фиксированным значением $k = 6$ против гипотетического детектора	Среднее	$-0,0364$	$0,0006$	$0,0232$	$-0,0095$
	с.к.о.	$0,0691$	$0,0018$	$0,0496$	$0,0411$
	Доля > 0	$34,5\%$	$82,4\%$	$83,1\%$	$51,7\%$
	Доля $> -0,05$	$68,4\%$	100%	$92,6\%$	$86,9\%$
Адаптивный детектор с фиксированным значением $k = 12$ против гипотетического детектора	Среднее	$-0,0262$	$0,0005$	$0,0195$	$-0,0055$
	с.к.о.	$0,0599$	$0,0018$	$0,0485$	$0,0361$
	Доля > 0	$38,1\%$	$79,3\%$	$79,7\%$	$54,2\%$
	Доля $> -0,05$	$73,7\%$	100%	$91,6\%$	$90,1\%$

Заклучение

В данной работе предложен алгоритм работы адаптивного детектора моментов мгновенной смены сцены в видеопотоке, использующий признаки тёмный / светлый, спокойный / динамичный, и предложен метод его обучения. Эксперимент на выборке из фрагментов художественных фильмов подтвердил работоспособность предложенного подхода. Направлениями дальнейших исследований можно назвать повышение чувствительности адаптивного детектора и изучение его свойств на выборках с более широким спектром содержимого (спортивные передачи, новости...).

ЛИТЕРАТУРА

1. Koprinska I., Carrato S. Temporal video segmentation: A survey // Signal Processing: Image Communication. 2001. V. 16. P. 477–500.
2. Adhikari P., Gargote N., Digge J. Abrupt Scene Change Detection // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2008. V. 2, № 6. P. 711–716.
3. Krulikowska L., Polec J., Hirner T. Fast Algorithm of Shot Cut Detection // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2012. V. 6, № 7. P. 633–636.
4. Amiri A., Fathy M. Video shot boundary detection using QR decomposition and Gaussian transition detection // EURASIP Journal on Advances in Signal Processing. 2009. Article ID 509438. P. 1–12.
5. Hanjalic A. Shot-boundary detection: unraveled and resolved? // IEEE Transactions on Circuits and Systems for Video Technology. 2002. V. 12, № 2. P. 90–105.
6. Ford R., Robson C., Temple D. Metrics for scene change detection in digital video sequences // International Conference on Multimedia Computing and Systems IEEE. 1997. P. 610–611.
7. Shukla D., Sharma M., Mithlesh C.S. Scene Change Detection using Block Processing Method // The Journal of Applied Sciences Research. 2015. V. 2 (2). P. 103–110.
8. Xiong W., Lee J.C.-M., Ip M.C. Net comparison: a fast and effective method for classifying image sequences // Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III. 1995. V. 2420. P. 318–328.

Богданов Александр Леонидович, канд. техн. наук, доцент. E-mail: bogdanov.al@mail.ru

Томский государственный университет

Богданова Юлия Вячеславовна, канд. физ.-мат. наук. E-mail: bogdanovaju@tpu.ru

Томский политехнический университет

Поступила в редакцию 25 апреля 2016 г.

Bogdanov Alexander L. Bogdanova Yulia V. (Tomsk State University, Tomsk Polytechnic University, Russian Federation).

An adaptive video-stream's scene boundaries detector and method of its learning based on the video stream's content characteristics: dark/light, slow/quick.

Keywords: temporal video segmentation; shot boundary detection; scene change detection; cut detection.

DOI: 10.17223/19988605/37/1

The problem of scene boundaries detection are considered. A scene is defined as an unbroken meaningful sequence of frames taken from one camera. There are two basic types of scene transitions: abrupt and gradual. This research is focused on abrupt transitions only. The aim of the project is to develop a detector (classifier) which every frame of the video-stream classify either to class 1 (first frame of the scene) or to class 0 (internal frame of the scene). To evaluate quality of the detector sensitivity, specificity, precision and F-score are used:

$$Se = \frac{TP}{P}, Sp = \frac{TN}{N}, Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}.$$

The proposed detector is based on using metrics and the frame partitioning into non-overlapping blocks. The metrics are the functions that calculate measure of similarity between frames. If the metric value for two consecutive frames is small, it means the frames are similar and, probably, belong to one scene. If the metric value is large, it means the frames are dissimilar and, probably, belong to different scenes. Splitting of the frame into blocks is used to improve the quality of the detector: a block is signaling about scene change if the metric value for the corresponding blocks of two successive frames exceeds a prespecified threshold value T_m ; the detector make a decision "scene change detected" in case portion of signaling blocks is above a prespecified threshold value T_b . Type of the metrics, number of the blocks and threshold values influence on detector's quality. It is discovered that optimal threshold values depend strongly on the video-stream's content. This fact makes harder to develop a fully automatic scene boundaries detector. To overcome this problem this work is dedicated.

It was found that the optimal threshold values depend on such characteristics of the video stream content, as dark/light and slow/quick. Methods for qualitative estimation of these characteristics for local fragments of the video-stream were developed. The algorithm for adaptive scene boundaries detector, which for each frame of the video-stream use optimal thresholds values based on data of nearby frames, was proposed. A method for estimating the parameters of threshold's dependency models was proposed.

To verify the developed algorithm and method of its learning the experiment on real data was conducted. Twenty fragments of popular movies was loaded from youtube.com. Twenty seven video-streams contained abrupt scene transitions only were prepared such that every source video-stream can be classified to one of the next four classes: dark-slow, light-slow, dark-quick, light-quick. The length of each video-stream was 1499 frames. Total number of scenes was 1098. One thousand experiments was conducted. In each experiment the source video-stream sample was divided into three part: learning, validation and test samples. Learning and validation samples were used to estimate parameters of adaptive detector, test sample was to evaluate its performance. Next mean values of sensitivity, specificity, precision and F-score were gained: $Se = 0.92$, $Sp = 0.99$, $Precision = 0.94$ и $F1 = 0.93$.

In comparison to fixed threshold's values detector with $T_b = 0.6$ и $T_m = 0.06$, adaptive detector has a higher classification precision and close values of other quality's measures. In comparison to the hypothetically best detector, adaptive detector has a little bit lower value of sensitivity and close values of other quality's measures. It must be noted, that obtained results were received on the sample of movies, the quality of the adaptive detector on video-streams with other types of content is subject for another research.

REFERENCES

1. Koprinska, I. & Carrato, S. (2001) Temporal video segmentation: A survey. *Signal Processing: Image Communication*. 16. pp. 477-500. DOI: 10.1016/S0923-5965(00)00011-4
2. Adhikari, P., Gargote, N. & Digge, J. (2008) Abrupt Scene Change Detection. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 2(6). pp. 711-716.
3. Krulikovska, L., Polec, J. & Hirner, T. (2012) Fast Algorithm of Shot Cut Detection. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 6(7). pp. 711-716.
4. Amiri, A. & Fathy, M. (2009) Video shot boundary detection using QR decomposition and Gaussian transition detection. *EURASIP Journal on Advances in Signal Processing*. 2009: 509438. DOI: 10.1155/2009/509438. pp. 1-12.
5. Hanjalic, A. (2002) Shot-boundary detection: unraveled and resolved. *IEEE Transactions on Circuits and Systems for Video Technology*. 12(2). pp. 90-105. DOI: 10.1109/76.988656
6. Ford, R., Robson, C. & Temple, D. (1997) Metrics for scene change detection in digital video sequences. *International Conference on Multimedia Computing and Systems IEEE*. pp. 610-611. DOI: 10.1109/MMCS.1997.609780
7. Shukla, D., Sharma, M. & Mithlesh, C.S. (2015) Scene Change Detection using Block Processing Method. *The Journal of Applied Sciences Research*. 2(2). pp. 103-110.
8. Xiong, W., Lee, J.C.-M. & Ip, M.C. (1995) Net comparison: a fast and effective method for classifying image sequences. *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III*. 2420. pp. 318-328. DOI: 10.1117/12.205298

Д.В. Ефанов

**К ВОПРОСУ СИНТЕЗА ГЕНЕРАТОРОВ МОДИФИЦИРОВАННЫХ КОДОВ
С СУММИРОВАНИЕМ ВЗВЕШЕННЫХ ИНФОРМАЦИОННЫХ РАЗРЯДОВ
С ПОСЛЕДОВАТЕЛЬНОСТЬЮ ВЕСОВЫХ КОЭФФИЦИЕНТОВ,
ОБРАЗУЮЩЕЙ НАТУРАЛЬНЫЙ РЯД ЧИСЕЛ**

Излагаются результаты, полученные автором в области синтеза генераторов контрольных векторов модифицированных кодов Бергера. Данные коды принадлежат к классу модифицированных модульно взвешенных кодов с суммированием. Приводятся общие структуры генераторов таких кодов, а также алгоритм построения генератора, позволяющий оптимизировать его структуру, сократив количество используемых функциональных элементов. Выводится формула подсчета общего количества двухвходовых логических элементов, необходимых для технической реализации генератора. Предложенный алгоритм синтеза генераторов является универсальным и может быть использован для построения генераторов любых модифицированных модульно взвешенных кодов с суммированием.

Ключевые слова: техническая диагностика; система функционального контроля; код Бергера; модифицированный код Бергера; взвешенный код с суммированием; наименьший неотрицательный вычет; модуль; генератор; полусумматор; полный сумматор.

В задачах технической диагностики дискретных систем часто применяют разнообразные коды с суммированием [1–17]. Данные коды принадлежат к классу систематических – в кодовых словах таких кодов выделяют информационный и контрольный векторы с длинами m и k соответственно. Введем обозначение для кодов с суммированием – (m, k) -коды.

Применение (m, k) -кодов в задачах технической диагностики, например при синтезе систем с обнаружением отказов [18–20], связано с внесением избыточности в структуру «базового» объекта. При этом способ построения кода напрямую определяет характеристики избыточности получаемого технического объекта, а также особенности идентифицируемых в нем неисправностей.

Наиболее простым среди известных кодов с суммированием является классический код Бергера [21]. Данный код имеет $k = \lceil \log_2(m+1) \rceil$ разрядов в контрольных векторах (запись $\lceil \dots \rceil$ обозначает целое сверху от вычисляемого значения), принимающих значения разрядов двоичных чисел, равных сумме единичных информационных разрядов (равных весу информационного вектора). Такие особенности построения кода Бергера приводят к крайне неравномерному распределению информационных векторов между контрольными векторами, что в конечном итоге сказывается на свойствах обнаружения им ошибок. Кодами Бергера не обнаруживается любая симметричная ошибка в информационном векторе (такая ошибка не нарушает веса информационного вектора и происходит при одинаковом количестве искажаемых нулевых и единичных разрядов [22]). Вне зависимости от длины информационного вектора кодами Бергера не обнаруживается одинаковый процент ошибок четной кратностью d от общего количества ошибок данной кратностью [23–25]. Это, к примеру, 50% двукратных и 37,5% четырехкратных ошибок в информационных векторах. С неравномерностью распределения информационных векторов между контрольными связано и большое общее количество необнаруживаемых ошибок в информационных векторах кодов Бергера. Тем не менее данные коды нашли широкое применение за счет их свойства обнаружения любых монотонных ошибок в кодовых словах [26–31].

Любые модификации классических кодов связаны с перераспределением информационных векторов между контрольными векторами, что позволяет уменьшать количество необнаруживаемых в информационных векторах ошибок. В [32] введено понятие оптимального по критерию минимума общего количества необнаруживаемых ошибок (m, k) -кода с установленными значениями длин информационных и контрольных векторов и показано, что классический код Бергера далек от оптимального кода.

В [33] разработан перспективный для задач технической диагностики модифицированный код Бергера, основанный на идее взвешивания разрядов информационных векторов и последующих операциях модификации значения суммарного веса информационного вектора. Данная работа посвящена изложению способа синтеза генератора, или кодера, данного оптимального кода с суммированием.

1. Модифицированный код Бергера

Как отмечалось выше, способ построения оптимального (m,k) -кода базируется на первоначальном установлении неравноправия между разрядами в информационном векторе путем приписывания им различных весовых коэффициентов [10, 13, 21, 34–36], а значения разрядов контрольных векторов получаются следующим образом.

Алгоритм 1. Правила построения оптимального кода с суммированием:

1. Устанавливается последовательность весовых коэффициентов разрядов информационного вектора, образующая натуральный ряд чисел, начиная с младшего разряда.

2. Вычисляется значение модуля $M = 2^{\lceil \log_2(m+1) \rceil - 1}$.

3. Определяется значение суммарного веса единичных разрядов информационного вектора:

$$W = \sum_{i=1}^m w_i f_i, \quad (1)$$

где $f_i = 0$ или $f_i = 1$ – значения разрядов информационного вектора.

4. Определяется наименьший неотрицательный вычет числа W по модулю M :

$$[W]_M = W \pmod{M}. \quad (2)$$

5. Вычисляется специальный поправочный коэффициент α как функция паритета разрядов информационного вектора, занимающих позиции с четными номерами, начиная с младшего разряда:

$$\alpha = \begin{cases} f_2 \oplus f_4 \oplus \dots \oplus f_m, & \text{если } m - \text{четное;} \\ f_2 \oplus f_4 \oplus \dots \oplus f_{m-1}, & \text{если } m - \text{нечетное.} \end{cases} \quad (3)$$

6. Формируется значение результирующего веса информационного вектора:

$$V = [W]_M + \alpha M. \quad (4)$$

7. Число V представляется в двоичном виде и записывается в разряды контрольного вектора.

Получаемые по алгоритму 1 коды с суммированием обозначим как $RWS(m,k)$ -коды. Они принадлежат к классу модифицированных модульно взвешенных кодов с суммированием и имеют такое же количество разрядов в контрольных векторах, как и классические коды Бергера, однако обладают улучшенными характеристиками обнаружения ошибок в информационных векторах.

Пользуясь алгоритмом 1, приведем пример построения $RWS(4,3)$ -кода (табл. 1).

Таблица 1

Кодовые слова $RWS(4,3)$ -кода

№ п/п	Информационный вектор				W	W_M	A	V	Контрольный вектор		
	f_4	f_3	f_2	f_1					g_3	g_2	g_1
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	1	0	1	0	0	1
2	0	0	1	0	2	2	1	6	1	1	0
3	0	0	1	1	3	3	1	7	1	1	1
4	0	1	0	0	3	3	0	3	0	1	1
5	0	1	0	1	4	0	0	0	0	0	0
6	0	1	1	0	5	1	1	5	1	0	1
7	0	1	1	1	6	2	1	6	1	1	0
8	1	0	0	0	4	0	1	4	1	0	0
9	1	0	0	1	5	1	1	5	1	0	1
10	1	0	1	0	6	2	0	2	0	1	0
11	1	0	1	1	7	3	0	3	0	1	1
12	1	1	0	0	7	3	1	7	1	1	1
13	1	1	0	1	8	0	1	4	1	0	0
14	1	1	1	0	9	1	0	1	0	0	1
15	1	1	1	1	10	2	0	2	0	1	0

Данный код не обнаруживает 8 двухкратных и 8 четырехкратных ошибок в информационных векторах (для сравнения: кодом Бергера при данной длине информационного вектора не обнаруживается 48 двухкратных и 6 четырехкратных ошибок [23]). Вообще $RWS(m,k)$ -кодами идентифицируется гораздо больше ошибок малых кратностей, чем кодами Бергера. Кроме того, данный класс кодов, так же как и коды Бергера, обнаруживает любые ошибки нечетных кратностей в информационных векторах.

2. Синтез генераторов модифицированных кодов Бергера

Поскольку при построении $RWS(m,k)$ -кодов используются операции суммирования весовых коэффициентов и значений разрядов информационных векторов, в качестве элементной базы для их генераторов удобно использовать стандартные схемы сумматоров по модулю два (XOR), полусумматоров (HA) и полных сумматоров (FA) [7, 37–39]. На рис. 1 приводятся условные обозначения данных функциональных элементов. При этом в условных обозначениях полусумматора и полного сумматора на входах и выходах указаны веса разрядов двоичных чисел.

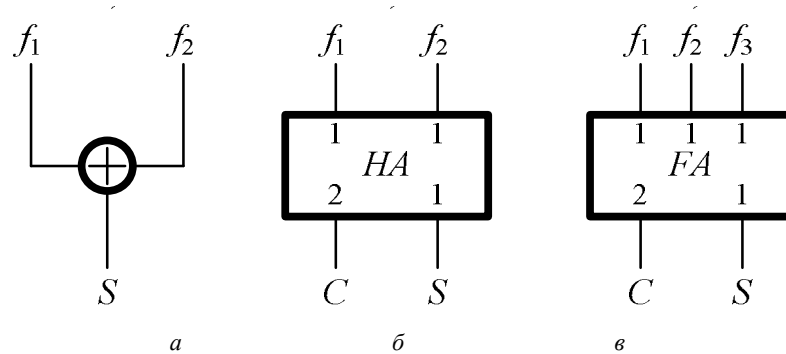


Рис. 1. Условные обозначения типовых функциональных элементов:
а – сумматора по модулю два; б – полусумматора; в – полного сумматора

Сумматор по модулю два реализует операцию сложения двух одноразрядных чисел без переноса и вычисляет функцию

$$S = f_1 \oplus f_2.$$

Схемы полусумматора и полного сумматора предназначены для сложения двух и трех двоичных чисел. Полусумматор имеет два входа и два выхода, на которых формируется двоичное число, равное суммарному количеству единиц на его входах:

$$\begin{cases} S = f_1 \oplus f_2; \\ C = f_1 f_2. \end{cases}$$

Аналогичную полусумматору функцию выполняет полный сумматор, имея при этом, однако, три входа и два выхода, описываемых следующими формулами:

$$\begin{cases} S = f_1 \oplus f_2 \oplus f_3; \\ C = f_1 f_2 \vee f_1 f_3 \vee f_2 f_3. \end{cases}$$

Согласно (1) генератор $RWS(m,k)$ -кода должен сначала подсчитать сумму весовых коэффициентов единичных разрядов информационного вектора, затем вычислить по формуле (2) наименьший неотрицательный вычет для полученного числа и прибавить к нему значение модуля при $\alpha = 1$ или сохранить неизменным при $\alpha = 0$. Таким образом, структурная схема генератора $RWS(m,k)$ -кода может иметь вид, представленный на рис. 2.

В генераторе выделяется четыре функциональных блока: B_1 предназначен для «взвешивания» разрядов; B_2 – устройство суммирования весовых коэффициентов; B_3 – схема определения наименьшего неотрицательного вычета суммарного веса единичных информационных разрядов; B_4 – формирователь старшего разряда контрольного вектора.

Схема генератора $RWS(m,k)$ -кода может быть построена и иным образом.

Преобразуем выражение (4), подставив в него выражения (1) и (2):

$$\begin{aligned}
 V &= [W]_M + \alpha M = W \pmod{M} + \alpha M = \left(\sum_{i=1}^m w_i f_i \right) \pmod{M} + \alpha M = \\
 &= (w_1 f_1 + w_2 f_2 + \dots + w_m f_m) \pmod{M} + \alpha M = \\
 &= w_1 f_1 \pmod{M} + w_2 f_2 \pmod{M} + \dots + w_m f_m \pmod{M} + \alpha M.
 \end{aligned} \tag{5}$$

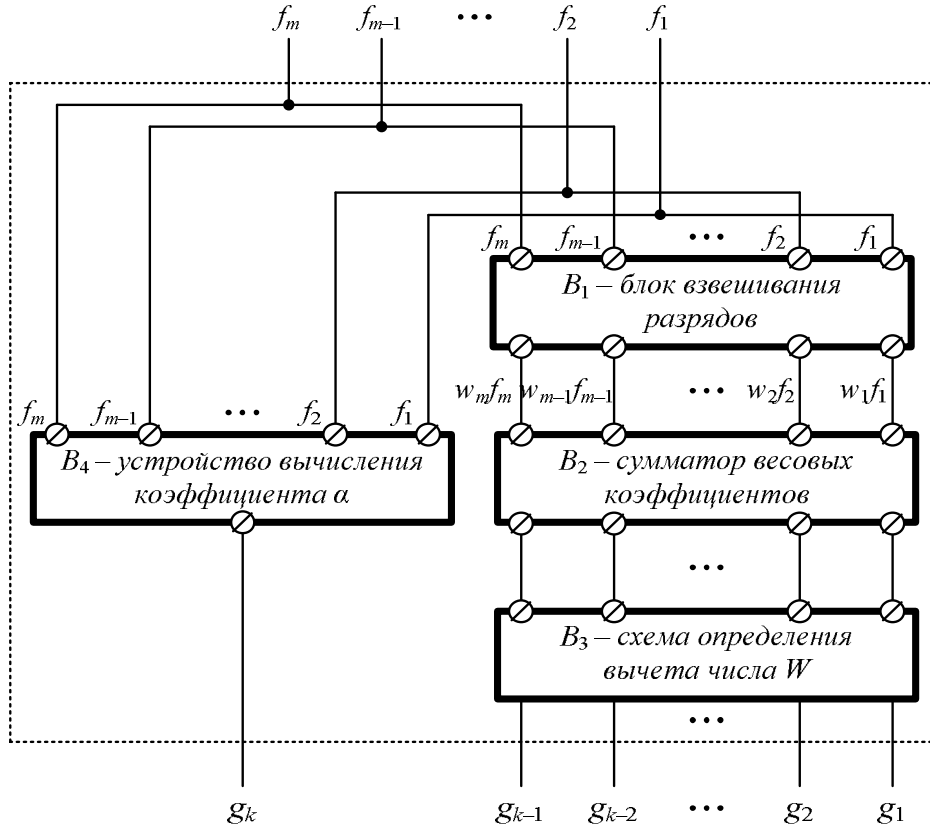


Рис. 2. Структурная схема первого типа генератора $RWS(m,k)$ -кода

Из формулы (5) следует, что перед тем, как определять суммарный вес единичных информационных разрядов, можно напрямую вычислять наименьшие неотрицательные вычеты для каждого весового коэффициента. Таким образом, структурная схема генератора $RWS(m,k)$ -кода примет вид, изображенный на рис. 3. В ней также выделяются четыре функциональных блока: B_1 предназначен для «взвешивания» разрядов; B_2 – устройство определения наименьших неотрицательных вычетов для каждого весового коэффициента; B_3 – схема суммирования наименьших неотрицательных вычетов весовых коэффициентов единичных разрядов по установленному модулю M ; B_4 – формирователь старшего разряда контрольного вектора.

Фактически блоки B_1 – B_3 могут быть построены путем каскадного соединения типовых схем сложения двоичных чисел по заданному модулю M . Такая типовая схема изображена на рис. 4. На ее входы подаются два двоичных числа $A = \langle a_1 a_2 \dots a_m \rangle$ и $B = \langle b_1 b_2 \dots b_m \rangle$, являющиеся двоичными представлениями наименьших неотрицательных вычетов по модулю M двух весовых коэффициентов. А на выходах формируется сумма чисел A и B за исключением переноса в старший разряд при переполнении – число $S = \langle s_1 s_2 \dots s_m \rangle$. В частности, на рис. 5 изображены типовые схемы сложения двоичных чисел по модулям $M = 4, 8$ и 16 , используемым при формировании $RWS(m,k)$ -кодов с длинами информационных векторов $m < 32$.

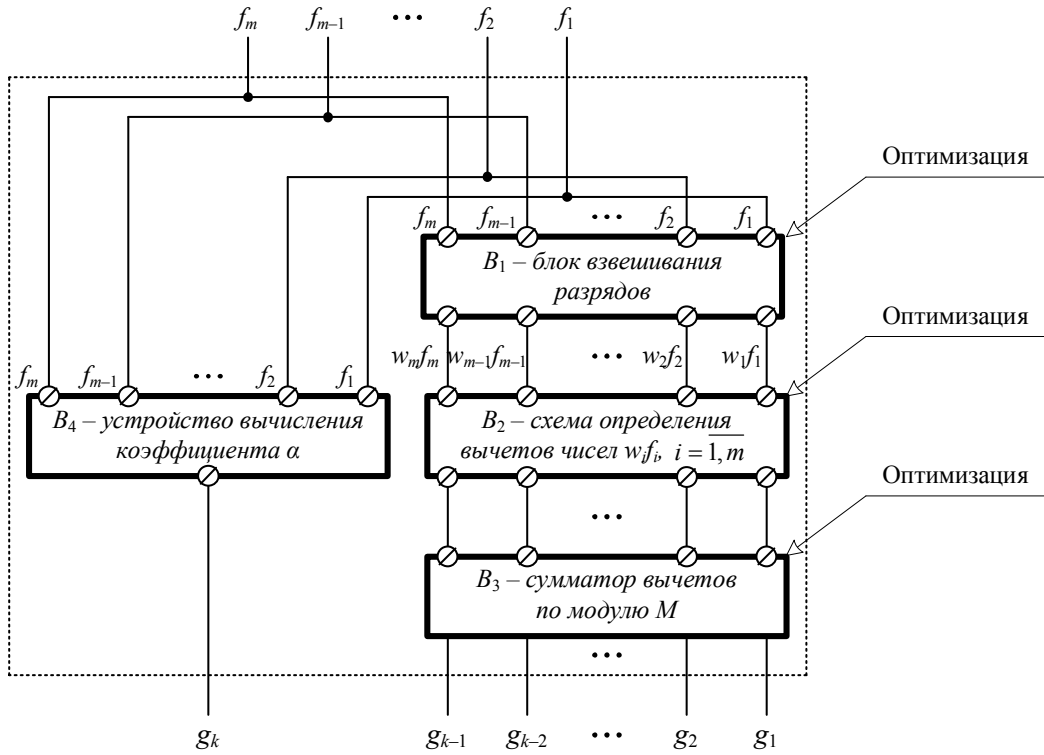


Рис. 3. Структурная схема второго типа генератора $RWS(m,k)$ -кода

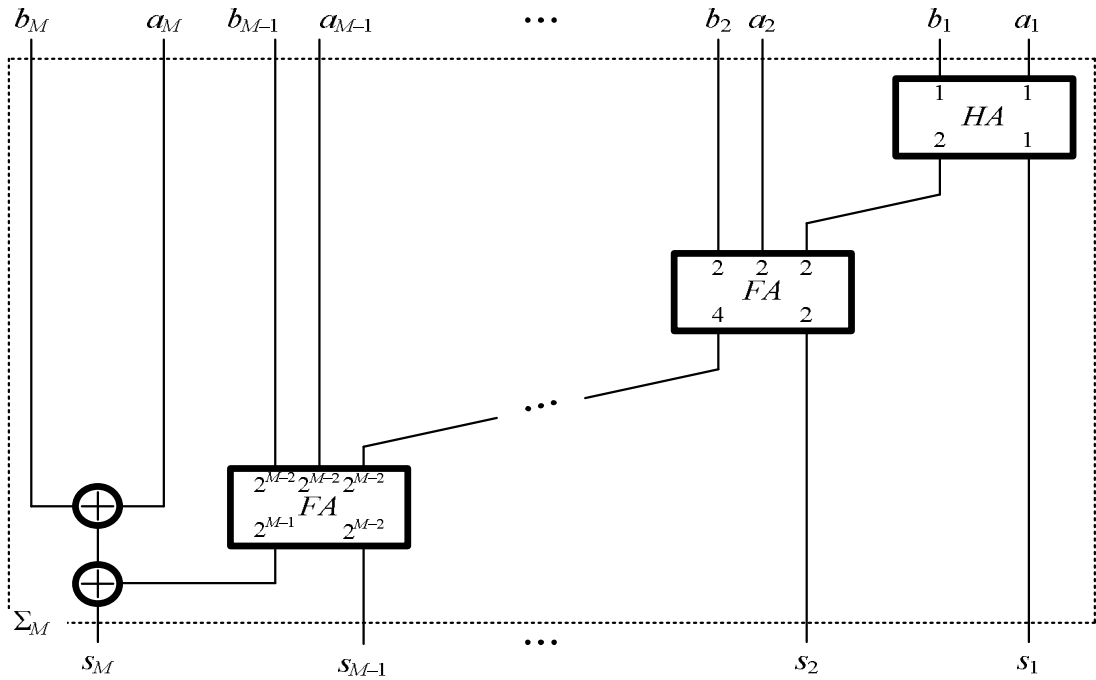


Рис. 4. Структурная схема функционального блока Σ_M

Целесообразно синтезировать генераторы по структурной схеме, приведенной на рис. 3, следующим способом.

Алгоритм 2. Правила синтеза генераторов $RWS(m,k)$ -кодов.

1. Устанавливается значение модуля M .
2. Определяется наименьший неотрицательный вычет по модулю M для каждого весового коэффициента w_i :

$$[w_i]_M = w_i \pmod{M}.$$

3. Каждое число $[w_i]_M$ представляется в двоичном виде.

4. Единичным разрядам двоичных чисел $[w_i]_M$ ставятся в соответствие значения разряда f_i .

5. Осуществляется разбиение двоичных чисел $[w_i]_M$ на пары с учетом оптимального заполнения рядов: числа $[w_i]_M$ разбиваются на $\left\lceil \frac{m}{2} \right\rceil$ групп, каждая из которых представляет собой пару двоичных чисел, полученных в п. 4 данного алгоритма, суммирование которых осуществляется без переносов.

6. Выполняется сложение чисел с учетом значений их разрядов и получение $\left\lceil \frac{m}{2} \right\rceil$ двоичных чисел.

7. С использованием функциональных блоков сложения двоичных чисел модулю M реализуется устройство суммирования $\left\lceil \frac{m}{2} \right\rceil$ двоичных чисел.

8. Реализуется блок вычисления старшего разряда контрольного вектора.

Из алгоритма 2, в частности, следует, что блоки B_1 и B_2 генератора не требуют функциональных элементов и содержат только провода, на которые подаются сигналы 0 или f_i . Более того, из п. 6 алгоритма следует, что фактически первым каскадом блока B_3 будут выступать только провода с поданными на них сигналами 0 или f_i . Таким образом, алгоритм 2 подразумевает некоторую оптимизацию структуры генератора $RWS(m,k)$ -кода.

Продemonстрируем работу алгоритма 2 на примере синтеза генератора $RWS(8,4)$ -кода.

Определим значения наименьших неотрицательных вычетов для всех весовых коэффициентов w_i и представим их в двоичном виде. После этого в каждом двоичном представлении вычета веса w_i все единичные разряды заменяем значениями соответствующего информационного разряда f_i , а все нулевые разряды оставляем заполненными значениями 0. Результат приведен в табл. 2.

Т а б л и ц а 2

Получение суммируемых двоичных чисел

Весовые коэффициенты							
w_8	w_7	w_6	w_5	w_4	w_3	w_2	w_1
8	7	6	5	4	3	2	1
$[w_i]_M$							
1000	0111	0110	0101	0100	0011	0010	0001
Двоичное число, соответствующее весу							
$f_8 000$	$0 f_7 f_7 f_7$	$0 f_6 f_6 0$	$0 f_5 0 f_5$	$0 f_4 0 0$	$0 0 f_3 f_3$	$0 0 f_2 0$	$0 0 0 f_1$

Далее полученные двоичные числа разбиваются на $\left\lceil \frac{8}{2} \right\rceil = 4$ группы. При этом числа в каждой группе должны складываться без переносов (см. рис. 6). Данные двоичные числа требуется сложить по модулю $M = 8$ с использованием функциональных блоков, изображенных на рис. 5, б. Так как получено четыре двоичных числа, то потребуются три функциональных блока.

Отдельно реализуется функция паритета $\alpha = f_2 \oplus f_4 \oplus f_6 \oplus f_8$, вычисляющая старший разряд контрольного вектора $RWS(8,4)$ -кода.

Синтезированный генератор рассматриваемого кода представлен на рис. 7. На каждой его линии указаны логические сигналы 0 и 1, соответствующие вычислению значений разрядов контрольного вектора при поступлении на входы генератора информационного вектора $\langle 10100110 \rangle$.

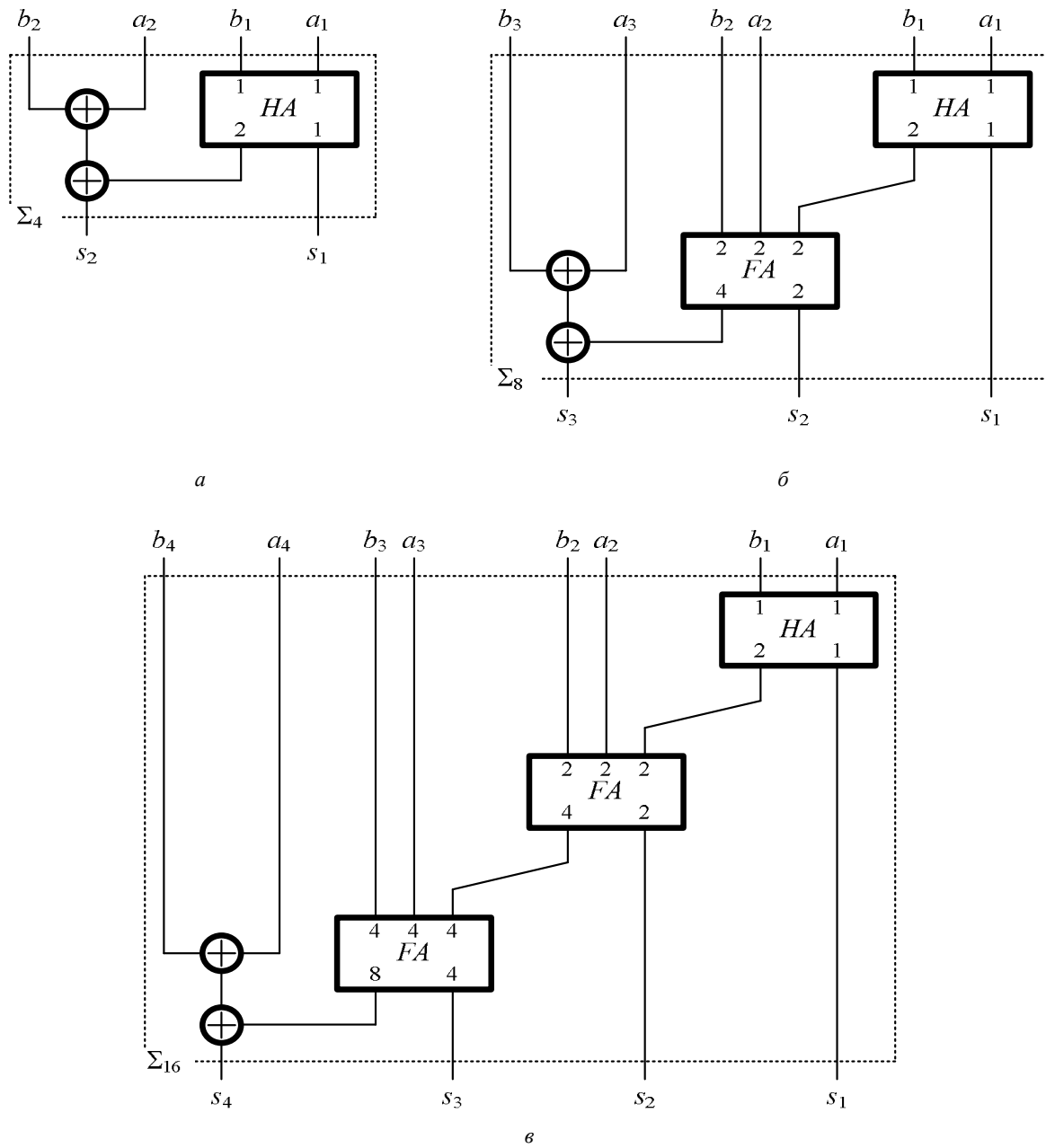


Рис. 5. Структурные схемы функциональных блоков Σ_M :
 а – сложения по модулю $M = 4$; б – сложения по модулю $M = 8$; в – сложения по модулю $M = 16$

$w_7 = \langle 0 f_7 f_7 f_7 \rangle$	$w_1 = \langle 0 0 0 f_1 \rangle$	$w_2 = \langle 0 0 f_2 0 \rangle$	$w_3 = \langle 0 0 f_3 f_3 \rangle$
+	+	+	+
$w_8 = \langle f_8 0 0 0 \rangle$	$w_6 = \langle 0 f_6 f_6 0 \rangle$	$w_5 = \langle 0 f_5 0 f_5 \rangle$	$w_4 = \langle 0 f_4 0 0 \rangle$
$w_7 + w_8 = \langle f_8 f_7 f_7 f_7 \rangle$	$w_1 + w_6 = \langle 0 f_6 f_6 f_1 \rangle$	$w_2 + w_5 = \langle 0 f_5 f_2 f_5 \rangle$	$w_3 + w_4 = \langle 0 f_4 f_3 f_3 \rangle$
$(w_7 + w_8)(\text{mod} 8) =$	$(w_1 + w_6)(\text{mod} 8) =$	$(w_2 + w_5)(\text{mod} 8) =$	$(w_3 + w_4)(\text{mod} 8) =$
$= \langle f_7 f_7 f_7 \rangle$	$= \langle f_6 f_6 f_1 \rangle$	$= \langle f_5 f_2 f_5 \rangle$	$= \langle f_4 f_3 f_3 \rangle$

Рис. 6. Получение двоичных чисел, суммируемых
 на входах функциональных блоков сложения по модулю $M = 8$

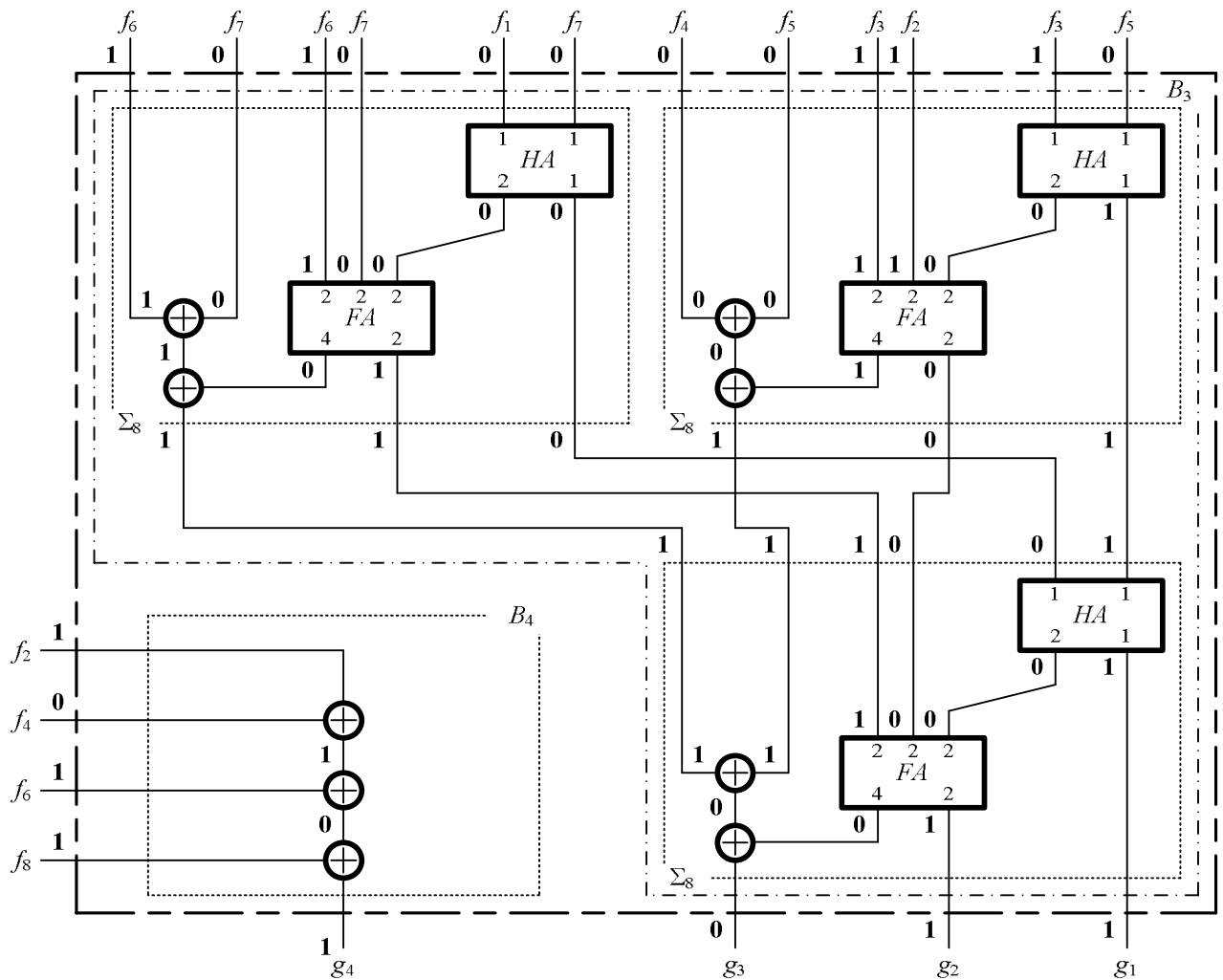


Рис. 7. Генератор разрядов контрольного вектора $RWS(8,4)$ -кода

Поскольку все весовые коэффициенты $RWS(m,k)$ -кода при использовании алгоритма 2 делятся на $\left\lfloor \frac{m}{2} \right\rfloor$ групп для их последующего суммирования, справедливо следующее положение.

Утверждение 1. Генератор $RWS(m,k)$ -кода содержит $\left\lfloor \frac{m}{2} \right\rfloor - 1$ функциональный блок Σ_M и блок формирования старшего контрольного разряда, включающий в себя $\left\lfloor \frac{m}{2} \right\rfloor - 1$ сумматоров по модулю два (запись $\left\lfloor \dots \right\rfloor$ обозначает целое снизу от вычисляемого значения).

Сформулированное выше положение о количестве функциональных элементов в структуре генератора модифицированного кода Бергера позволяет дать оценку сложности его технической реализации в зависимости от длины информационного вектора. Одним из таких показателей является количество двухвходовых элементарных логических элементов, требующихся для реализации генератора [6]. Данный показатель известен для сумматора по модулю два, полусумматора и полного сумматора [40]: $L_{XOR} = 3$, $L_{HA} = 3$, $L_{FA} = 7$. Зная эти данные, нетрудно оценить сложность технической реализации функциональных блоков Σ_M : $L_{\Sigma_4} = 9$, $L_{\Sigma_8} = 16$ и $L_{\Sigma_{16}} = 23$. В табл. 3 приводятся рассчитанные значения показателя сложности технической реализации генераторов $RWS(m,k)$ -кодов при длинах информационных векторов $m = 3 \div 20$, а на рис. 8 иллюстрируется зависимость сложности технической реализации генераторов от длины информационного вектора. Она носит линейный характер.

Формализуем процедуру оценки сложности технической реализации генераторов $RWS(m,k)$ -кодов.

Из утверждения 1 ясно, что сложность технической реализации генератора определяется суммой

$$L_{RWS} = \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) L_{\Sigma_M} + \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) L_{XOR}. \quad (6)$$

Т а б л и ц а 3

Показатели сложности технической реализации генераторов $RWS(m,k)$ -кодов

m	Формула генератора	Количество двухвходовых элементов в структуре генератора
3	$2XOR$	6
4	$1XOR+1\Sigma_4$	12
5	$1XOR+2\Sigma_4$	21
6	$2XOR+2\Sigma_4$	24
7	$2XOR+3\Sigma_4$	33
8	$3XOR+3\Sigma_8$	57
9	$3XOR+4\Sigma_8$	73
10	$4XOR+4\Sigma_8$	76
11	$4XOR+5\Sigma_8$	92
12	$5XOR+5\Sigma_8$	95
13	$5XOR+6\Sigma_8$	111
14	$6XOR+6\Sigma_8$	114
15	$6XOR+7\Sigma_8$	130
16	$7XOR+7\Sigma_{16}$	182
17	$7XOR+8\Sigma_{16}$	205
18	$8XOR+8\Sigma_{16}$	208
19	$8XOR+9\Sigma_{16}$	231
20	$9XOR+9\Sigma_{16}$	234

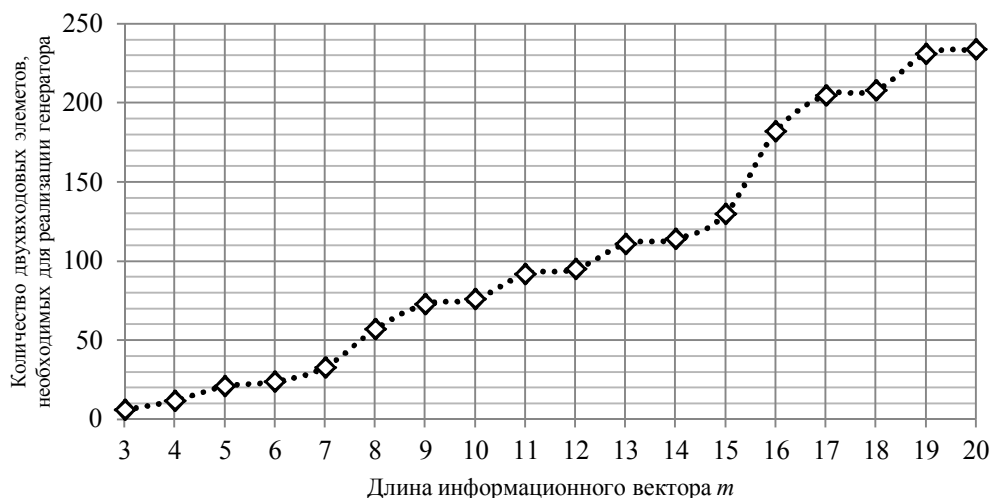


Рис. 8. Зависимость показателя сложности технической реализации генератора $RWS(m,k)$ -кода от длины информационного вектора

Величина L_{Σ_M} в формуле (6) зависит от количества разрядов в контрольных векторах. Каждый такой блок, за исключением случая $M = 2$, содержит один элемент HA , $\lceil \log_2(m+1) \rceil - 3$ элемента FA и два элемента XOR (см. рис. 4 и 5). Это объясняется тем, что на полусумматоре реализуется функция вычисления младшего разряда контрольного вектора $RWS(m,k)$ -кода, на $k-3$ полных сумматорах реализуются контрольные разряды $g_2, g_3, \dots, g_{k-2}$, а на двух сумматорах по модулю два – контрольный разряд g_{k-1} . Таким образом, сложность функционального блока Σ_M вычисляется по формуле

$$L_{\Sigma_M} = L_{HA} + (\lceil \log_2(m+1) \rceil - 3) L_{FA} + 2L_{XOR}. \quad (7)$$

Подставим (7) в (6), а также внесем туда значения $L_{XOR} = 3$, $L_{HA} = 3$, $L_{FA} = 7$.

Утверждение 2. Количество двухвходовых элементов в структуре генератора $RWS(m,k)$ -кода определяется выражением

$$\begin{aligned} L_{RWS(m,k)} &= \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) \left(L_{HA} + (\lceil \log_2(m+1) \rceil - 3) L_{FA} + 2L_{XOR} \right) + \left(\left\lfloor \frac{m}{2} \right\rfloor - 1 \right) L_{XOR} = \\ &= \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) (7 \cdot (\lceil \log_2(m+1) \rceil - 3) + 9) + 3 \cdot \left(\left\lfloor \frac{m}{2} \right\rfloor - 1 \right) = \\ &= 7 \lceil \log_2(m+1) \rceil \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) - 12 \left\lceil \frac{m}{2} \right\rceil + 3 \left\lfloor \frac{m}{2} \right\rfloor + 9. \end{aligned} \quad (8)$$

Использование формулы (8) для рассмотренного выше $RWS(8,4)$ -кода дает следующий результат:

$$L_{RWS(8,4)} = 7 \lceil \log_2(8+1) \rceil \left(\left\lceil \frac{8}{2} \right\rceil - 1 \right) - 12 \left\lceil \frac{8}{2} \right\rceil + 3 \left\lfloor \frac{8}{2} \right\rfloor + 9 = 7 \lceil \log_2(9) \rceil \left(\left\lceil \frac{8}{2} \right\rceil - 1 \right) - 12 \left\lceil \frac{8}{2} \right\rceil + 3 \left\lfloor \frac{8}{2} \right\rfloor + 9 = 57.$$

Заключение

Предложенный в настоящей работе алгоритм синтеза генераторов модифицированных кодов Бергера позволяет строить данные устройства с простой структурой на основе каскадного соединения функциональных блоков сложения по модулю M . Сложность структур генераторов с увеличением длины информационного вектора возрастает линейно.

Описанный способ синтеза генераторов является универсальным и может быть применен для построения генераторов как кодов с суммированием взвешенных информационных разрядов [10, 34, 40–44], так и кодов с суммированием взвешенных переходов между разрядами, занимающими соседние позиции в информационных векторах [13, 35, 45–49]. Однако возможность оптимизации структур генераторов будет напрямую определяться значениями весовых коэффициентов кода.

ЛИТЕРАТУРА

1. Слабаков Е.В. Построение полностью самопроверяемых комбинационных устройств с использованием остаточных кодов // Автоматика и телемеханика. 1979. № 10. С. 133–141.
2. Слабаков Е.В., Согомонян Е.С. Самопроверяемые вычислительные устройства и системы (обзор) // Автоматика и телемеханика. 1981. № 11. С. 147–167.
3. McCluskey E.J. Logic Design Principles: With Emphasis on Testable Semicustom Circuits. N.J.: Prentice Hall PTR, 1986. 549 p.
4. Сагалович Ю.Л. Кодовая защита оперативной памяти ЭВМ от ошибок // Автоматика и телемеханика. 1991. № 5. С. 3–45.
5. Согомонян Е.С., Слабаков Е.В. Самопроверяемые устройства и отказоустойчивые системы. М.: Радио и связь, 1989. 208 с.
6. Сапожников В.В., Сапожников Вл.В. Самопроверяемые дискретные устройства. СПб.: Энергоатомиздат, 1992. 224 с.
7. Piestrak S.J. Design of Self-Testing Checkers for Unidirectional Error Detecting Codes. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 1995. 111 p.
8. Pradhan D.K. Fault-Tolerant Computer System Design. New-York: Prentice Hall, 1996. 560 p.
9. Das D., Toubia N.A. Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose-Lin Codes // Journal of Electronic Testing: Theory and Applications. 1999. V. 15, Issue 1–2. P. 145–155.
10. Das D., Toubia N.A., Seuring M., Gossel M. Low Cost Concurrent Error Detection Based on Modulo Weight-Based Codes // Proceedings of IEEE 6th International On-Line Testing Workshop (IOLTW), Spain, Palma de Mallorca, July 3–5. 2000. P. 171–176.
11. Matrosova A.Yu., Levin I., Ostanin S.A. Self-Checking Synchronous FSM Network Design with Low Overhead // VLSI Design. 2000. V. 11, Issue 1. P. 47–58.
12. Fujiwara E. Code Design for Dependable Systems: Theory and Practical Applications. John Wiley & Sons, 2006. 720 p.
13. Mehov V., Sapozhnikov V., Sapozhnikov V.I., Urganskov D. Concurrent Error Detection Based on New Code with Modulo Weighted Transitions between Information Bits // Proceedings of 7th IEEE East-West Design & Test Workshop (EWDWTW 2007), Erevan, Armenia, September 25–30. 2007. P. 21–26.
14. Блюдов А.А., Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. Коды с суммированием для организации контроля комбинационных схем // Автоматика и телемеханика. 2013. № 6. С. 153–164.
15. Блюдов А.А., Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. О кодах с суммированием единичных разрядов в системах функционального контроля // Автоматика и телемеханика. 2014. № 8. С. 131–145.
16. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В. Применение кодов с суммированием при синтезе систем железнодорожной автоматики и телемеханики на программируемых логических интегральных схемах // Автоматика на транспорте. 2015. Т. 1, № 1. С. 84–107.

17. Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. Применение модульных кодов с суммированием для построения систем функционального контроля комбинационных логических схем // Автоматика и телемеханика. 2015. № 10. С. 152–169.
18. Nicolaidis M., Zorian Y. On-Line Testing for VLSI – A Compendium of Approaches // Journal of Electronic Testing: Theory and Applications. 1998. Issue 12. P. 7–20.
19. Mitra S., McClaskey E.J. Which Concurrent Error Detection Scheme to Choose? // Proceedings of International Test Conference, 2000, USA, Atlantic City, NJ, 03–05 October 2000. P. 985–994.
20. Ефанов Д.В., Грошев Г.М., Маликов О.Б. Способы организации систем функционального диагностирования логических схем без памяти // Электротехника. 2016. № 5. С. 53–56.
21. Berger J.M. A Note on Error Detecting Codes for Asymmetric Channels // Information and Control. 1961. V. 4, Issue 1. P. 68–73.
22. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В. Классификация ошибок в информационных векторах систематических кодов // Известия вузов. Приборостроение. 2015. Т. 58, № 5. С. 333–343.
23. Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. О свойствах кода с суммированием в схемах функционального контроля // Автоматика и телемеханика. 2010. № 6. С. 155–162.
24. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В. Предельные свойства кода с суммированием // Известия Петербургского университета путей сообщения. 2010. № 3. С. 290–299.
25. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Черепанова М.Р. Модульные коды с суммированием в системах функционального контроля. I. Свойства обнаружения ошибок кодами в информационных векторах // Электронное моделирование. 2016. Т. 38, № 2. С. 27–48.
26. Busaba F.Y., Lala P.K. Self-Checking Combinational Circuit Design for Single and Unidirectional Multibit Errors // Journal of Electronic Testing: Theory and Applications. 1994. Issue 5. P. 19–28.
27. Гессель М., Морозов А.А., Сапожников В.В., Сапожников Вл.В. Построение комбинационных самопроверяемых устройств с монотонно независимыми выходами // Автоматика и телемеханика. 1994. № 7. С. 148–160.
28. Гессель М., Морозов А.А., Сапожников В.В., Сапожников Вл.В. Исследование комбинационных самопроверяемых устройств с независимыми и монотонно независимыми выходами // Автоматика и телемеханика. 1997. № 2. С. 180–193.
29. Morosow A., Sapozhnikov V.V., Sapozhnikov V.I., Goessel M. Self-Checking Combinational Circuits with Unidirectionally Independent Outputs // VLSI Design. 1998. V. 5, Issue 4. P. 333–345.
30. Saposhnikov V.V., Morosov A., Saposhnikov V.I., Göessel M. A New Design Method for Self-Checking Unidirectional Combinational Circuits // Journal of Electronic Testing: Theory and Applications. 1998. V. 12, Issue 1–2. P. 41–53.
31. Matrosova A., Levin I., Ostanin S. Survivable Self-Checking Sequential Circuits // Proceedings of 2001 IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT 2001). October 24–26, San Francisco, CA, 2001. P. 395–402.
32. Блюдов А.А., Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. Построение модифицированного кода Бергера с минимальным числом необнаруживаемых ошибок информационных разрядов // Электронное моделирование. 2012. Т. 34, № 6. С. 17–29.
33. Efanov D., Sapozhnikov V., Sapozhnikov V.I., Nikitin D. Sum Code Formation with Minimum Total Number of Undetectable Errors in Data Vectors // Proceedings of 13th IEEE East-West Design & Test Symposium (EWDTS'2015), Batumi, Georgia, September 26–29, 2015. P. 141–148.
34. Das D., Touba N.A. Weight-Based Codes and Their Application to Concurrent Error Detection of Multilevel Circuits // Proceedings of 17th IEEE Test Symposium, USA, California, 1999. P. 370–376.
35. Мехов В.Б., Сапожников В.В., Сапожников Вл.В. Контроль комбинационных схем на основе модифицированных кодов с суммированием // Автоматика и телемеханика. 2008. № 8. С. 153–165.
36. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В. Взвешенные коды с суммированием для организации контроля логических устройств // Электронное моделирование. 2014. Т. 36, № 1. С. 59–80.
37. Marouf M.A., Friedman A.D. Design of Self-Checking Checkers for Berger Codes // Proceedings of 8th Annual International Conference on Fault-Tolerant Computing, Toulouse, France, 1978. P. 179–183.
38. Lala P.K. Principles of Modern Digital Design. New Jersey : John Wiley & Sons, 2007. 436 p.
39. Blyudov A.A. On the Synthesis of Test Equipment for Modulo Codes with Summation // Proceedings of Petersburg Transport University. 2013. Issue 1. P. 53–58.
40. Дрозд А.В., Харченко В.С., Антошук С.Г., Дрозд Ю.В., Дрозд М.А., Сулима Ю.Ю. Рабочее диагностирование безопасных информационно-управляющих систем / под ред. А.В. Дрозда, В.С. Харченко. Харьков : Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», 2012. 614 с.
41. Favalli M., Metra C. Optimization of Error Detecting Codes for the Detection of Crosstalk Originated Errors // Design, Automation and Test in Europe (DATE), March 13–16, 2001. P. 290–296.
42. Ghosh S., Lai K.W., Jone W.B., Chang S.C. Scan Chain Fault Identification Using Weight-Based Codes for SoC Circuits // Proceedings of 13th Asian Test Symposium, 15–17 November. 2004. P. 210–215.
43. Sapozhnikov V., Sapozhnikov V.I., Efanov D., Nikitin D. Combinational Circuits Checking on the Base of Sum Codes with One Weighted Data Bit // Proceedings of 12th IEEE East-West Design & Test Symposium (EWDTS'2014), Kyev, Ukraine, September 26–29, 2014. P. 126–136. DOI 10.1109/EWDTS.2014.7027064.
44. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Никитин Д.А. Модульно взвешенный код с суммированием для систем технического диагностирования // Информатика и системы управления. 2015. № 3. С. 53–62.
45. Saposhnikov V., Saposhnikov V.I. New Code for Fault Detection in Logic Circuits // Proceedings of 4th International conference on Unconventional Electromechanical and Electrical Systems, St. Petersburg, Russia, June 21–24. 1999. P. 693–696.

46. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Дмитриев В.В. Свойства кодов с суммированием взвешенных переходов с прямой последовательностью весовых коэффициентов // Информатика и системы управления. 2014. № 4. С. 77–88.
47. Sapozhnikov V., Sapozhnikov Vl., Efanov D., Dmitriev V., Cherepanova M. Optimum Sum Codes, That Effectively Detect the Errors of Low Multiplicities // RadioElectronics & Informatics. 2015. No. 1. P. 17–22.
48. Дмитриев В.В. О двух способах взвешивания и их влиянии на свойства кодов с суммированием взвешенных переходов в системах функционального контроля логических схем // Известия Петербургского университета путей сообщения. 2015. № 3. С. 119–129.
49. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Дмитриев В.В., Черепанова М.Р. Способ построения кода с суммированием с улучшенными показателями обнаружения ошибок в информационных векторах // Автоматика на транспорте. 2016. Т. 2, № 1. С. 95–123.

Ефанов Дмитрий Викторович, канд. техн. наук, доцент. E-mail: TrES-4b@yandex.ru

Петербургский государственный университет путей сообщения Императора Александра I (г. Санкт-Петербург)

Поступила в редакцию 29 июня 2016 г.

Efanov Dmitry V. (Petersburg State Transport University, St. Petersburg, Russian Federation).

On the problem of modified codes with summation of weighted data bits with natural numbers sequence of weight indexes generators synthesis.

Keywords: technical diagnostics; concurrent error detection system (CED system); Berger code; modified Berger code; weight-based sum code; least nonnegative residue; modulus; generator; half-adder; full-adder.

DOI: 10.17223/19988605/37/2

Different codes with summation (most of which are Berger code modifications) are often used in reliable discrete devices design. Modified code with summation of weighted data bits refers to such modifications. It is systematic code (the same as Berger code) and has $k = \lceil \log_2(m+1) \rceil$ check bits (m – data vector length). This code formation involves attribution of natural numbers weight indexes, calculation of total active bits weight and its least nonnegative residue by modulo $M = 2^{\lceil \log_2(m+1) \rceil - 1}$. This value is placed in low bits of check vector. High bit is calculated as modulo two sum of data bits with even positions (that have even weight indexes). This code has better data bits error detection properties comparing with Berger code. Also it is the code with minimum total number of undetectable data bits errors for given length of data and check vectors.

The paper describes the method of modified Berger code generator synthesis. Component base list contains standard circuits of modulo two, half and full adders. They compose functional blocks of $M = 2^{\lceil \log_2(m+1) \rceil - 1}$ sum. Generator itself is formed by cascade connection of such blocks.

Two structural diagrams of modified Berger code generators are given; also the generator synthesis algorithm is described that optimizes its structure by separation of weight indexes by groups of summed numbers without transfers. These numbers are sent on inputs of counters of $M = 2^{\lceil \log_2(m+1) \rceil - 1}$ sum. High bit of check vector is formed on the output of the block that calculate modulo two sum of data bits with even positions. The algorithm is illustrated on the example synthesis of the generator for modified Berger code with $m = 8$.

The paper shows that of modified Berger code includes $\left\lceil \frac{m}{2} \right\rceil - 1$ functional blocks of sum by $M = 2^{\lceil \log_2(m+1) \rceil - 1}$ modulo and block of high bit formation that consists of $\left\lceil \frac{m}{2} \right\rceil - 1$ modulo two adders. Dependence between complexity and data vector length is stated. It is shown that number of two-input gates in generator structure is calculated with the following expression:

$$7 \lceil \log_2(m+1) \rceil \left(\left\lceil \frac{m}{2} \right\rceil - 1 \right) - 12 \left\lceil \frac{m}{2} \right\rceil + 3 \left\lfloor \frac{m}{2} \right\rfloor + 9.$$

The described way of generator synthesis is universal and could be used for the formation of generators for codes with summation of weighted data bits and for codes with summation of weighted transitions between adjacent bits as well.

REFERENCES

1. Slabakov, E.V. (1979) Design of Completely Self-Checking Combinational Devices with the use of Residual Codes. *Automation and Remote Control*. 10. pp. 133-141.
2. Slabakov, E.V. & Sogomonyan, E.S. (1981) Self-Checking Computing Devices and Systems (Review). *Automation and Remote Control*. 11. pp. 147-167.
3. McCluskey, E.J. (1986) *Logic Design Principles: With Emphasis on Testable Semicustom Circuits*. New Jersey: Prentice Hall PTR.
4. Sagalovich, Yu.L. (1991) Error-Correcting Codes for Computer Memories. *Automation and Remote Control*. 5. pp. 3-45.
5. Sogomonyan, E.S. & Slabakov, E.V. (1989) *Samoproveryaemye ustroystva i otkazoustoychivye sistemy* [Self-Checking Devices and Fault-Tolerant Systems]. Moscow: Radio i svyaz'.
6. Sapozhnikov, V.V. & Sapozhnikov Vl.V. (1992) *Samoproveryaemye diskretnye ustroystva* [Self-Checking Digital Devices]. St. Petersburg: Energoatomizdat. 224 p.
7. Piestrak, S.J. (1995) *Design of Self-Testing Checkers for Unidirectional Error Detecting Codes*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.

8. Pradhan, D.K. (1996) *Fault-Tolerant Computer System Design*. New-York: Prentice Hall.
9. Das, D. & Touba, N.A. (1999) Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose-Lin Codes. *Journal of Electronic Testing: Theory and Applications*. 15(1-2). pp. 145-155. DOI: 10.1023/A:1008344603814
10. Das, D., Touba, N.A., Seuring, M. & Gossel, M. (2000) Low Cost Concurrent Error Detection Based on Modulo Weight-Based Codes. *Proceedings of IEEE 6th International On-Line Testing Workshop (IOLTW)*. Spain, Palma de Mallorca. July 3-5. pp. 171-176. DOI: 10.1109/OLT.2000.856633
11. Matrosova, A.Yu., Levin, I. & Ostanin, S.A. (2000) Self-Checking Synchronous FSM Network Design with Low Overhead. *VLSI Design*. 11(1). pp. 47-58.
12. Fujiwara, E. (2006) *Code Design for Dependable Systems: Theory and Practical Applications*. John Wiley & Sons.
13. Mehov, V., Sapozhnikov, V., Sapozhnikov, V.I. & Urganskov, D. (2007) Concurrent Error Detection Based on New Code with Modulo Weighted Transitions between Information Bits. *Proceedings of 7th IEEE East-West Design & Test Workshop (EWDTS'2007)*. Erevan, Armenia. September 25-30. pp. 21-26.
14. Blyudov, A.A., Efanov, D.V., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (2013) Summation Codes for Organization of Control of Combinational Circuits. *Automation and Remote Control*. 6. pp. 153-164. DOI: 10.1134/S0005117913060118
15. Blyudov, A.A., Efanov, D.V., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (2014) On Summation Code of Unit Bits in Concurrent Error Detection Systems. *Automation and Remote Control*. 8. pp. 131-145. DOI: 10.1134/S0005117914080098
16. Sapozhnikov, V.V., Sapozhnikov, V.I.V. & Efanov, D.V. (2015) Primenenie kodov s summirovaniem pri sinteze sistem zheleznodorozhnoy avtomatiki i telemekhaniki na programmiruemykh logicheskikh integral'nykh skhemakh [Application of Codes with Summation for the Synthesis of Railway Automation and Remote Control Systems Based on Field-Programmable Gate Arrays]. *Avtomatika na transporte – Automation on Transport*. 1(1). pp. 84-107.
17. Efanov, D.V., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (2015) Applications of Modular Summation Codes to Concurrent Error Detection Systems for Combinational Boolean Circuits. *Automation and Remote Control*. 10. pp. 152-169. DOI: 10.1134/S0005117915100112
18. Nicolaidis, M. & Zorian, Y. (1998) On-Line Testing for VLSI – A Compendium of Approaches. *Journal of Electronic Testing: Theory and Applications*. 12. pp. 7-20. DOI: 10.1023/A:1008244815697
19. Mitra, S. & McClaskey, E.J. (2000) Which Concurrent Error Detection Scheme to Choose? *Proceedings of International Test Conference*. USA, Atlantic City, NJ, October 3-5. pp. 985-994. DOI: 10.1109/TEST.2000.894311
20. Efanov, D.V., Groshev, G.M. & Malikov, O.B. (2016) Sposoby organizatsii sistem funktsional'nogo diagnostirovaniya logicheskikh skhem bez pamyati [Ways of Organization of Concurrent Error Detection Systems for Combinational Circuits]. *Elektrotehnika – Electrical Engineering*. 5. pp. 53-56.
21. Berger, J.M. (1961) A Note on Error Detecting Codes for Asymmetric Channels. *Information and Control*. 4(1). pp. 68-73. DOI: 10.1016/S0019-9958(61)80037-5
22. Sapozhnikov, V.V., Sapozhnikov, V.I.V. & Efanov, D.V. (2015) Errors classification in information vectors of systematic codes. *Izvestiya Vysshikh Uchebnykh Zavedeniy. Priborostroenie – Journal of Instrument Engineering*. 58(5). pp. 333-343. DOI: 10.17586/0021-3454-2015-58-5-333-343. (In Russian).
23. Efanov, D.V., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (2010) On Summation Code Properties in Functional Control Circuits. *Automation and Remote Control*. 6. pp. 155-162. DOI: 10.1134/S0005117910060123
24. Sapozhnikov, V.V., Sapozhnikov, V.I.V. & Efanov, D.V. (2010) Limit Property of Berger Code. *Izvestiya Peterburgskogo universiteta putey soobshcheniya – Proceedings of Petersburg Transport University*. 3. pp. 290-299. (In Russian).
25. Sapozhnikov, V.V., Sapozhnikov, V.I.V., Efanov, D.V. & Cherepanova, M.R. (2016) Modulo Codes with Summation in Concurrent Error Detection Systems. I. Ability to Detect Errors by Modulo Codes in Data Vectors. *Elektronnoe modelirovanie – Electronic Modeling*. 38(2). pp. 27-48. (In Russian).
26. Busaba, F.Y. & Lala, P.K. (1994) Self-Checking Combinational Circuit Design for Single and Unidirectional Multibit Errors. *Journal of Electronic Testing: Theory and Applications*. 5. pp. 19-28. DOI: 10.1007/BF00971960
27. Goessel, M., Morozov, A.A., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (1994) Postroenie kombinatsionnykh samoproveryaemykh ustroystv s monotonno nezavisimymi vykhodami [Construction of Combinational Self-Checking Devices with Monotonically Independent Outputs]. *Avtomatika i telemekhanika – Automation and Remote Control*. issue 7. pp. 148–160.
28. Goessel, M., Morozov, A.A., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (1997) Issledovanie kombinatsionnykh samoproveryaemykh ustroystv s nezavisimymi i monotonno nezavisimymi vykhodami [Investigation of Combination Self-Testing Devices Having Independent and Monotone Independent Outputs]. *Avtomatika i telemekhanika – Automation and Remote Control*. 2. pp. 180-193.
29. Morosow, A., Sapozhnikov, V.V., Sapozhnikov, V.I.V. & Goessel, M. (1998) Self-Checking Combinational Circuits with Unidirectionally Independent Outputs. *VLSI Design*. 5(4). pp. 333-345.
30. Sapozhnikov, V.V., Morosov, A., Sapozhnikov, V.I.V. & Goessel, M. (1998) A New Design Method for Self-Checking Unidirectional Combinational Circuits. *Journal of Electronic Testing: Theory and Applications*. 12(1-2). pp. 41-53. DOI: 10.1007/978-1-4757-6069-9_4
31. Matrosova, A., Levin, I. & Ostanin, S. (2001) Survivable Self-Checking Sequential Circuits. *Proceedings of 2001 IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT 2001)*. October 24-26, San Francisco, CA. pp. 395-402.
32. Blyudov, A.A., Efanov, D.V., Sapozhnikov, V.V. & Sapozhnikov, V.I.V. (2012) Formation of the Berger Modified Code with Minimum Number of Undetectable Errors of Informational Bits. *Elektronnoe modelirovanie – Electronic Modeling*. 34(6). pp. 17-29. (In Russian).
33. Efanov, D., Sapozhnikov, V., Sapozhnikov, V.I. & Nikitin, D. (2015) Sum Code Formation with Minimum Total Number of Undetectable Errors in Data Vectors. *Proceedings of 13th IEEE East-West Design & Test Symposium (EWDTS'2015)*. Batumi, Georgia, September 26-29. pp. 141-148.

34. Das, D. & Touba, N.A. (1999) Weight-Based Codes and Their Application to Concurrent Error Detection of Multilevel Circuits. *Proceedings of 17th IEEE Test Symposium*. USA, California. pp. 370-376. DOI: 10.1109/43.644041
35. Mekhov, V.B., Sapozhnikov, V.V. & Sapozhnikov, VI.V. (2008) Checking of Combinational Circuits Basing on Modification Sum Codes. *Automation and Remote Control*. 8. pp. 153-165. DOI: 10.1134/S0005117908080134
36. Sapozhnikov, V.V., Sapozhnikov, VI.V. & Efanov, D.V. (2014) Weighted Codes with Summation for Organization of the Check of Logic Units. *Elektronnoe modelirovanie – Electronic Modeling*. 36(1). pp. 59-80. (In Russian).
37. Marouf, M.A. & Friedman, A.D. (1978) Design of Self-Checking Checkers for Berger Codes. *Proceedings of 8th Annual International Conference on Fault-Tolerant Computing*. Toulouse, France. pp. 179-183.
38. Lala, P.K. (2007) *Principles of Modern Digital Design*. New-Jersey: John Wiley & Sons.
39. Blyudov, A.A. (2013) On the Synthesis of Test Equipment for Modulo Codes with Summation. *Proceedings of Petersburg Transport University*. 1. pp. 53-58. (In Russian)
40. Drozd, A.V., Kharchenko, V.S., Antoshchuk, S.G., Drozd, Ju.V., Drozd, M.A. & Sulima, Yu.Yu. (2012) *Objects and Methods of On-Line Testing for Safe Instrumentation and Control Systems*. Kharkov: National Aerospace University. (In Russian)
41. Favalli, M. & Metra, C. (2001) Optimization of Error Detecting Codes for the Detection of Crosstalk Originated Errors. *Design, Automation and Test in Europe (DATE)*. March 13-16. pp. 290-296. DOI: 10.1109/DATE.2001.915039
42. Ghosh, S., Lai, K.W., Jone, W.B. & Chang, S.C. (2004) Scan Chain Fault Identification Using Weight-Based Codes for SoC Circuits. *Proceedings of 13th Asian Test Symposium*. November 15-17. pp. 210-215. DOI: 10.1109/ATS.2004.77
43. Sapozhnikov, V., Sapozhnikov, VI., Efanov, D. & Nikitin, D. (2014) Combinational Circuits Checking on the Base of Sum Codes with One Weighted Data Bit. *Proceedings of 12th IEEE East-West Design & Test Symposium (EWDTS'2014)*. Kyev, Ukraine. September 26-29. pp. 126-136. DOI: 10.1109/EWDTS.2014.7027064.
44. Sapozhnikov, V.V., Sapozhnikov, VI.V., Efanov, D.V. & Nikitin, D.A. (2015) Modulo Weighted Code with Summation for Technical Diagnostics Systems. *Informatika i sistemy upravleniya – Information Science and Control Systems*. 3. pp. 53-62.
45. Sapozhnikov, V. & Sapozhnikov, VI. (1999) New Code for Fault Detection in Logic Circuits. *Proceedings of 4th International conference on Unconventional Electromechanical and Electrical Systems*. St. Petersburg, Russia. June 21-24. pp. 693-696. (In Russian).
46. Sapozhnikov, V.V., Sapozhnikov, VI.V., Efanov, D.V. & Dmitriev, V.V. (2014) Properties of Codes with Summation of Transitions Between Data Bits With Direct Sequence of Weighting Coefficient. *Information Science and Control Systems*. issue 4. pp. 77-88.
47. Sapozhnikov, V., Sapozhnikov, VI., Efanov, D., Dmitriev, V. & Cherepanova, M. (2015) Optimum Sum Codes, That Effectively Detect the Errors of Low Multiplicities. *RadioElectronics & Informatics*. 1. pp. 17-22. (In Russian).
48. Dmitriev, V.V. (2015) On the Two Weighing Methods and Their Influence on Properties of Sum Codes of Weighted Transitions in Functional Logical Scheme Control Systems. *Izvestiya Peterburgskogo universiteta putey soobshcheniya – Proceedings of Petersburg Transport University*. 3. pp. 119-129. (In Russian).
49. Sapozhnikov, V.V., Sapozhnikov, VI.V., Efanov, D.V., Dmitriev, V.V. & Cherepanova, M.R. (2016) Method of Formation of Sum Code with Improved Efficiency of Error Detection in Data Vectors. *Avtomatika na transporte – Automation on Transport*. 2(1). pp. 95-123. (In Russian).

УДК 519.95

DOI: 10.17223/19988605/37/3

Н.А. Игнатьев**ИНДЕКСИРОВАНИЕ ОБЪЕКТОВ ПО ИНДИВИДУАЛЬНЫМ НАБОРАМ
ИНФОРМАТИВНЫХ ПРИЗНАКОВ**

Рассматриваются критерии для оценивания (индексирования) объекта в задачах распознавания с учителем. Значение оценки вычисляется как экстремум критерия по индивидуальному набору информативных признаков объекта. Проверяется истинность гипотезы, что в окрестности каждого объекта существует устойчивая логическая закономерность.

Ключевые слова: распознавание образов; индекс объекта; логические закономерности; информативные признаки объекта.

Потребность в индексировании объектов возникает при принятии решений в различных предметных областях. Значения индексов востребованы для мониторинга рынка купли-продажи ценных бумаг, экологического состояния окружающей среды, уровня террористической угрозы, оценки степени социального благополучия общества, цитируемости научных публикаций и т.д. Вычисление значений, как правило, производится по строго фиксированным наборам показателей.

Информативные признаки (показатели), определяемые для всей обучающей выборки, не отражают специфику закономерностей, присущих той или иной области признакового пространства. В [1] утверждается, что для каждого объекта существует своя логическая закономерность, для обнаружения которой предлагалось использовать локальные метрики. Применение локальных метрик основано на эвристиках, так как четких критериев их выбора не существует. Интерес представляет разработка методов отбора информативных признаков инвариантных к масштабам измерений данных, комбинаторная сложность реализации которых позволяет получать результаты за приемлемое время.

Самую простую и легко интерпретируемую структуру, задаваемую отношениями на элементах непустого множества, представляет линейный порядок. Как правило, первыми кандидатами для включения в информативный набор являются независимые признаки. Примером использования линейного порядка является отбор наборов признаков с максимально выраженной независимостью, который применялся при синтезе искусственных нейронных сетей с минимальной конфигурацией в [2].

Потребность в использовании индивидуального набора информативных признаков объекта для принятия решений возникает при постановке диагноза болезни, разработке мер по предотвращению техногенных катастроф на конкретной территории. В медицинской практике при одном и том же диагнозе у двух человек причинами (диагностическими признаками) болезни могут быть разные симптомы и синдромы.

Метод отбора индивидуальных информативных наборов признаков с помощью локальных метрик объектов описан в [3]. Для отбора использовался критерий на основе максимальной разницы частот встречаемости представителей (объектов) двух классов K_1 и K_2 в последовательности, упорядоченной по локальной метрике объекта. Экстремальное значение критерия, полученное по медицинским данным с описанием состояния больных и практически здоровых индивидуумов, предлагалось интерпретировать как индекс здоровья.

Индивидуальный набор информативных признаков допустимого объекта позволяет:

- выделять логические закономерности в его окрестности;
- объяснять процесс принятия решения при распознавании;
- определять принадлежность к шумовым (аномальным) объектам классов;
- производить выбор опорных множеств признаков в моделях алгоритмов распознавания.

В данной работе для отбора индивидуальных наборов информативных признаков предложено два новых критерия, отличных от описанного в [3]. Также как и в [3], при вычислении значений по этим критериям используются функции близости по определяемым наборам признаков. Упорядочение объектов по значениям функции близости в зависимости от поставленных целей позволяет определять:

- устойчивость логической закономерности относительно исследуемого объекта;
- границу между представителями двух классов, степень истинности гипотезы о компактности при которой максимальна.

Свойство инвариантности к масштабам шкал измерений является атрибутом критерия нелинейного отображения групп исходных (сырых) разнотипных (номинальных и количественных) признаков на числовую ось. По аналогии с методом локальной геометрии [1] начало координат размещается в исследуемом объекте. При нелинейном отображении наряду с синтезом латентных признаков по критерию группировки происходит их упорядочение по степени информативности. Существует вывод аналитического представления (формул) для вычисления значений латентных признаков из исходных [4].

1. Критерии отбора индивидуальных наборов информативных признаков

Рассматривается задача распознавания в стандартной постановке. Объекты обучения заданы через множество $E_0 = \{S_1, \dots, S_m\}$, разделённое на два непересекающихся подмножества (класса) K_1 и K_2 , $E_0 = K_1 \cup K_2$. Описание объектов производится с помощью набора из n разнотипных признаков $X(n) = (x_1, \dots, x_n)$, ξ из которых измеряются в интервальных шкалах, $n - \xi$ – в номинальной.

Обозначим через I, J множество индексов соответственно количественных и номинальных признаков. Считается, что заданы критерии для отбора информативных признаков объекта $S \in E_0$. Требуется по каждому критерию для указанного объекта $S \in E_0$ определить:

- информативный набор признаков $X(k) = \{x_i\}_{i \in I \cup J}, k \geq 1$;
- оценку объекта S как экстремальное значение критерия на информативном наборе $X(k)$.

Описание допустимого объекта в рамках собственного пространства из информативных признаков необходимо для нахождения индивидуальной меры сходства (различия) с другими объектами. Эта мера отражает отношения между объектами и служит средством для принятия решения.

Для унификации масштабов измерений значения количественных признаков дробно-линейным преобразованием отображаются в $[0, 1]$. В качестве меры близости между объектами $S_a = (x_{a1}, \dots, x_{an})$ и $S_b = (x_{b1}, \dots, x_{bn})$ используется метрика Журавлёва

$$\rho(S_a, S_b) = \sum_{i \in I} |x_{ai} - x_{bi}| + \sum_{i \in J} \begin{cases} 1, x_{ai} \neq x_{bi}, \\ 0, x_{ai} = x_{bi}. \end{cases}$$

Положим, что для объекта $S_d \in K_p$ по набору признаков $X(k), k \leq n$, построена

$$S_{d_0}, \dots, S_{d_{m-1}}, S_{d_m} = S_d, \quad (1)$$

упорядоченная последовательность объектов E_0 , отношения между которыми определяются неравенствами вида $\rho(S_{d_i}, S_d) \leq \rho(S_{d_{i+1}}, S_d)$. Для оценки объекта $S_d \in K_p$ по (1) используется функционал

$$F(S_d, X(k)) = \max_{0 \leq i \leq m-1} \left(\frac{z_p(i)}{|K_p \cap E_0|} - \frac{z_{3-p}(i)}{|K_{3-p} \cap E_0|} \right), \quad (2)$$

где $z_p(i), z_{3-p}(i)$ – число объектов в $\{S_{d_0}, \dots, S_{d_i}\} \subset E_0$, определяемых по (1) соответственно из класса K_p и K_{3-p} . Множество допустимых значений (2) принадлежит интервалу $(0, 1]$.

Как отдельную задачу можно рассматривать отбор для $S_d \in K_p$ набора информативных признаков $X(\mu), \mu \leq n$, на котором

$$F(S_d, X(\mu)) = \max_{1 \leq k \leq n} \max_{\{X(k)\}} F(S_d, X(k)). \quad (3)$$

Значение (3) для допустимого объекта S_d по набору медицинских показателей $X(\mu)$ в [3] интерпретировалось как индекс здоровья по классу $K_p, p = 1, 2$.

В [3] для нахождения экстремума (3) использовались эвристические пошаговые алгоритмы отбора. Было показано, что различные схемы отбора (последовательным удалением малоинформативных либо последовательным включением наиболее информативных) признаков не давали схожих результатов.

Существенным препятствием для эффективного использования (3) в эвристических алгоритмах отбора является большая размерность признакового пространства. Значения близости между объектами становятся размытыми, экспоненциально растёт сложность вычислений. Для уменьшения комбинаторной сложности вычислений предлагается применять предобработку данных.

Идея использования порядка следования разнотипных признаков по степени их независимости для синтеза моделей искусственных нейронных сетей с минимальной конфигурацией описана в [2]. С этой целью формировалась матрица парных близостей (различий) между признаками. Для унификации шкал измерений использовалось преобразование количественных признаков в номинальные по специальному критерию. Порядок следования признаков определялся по матрице парных близостей (различий).

В отличие от [2] в данной работе предлагается использовать значения матрицы близости для пар признаков $(x_i, x_j) \in X(n)$ по метрике Журавлёва без унификации шкал (сведения к одной шкале) измерений. Элементы матрицы близости $B(S) = \{b_{ij}\}_{n \times n}$ объекта $S \in K_t$ вычисляются как

$$b_{ij} = \begin{cases} \frac{1}{2|K_{3-t}|} \sum_{S_u \in K_{3-t}} \rho(S, S_u) - \frac{1}{2|K_t|-1} \sum_{S_u \in K_t} \rho(S, S_u), & x_i, x_j \in X(n), i \neq j, \\ 0, & i = j. \end{cases} \quad (4)$$

Очевидно, что $|b_{ij}| \leq 1$.

Обозначим через P множество, значениями элементов которого являются номера исходных признаков. Для выбора по элементам (4) матрицы $B(S) = \{b_{ij}\}_{n \times n}$ упорядоченного набора $X(k) = (x_1, \dots, x_k)$, $2 \leq k \leq n$, используется рекурсивная процедура построения последовательности признаков

$$x_{S_1}, x_{S_2}, \dots, x_{S_n}. \quad (5)$$

Положим $P = \emptyset$. Из матрицы $B(S)$ выделяется пара (x_i, x_j) с наибольшим значением b_{ij} и включается (слева направо) в (5). Номера выделенных признаков фиксируются в $P = PU\{i, j\}$. Порядок следования в (x_i, x_j) выбирается из условия $\max_{\mu \in P} b_{i\mu} \geq \max_{\mu \in P} b_{j\mu}$. Каждая следующая пара признаков из $\{1, \dots, n\} \setminus P$ для (5) по аналогичному принципу определяется из $B(S)$ после удаления в ней строк и столбцов с номерами i и j .

С целью сокращения комбинаторной сложности алгоритмов из (5) удаляется (справа налево) определяемое число r ($0 < r < n$) элементов. Набор $x_{S_1}, \dots, x_{S_{n-r}}$ является исходным для начала процесса отбора информативных признаков по (3).

Предлагается ещё один, отличный от (3), способ использования последовательности (1) для отбора информативных признаков. Пусть u_i^1, u_i^2 – количество объектов класса K_i , $i = 1, 2$, соответственно в интервалах $[c_1, c_2], (c_2, c_3]$, η – порядковый номер из последовательности (1), $c_1 = 0$, $c_2 = \rho(S_d, S_{d_{\eta}})$, $c_3 = \rho(S_d, S_{d_{m-1}})$.

Критерий для определения границы c_2 основывается на проверке гипотезы (утверждения) о том, что каждый из двух интервалов $[c_1, c_2], (c_2, c_3]$ содержит значения расстояния $\rho(x, y)$ до объектов только одного класса. Экстремальное значение критерия на наборе $X(k)$, $2 \leq k \leq n$, вычисляется как

$$R(S_d, X(k)) = \left(\frac{\sum_{i=1}^2 u_i^1(u_i^1 - 1) + u_i^2(u_i^2 - 1)}{\sum_{i=1}^2 |K_i|(|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i}| - u_{3-i}^d)}{2|K_1||K_2|} \right) \rightarrow \max_{c_1 < c_2 < c_3}, \quad (6)$$

а множество его допустимых значений принадлежит $(0, 1]$. Выражение в левых скобках (6) представляет внутрикласовое сходство, в правых – межклассовое различие. Информативный набор признаков $X(\mu)$ по (6) определяется как

$$R(S_d, X(\mu)) = \max_{\{X(k)\}} R(S_d, X(k)).$$

Обозначим $\lambda_1(t) = |\{S_a \in K_t \mid \rho(S_d, S_a) \in [c_1, c_2]\}|$, $\lambda_2(t) = |\{S_a \in K_{3-t} \mid \rho(S_d, S_a) \in [c_1, c_2]\}|$, $\theta_1(t) = \lambda_1(t)/|K_t|$, $\theta_2(t) = \lambda_2(t)/|K_{3-t}|$, где интервал $[c_1, c_2]$ получен по набору признаков $X(k)$ по (6). Оценка (устойчивость) объекта $S_d \in K_t$ по набору $X(k)$ вычисляется как $U(S_d, X(k)) = \theta_1(t)(1 - \theta_2(t))$ и

$$U(S_d, X(\mu)) = \max_{\{X(k)\}} U(S_d, X(k)), \quad (7)$$

где $X(\mu)$ – информативный набор признаков. Вычисление значения оценки по (7) характеризует вид критерия как мультипликативный, а аналогично (3) – как аддитивный.

Логическая закономерность в форме гипершара с центром $S_d \in K_p$ по набору $X(k)$ определяется множеством

$$\phi(S_d, X(k)) = \{S_a \in K_p \mid \rho(S_d, S_a) < \rho(S_d, S_b)\},$$

где $S_b \in K_{3-p}$ – ближайший к S_d объект из противоположного класса. Для выбора и интерпретации информативного набора признаков объекта S_d и его оценки предлагается использовать экстремум критерия устойчивости логической закономерности в форме гипершара

$$F(S_d, X(\mu)) = \max_{\{X(k)\}} \frac{|\phi(S_d, X(k))|}{|K_p|}. \quad (8)$$

2. Выбор латентных признаков для описания объекта

Рассматривается стандартная постановка задачи распознавания, аналогичная описанной в п. 1. Производится выбор собственного признакового пространства объекта $S_d \in E_0$, $d = 1, \dots, m$, с помощью алгоритма иерархической агломеративной группировки [4]. Алгоритм группировки разбивает набор признаков $X(n)$ на непересекающиеся группы $X(k_1), \dots, X(k_t)$, $k_1 + \dots + k_t \leq n$. Нелинейное отображение представителей каждой группы на числовую ось образует новый латентный признак в описании объекта. Считается, что задан критерий для проверки истинности гипотезы о компактности по значениям латентного признака через произведение внутриклассового сходства и межклассового различия. Требуется определить признак с максимальным значением критерия.

Для выбора латентных признаков в собственном пространстве объекта $S_d \in E_0$, $S_d = (a_{d1}, \dots, a_{dn})$, произведём предобработку данных следующим образом. Значения признаков объекта $S = (b_1, \dots, b_n)$, $S \in E_0$, преобразуем как

$$b_i = \begin{cases} |a_{di} - b_i|, i \in I, \\ 1, a_{di} = b_i, i \in J, \\ 0, a_{di} \neq b_i, i \in J. \end{cases} \quad (9)$$

Преобразованные по (9) признаки считаются измеренными в количественной шкале измерений, множество номеров которых идентифицируются как $I = \{1, \dots, n\}$. Для вычисления значений латентных признаков используются правила иерархической агломеративной группировки. Латентные признаки, полученные на p -м шаге группировки, обозначаются как x_j^p , $j \in I$, $p \geq 0$. При $p = 0$, $|I| = n$. Упорядоченное множество значений признака x_j^p объектов из E_0 аналогично (6) разделим на два интервала $[c_1^{jp}, c_2^{jp}]$, $(c_2^{jp}, c_3^{jp}]$, каждый из которых рассматривается как градация номинального признака.

Пусть u_i^1, u_i^2 – количество значений признака x_j^p , $j \in I$, класса K_i , $i = 1, 2$, соответственно в интервалах $[c_1^{jp}, c_2^{jp}]$, $(c_2^{jp}, c_3^{jp}]$, $|K_i| > 1$; v – порядковый номер элемента упорядоченной по возрастанию последовательности $r_{j_1}, \dots, r_{j_v}, \dots, r_{j_m}$ значений x_j^p у объектов из E_0 , определяющий границы интервалов как $c_1^{jp} = r_{j_1}$, $c_2^{jp} = r_{j_v}$, $c_3^{jp} = r_{j_m}$. Аналогичный (6) критерий

$$\left(\frac{\sum_{i=1}^2 u_i^1 (u_i^1 - 1) + u_i^2 (u_i^2 - 1)}{\sum_{i=1}^2 |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i} - u_{3-i}^d|)}{2|K_1||K_2|} \right) \rightarrow \max_{c_1^{jp} < c_2^{jp} < c_3^{jp}} \quad (10)$$

позволяет вычислять оптимальное значение границы c_2^{jp} для интервалов $[c_1^{jp}, c_2^{jp}]$ и $(c_2^{jp}, c_3^{jp}]$.

Экстремум критерия (10) используется в качестве веса w_j^p ($0 \leq w_j^p \leq 1$) признака x_j^p . При $w_j^p = 1$ значения признака x_j^p у объектов из классов K_1 и K_2 не пересекаются между собой.

Значение комбинации b_{rij}^p по паре признаков (x_i^p, x_j^p) , $0 \leq p < n$, $i, j \in I$, $i \neq j$, объекта $S_r = \{a_{ru}^p\}_{u \in I}$, $S_r \in E_0$, вычисляется как

$$b_{rij}^p = \eta_{ij} \left(t_i w_i^p (a_{ri}^p - c_2^{ip}) / (c_3^{ip} - c_1^{ip}) + t_j w_j^p (a_{rj}^p - c_2^{jp}) / (c_3^{jp} - c_1^{jp}) \right) + \\ + (1 - \eta_{ij}) t_{ij} w_{ij}^p (a_{ri}^p a_{rj}^p - c_2^{ijp}) / (c_3^{ijp} - c_1^{ijp}), i, j \in I, t_{ij}, t_i, t_j \in \{-1, 1\}, \eta_{ij} \in [0, 1],$$

где w_i^p, w_j^p, w_{ij}^p – веса признаков, определяемые по (10) соответственно по множеству значений x_i^p, x_j^p и их произведения $x_i^p x_j^p$ на E_0 , значения $t_{ij}, t_i, t_j \in \{-1, 1\}$, $\eta_{ij} \in [0, 1]$ выбираются по экстремуму функционала

$$\Phi(p, i, j) = \frac{\min_{S_r \in K_1} b_{rij}^p - \max_{S_r \in K_2} b_{rij}^p}{\max_{S_r \in E_0} b_{rij}^p - \min_{S_r \in E_0} b_{rij}^p} = \max_{t_{ij}, t_i, t_j \in \{-1, 1\}, \eta_{ij} \in [0, 1]}. \quad (11)$$

Экстремум функционала (11) интерпретируется как отступ между объектами классов K_1 и K_2 по множеству значений по паре признаков (x_i^p, x_j^p) , $0 \leq p < n$, $i, j \in I$, $i \neq j$.

Обозначим через $\{z_{ij}^p\}_{i, j \in I}$, $p \geq 0$, квадратную матрицу размера $(n-p) \times (n-p)$, значение элемента z_{ij}^p которой при $p = 0$ определяется как

$$z_{ij}^p = \begin{cases} w_i^p, i = j, \\ \text{значению (10) по } \{b_{rij}^p\}_{r=1}^m, i \neq j, \end{cases} \quad (12)$$

через Γ_η , $\eta > 0$, – подмножество номеров признаков из $X(n)$. Пошаговая реализация алгоритма иерархической агломеративной группировки будет такой:

1-й шаг: $p = 0$, $\lambda c = 0$, $\eta = 1$. Выполнять $\Gamma_\eta = \{\eta\}$, $\text{Margin}_\eta = -2$, $\eta = \eta + 1$, пока $\eta \leq n$;

2-й шаг: Вычислить значения элементов матрицы $\{z_{ij}^p\}_{i, j \in I}$ по (12);

3-й шаг: Выделить $\Phi = \{z_{uv}^p \mid z_{uv}^p \geq \max(w_u^p, w_v^p) \text{ and } u \neq v, u, v \in I\}$. Если $\Phi = \emptyset$, то идти 9;

4-й шаг: Вычислить $\lambda n = \max_{z_{uv}^p \in \Phi} z_{uv}^p$. Выделить $\Delta = \{(s, t), s, t \in I \mid z_{st}^p = \lambda n \text{ and } s < t\}$. Определить пару

$\{i, j\}$, $i < j$, как

$$\{i, j\} = \begin{cases} \Delta, |\Delta| = 1, \\ \{s, t\}, (s, t) \in \Delta \quad \text{and} \quad \Phi(p, s, t) > \max_{(u, v) \in \Delta \setminus (s, t)} \Phi(p, u, v); \end{cases}$$

5-й шаг: Если $\lambda n > \lambda c$ или $\lambda n = \lambda c$ и $\text{Margin}_i < \Phi(p, i, j)$, то $\Gamma_i = \Gamma_i \cup \Gamma_j$, $\Gamma_j = \emptyset$, $\text{Margin}_i = \Phi(p, i, j)$, идти 7;

6-й шаг: Вывод номеров признаков из Γ_i , $\Gamma_i = \emptyset$, $I = I \setminus \{i\}$, идти 3;

7-й шаг: $p = p + 1$, $I = I \setminus \max(i, j)$, $k = \min(i, j)$, $\lambda c = \lambda n$. Заменить значения признаков в описании объекта $S_r = \{a_{ru}^{p-1}\}_{u \in I}$, $r = 1, \dots, m$ на

$$a_{ru}^p = \begin{cases} a_{ru}^{p-1}, u \in I \setminus \{k\}, \\ b_{rij}^p, u = k; \end{cases}$$

8-й шаг: Для каждой пары (u, v) , $u, v \in I$, определить значение

$$z_{uv}^p = \begin{cases} z_{uv}^{p-1}, u \in I \setminus \{k\}, v \in I, \\ \text{значению (10) на } \{a_{rv}^p\}_{r=1}^m, u = k, v \in I. \end{cases}$$

Если $n-p > 1$, то идти 3;

9-й шаг: Конец.

Реализация изложенного выше алгоритма представляет один из способов решения задачи компактного описания (агрегирования) данных через поиск функциональных зависимостей между признаками. Агрегирование данных выражается в формировании нового набора из латентных признаков в описании объекта, позволяющего, не прибегая к перебору, относительно легко обнаруживать устойчивые логические закономерности.

Пусть на базе группы Γ_r , $1 \leq r$, получен латентный признак $z(S)$ с максимальным значением (10), определены границы его интервалов $[c_1, c_2]$, $(c_2, c_3]$. Будем считать элементы Γ_r номерами информативного набора исходных признаков при выборе нового пространства в описании объекта $S_d \in K_t$. Значение оценки объекта $S_d \in K_t$ по латентному признаку $z(S)$ вычисляется как

$$\Omega(z(S_d)) = \theta_1(1-\theta_2), \quad (13)$$

где

$$\theta_1 = \frac{|\{S_i \in K_t \mid z(S_i) \in [c_1, c_2]\}|}{|K_t|}, \quad \theta_2 = \frac{|\{S_i \in K_{3-t} \mid z(S_i) \in [c_1, c_2]\}|}{|K_{3-t}|}.$$

3. Вычислительный эксперимент

Для вычислительного эксперимента использовалась выборка данных по пневмококковому и серозному менингиту [5]. Каждый из 64 объектов выборки описывался 3 количественными и 18 номинальными признаками. Первый класс K_1 (пневмококковый менингит) представлен 35 объектами, второй класс K_2 (серозный менингит) – 29 объектами.

Наборы признаков для ряда объектов выборки E_0 , формируемые алгоритмом пошагового включения информативных признаков по аддитивному (3) и мультипликативному (7) критериям, приводятся в табл. 1.

Т а б л и ц а 1

Информативные наборы признаков объектов

№ объекта (класс)	Информативные признаки по критерию	
	аддитивному (3)	мультипликативному (7)
1(1)	$x_3, x_4, x_{12}, x_{15}, x_{16}, x_{20}$	$x_3, x_4, x_6, x_{12}, x_{15}, x_{16}, x_{17}, x_{20}$
6(1)	x_2, x_3, x_{16}, x_{20}	x_2, x_3, x_{16}, x_{20}
15(1)	x_4, x_5, x_{14}	$x_1, x_6, x_7, x_{19}, x_{21}$
37(2)	$x_3, x_7, x_{12}, x_{16}, x_{20}$	$x_2, x_3, x_7, x_{12}, x_{16}$
54(2)	x_2, x_6, x_7, x_{12}	x_2, x_6, x_7, x_{12}
57(2)	$x_2, x_3, x_4, x_6, x_7, x_{12}, x_{20}$	$x_6, x_7, x_8, x_{12}, x_{16}, x_{20}$

Как видно из табл. 1, количество признаков в наборах и их состав, определяемые по (3) и (7), сильно не отличаются друг от друга. Насколько значения оценок объектов, полученные по экстремальным значениям трёх критериев (3), (7), (8), близки друг к другу, показано в табл. 2.

Результаты экспериментов наглядно демонстрируют наличие закономерностей, выражаемых через сходство их составов информативных наборов (см. табл. 1) и близость оценок признаков объектов (см. табл. 2). Низкие показатели оценок объекта № 15 указывают на несоответствие (аномальность) значений признаков в его описании относительно представителей класса K_1 .

Эффект от предобработки данных можно получить путём отказа от просмотра вариантов, не ведущих к оптимальному, с точки зрения используемого критерия, результату. Требуется проверить зави-

симость оценок объекта с учётом предобработки данных по (4). Результаты вычисления оценок по аддитивному критерию (3) на $X(n-r)$, $0 \leq r \leq n-2$, после удаления (справа налево) r признаков из (5) демонстрируются в табл. 3.

Т а б л и ц а 2

Оценки объектов по критериям

№ объекта (класс)	Критерий		
	аддитивный (3)	мультипликативный (7)	устойчивости лог. закономерности (8)
1(1)	0,8512	0,8276	0,8000
6(1)	0,7941	0,8000	0,7143
15(1)	0,3586	0,0985	0,2571
37(2)	0,8798	0,8778	0,6897
54(2)	0,9025	0,9044	0,5172
57(2)	0,9429	0,9103	0,6897

Т а б л и ц а 3

Оценки объектов по (3) с учётом предобработки

№ объекта (класс)	Число удаляемых признаков $r =$		
	5	10	15
1(1)	0,8571	0,8571	0,8227
6(1)	0,7941	0,7941	0,7941
15(1)	0,3586	0,3586	0,3419
37(2)	0,8798	0,8798	0,8453
54(2)	0,9025	0,9025	0,9143
57(2)	0,9429	0,9429	0,9143

Анализ результатов отбора информативных признаков и соответствующих им оценок по (3) на $X(21)$ (см. табл. 2) и с учётом предобработки на $X(16)$, $X(11)$ (см. табл. 3) показывает целесообразность поиска закономерностей в данных для сокращения комбинаторной сложности алгоритмов. Попытки применить подобный способ предобработки для вычисления информативных наборов признаков и оценок объектов по (7) и (8) оказались неэффективными.

Эффект от использования нелинейных преобразований признаков рассмотрим на примере вычисления оценок по (13). Вычисление проводилось по латентному признаку с максимальным значением (10) и соответствующему ему набору исходных признаков. Результаты представлены в табл. 4.

Т а б л и ц а 4

Нелинейные отображения групп признаков на числовую ось

№ объекта (класс)	Число групп	Латентный признак получен из набора	Значение критерия (13)
1(1)	4	$x_3, x_5, x_6, x_7, x_8, x_9, x_{12}, x_{15}, x_{16}, x_{20}$	0,9103
6(1)	6	x_3, x_7, x_{16}	0,8246
15(1)	5	$x_2, x_4, x_5, x_6, x_{10}$	0,8374
37(2)	6	$x_4, x_7, x_8, x_{12}, x_{17}, x_{20}$	0,8571
54(2)	5	x_2, x_6, x_7, x_{10}	0,8621
57(2)	5	$x_2, x_4, x_5, x_6, x_7, x_{12}, x_{20}$	0,9429

Неожиданный результат как эффект от использования методов интеллектуального анализа данных получен для объекта № 15. Высокая относительно критериев (3), (7), (8) оценка 0,8374 за объект по (13) свидетельствует о наличии скрытых закономерностей, обнаружить которые удастся лишь с учётом нелинейности. Существует возможность аналитического описания закономерностей по результатам нелинейного отображения по группе признаков. Последовательность формирования латентного признака алгоритмом агломеративной иерархической группировки на примере объекта № 54 (см. табл. 4) такова:

$$x_2^1 = 0,3(0,0051(x_2^0 - 3) - 0,536x_6^0) + 0,007x_2^0x_6^0;$$

$$x_2^2 = 0,1(1,6813(x_2^1 + 0,0175) - 0,2426x_{10}^1) + 1,7133(x_2^1 x_{10}^1 + 0,0175);$$

$$x_2^3 = 1,0118(x_2^2 + 0,0242) - 0,2144x_7^2.$$

Заключение

Предложены критерии для оценки объектов по индивидуальным информативным наборам признаков. С помощью вычислительного эксперимента доказано, что, несмотря на различие по составам наборов признаков, значения оценок по различным критериям оказались близки друг к другу. Результаты вычислений могут быть использованы для наполнения баз знаний и построения информационных моделей в слабоформализованных предметных областях.

ЛИТЕРАТУРА

1. Дюк В.А. Методология поиска логических закономерностей в предметной области с нечеткой системологией: На примере клинико-экспериментальных исследований : дис. ... д-ра тех. наук. СПб., 2005. 309 с.
2. Згуральская Е.Н. Выбор информативных признаков для решения задач классификации с помощью искусственных нейронных сетей // Нейрокомпьютеры: разработка, применение. 2012. № 2. С. 20–27.
3. Ignat'ev N.A., Mirzaev A.I. The Intelligent Health Index Calculation System // Pattern Recognition and Image Analysis. 2016. V. 26, No. 1. P. 73–77.
4. Игнат'ев Н.А. Вычисление обобщённых оценок объектов и иерархическая группировка признаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 4 (33). С. 31–37.
5. Касымова Р.И. Клинико-лабораторные особенности острых гнойных и серозных менингитов в зависимости от этиологии : дис. ... канд. мед. наук. Ташкент, 2009. 145 с.

Игнат'ев Николай Александрович, д-р физ.-мат. наук, профессор. E-mail: ignatev@rambler.ru
Национальный университет Узбекистана. Ташкент.

Поступила в редакцию 2 июля 2016 г.

Ignat'ev Nikolay A. (National University of Uzbekistan. Republic of Uzbekistan).

Indexing of objects on the individual sets of informing features.

Keywords: patterns recognition; index of object; logical regularity; informing features of object.

DOI: 10.17223/19988605/37/3

Criteria are investigated for the estimation (indexing) of object on the individual set of different-type features. It is considered, that a training sets of $E_0 = \{S_1, \dots, S_m\}$ divided into of disjoint subset (classes) K_1, K_2 . The objects of selection are described by a set of different-type features of $X(n)$, distances between objects are calculated on the metric of Juravlev. The individual set of informing features of possible object allows:

- to distinguish logical regularities in his neighborhood and calculate their stability;
- to explain a decision-making process at recognition;
- to determine belonging to the noise objects of classes;
- to produce the choice of supporting sets of features in the models of algorithms of recognition.

For the search of informing set of features of $X(k) \subset X(n)$, $k < n$ is used criteria: additive, multiplicative and to stability of logical regularity in the neighborhood of object of $S \in E_0$. The value of estimation of object of $S \in E_0$ is calculated as extremum of criterion on the individual set of $X(k)$. A heuristic allowing to decrease combinatory complication of algorithm of extraction of features on an additive criterion offers.

The method of synthesis of own space is worked out from latent features for description of object of $S \in E_0$ on the basis of hierarchical clustering. Every latent feature is result of nonlinear mapping of group of initial features on the real axis. The sequence of forming of latent feature the algorithm of hierarchical clustering is brought around to an example:

$$x_2^1 = 0,3(0,0051(x_2^0 - 3) - 0,536x_6^0) + 0,007x_2^0x_6^0;$$

$$x_2^2 = 0,1(1,6813(x_2^1 + 0,0175) - 0,2426x_{10}^1) + 1,7133(x_2^1 x_{10}^1 + 0,0175);$$

$$x_2^3 = 1,0118(x_2^2 + 0,0242) - 0,2144x_7^2.$$

The estimations of objects can be used for the construction of models in weakly formalizing subject domains. A requirement in using on the individual set of informing features of object for making decision arises up at raising of diagnosis of illness, to development of measures on prevention of technogenic catastrophes on concrete territory. Values of estimations in $[0,1]$ are easily interpreted in terms of fuzzy logic and can be used for verbalization of knowledge.

REFERENCES

1. Duyk, V.A. (2005) *Metodologiya poiska logicheskikh zakonomernostey v predmetnoy oblasti s nechetkoy sistemologiyey: Na primere kliniko-eksperimental'nykh issledovaniy* [Methodology of search of logical regularity in a problem domain with fuzzy systemology: On the example of clinic – experimental researches]. Engineering Doc. Diss. St. Petersburg.
2. Zguralskaya, E.N. (2012) Selecting informative features for solving problems of classification using artificial neural networks. *Neyrokompyutery: razrabotka, primeneniye*. 2. pp. 20-27. (In Russian).
3. Ignatiev, N.A. & Mirzaev, A.I. (2016) The Intelligent Health Index Calculation System. *Pattern Recognition and Image Analysis*. 26(1). pp. 73-77. DOI: 10.1134/S1054661816010089
4. Ignatiev, N.A. (2015) Computation generalized estimates of objects and hierarchical clustering of features. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(33). pp. 31-37. (In Russian).
5. Kasimova, R.I. (2009) *Kliniko-laboratornye osobennosti ostrykh gnoynykh i seroznykh meningitov v zavisimosti ot etiologii* [Clinic - laboratorial peculiarities of sharp, festering and serosal meningitises depending on etiology]. Medicine Cand. Diss. Tashkent.

УДК 519.872

DOI: 10.17223/19988605/37/4

В.А. Лохвицкий, А.В. Уланов

ЧИСЛЕННЫЙ АНАЛИЗ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ГИПЕРЭКСПОНЕНЦИАЛЬНЫМ «ОХЛАЖДЕНИЕМ»

Рассматривается многоканальная система массового обслуживания, в которой после ухода последней заявки периода непрерывной занятости на некоторое время наступает режим «охлаждения». Приходящие за время охлаждения заявки накапливаются и ожидают, пока система не перейдет в режим обслуживания. Получено распределение числа заявок и преобразование Лапласа–Стилтьеса времени ожидания для системы с простейшим входящим потоком, экспоненциальным обслуживанием и гиперэкспоненциальным (с произвольным типом параметров) охлаждением.

Ключевые слова: системы массового обслуживания с «охлаждением»; немарковские системы массового обслуживания; гиперэкспоненциальное распределение; численные методы.

В ряде систем массового обслуживания (СМО), в особенности при относительно малой загрузке, оказывается целесообразным введение порога включения, когда обслуживание начинается при скоплении в системе некоторого количества заявок и заканчивается при полном освобождении системы. Такой режим увеличивает как период непрерывной занятости, так и время, в течение которого обслуживание не ведется. Это на относительно длительные периоды позволяет переводить автоматическую аппаратуру (в целях экономии ресурса и электроэнергии) в облегченный (дежурный) режим. В системах с участием человека появляется возможность полностью выключать значительную часть техники и переводить оператора на решение других задач.

Очевидно, что обслуживание первой заявки в этих условиях сопряжено с выполнением некоторых дополнительных операций («разогрев» системы) и в среднем будет продолжаться дольше, чем обслуживание прочих заявок. СМО с «разогревом» рассматриваются в работах [1, 2].

Возможен и другой случай, когда после полного освобождения системы требуется проведение мероприятий, связанных с восстановлением работоспособности, проведением технического обслуживания, перерывами в работе и т.д. В этом случае уместно говорить, что СМО переходит в режим «охлаждения». Начало обслуживания вновь прибывшей заявки не начнется, пока не завершатся все операции охлаждения. В отличие от разогрева, процесс охлаждения системы не зависит от прибытия первой заявки периода занятости. Если система успеет «охладиться» до прихода заявки, обслуживание начнется без дополнительных задержек.

Из работ по исследованию систем с охлаждением авторам встречалась лишь статья [3], в которой рассматривалась полностью марковская система $M/M/M/n$. Ниже представлены модель и методика расчета вероятностно-временных характеристик системы с простейшим входящим потоком, марковским обслуживанием и немарковским охлаждением. Время охлаждения аппроксимируется гиперэкспоненциальным распределением второго порядка (H_2) с возможностью комплексного типа параметров, что обеспечивает достаточную для инженерных расчетов точность [4–8]. Модель системы $M/M/H_2/n$ представляет собой дискретный марковский процесс с непрерывным временем.

1. Модель системы $M/M/H_2/n$

Рассмотрим модель многоканальной СМО с «охлаждением» типа $M/M/H_2/n$. Напомним, что гиперэкспоненциальное распределение второго порядка относится к распределениям фазового типа и предполагает выбор случайным процессом одной из двух альтернативных фаз [2]. С вероятностью y_1 про-

цесс охлаждения попадает в первую фазу и задерживается в ней случайное время, распределенное по экспоненциальному закону с параметром μ_1^c . С вероятностью $y_2 = 1 - y_1$ процесс попадает во вторую фазу, где экспоненциальная задержка имеет параметр μ_2^c . Диаграмма переходов между состояниями марковского процесса, описывающего систему $M/M/H_2/n$, представлена на рис. 1.

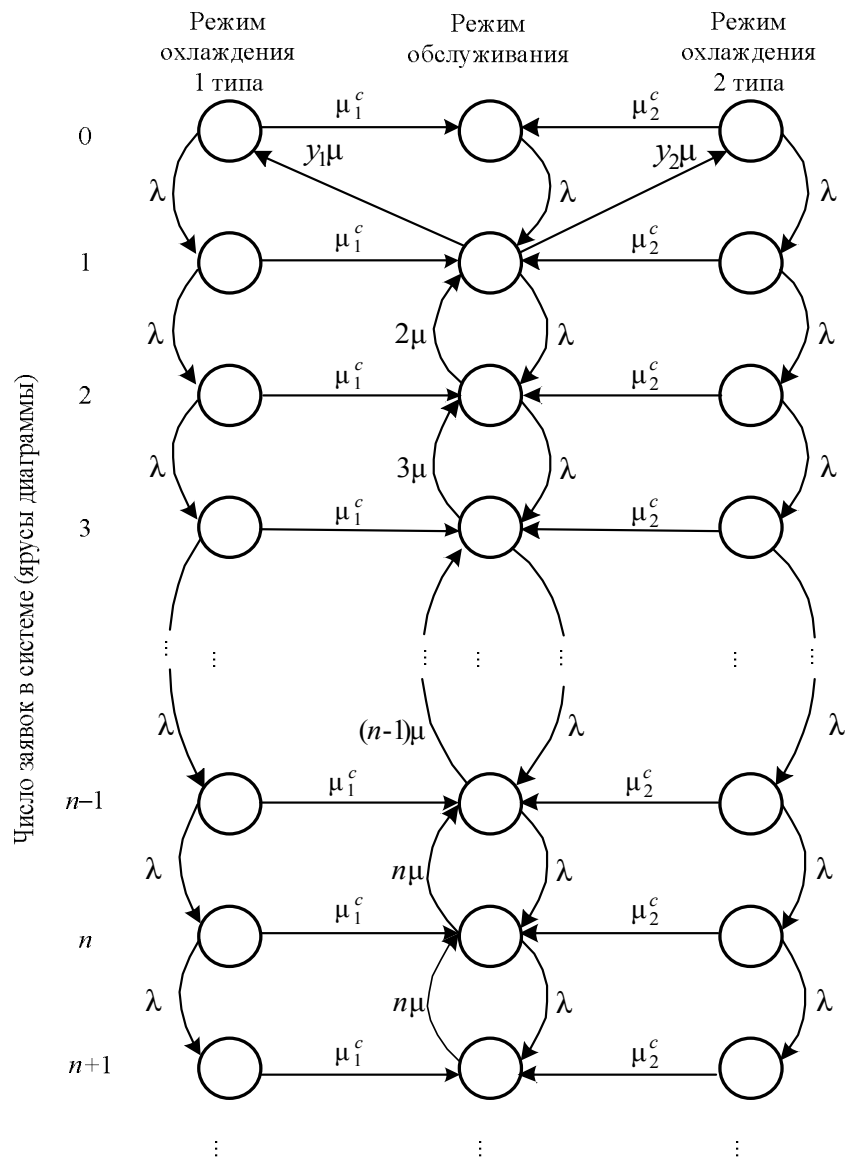


Рис 1. Диаграмма переходов для системы с охлаждением $M/M/H_2/n$

Крайний левый столбец на диаграмме показывает текущее число заявок в системе и указывает на номер яруса диаграммы. На каждом ярусе система находится в одном из трех состояний: обслуживание и охлаждение одного из двух типов. Состояние охлаждения характеризует некоторую работу, которая выполняется системой после окончания обслуживания последней заявки. В данной модели длительность охлаждения системы характеризуется двухфазным гиперэкспоненциальным распределением. После обслуживания последней заявки система переходит в режим (фазу) охлаждения 1-го типа с вероятностью y_1 и в режим охлаждения 2-го типа с вероятностью y_2 . Каждая из фаз охлаждения характеризуется своей интенсивностью — μ_1^c и μ_2^c соответственно. Обслуживание заявок в системе осуществляется с интенсивностью μ , умноженной на число занятых каналов обслуживания. При полной занятости (на ярусах с номерами больше n) диаграмма стабилизируется на всех нижележащих ярусах.

На основе диаграммы переходов марковизированной СМО строятся матрицы интенсивностей переходов и решаются векторно-матричные уравнения баланса переходов между микросостояниями.

2. Методика расчета вероятно-временных характеристик

Обозначим через S_j множество всех возможных микросостояний системы, при которых на обслуживании находится ровно j заявок, а через σ_j – количество элементов в S_j . Далее, в соответствии с диаграммой переходов построим матрицы интенсивностей инфинитезимальных переходов:

- $A_j[\sigma_j \times \sigma_{j+1}]$ – в S_{j+1} (по прибытию заявок);
- $B_j[\sigma_j \times \sigma_{j-1}]$ – в S_{j-1} (по завершению обслуживания);
- $C_j[\sigma_j \times \sigma_j]$ – в S_j (по завершению охлаждения);
- $D_j[\sigma_j \times \sigma_j]$ – ухода из микросостояний j -го яруса (диагональная матрица).

В квадратных скобках здесь и далее указывается размер матриц. Элемент (i, k) любой из этих матриц представляет интенсивность перехода из i -го состояния j -го яруса в k -е состояние смежного (по переходам рассматриваемого типа) яруса.

Для СМО $M/M/H_2/n$ матрицы переходов будут иметь следующий вид:

$$A_j = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \quad j = \overline{0, N-1};$$

$$B_1 = \begin{bmatrix} 0 & 0 & 0 \\ \mu y_1 & 0 & \mu y_2 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2\mu & 0 \\ 0 & 0 & 0 \end{bmatrix} \dots B_j = \begin{bmatrix} 0 & 0 & 0 \\ 0 & n\mu & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad j = \overline{n, N};$$

$$C_j = \begin{bmatrix} 0 & \mu_1^c & 0 \\ 0 & 0 & 0 \\ 0 & \mu_2^c & 0 \end{bmatrix}, \quad D_j = \begin{bmatrix} \lambda + \mu_1^c & 0 & 0 \\ 0 & \lambda + \min(n, j)\mu & 0 \\ 0 & 0 & \lambda + \mu_2^c \end{bmatrix}, \quad j = \overline{0, N}.$$

Очевидно, матрицы A и C имеют одинаковый вид для всех ярусов диаграммы, а матрицы B и D зависят от номера яруса, но при $j > n$ стабилизируются.

Введем векторы-строки $\gamma_j = [\gamma_{j,1}, \gamma_{j,2}, \gamma_{j,3}]$ нахождения СМО в микросостояниях j -го яруса. Теперь запишем векторно-матричные уравнения баланса переходов между состояниями [2], указанными на диаграмме:

$$\gamma_0 D_0 = \gamma_0 C_0 + \gamma_1 B_1,$$

$$\gamma_j D_j = \gamma_{j,1} A_{j,1} + \gamma_j C_j + \gamma_{j+1} B_{j+1}, \quad j = \overline{1, N},$$

где N – число обсчитываемых ярусов диаграммы.

Для решения данной системы уравнений воспользуемся итерационным методом, предложенным в [8] и доработанным в [9]. Направление прогонки при расчёте определим сверху вниз – в сторону увеличения номера яруса. В результате расчета получим отношения смежных вероятностей числа заявок в системе $x_j = p_{j+1}/p_j$, $j = \overline{1, N}$.

После окончания итераций, используя значения x_j , осуществим переход к вероятностям состояний системы по следующему алгоритму:

- вероятность свободного состояния p_0 положить равной 1;
- рассчитать $p_{j+1} = p_j x_j$, $j = \overline{1, N}$;
- вычислить сумму $S = \sum_{j=0}^N p_j$;
- нормировать полученные значения: $p_j = p_j/S$, $j = \overline{0, N}$.

Расчет возможен и для систем без ограничения буфера. В этом случае предполагается, что «хвост» распределения числа заявок представляет собой бесконечно убывающую геометрическую прогрессию, знаменатель которой равен отношению двух последних рассчитанных вероятностей.

Время ожидания вновь прибывшей заявки определяется состоянием системы непосредственно перед ее прибытием. В соответствии с теоремой PASTA (Poisson Arrival See Time Average) [10, 11] рас-

пределение числа заявок перед прибытием очередной совпадает со стационарным. Перед прибытием очередной заявки система находится в одном из трех микросостояний – охлаждение 1-го или 2-го типов или обслуживание.

Пусть перед прибытием очередной заявки в системе уже находилось j заявок. Вероятности соответствующих микросостояний записаны в векторе-строке $\gamma_j = [\gamma_{j,1}, \gamma_{j,2}, \gamma_{j,3}]$. Если система находилась в первом микросостоянии j -го яруса (см. рис. 1), вновь прибывшая заявка будет ожидать завершения охлаждения первого типа плюс $(j - n + 1)$ продвижений очереди. Преобразование Лапласа–Стилтьеса (ПЛС) соответствующего времени

$$\omega_{j,1}(s) = \frac{\mu_1^c}{\mu_1^c + s} \left(\frac{n\mu}{n\mu + s} \right)^{j-n+1}.$$

В третьем микросостоянии происходит охлаждение второго типа. ПЛС соответствующего ожидания

$$\omega_{j,3}(s) = \frac{\mu_2^c}{\mu_2^c + s} \left(\frac{n\mu}{n\mu + s} \right)^{j-n+1}.$$

Второе микросостояние соответствует режиму обслуживания. ПЛС времени ожидания вновь прибывшей заявки составит

$$\omega_{j,2}(s) = \left(\frac{n\mu}{n\mu + s} \right)^{j-n+1}.$$

Введем вектор-столбец $\omega_j(s) = [\omega_{j,1}(s), \omega_{j,2}(s), \omega_{j,3}(s)]^T$. Тогда ПЛС искомого времени ожидания начала обслуживания

$$\omega(s) = \sum_{j=n}^{N-1} \gamma_j \omega_j(s). \quad (1)$$

Для получения начальных моментов времени ожидания можно выполнить дифференцирование в нуле выражения (1) либо заменить все входящие в него ПЛС наборами начальных моментов, а их произведения – свертками в моментах [2].

3. Результаты расчетов

Представленные диаграмму переходов и уравнения баланса переходов между ее микросостояниями принято считать справедливыми для системы $M/M/H_2/n$ лишь в том случае, когда параметры гиперэкспоненты вещественны. Это соответствует коэффициенту вариации времени охлаждения $v > 1$. Когда же коэффициент вариации времени охлаждения $v < 1$, подбираемые по методу моментов параметры H_2 -распределения уходят либо в комплексную, либо в «парадоксальную» область [8]. Строго говоря, в этом случае мы имеем дело уже не с гиперэкспонентой, а с некоторой трехпараметрической функцией

$$R(x) = y_1 e^{-\mu_1 x} + y_2 e^{-\mu_2 x}; \quad y_1 + y_2 = 1,$$

параметры которой могут принимать комплексные значения. Очевидно, H_2 -распределение является частным случаем функции $R(x)$ с вещественными параметрами.

Соответственно, при использовании R -аппроксимации с комплексными параметрами элементы матриц интенсивностей и параметры уравнений баланса также принимают комплексные значения. Тем не менее многочисленные вычислительные эксперименты [2, 6–8] показывают, что при последующем решении уравнений баланса указанная патология проявляется лишь в промежуточных результатах – вероятностях «фиктивных» микросостояний, на которые расщепляются «фактические» состояния СМО. На этапе суммирования вероятностей микросостояний комплексные части аннигилируются и результат расчета – распределение числа заявок в системе – становится вещественным.

Выполним расчет системы $M/M/H_2/n$ на имитационной модели и численным методом через H_2 -аппроксимацию для следующего набора исходных данных: число каналов обслуживания $n = 3$, средняя интенсивность входящего потока $\lambda = 2,5$, среднее время обслуживания заявок $b_1 = 1$, среднее время

охлаждения системы $c_1 = 2$. Параметры H_2 -распределения подбирались по трем моментам различных исходных распределений длительности охлаждения:

- вырожденного D (коэффициент вариации $v = 0$);
- равномерного U на интервале $[0; 4]$ ($v \approx 0,577$);
- экспоненциального M ($v = 1$);
- гамма Γ с параметром формы 0,5 ($v \approx 1,41$);
- Вейбулла W с параметром формы 0,46 ($v = 2,5$).

Результаты расчета распределения числа заявок в системе для D , M и W распределений приведены на рис. 2. Штриховкой показаны графики, полученные с помощью имитационной модели, сплошными линиями – на основе численного метода расчета.

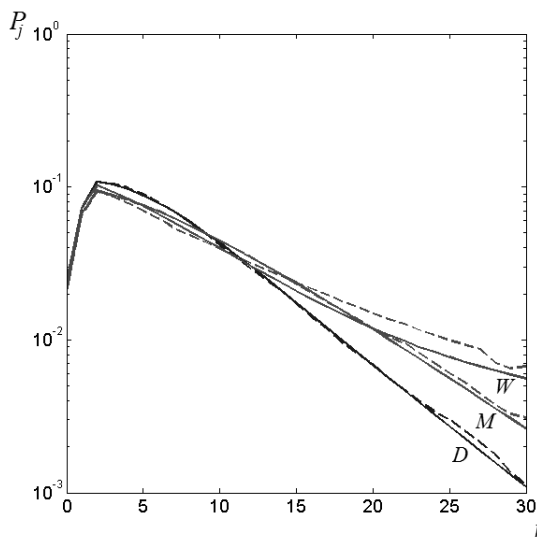


Рис 2. Распределение числа заявок в системе с охлаждением $M/M/H_2/n$

Из графиков видно согласие результатов даже в области комплексных и парадоксальных параметров гиперэкспоненты. Расстояние Колмогорова между результатами, полученными методом имитационного моделирования и с помощью R -аппроксимации D , U , M , Γ и W распределений времени охлаждения, составило $\{0,006; 0,005; 0,006; 0,004; 0,043\}$ соответственно, что говорит о приемлемой точности R -аппроксимации. Заметим, что для D , U распределений параметры R -функции принимали комплексные значения, что, тем не менее, не повлияло на корректность конечного результата.

Увеличение расстояния Колмогорова для W -распределения с параметром формы 0,46 может быть обусловлено тем, что оно относится к классу распределений с «толстым хвостом» и, следовательно, описываемая им случайная величина имеет очень большой разброс. В этой связи возрастают требования к качеству используемых в имитационной модели программных датчиков случайных чисел.

В таблице представлены результаты расчета начальных моментов распределения времени ожидания $\{w_i\}$, $i = \overline{1, 3}$, полученные численно (числ.) и с помощью имитационной модели (ИМ).

Начальные моменты времени ожидания

$B(t)$	w_1		w_2		w_3	
	Числ.	ИМ	Числ.	ИМ	Числ.	ИМ
D	1,788	1,709	6,674	6,495	37,19	34,16
U	1,971	1,870	7,864	7,230	40,46	37,22
M	2,272	2,225	1,035	1,035	62,88	64,02
$\Gamma_{0,5}$	2,582	2,494	1,406	1,356	108,0	103,9
W	2,600	2,802	2,586	2,476	767,0	687,7

Из представленных результатов следует, что с увеличением коэффициента вариации распределения времени «охлаждения» среднее время ожидания возрастает. Сходство результатов, полученных численно и с помощью имитационного моделирования, позволяет сделать вывод о корректности предложенной методики численного расчета вероятностно-временных характеристик СМО с «охлаждением».

Заключение

Представленные численные результаты свидетельствуют о существенном влиянии коэффициента вариации времени охлаждения на вероятностно-временные характеристики СМО с «охлаждением». Используемая для аппроксимации немарковской компоненты СМО R -функция позволяет с приемлемой точностью проводить анализ систем массового обслуживания с произвольным коэффициентом вариации распределения времени охлаждения.

Областью потенциальных приложений рассматриваемого класса СМО являются случаи временных задержек, наступающих после периода непрерывной занятости и связанных, например, со следующими причинами:

- восстановлением работоспособности системы;
- проведением технического обслуживания;
- восстановлением ресурсов, расходуемых при обслуживании;
- необходимостью передачи информации о количестве обслуженных за период непрерывной занятости заявок.

Дальнейшим направлением развития предложенной методики расчета может быть ее расширение на случай немарковского обслуживания и (или) рекуррентного входящего потока заявок, а также расчет сети массового обслуживания, узлы которой представляют собой СМО с «охлаждением».

ЛИТЕРАТУРА

1. Гиндин С.И., Хомоненко А.Д., Ададуров С.Е. Численный расчет многоканальной системы массового обслуживания с рекуррентным входящим потоком и «разогревом» // Известия Петербургского университета путей сообщения. 2013. № 4 (37). С. 92–101.
2. Рыжиков Ю.И. Алгоритмический подход к задачам массового обслуживания. СПб. : ВКА им. А.Ф. Можайского, 2013. 496 с.
3. Хомоненко А.Д., Халиль М.М., Гиндин С.И. Моделирование облачных вычислений с использованием многоканальной системы массового обслуживания с «охлаждением» // Материалы XIX Международной конференции по мягким вычислениям и измерениям (SCM-2016). 25–27 мая 2016 г. СПб. : Изд-во СПбГЭТУ «ЛЭТИ», 2016. Т. 1. С. 247–251.
4. Кендалл М.Дж., Стюарт А. Теория распределений: пер. с англ. М. : Наука, 1966. 587 с.
5. Cox D.R. A Use of Complex Probabilities in the Theory of Stochastic Process // Proc. of the Cambridge Phil. Soc. 1955. P. 313–323.
6. Рыжиков Ю.И., Уланов А.В. Расчет гиперэкспоненциальной системы $M/H_2/n-H_2$ с заявками, нетерпеливыми в очереди // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 2 (27). С. 47–53.
7. Рыжиков Ю.И., Лохвицкий В.А. Расчёт узла сети с отрицательными заявками // Сборник трудов по материалам российской конференции «Информационные технологии в управлении» (ИТУ-2012). СПб., 2012. С. 354–359.
8. Рыжиков Ю.И., Уланов А.В. Применение гиперэкспоненциальной аппроксимации в задачах расчета немарковских систем массового обслуживания // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2016. № 3 (36). С. 60–65.
9. Takahashi Y., Takami Y. A Numerical Method for the Steady-State Probabilities of a GI/G/c Queuing System in a General Class // J. of the Operat. Res. Soc. of Japan. 1976. V. 19, No. 2. P. 147–157.
10. Рыжиков Ю.И., Хомоненко А.Д. Итерационный метод расчёта многоканальных систем с произвольным законом обслуживания // Проблемы управления и теории информации. 1980. Т. 9, № 3. С. 32–38.
11. Serfozo R. Basics of Applied Stochastic Processes. Berlin : Springer, 2009. 443 p.

Lokhvitsky Vladimir A., Ulanov Alexander V. (Mozhaisky Military Space Academy, Russian Federation, St. Petersburg).

The numerical analyses of queuing system with hyperexponential distribution of cooling time.

Keywords: queuing systems with cooling; non-markovian queuing systems; hyperexponential distribution; numerical methods.

DOI: 10.17223/19988605/37/4

The numerical analysis of QS with cooling is considered. Cooling is understood as the temporary delays coming after the period of continuous employment. In practice such delays arise when after service of the last request in system carrying out the actions connected with maintenance, carrying out maintenance, breaks in work, etc. is required. In this case pertinently to say that QS passes into the cooling mode. The beginning of service of again arrived request won't begin, all operations of cooling won't come to the end yet. Unlike a warming up, process of cooling of system doesn't depend on arrival of the first request of the period of employment. If the system manages to be cooled before arrival of this request, service will begin without additional delays.

In this paper is presented the technique of the numerical analysis of multichannel QS with the Poisson arrival, exponential service and not-exponential distribution of cooling time.

For approximation of cooling time it is offered to use the function $R(x)$ which is formally coinciding with function of the hyperexponential distribution of the second order (H_2), but allowing unlike hyper exhibitors complex type of parameters

$$R(x) = y_1 e^{-\mu_1 x} + y_2 e^{-\mu_2 x}; y_1 + y_2 = 1.$$

Considering that prototype of function $R(x)$ – the hyperexponential H_2 distribution belongs to distributions of phase type, conditions of non-Markov QS and transitions between them can be presented in the form of discrete Markov process with continuous time. Possibilities of R -approximation in that case, when coefficient of a variation of initial distribution $v < 1$ are shown in article. Function parameters thus accept complex values. Nevertheless, at calculation of QS with application of R -approximation in the field of complex values this pathology is shown only in intermediate results – probabilities of fictitious microstates of the chart of transitions on which physical conditions of QS are split. At a stage of summation of probabilities of microstates of circles their complex parts are annihilated also result of calculation – probability of number of requests in system – becomes material. Distribution of number of applications and Laplace transformation of a waiting time is received.

In work the results of calculation of QS received by means of R -approximation and by means of imitating modeling are compared. Various initial distributions of cooling – determined uniform, scale with form parameter 0,5, exponential and Weibull with form parameter 0,46 are considered. It is shown that the accuracy of calculation of stationary distribution of number of requests and the first three moments waiting time is well. The maximum distance of Kolmogorov (for Weibull distribution) made 0,043. It is shown that when coefficient of a variation of cooling time distribution is increased, the average time of expectation is increased too.

Scope of the offered calculation procedure is the analysis of functioning of organizational and technical systems with queues in which after end of the period of continuous employment there comes the temporary delay connected with the following reasons:

- maintenance of system;
- carrying out maintenance;
- restoration of the material resources spent at service;
- need of transfer to a higher body of information directorate about number of the requests served during continuous employment.

Its expansion on a case of non-Markov service and the recurrent arrival, and also calculation of a queuing network which nodes is QS with "cooling" can be the further direction of development of the offered calculation procedure.

REFERENCES

1. Khomonenko, A.D., Adadurov, S.E. & Gindin, S.I. (2013) Numerical calculations of multichannel queuing system with recurrent input and "warm up". *Izvestiya Peterburgskogo universiteta putey soobscheniya – Proceedings of Petersburg Transport University*. 4(37). pp. 92-101. (In Russian).
2. Ryzhikov, Yu.I. (2013) *Algoritmicheskiy podkhod k zadacham massovogo obsluzhivaniya* [A use of algorithm approach in the queuing theory]. St. Petersburg: Mozhaisk Military Space Academy.
3. Khomonenko, A.D., Khalil, M.M. & Gindin, S.I. (2016) [Modelling of cloud computing using multichannel queuing system with a cooling]. *Proc. of The IXI Intern. Conference on Soft Computing and Measurements. St. Petersburg. SCM'2016*. St. Petersburg. May 25–27, 2016. pp. 247-251. (In Russian).
4. Kendall, M. & Stewart, A. (1966) *Teoriya raspredeleniy* [The advanced theory of statistics. Distribution theory]. Translated from English. Moscow: Nauka.
5. Cox, D.R. (1955) A Use of Complex Probabilities in the Theory of Stochastic Process. *Proc. of the Cambridge Phil. Soc.* pp. 313-323. DOI: 10.1017/S0305004100030231

6. Ryzhikov, Yu.I. & Ulanov, A.V. (2014) The method of calculating M/H₂/n-H₂ queuing system with impatient customers. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(27). pp. 47-53. (In Russian).
7. Ryzhikov, Yu.I. & Lokhvitsky, V.A. (2012) [The method of calculating of G-network]. *Informatsionnye tekhnologii v upravlenii* [Information technology in management]. Proc. of the Conference. St. Petersburg. pp. 354-359. (In Russian).
8. Ryzhikov, Yu.I. & Ulanov, A.V. (2016) A use of hyperexponential distribution in non-markovian queuing systems analyses. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(36). pp. 60-65. (In Russian). DOI: 10.17223/19988605/36/6.
9. Takahashi, Y. & Takami, Y. (1976) A Numerical Method for the Steady-State Probabilities of a GI/G/c Queuing System in a General Class. *J. of the Operat. Res. Soc. of Japan*. 19(2). pp. 147-157.
10. Ryzhikov, Yu.I. & Khomonenko, A.D. (1980) Iteratsionnyy metod rascheta mnogokanal'nykh sistem s proizvol'nym zakonom obsluzhivaniya [Iterative methods of multichannel queuing system calculation with arbitrary service time distribution]. *Problemy upravleniya i teorii informatsii – Problems of Control and Information Theory*. 9(3). pp. 203-213.
11. Serfozo, R. (2009) *Basics of Applied Stochastic Processes*. Berlin: Springer.

УДК 336:51

DOI: 10.17223/19988605/37/5

Г.А. Медведев, Д.А. Павлив

О КВАДРАТИЧНЫХ МОДЕЛЯХ ДОХОДНОСТИ В РИСК-НЕЙТРАЛЬНОЙ СРЕДЕ

Представлено математически эквивалентное, но более компактное описание обычно встречающейся квадратичной модели доходности. Получены уравнения для функций временной структуры и приведены общие свойства их решений. Основным содержанием статьи является рассмотрение случая, когда вероятностные свойства модели подчиняются нейтральной к риску вероятностной мере, что позволяет получить аналитические решения задачи о форме кривых доходности и форвардных кривых в явном виде. Уравнения кривых представлены в компактной форме с использованием гиперболических функций. Найдены семейства кривых доходности и форвардных кривых, которые для фиксированной краткосрочной ставки доходности обеспечивают получение одной и той же предельной долгосрочной доходности. Полученные результаты иллюстрируются численным примером.

Ключевые слова: временные структуры доходности; квадратичные модели; нейтральная к риску вероятностная мера.

Обычно отправным пунктом при анализе квадратичных моделей временной структуры являются следующие предположения. Считается, что существует некоторое пространство состояний финансового рынка, описываемое n -мерным вектором x , в котором состояние эволюционирует согласно диффузионному случайному процессу $x(t)$, описываемому стохастическим дифференциальным уравнением

$$dx(t) = \xi(x(t)) dt + \sigma(x(t)) dW(t), \quad t > t_0, \quad x(t_0) = x_0,$$

с векторной функцией дрейфа $\xi(x)$ и матрицей волатильности $\sigma(x)$, $W(t)$ – n -мерный вектор независимых винеровских процессов. Предполагается, что краткосрочная ставка доходности $r(t)$ связана с состоянием финансового рынка $x(t)$ соотношением (верхний индекс T означает транспонирование):

$$r(t) = \alpha + x(t)^T \Psi x(t) + x(t)^T \psi, \quad x(t) \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}, \quad \Psi \in \mathbb{R}^{n \times n}, \quad \psi \in \mathbb{R}^{n \times 1}. \quad (1)$$

При этом считается, что состояние финансового рынка (вектор $x(t)$) непосредственно не наблюдается, а процентная ставка $r(t)$ – наблюдаемый процесс. Так что $x(t)$ – совокупность неких латентных переменных, которые сами не наблюдаются, но влияют на временную структуру доходности активов, стоимость которых определяется процентной ставкой. Заметим, что поскольку по экономическому смыслу процентная ставка $r(t)$ принимает положительные значения для любых ненулевых $x(t)$, то естественно считать матрицу Ψ положительно определенной. Кроме того, поскольку $x(t)^T \Psi x(t) = x(t)^T \Psi^T x(t)$, для любых $x(t)$, не нарушая общности, можно считать матрицу Ψ также симметрической. Заметим, что согласно (1) процентная ставка принимает свое минимальное значение, когда $2\Psi x(t) + \psi = 0$, т.е. при $x(t) = x_{\min} \equiv -\Psi^{-1}\psi/2$, и минимальное значение процентной ставки в этой модели равно $r_{\min} \equiv \alpha - \psi^T \Psi^{-1} \psi/4$.

Модель (1)–(2), называемая квадратичной, составляет некоторую конкуренцию аффинным моделям временной структуры доходности, которые допускают аналитическое представление решений, но довольно неточно описывают реальные временные структуры доходности. Считается, что квадратичная модель будет приспособливаться к реальным временным структурам более точно. Попытаемся рассмотреть этот вопрос для класса диффузионных процессов $x(t)$ с линейной функцией дрейфа $\xi(x) = k(\vartheta - x)$ и постоянной матрицей волатильности $\sigma(x) = \sigma$ [1]:

$$dx(t) = k(\vartheta - x(t)) dt + \sigma dW(t), \quad t > t_0, \quad x(t_0) = x_0. \quad (2)$$

Как это принято, временную структуру доходности находят с помощью уравнения временной структуры [2], или, как его иногда называют [1], фундаментального уравнения с частными производными для цен бескупонных облигаций $P(r(x), t; T)$:

$$\frac{\partial P}{\partial t} + \frac{\partial P}{\partial x} k(\vartheta - x) + \frac{1}{2} \text{tr} \left(\frac{\partial^2 P}{\partial x^2} \sigma \sigma^T \right) - r(x) P(r(x), t; T) = \frac{\partial P}{\partial x} \sigma \lambda(x), \quad P(r(x), t; T) = 1. \quad (3)$$

Здесь T – дата погашения облигации, $\lambda(x)$ – вектор-функция рыночной цены риска. Здесь цена облигации $P(r(x), t; T)$ определяется как условное математическое ожидание по объективной вероятностной мере при фиксированном текущем состоянии, т.е. при фиксированных t и $r(x)$, а следовательно, и $x(t)$. Рыночная цена риска задается вектором [1]: $\lambda(x) = \lambda_0 + \lambda_1 x$, где λ_0 и λ_1 – вектор и матрица постоянных коэффициентов. Решение уравнения (3) имеет достаточно простой вид

$$P(r(x), t; t + \tau) = \exp[-x^T a(\tau)x - x^T b(\tau) - c(\tau)], \quad (4)$$

где для удобства введен символ $\tau = T - t$ для обозначения срока до погашения облигации, а также $(n \times n)$ матрица $a(\tau)$, n -вектор $b(\tau)$ и скалярная функция $c(\tau)$. Заметим, что с возрастанием компонент вектора x (точнее, с возрастанием $x^T \Psi x$) процентная ставка r увеличивается. С другой стороны, по экономическому смыслу с увеличением процентной ставки цена актива $P(r(x), t; t + \tau)$ должна уменьшаться. Следовательно, с увеличением $x^T \Psi x$ величина $x^T a(\tau)x$ тоже должна увеличиваться. Кроме того, для любых x справедливо равенство $x^T a(\tau)x = x^T a(\tau)^T x$. Поэтому без ограничения общности можно считать, что матрица $a(\tau)$ является положительно определенной и симметрической. Функции $a(\tau)$, $b(\tau)$ и $c(\tau)$ находятся из следующей системы обыкновенных нелинейных дифференциальных уравнений

$$\begin{aligned} \frac{da(\tau)}{d\tau} + 2a(\tau) \sigma \sigma^T a(\tau) + a(\tau)(k + \sigma \lambda_1) + (k + \sigma \lambda_1)^T a(\tau) - \Psi &= 0, \\ \frac{db(\tau)}{d\tau} + 2a(\tau) \sigma \sigma^T b(\tau) + (k + \sigma \lambda_1)^T b(\tau) - 2a(\tau)(k\vartheta - \sigma \lambda_0) - \psi &= 0, \\ \frac{dc(\tau)}{d\tau} - \text{tr}[\sigma \sigma^T a(\tau)] + \frac{1}{2} b(\tau)^T \sigma \sigma^T b(\tau) - b(\tau)^T (k\vartheta - \sigma \lambda_0) - \alpha &= 0 \end{aligned} \quad (5)$$

с начальными условиями $a(0) = 0$, $b(0) = 0$, $c(0) = 0$.

Чаще всего для представления временной структуры процентных ставок доходности используют кривую доходности $y(\tau, x)$ или форвардную кривую $f(\tau, x)$, которые, согласно определениям, выражаются через цену бескупонной облигации по формулам

$$y(\tau, x) = -\frac{\ln P(x, t, t + \tau)}{\tau} = \frac{x^T a(\tau)x + b(\tau)^T x + c(\tau)}{\tau}, \quad (6)$$

$$f(\tau, x) = -\frac{\partial \ln P(x, t, t + \tau)}{\partial \tau} = x^T \frac{da(\tau)}{d\tau} x + \left(\frac{db(\tau)}{d\tau} \right)^T x + \frac{dc(\tau)}{d\tau}. \quad (7)$$

К сожалению, система уравнений (5) в общем случае не имеет аналитического решения, поэтому особенности кривых доходности $y(\tau, x)$ и $f(\tau, x)$ не удастся исследовать аналитическим путем, хотя численный анализ не представляет особого труда, если заданы параметры k , ϑ , σ , α , ψ , Ψ , λ_0 и λ_1 . Тем не менее в некоторых частных случаях найти аналитические решения и провести некоторый анализ удастся. Рассмотрим эти случаи.

Однако сначала представим математически эквивалентное, но более компактное описание рассматриваемой квадратичной модели. Поскольку матрица Ψ положительно определенная и симметрическая, ее собственные числа положительные, а она имеет диагональную форму $\Psi = L^T \Phi L$, где Φ – диагональная матрица, по диагонали которой расположены положительные собственные числа $\{\phi\}$, L – матрица нормированных собственных векторов, которая обладает следующими свойствами: $L^T = L^{-1}$, $L^T L = L L^T = I$, где I – единичная матрица. Преобразуем пространство состояний $\{x\}$ в пространство состояний $\{X\}$ с помощью линейного преобразования $X = Lx - L\Psi^{-1}\psi/2$. Тогда вместо модели процентной ставки (1)–(2) можно получить эквивалентное описание процентной ставки соотношениями

$$dX(t) = K(\theta - X(t)) dt + S dW(t), \quad t > t_0, \quad X(t_0) = X_0, \quad (8)$$

$$r(t) = r_{\min} + X(t)^T \Phi X(t), \quad X(t) \in \mathbb{R}^n, \quad r \in \mathbb{R}^1, \quad \Phi \in \mathbb{R}^{n \times n}, \quad (9)$$

где обозначено $K = LkL^T$, $\theta = L\vartheta + L\Psi^{-1}\psi/2$, $S = L\sigma$, $X_0 = Lx_0 - L\Psi^{-1}\psi/2$. Модель (8)–(9) отличается от модели (1)–(2) более простым, но эквивалентным представлением процентной ставки $r(t)$ через другие ла-

тентные переменные $\{X\}$, взаимно однозначно связанные с исходными латентными переменными $\{x\}$. Свойства диффузионных процессов $X(t)$ и $x(t)$ совпадают с точностью до значений факторов уравнений (2) и (8). Модель (8)–(9) приводит к уравнению (3) для функции $P(r(X), t; t + \tau)$, в котором вместо k , ϑ и σ используются K , θ и S соответственно, а также вместо $\lambda(x)$ – функция рыночной цены риска $\Lambda(X) = \Lambda_0 + \Lambda_1 X$, где вектор Λ_0 и матрица Λ_1 постоянных коэффициентов определяются формулами $\Lambda_0 = \lambda_0 + \lambda_1 \Psi^{-1} \psi / 2$, $\Lambda_1 = \lambda_1 L^T$. Решение этого уравнения имеет вид, аналогичный выражению (4):

$$P(r(X), t; t + \tau) = \exp[-X^T A(\tau) X - X^T B(\tau) - C(\tau)], \quad (10)$$

где функции $A(\tau)$, $B(\tau)$ и $C(\tau)$ находятся из системы

$$\begin{aligned} \frac{dA(\tau)}{d\tau} + 2A(\tau)SS^T A(\tau) + A(\tau)(K + S\Lambda_1) + (K + S\Lambda_1)^T A(\tau) - \Phi &= 0, \\ \frac{dB(\tau)}{d\tau} + 2A(\tau)SS^T B(\tau) + (K + S\Lambda_1)^T B(\tau) - 2A(\tau)(K\theta - S\Lambda_0) &= 0, \\ \frac{dC(\tau)}{d\tau} - \text{tr}[SS^T A(\tau)] + \frac{1}{2}B(\tau)^T SS^T B(\tau) - B(\tau)^T (K\theta - S\Lambda_0) - r_{\min} &= 0 \end{aligned} \quad (11)$$

с начальными условиями $A(0) = 0$, $B(0) = 0$, $C(0) = 0$. Эта система, так же как и система (5), в общем случае в аналитическом виде неразрешима. Поэтому прежде чем переходить к частным случаям, которые разрешимы, попытаемся обсудить в общих чертах свойства решений системы (11), если они существуют.

В литературе достаточно много внимания уделено свойствам кривых доходности и форвардных кривых для аффинных моделей временных структур [3–9]. В этом классе моделей кривые $y(\tau)$ и $f(\tau)$ стартуют из одной точки (это общее свойство по определению этих кривых, оно основано на том, что краткосрочная доходность равна процентной ставке r) и при $\tau \rightarrow \infty$ стремятся к одному и тому же постоянному пределу. Достаточным условием последнего может служить тот факт, что при $\tau \rightarrow \infty$ функции $A(\tau)$ и $B(\tau)$ имеют конечные пределы $A(\infty)$ и $B(\infty)$, что обеспечивает независимость предельных значений $y(\infty)$ и $f(\infty)$ от текущих (начальных) значений переменных состояния рынка $X(t)$ и обеспечивает предельное равенство

$$\lim_{\tau \rightarrow \infty} \frac{dA(\tau)}{d\tau} = \lim_{\tau \rightarrow \infty} \frac{A(\tau)}{\tau} = 0, \quad \lim_{\tau \rightarrow \infty} \frac{dB(\tau)}{d\tau} = \lim_{\tau \rightarrow \infty} \frac{B(\tau)}{\tau} = 0.$$

При этом предельные значения $A = A(\infty)$ и $B = B(\infty)$ могут быть найдены из системы алгебраических матричных уравнений

$$\begin{aligned} 2A SS^T A + A(K + S\Lambda_1) + (K + S\Lambda_1)^T A &= \Phi, \\ 2A SS^T B + (K + S\Lambda_1)^T B - 2A(K\theta - S\Lambda_0) &= 0. \end{aligned}$$

Наконец, предельное свойство функции $C(\tau)$ находится из третьего уравнения системы

$$\lim_{\tau \rightarrow \infty} \frac{dC(\tau)}{d\tau} = \lim_{\tau \rightarrow \infty} \frac{C(\tau)}{\tau} = \text{tr}[SS^T A] - \frac{1}{2}B^T SS^T B + B^T (K\theta - S\Lambda_0) + r_{\min}.$$

Используя перечисленные свойства функций временной структуры $A(\tau)$, $B(\tau)$ и $C(\tau)$, можно установить соответствующие свойства кривых $y(\tau)$ и $f(\tau)$: при $\tau = 0$ они стартуют из одной общей точки $y(0) = f(0) = r(X)$ и при $\tau \rightarrow \infty$ стремятся к одному общему пределу

$$y(\infty) = f(\infty) = \text{tr}[SS^T A(\infty)] - \frac{1}{2}B(\infty)^T SS^T B(\infty) + B(\infty)^T (K\theta - S\Lambda_0) + r_{\min}.$$

Характер поведения кривых между этими предельными точками существенно зависит от стартового значения процентной ставки $r(X)$, как это имеет место в аффинных моделях [3–10]. Однако поскольку некоторому фиксированному значению ставки $r(X) = r_0$ соответствует некоторое множество векторов X , вполне возможна зависимость кривых и от величины векторов из этого множества. Для того чтобы выяснить этот вопрос, рассмотрим частный случай, допускающий аналитическое решение системы (11).

Сделаем следующие предположения:

Предположение 1. Процесс латентных переменных $X(t)$, порождаемый уравнением (8), является нормальным процессом со стационарным математическим ожиданием θ . Предположим, что $\theta = 0$. Заметим, что такое предположение принимается довольно часто, например в моделях [11, 12].

Предположение 2. Рассмотрим случай, когда вероятностные свойства процентной ставки $r(X)$ подчиняются нейтральной к риску вероятностной мере. Это означает, что средняя ставка доходности актива совпадает с краткосрочной процентной ставкой, иначе говоря, пусть $\Lambda(X) = \Lambda_0 + \Lambda_1 X = 0$, т.е. $\Lambda_0 = 0$ и $\Lambda_1 = 0$. Такое предположение упрощает анализ, но идеализирует рыночные отношения. Получающиеся при этом результаты отражают основные закономерности, а введение в расчеты рыночных цен риска обычно только изменяет численные значения, не меняя закономерностей (см., например рис. 3.6, 3.7, 3.8 в [10] для аффинных моделей).

При этих предположениях система (11) упрощается и имеет вид

$$\frac{dA(\tau)}{d\tau} + 2A(\tau)SS^T A(\tau) + A(\tau)K + K^T A(\tau) - \Phi = 0, \quad A(0) = 0,$$

$$\frac{dB(\tau)}{d\tau} + 2A(\tau)SS^T B(\tau) + K^T B(\tau) = 0, \quad B(0) = 0,$$

$$\frac{dC(\tau)}{d\tau} - \text{tr}[SS^T A(\tau)] + \frac{1}{2}B(\tau)^T SS^T B(\tau) - r_{\min} = 0, \quad C(0) = 0.$$

Выясняется, что уравнение для $B(\tau)$ оказывается линейным однородным дифференциальным уравнением с нулевыми начальными условиями, а это приводит к тому, что функция $B(\tau)$ равна нулю для всех значений $\tau > 0$. Таким образом, система (11) становится еще проще:

$$\frac{dA(\tau)}{d\tau} = -2A(\tau)SS^T A(\tau) - A(\tau)K - K^T A(\tau) + \Phi, \quad A(0) = 0, \quad (12)$$

$$\frac{dC(\tau)}{d\tau} = \text{tr}[SS^T A(\tau)] + r_{\min}, \quad C(0) = 0. \quad (13)$$

При этом собственно уравнением является только первое для матрицы $A(\tau)$, а второе – это формула для производной функции $C(\tau)$.

Предположение 3. Латентные переменные, составляющие вектор $X(t)$, являются независимыми случайными процессами. Это достигается в том случае, когда матрицы K и S являются положительно определенными диагональными. Такое предположение вовсе не означает, что исходные переменные состояния в уравнении (2) независимы, поскольку диагональность матриц K и S в уравнении (8) не означает, что матрицы k и σ в уравнении (2) являются диагональными. Они связаны соотношениями $K = LkL^T$, $S = L\sigma$, где L – матрица нормированных собственных векторов, которая обычно диагональной не является.

Пусть k_i – элемент матрицы K , стоящий на i -м месте главной диагонали, s_i – элемент матрицы S , стоящий на i -м месте главной диагонали, ϕ_i – i -е собственное число матрицы Ψ , занимающее i -е место главной диагонали матрицы Φ . Если использовать предположение 3 в матричном уравнении (12) и рассмотреть систему скалярных дифференциальных уравнений для элементов матрицы $A(\tau)$, то можно заметить, что для каждого элемента $A_{jk}(\tau)$, $j \neq k$, получается однородное линейное дифференциальное уравнение с нулевым начальным условием. Это означает, что решения всех этих уравнений нулевые, т.е. $A_{jk}(\tau) = 0$, $j \neq k$, для всех $\tau > 0$. Учитывая этот факт можно получить для каждого элемента матрицы $A(\tau)$, стоящего на главной диагонали, уравнение в следующем виде:

$$\frac{dA_{ii}(\tau)}{d\tau} = -2s_i^2 A_{ii}(\tau)^2 - 2k_i A_{ii}(\tau) + \phi_i, \quad A_{ii}(0) = 0, \quad 1 \leq i \leq n. \quad (14)$$

Таким образом, при сделанных предположениях матричное уравнение (12) распадается на n независимых уравнений Риккати (14) для диагональных элементов матрицы $A(\tau)$, а недиагональные элементы оказываются нулевыми, т.е. матрица $A(\tau)$ – диагональная. Поскольку коэффициенты уравнения (14) являются константами, используя разделение переменных, можно привести это уравнение к симметричной форме (для краткости индексы здесь опускаем):

$$\frac{dA}{2s^2 A^2 + 2kA - \phi} = -d\tau,$$

что с учетом начального условия $A(0) = 0$ дает следующее решение:

$$A_{ii}(\tau) = \frac{\phi_i [1 - \exp(-2\tau\sqrt{k_i^2 + 2s_i^2\phi_i})]}{k_i + \sqrt{k_i^2 + 2s_i^2\phi_i} + (\sqrt{k_i^2 + 2s_i^2\phi_i} - k_i) \exp(-2\tau\sqrt{k_i^2 + 2s_i^2\phi_i})}, \quad 1 \leq i \leq n. \quad (15)$$

Таким образом, для положительных k_i , s_i и ϕ_i функция $A_{ii}(\tau)$ является монотонно возрастающей функцией от $A_{ii}(0) = 0$ при $\tau = 0$ до конечного предела

$$A_{ii}(\infty) = \frac{\phi_i}{k_i + \sqrt{k_i^2 + 2s_i^2\phi_i}}, \quad 1 \leq i \leq n,$$

при $\tau \rightarrow \infty$. В дальнейшем для краткости будем обозначать $v_i \equiv \sqrt{k_i^2 + 2s_i^2\phi_i}$. Для функции $C(\tau)$ из равенства (13) получаем соотношение

$$\frac{dC(\tau)}{d\tau} = r_{\min} + \sum_{i=1}^n \frac{s_i^2 \phi_i [1 - \exp(-2v_i \tau)]}{k_i + v_i + (v_i - k_i) \exp(-2v_i \tau)}. \quad (16)$$

Поэтому

$$C(\tau) = \tau r_{\min} + \sum_{i=1}^n \frac{1}{2} \left(\ln \left(\frac{1 + \omega_i \exp(2v_i \tau)}{1 + \omega_i} \right) - (v_i + k_i) \tau \right), \quad \omega_i \equiv \frac{v_i + k_i}{v_i - k_i}. \quad (17)$$

Таким образом, в рассматриваемом иллюстративном частном случае кривую доходности $y(\tau, X)$ и форвардную кривую $f(\tau, X)$ можно выразить в явной аналитической форме

$$y(\tau, X) = \frac{1}{\tau} \left(\sum_{i=1}^n A_{ii}(\tau) X_i^2 + C(\tau) \right),$$

где функции $A_{ii}(\tau)$ и $C(\tau)$ вычисляются по формулам (15) и (17),

$$f(\tau, X) = \sum_{i=1}^n \frac{dA_{ii}(\tau)}{d\tau} X_i^2 + \frac{dC(\tau)}{d\tau},$$

а производные функций $A_{ii}(\tau)$ и $C(\tau)$ вычисляются по формулам (14), (15) и (16).

Кривые доходности $y(\tau, X)$ и $f(\tau, X)$ определяются выражениями, из которых видно, что кривые зависят не от компонент вектора X , а от их квадратов. Кроме того они явно зависят не от собственных чисел $\{\phi_i\}$, а от чисел $\{v_i\}$, взаимнооднозначно связанных с ϕ_i . Поэтому для удобства рассуждений в дальнейшем введем в рассмотрение вместо переменных $\{X_i^2, \phi_i\}$ переменные $\{z_i, v_i\}$ согласно формулам преобразования

$$z_i \equiv \phi_i X_i^2 \geq 0, \quad v_i \equiv \sqrt{k_i^2 + 2s_i^2\phi_i} \geq k_i, \quad 1 \leq i \leq n.$$

$$X_i^2 \equiv \frac{z_i}{\phi_i}, \quad \phi_i \equiv \frac{v_i^2 - k_i^2}{2s_i^2} \geq 0, \quad 1 \leq i \leq n.$$

Равенства (15)–(17) можно представить в более компактной форме с помощью гиперболических функций, что дает более компактные выражения для кривых $y(\tau, X)$ и $f(\tau, X)$ в виде

$$Y(\tau|z, \phi) = r_{\min} + \sum_{i=1}^n \frac{1}{2\tau} \left(\frac{2z_i}{v_i \operatorname{cth}(v_i \tau) + k_i} + \ln \left(\operatorname{ch}(v_i \tau) + \frac{k_i}{v_i} \operatorname{sh}(v_i \tau) \right) - k_i \tau \right), \quad (18)$$

$$F(\tau|z, \phi) = r_{\min} + \sum_{i=1}^n \left(\frac{v_i^2 z_i}{(v_i \operatorname{ch}(v_i \tau) + k_i \operatorname{sh}(v_i \tau))^2} + \frac{1}{2} \frac{v_i^2 - k_i^2}{v_i \operatorname{cth}(v_i \tau) + k_i} \right). \quad (19)$$

Используя выражения (9), (14)–(17), можно убедиться, что имеют место следующие предельные соотношения:

$$\lim_{\tau \rightarrow 0} \frac{dA_{ii}(\tau)}{d\tau} = \lim_{\tau \rightarrow 0} \frac{A_{ii}(\tau)}{\tau} = \phi_i, \quad \lim_{\tau \rightarrow 0} \frac{dC(\tau)}{d\tau} = \lim_{\tau \rightarrow 0} \frac{C(\tau)}{\tau} = r_{\min},$$

$$\lim_{\tau \rightarrow 0} y(\tau, X) = \lim_{\tau \rightarrow 0} f(\tau, X) = r = r_{\min} + \sum_{i=1}^n \phi_i X_i^2 = r_{\min} + \sum_{i=1}^n z_i; \quad (20)$$

$$\begin{aligned} \lim_{\tau \rightarrow \infty} \frac{dA_{ii}(\tau)}{d\tau} &= \lim_{\tau \rightarrow \infty} \frac{A_{ii}(\tau)}{\tau} = 0, \\ \lim_{\tau \rightarrow \infty} \frac{dC(\tau)}{d\tau} &= \lim_{\tau \rightarrow \infty} \frac{C(\tau)}{\tau} = r_{\min} + \sum_{i=1}^n \frac{s_i^2 \phi_i}{k_i + v_i} = r_{\min} + \sum_{i=1}^n \frac{v_i - k_i}{2}; \\ \lim_{\tau \rightarrow \infty} y(\tau, X) &= \lim_{\tau \rightarrow \infty} f(\tau, X) = r_{\min} + \sum_{i=1}^n \frac{v_i - k_i}{2} \equiv y(\infty). \end{aligned} \quad (21)$$

Таким образом, обе кривые $Y(\tau|z, \phi)$ и $F(\tau|z, \phi)$ при изменении τ от нуля до ∞ , стартуя от значения процентной ставки r , определяемой текущими значениями переменных состояния, стремятся к общему пределу $y(\infty)$, не зависящему от текущего состояния, а зависящему только от параметров модели.

Заметим, что функции (18) и (19) для малых τ имеют следующие производные:

$$\left. \frac{dY(\tau|\phi, X)}{d\tau} \right|_{\tau=0} = \frac{1}{2} \left. \frac{dF(\tau|\phi, X)}{d\tau} \right|_{\tau=0} = \frac{1}{2} \sum_{i=1}^n \phi_i (s_i^2 - 2k_i X_i^2) = \frac{1}{4} \sum_{i=1}^n (v_i^2 - k_i^2 - 4k_i z_i^2).$$

Это говорит о том, что при малых τ , во-первых, форвардная кривая изменяется в два раза быстрее, чем кривая доходности, и, во-вторых, вектор латентных переменных X может определять знак производных. То есть при фиксированной стартовой процентной ставке r в зависимости от вектора X кривые $Y(\tau|z, \phi)$ и $F(\tau|z, \phi)$ для малых τ могут как возрастать, так и убывать.

Анализ временных структур аффинного класса показал, что характер кривых доходностей существенно зависит от текущей (стартовой) процентной ставки r , хотя предельное значение долгосрочной доходности от текущей ставки не зависит и определяется только параметрами модели. В квадратичных моделях текущая процентная ставка r при построении кривых доходности в явном виде не используется, вместо нее используются значения латентных переменных X , которые при известных параметрах модели однозначно определяют процентную ставку r . Однако в квадратичных моделях одно и то же значение процентной ставки r может быть получено для некоторого набора различных переменных состояния X . Интересно выяснить, каково разнообразие кривых доходности при фиксированной процентной ставке, определяемой разнообразием переменных состояния, соответствующих этой фиксированной процентной ставке. Точнее, какова ширина полосы, в которой лежат все возможные кривые доходности (или форвардные кривые), соответствующие некоторой фиксированной ставке r и различным стартовым векторам X .

Рассмотрим множество кривых доходности, стартовых от процентной ставки r и имеющих предельную доходность $y(\infty)$. Процентная ставка в скалярной форме определяется с помощью соотношения (9) выражением

$$r = r_{\min} + X^T \Phi X = r_{\min} + \sum_{i=1}^n \phi_i X_i^2 = r_{\min} + \sum_{i=1}^n z_i. \quad (22)$$

С помощью (18) запишем

$$Y(\tau|z, \phi) = r_{\min} + \sum_{i=1}^n \left[\left(\frac{1}{\tau} \frac{z_i}{v_i \operatorname{cth}(v_i \tau) + k_i} \right) + \left(\frac{1}{2\tau} \ln \left(\operatorname{ch}(v_i \tau) + \frac{k_i}{v_i} \operatorname{sh}(v_i \tau) \right) - \frac{k_i}{2} \right) \right]. \quad (23)$$

Рассмотрим сначала случай, когда матрица Φ (т.е. набор собственных чисел $\{\phi_i\}$) фиксирована. Будем называть кривую доходности $Y(\tau|z, \phi)$ (форвардную кривую) допустимой, если $Y(0|z, \phi) = r$ и $Y(\infty|z, \phi) = y(\infty)$. Ширина полосы, в которой лежат все возможные допустимые кривые доходности, определяется разностью $Y_{\max}(\tau, z) - Y_{\min}(\tau, z)$, где $Y_{\max}(\tau, z)$ – максимум $Y(\tau|z, \phi)$ по z при ограничении (22), а $Y_{\min}(\tau, z)$ – аналогичный минимум. Переменные z_i – неотрицательные величины, которые принимают значения в интервале $(0, r - r_{\min})$. Как следует из (23) доходность определяется как сумма n положительных слагаемых, каждое из которых, в свою очередь, распадается на два положительных слагаемых, из которых

только первое зависит от z_i . Тогда максимизация (минимизация) доходности $Y(\tau|z, \Phi)$ достигается просто: нужно найти максимальный (минимальный) коэффициент при z_i и назначить максимальное значение для соответствующего ему параметра z_i . Обозначим символом M индекс i , для которого $v_i \operatorname{cth}(v_i \tau) + k_i \leq v_j \operatorname{cth}(v_j \tau) + k_j$ для всех j . Аналогично обозначим символом m индекс i , для которого $v_i \operatorname{cth}(v_i \tau) + k_i \geq v_j \operatorname{cth}(v_j \tau) + k_j$ для всех j . Тогда верхняя и нижняя границы рассматриваемой полосы при заданной матрице Φ определяются соотношениями

$$Y_{\max}(\tau, z) = r_{\min} + \frac{1}{\tau} \frac{r - r_{\min}}{v_M \operatorname{cth}(v_M \tau) + k_M} + \sum_{i=1}^n \left(\frac{1}{2\tau} \ln \left(\operatorname{ch}(v_i \tau) + \frac{k_i}{v_i} \operatorname{sh}(v_i \tau) \right) - \frac{k_i}{2} \right),$$

$$Y_{\min}(\tau, z) = r_{\min} + \frac{1}{\tau} \frac{r - r_{\min}}{v_m \operatorname{cth}(v_m \tau) + k_m} + \sum_{i=1}^n \left(\frac{1}{2\tau} \ln \left(\operatorname{ch}(v_i \tau) + \frac{k_i}{v_i} \operatorname{sh}(v_i \tau) \right) - \frac{k_i}{2} \right).$$

Таким образом, ширина полосы, в которой лежат все возможные допустимые кривые доходности $Y(\tau|z, \Phi)$ при фиксированной процентной ставке r и заданной матрице Φ , равна

$$\Delta_Y = Y_{\max}(\tau, z) - Y_{\min}(\tau, z) = \left(\frac{r - r_{\min}}{v_M \operatorname{cth}(v_M \tau) + k_M} - \frac{r - r_{\min}}{v_m \operatorname{cth}(v_m \tau) + k_m} \right) \frac{1}{\tau}.$$

Аналогично решается проблема определения верхней $F_{\max}(\tau, z)$ и нижней границ $F_{\min}(\tau, z)$ для допустимых форвардных кривых:

$$F(\tau|z, \Phi) = r_{\min} + \sum_{i=1}^n \left[\left(\frac{v_i^2 z_i}{(v_i \operatorname{ch}(v_i \tau) + k_i \operatorname{sh}(v_i \tau))^2} \right) + \left(\frac{1}{2} \frac{v_i^2 - k_i^2}{v_i \operatorname{cth}(v_i \tau) + k_i} \right) \right].$$

Пусть символом H обозначается индекс i , для которого $\operatorname{ch}(v_i \tau) + (k_i / v_i) \operatorname{sh}(v_i \tau) \leq \operatorname{ch}(v_j \tau) + (k_j / v_j) \operatorname{sh}(v_j \tau)$ для всех j . Аналогично обозначим символом h индекс i , для которого $\operatorname{ch}(v_i \tau) + (k_i / v_i) \operatorname{sh}(v_i \tau) \geq \operatorname{ch}(v_j \tau) + (k_j / v_j) \operatorname{sh}(v_j \tau)$ для всех j . Тогда верхняя и нижняя границы рассматриваемой полосы при заданной матрице Φ определяются соотношениями

$$F_{\max}(\tau, z) = r_{\min} + \frac{v_H^2 (r - r_{\min})}{(v_H \operatorname{ch}(v_H \tau) + k_H \operatorname{sh}(v_H \tau))^2} + \sum_{i=1}^n \frac{1}{2} \frac{v_i^2 - k_i^2}{v_i \operatorname{cth}(v_i \tau) + k_i},$$

$$F_{\min}(\tau, z) = r_{\min} + \frac{v_h^2 (r - r_{\min})}{(v_h \operatorname{ch}(v_h \tau) + k_h \operatorname{sh}(v_h \tau))^2} + \sum_{i=1}^n \frac{1}{2} \frac{v_i^2 - k_i^2}{v_i \operatorname{cth}(v_i \tau) + k_i}.$$

Ширина полосы, в которой лежат все возможные допустимые форвардные кривые $F(\tau|z, \Phi)$ при фиксированной процентной ставке r и заданной матрице Φ , равна

$$\Delta_F = F_{\max}(\tau, z) - F_{\min}(\tau, z) = \left(\frac{v_H^2 (r - r_{\min})}{(v_H \operatorname{ch}(v_H \tau) + k_H \operatorname{sh}(v_H \tau))^2} - \frac{v_h^2 (r - r_{\min})}{(v_h \operatorname{ch}(v_h \tau) + k_h \operatorname{sh}(v_h \tau))^2} \right) \frac{1}{\tau}.$$

Теперь исследуем влияние матрицы Φ на исследуемое множество допустимых кривых доходности (форвардных кривых), составляющих рассматриваемую полосу. Элементы этой матрицы определяют не только процентную ставку (22), но и предельную доходность $y(\infty)$:

$$y(\infty) = r_{\min} + \sum_{i=1}^n \frac{v_i - k_i}{2}. \quad (24)$$

Поскольку матрица Φ задается набором собственных чисел $\{\phi_i\}$, то и допустимый набор матриц может определяться в пространстве положительных собственных чисел. Однако проблема выбора матриц Φ будет сложнее, так как эти матрицы определяют не только стартовую процентную ставку r , но и предельную долгосрочную доходность $y(\infty)$. Поэтому обсуждаемое разнообразие матриц (или, что одно

и то же, разнообразие наборов чисел $\{v_i\}$ должно удовлетворять не только равенству (22), но и соотношению (24) при фиксированных r и $y(\infty)$.

Заметим, что из (24) следует, что $k_i \leq v_i \leq 2(y(\infty) - r_{\min}) + k_i$, $1 \leq i \leq n$. В первом слагаемом, зависящем от z_i в выражении (23) для кривой доходности $Y(\tau|z, \Phi)$, знаменатель $v_i \operatorname{cth}(v_i \tau) + k_i$ монотонно возрастает с увеличением v_i . Следовательно, минимальное по v_i значение знаменателя $k_i (\operatorname{cth}(k_i \tau) + 1)$ достигается при $v_i = k_i$, а минимальное по v_i и k_i значение знаменателя достигается при минимальном $k_i = k_0$. Второе слагаемое в выражении (23) оказывается равным нулю, когда $v_i = k_i$, т.е. принимает свое минимальное значение. С ростом v_i оно монотонно возрастает и достигает своего наибольшего значения при $v_i = 2(y(\infty) - r_{\min}) + k_i$. Поэтому наибольшее значение второго слагаемого достигается при максимальном $k_i = k^0$, для которого $v_i = v^0 = 2(y(\infty) - r_{\min}) + k^0$. Так что границы множества допустимых кривых доходности $Y(\tau|z, \Phi)$ при варьировании по $\{z, v\}$ имеют вид

$$Y_{\max}(\tau|z, \Phi) = r_{\min} + \frac{1}{\tau} \frac{r - r_{\min}}{k_0 \operatorname{cth}(k_0 \tau) + k_0} + \frac{1}{2\tau} \ln \left(\operatorname{ch}(v^0 \tau) + \frac{k^0}{v^0} \operatorname{sh}(v^0 \tau) \right) - \frac{k^0}{2},$$

$$Y_{\min}(\tau|z, \Phi) = r_{\min} + \frac{1}{\tau} \frac{r - r_{\min}}{v^0 \operatorname{cth}(v^0 \tau) + k^0} + \frac{1}{2\tau} \ln \left(\operatorname{ch}(v_0 \tau) + \frac{k_0}{v_0} \operatorname{sh}(v_0 \tau) \right) - \frac{k_0}{2}.$$

Аналогичный анализ форвардных кривых $F(\tau|z, \Phi)$ дает следующие границы множества допустимых форвардных кривых при варьировании по $\{z, v\}$:

$$F_{\max}(\tau|z, \Phi) = r_{\min} + \frac{r - r_{\min}}{\exp(2k_0 \tau)} + \frac{1}{2} \frac{(v^0)^2 - (k^0)^2}{v^0 \operatorname{cth}(v^0 \tau) + k^0},$$

$$F_{\min}(\tau|z, \Phi) = r_{\min} + \frac{r - r_{\min}}{(\operatorname{ch}(v^0 \tau) + (k^0 / v^0) \operatorname{sh}(v^0 \tau))^2} + \frac{1}{2} \frac{(v_0)^2 - (k_0)^2}{v_0 \operatorname{cth}(v_0 \tau) + k_0}.$$

Заметим, что вычисление границ множества допустимых кривых при варьировании по $\{z, v\}$ при максимизации первых слагаемых требует предположения $v_i = k_i$, что, в свою очередь, подразумевает равенство нулю, по крайней мере, одного собственного числа матрицы Φ . А это противоречит предположению о положительной определенности этой матрицы. Поэтому такие границы являются реально недостижимыми. Реальная граница будет настолько близкой к этой недостижимой, насколько близким к нулю будет минимальное собственное число матрицы Φ .

Проиллюстрируем сказанное численным примером для простейшего случая $n = 2$. Пусть $r_{\min} = 0$ и

$$K = \begin{pmatrix} 0,6 & 0 \\ 0 & 0,13 \end{pmatrix}, S = \begin{pmatrix} 0,07 & 0 \\ 0 & 0,08 \end{pmatrix}.$$

Для определения базовых кривых будем использовать следующую пару (Φ, X) :

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}, X = \begin{pmatrix} 0,15 \\ 0,08 \end{pmatrix}.$$

В этом случае значение стартовой процентной ставки $r = 0,0481$, значение предельной доходности $y(\infty) = 0,06954$. На рис. 1 представлены кривая доходности $Y_0(u)$ и форвардная кривая $F_0(u)$ для этого случая. Чтобы представить эти кривые «целиком», для всего интервала значений сроков до погашения $\tau \in (0, \infty)$ использовано нелинейное преобразование сроков до погашения $u = 1 - e^{-\rho \tau}$, которое отображает положительную полуось $(0, \infty)$ в единичный интервал $(0, 1)$. Принятое при расчетах численное значение $\rho = \ln 10 / 30 = 0,07675$ соответствует тому, что сроки до погашения от 0 до 30 отображаются в интервал $(0, 0,9)$. Так что $Y_0(u) = y(\tau, X)$, $F_0(u) = f(\tau, X)$, где $\tau = -\ln(1 - u) / \rho$. По обычно принятой классификации полученные кривые относятся к типу «нормальных кривых», т.е. монотонно возрастающих до конечного предела на всем интервале изменения τ .

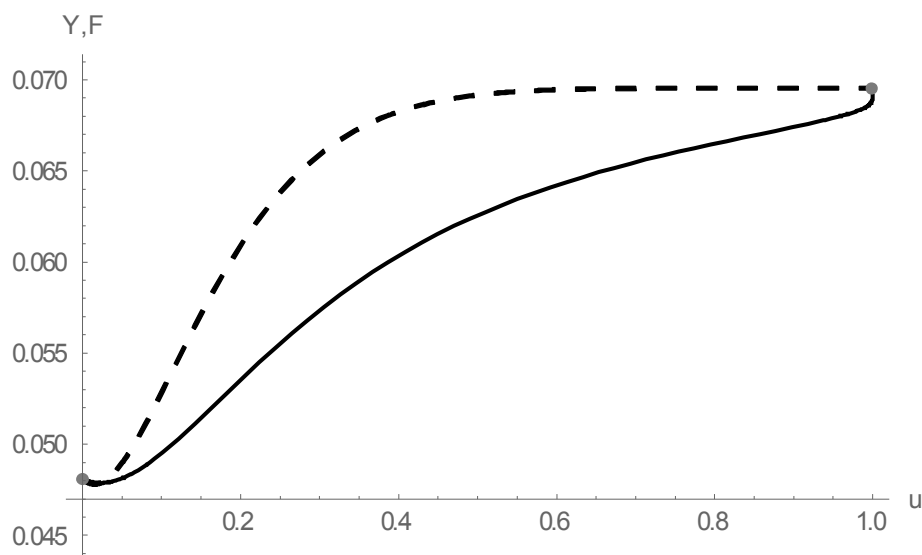


Рис. 1. Кривая доходности $Y_0(u)$ (сплошная) и форвардная кривая $F_0(u)$ (пунктирная) для базового варианта (параметры $\phi_1 = 1$, $\phi_2 = 4$, $X_1 = 0,15$, $X_2 = 0,08$, соответствуют значениям $r = 0,0481$, $y(\infty) = 0,06954$)

На рис. 2. изображены кривая доходности $Y_0(u)$ (сплошная) и верхняя и нижняя границы $Y_{\max}(u)$ и $Y_{\min}(u)$ той полосы, в которой лежат все возможные кривые доходности, соответствующие допустимым векторам переменных состояния. Допустимыми векторами считаются те, которые обеспечивают значения $r = 0,0481$, $y(\infty) = 0,06954$ при фиксированной (базовой) матрице Φ . При этих же условиях на рис. 3 представлена форвардная кривая $F_0(u)$ (сплошная) и верхняя и нижняя границы $F_{\max}(u)$ и $F_{\min}(u)$ полосы форвардных кривых, соответствующих допустимым векторам переменных состояния X .

Те же кривые, что и на рис. 2 и 3, показаны на рис. 4 и 5: кривая доходности $Y_0(u)$ (сплошная), форвардная кривая $F_0(u)$ (сплошная) и верхняя и нижняя границы (пунктирными линиями) полос соответствующих допустимых кривых, но уже для случая, когда могут варьироваться не только допустимые вектора переменных состояния X , но и допустимые матрицы Φ , что, как видно, существенно расширяет полосу возможных кривых доходности.

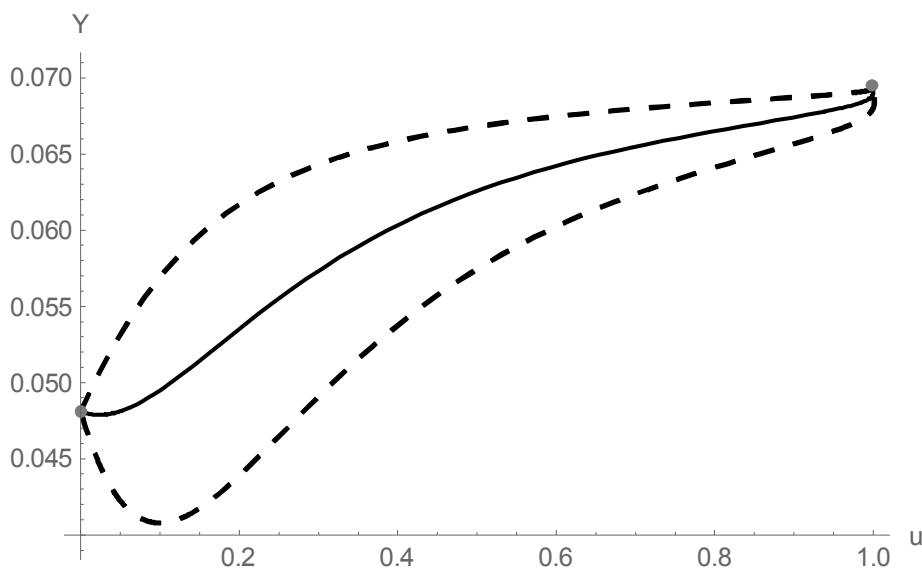


Рис. 2. Кривая доходности $Y_0(u)$ (сплошная) и верхняя и нижняя границы $Y_{\max}(u)$ и $Y_{\min}(u)$ полосы, в которой лежат все возможные кривые доходности, соответствующие допустимым векторам переменных состояния, обеспечивающим значения $r = 0,0481$, $y(\infty) = 0,06954$ при фиксированной (базовой) матрице Φ

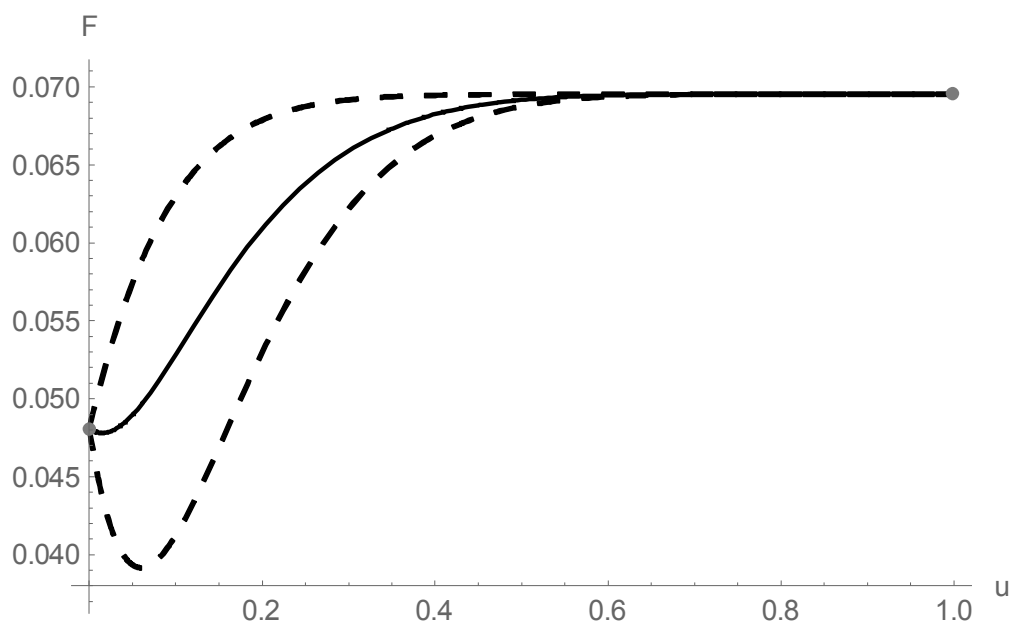


Рис. 3. Форвардная кривая $F_0(u)$ (сплошная) и верхняя и нижняя границы $F_{\max}(u)$ и $F_{\min}(u)$ (пунктирные) полосы форвардных кривых, соответствующих допустимым векторам переменных состояния X при фиксированной (базовой) матрице Φ

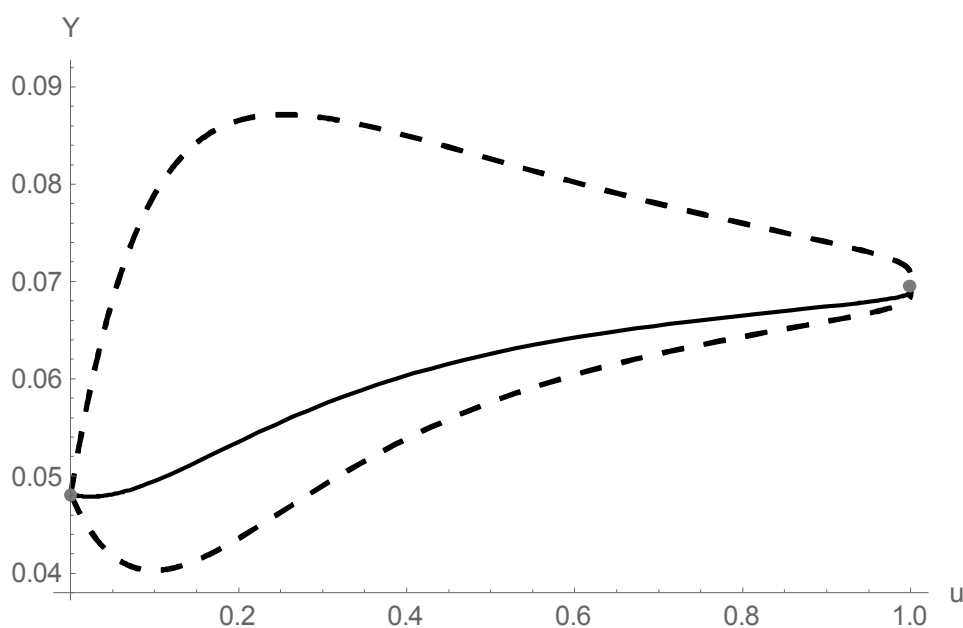


Рис. 4. Кривая доходности $Y_0(u)$ (сплошная) и верхняя и нижняя границы $Y_{\max}(u)$ и $Y_{\min}(u)$ (пунктирные) полосы, в которой лежат все возможные кривые доходности, соответствующие допустимым векторам переменных состояния X и матрицам Φ , обеспечивающим значения $r = 0,0481$, $y(\infty) = 0,06954$

Изменение ширины полос допустимых кривых в зависимости от срока до погашения для кривых, изображенных на рис. 2–5, показано на рис. 6. Для удобства сравнения на этом рисунке приведены сведения для обоих случаев: 1) допустимые кривые соответствуют различным векторам X при фиксированной матрице Φ ; 2) допустимые кривые определяются изменением как векторов X , так и матриц Φ .

В заключение отметим еще одну особенность квадратичных моделей временных структур на основе рассмотренного численного примера. При принятых исходных данных базовые кривые $Y_0(u) = y(\tau, X)$ и $F_0(u) = f(\tau, X)$ относятся к типу «нормальных кривых», т.е. монотонно возрастают на всем интервале изменения τ от стартовой процентной ставки r до конечного предела $y(\infty)$ долгосрочной доходности.

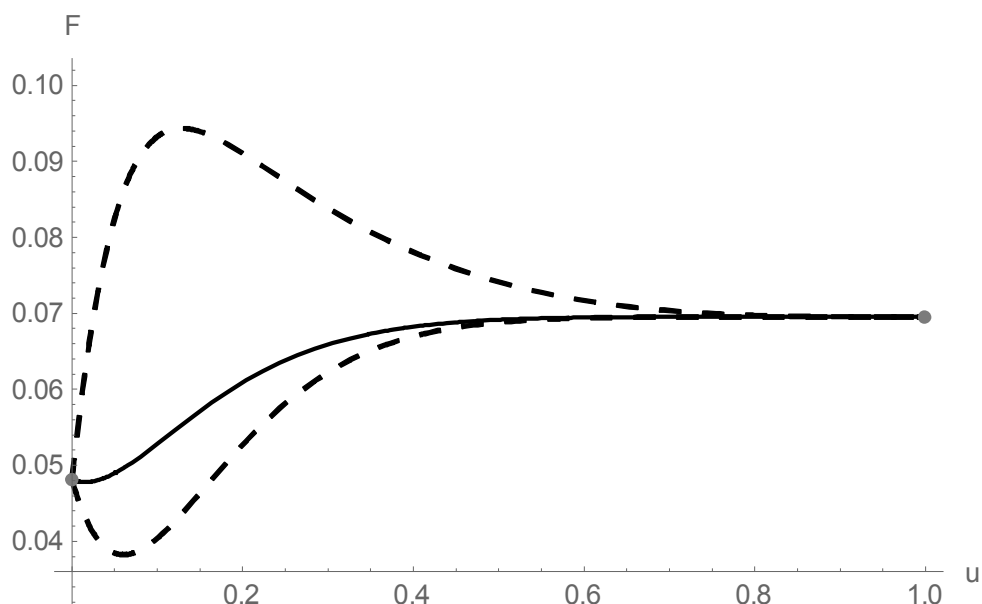


Рис. 5. Форвардная кривая $F_0(u)$ (сплошная) и верхняя и нижняя границы $F_{\max}(u)$ и $F_{\min}(u)$ полосы, в которой лежат все возможные форвардные кривые, соответствующие допустимым векторам переменных состояния X и матрицам Φ , обеспечивающим значения $r = 0,0481$, $y(\infty) = 0,06954$

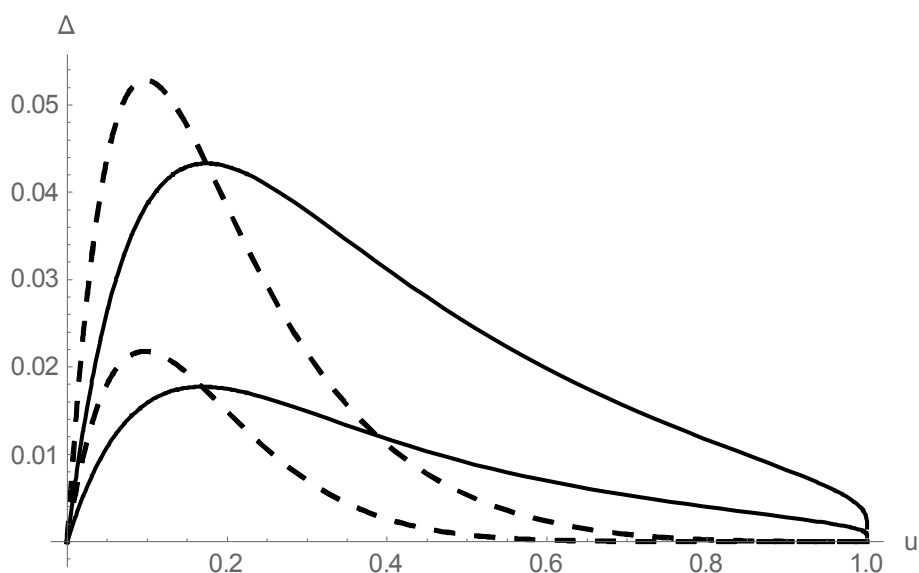


Рис. 6. Ширины Δ полос, представленных на рис. 2–5, для кривых доходности (сплошные линии) и форвардных кривых (пунктирные линии) при варьировании только вектором X (нижние кривые) и при варьировании парой (X, Φ) , верхние кривые

В аффинной модели временной структуры это будет единственным решением. В квадратичной модели мы получаем целое семейство кривых, обладающих свойством $Y(0|z, \phi) = r$ и $Y(\infty|z, \phi) = y(\infty)$. При этом в этом семействе имеются кривые, относящиеся как к типу «нормальных кривых», так и к другим типам, предполагающим наличие максимума или минимума.

Заключение

Представленные результаты иллюстрируют тот факт, что квадратичные модели временной структуры доходности позволяют получать более богатый спектр допустимых кривых доходности и форвардных кривых, чем аффинные модели, которые для фиксированной стартовой процентной ставки позволяют получить только одну кривую, в то время как квадратичные модели при условиях фиксиро-

ванных стартовой и долгосрочной предельной ставок r и $y(\infty)$ могут давать довольно широкий спектр кривых доходности. Следовательно, они обладают большими возможностями для подгонки теоретических кривых к реальным рыночным временным структурам. Однако это будет реально осуществимо, если существует возможность контролировать используемые в квадратичной модели латентные переменные. До сих пор в литературе отсутствует какое-либо описание этих переменных в экономическом смысле.

ЛИТЕРАТУРА

1. Ahn D.-H., Dittmar R.F., Gallant A.R. Quadratic term structure models: Theory and evidence // Review of Financial Studies. 2001. V. 15. P. 243–288.
2. Vasicek O. An equilibrium characterization of the term structure // Journal of Financial Economics. 1977. V. 5. P. 177–188.
3. Медведев Г.А. О временной структуре доходности. 1. Модель Васичека // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 1 (18). С. 102–111.
4. Медведев Г.А. О временной структуре доходности. 2. Модель Кокса–Ингерсолла–Росса // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2 (19). С. 102–111.
5. Медведев Г.А. О временной структуре доходности. 3. Однофакторная модель Даффи–Кана // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 3 (20). С. 71–80.
6. Медведев Г.А. О временной структуре доходности. 4. Двухфакторные модели Даффи–Кана // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4 (21). С. 89–99.
7. Медведев Г.А. О временной структуре доходности. 5. Двухфакторные модели Даффи–Кана (продолжение) // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 64–74.
8. Медведев Г.А. О временной структуре доходности. 6. Трехфакторные модели // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 3 (24). С. 113–122.
9. Медведев Г.А. О временной структуре доходности. 7. Новая версия // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 4 (25). С. 71–83.
10. Медведев Г.А. Стохастические процессы финансовой математики. Минск : БГУ, 2005. 243 с.
11. Constantinides G. A theory of the nominal structure of interest rates // Review of Financial Studies. 1992. V. 5. P. 531–552.
12. Leippold M., Wu L. Design and Estimation of Quadratic Term Structure Models // European Finance Review. 2003. V. 7. P. 47–73.

Медведев Геннадий Алексеевич – д-р. физ.-мат. наук, профессор. E-mail: MedvedevGA@bsu.by

Павлив Дмитрий Александрович. E-mail: simoco@yandex.ru

Белорусский государственный университет. Минск, Республика Беларусь

Поступила в редакцию 30 июня 2016 г.

Medvedev Gennady A., Pavliv Dmitriy A. (Belarusian State University, Belarus).

On quadratic models of yield in risk-neutral world.

Keywords: yield term structure; quadratic model; risk neutral probability measure.

DOI: 10.17223/19988605/37/5

Instead of "usual used" quadratic models of the interest rate it is possible to obtain the equivalent description of the interest rate through process $X(t)$ unobserved (latent) state variables by relations

$$dX(t) = K(\theta - X(t)) dt + S dW(t), t > t_0, X(t_0) = X_0, r(t) = r_{\min} + X(t)^T \Phi X(t), X(t) \in \mathbb{R}^n, r \in \mathbb{R}^1, \Phi \in \mathbb{R}^{n \times n},$$

Where K, θ, S, X_0 are corresponding constant vectors and matrixes, and Φ is a diagonal matrix on which diagonal positive eigenvalues of initial matrix of "usual used" model are located. The condition of absence of arbitrage leads to the equation in partial derivatives for the price of zero coupon bond which is obtained in form $P(r(X), \tau) = \exp[-X^T A(\tau) X - X^T B(\tau) - C(\tau)]$. For functions of term structure $A(\tau), B(\tau)$ and $C(\tau)$ there is a system of the nonlinear ordinary differential equations which generally in an explicit analytical form is unsolved.

In paper assumptions are made: 1) process of latent variables $X(t)$ is normal process with a stationary expectation $\theta = 0$. 2) probability properties of the interest rate $r(X)$ submit to a risk neutral probability measure. At these assumptions it is found out that function $B(\tau) = 0$ for all values τ . 3) the latent variables constituting vector $X(t)$, are independent stochastic processes.

At these assumptions the functions of term structure $A(\tau)$ and $C(\tau)$ are obtained in the analytical form and the yield curve $Y(\tau)$ and the forward rate curve $F(\tau)$ are derived in an explicit form for $z_i \equiv \phi_i X_i^2 \geq 0, v_i \equiv \sqrt{k_i^2 + 2s_i^2 \phi_i} \geq k_i, 1 \leq i \leq n$:

$$Y(\tau) = r_{\min} + \sum_{i=1}^n \frac{1}{2\tau} \left(\frac{2z_i}{v_i \operatorname{cth}(v_i \tau) + k_i} + \ln \left(\operatorname{ch}(v_i \tau) + \frac{k_i}{v_i} \operatorname{sh}(v_i \tau) \right) - k_i \tau \right),$$

$$F(\tau) = r_{\min} + \sum_{i=1}^n \left(\frac{v_i^2 z_i}{(v_i \operatorname{ch}(v_i \tau) + k_i \operatorname{sh}(v_i \tau))^2} + \frac{1}{2} \frac{v_i^2 - k_i^2}{v_i \operatorname{cth}(v_i \tau) + k_i} \right).$$

Both curves $Y(\tau)$ and $F(\tau)$ at change τ from 0 to τ or 0 до ∞ , starting from the fixed value of the interest rate r , determined by current values of state variables $X(t)$, converge to a general limit $y(\infty) = r_{\min} + \sum_{i=1}^n \frac{v_i - k_i}{2}$ that is independent on current state $X(t)$ and is depending only from model parameters.

As seen, in quadratic models the current interest rate r at yield curve deriving in an explicit form is not used, instead of it values of latent variables X which at known parameters models uniquely determine the interest rate r are used. However in quadratic models the same value of the interest rate r can be derived for some set of various state variables X . In paper it is found out, how is the family of yield curve at a fixed interest rate determined by a variety of state variables, corresponding to this fixed interest rate. Besides, the width of a strip in which lie all possible yield curves (or forward curves), corresponding to various starting vectors X and to matrixes Φ , that determine the rate r , is found.

All received analytical results are illustrated by a numerical example.

REFERENCES

1. Ahn, D.-H., Dittmar, R. F. & Gallant, A.R. (2001) Quadratic term structure models: Theory and evidence. *Review of Financial Studies*. 15. pp. 243-288. DOI: 10.1093/rfs/15.1.243
2. Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics*. 5. pp. 177-188. DOI: 10.1002/9781119186229.ch6.
3. Medvedev, G.A. (2012) On term structure of yield rates. 1. Vasiček model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(18). pp. 102-111. (In Russian).
4. Medvedev, G.A. (2012) On term structure of yield rates. 2. The Cox – Ingersoll – Ross model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(19). pp. 102-111. (In Russian).
5. Medvedev, G.A. (2012) On term structure of yield rates. 3. The Duffie – Kan one-factor model. *Tomsk Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(20). pp. 71-80. (In Russian).
6. Medvedev, G.A. (2012) On term structure of yield rates. 4. The Duffie – Kan two factor model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(21). pp. 89-99. (In Russian).
7. Medvedev, G.A. (2013) On term structure of yield rates. 5. The Duffie-Kan two factor model (continuation). *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 64-74. (In Russian).
8. Medvedev, G.A. (2013) On term structure of yield rates. 6. The three factor model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(24). pp. 113-122. (In Russian).
9. Medvedev, G.A. (2013) On term structure of yield rates. 7. The new version. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(25). pp. 71-83. (In Russian).
10. Medvedev, G.A. (2005) *Stokhasticheskie protsessy finansovoy matematiki* [Stochastic processes of financial mathematics]. Minsk: BSU.
11. Constantinides, G. (1992) A theory of the nominal structure of interest rates. *Review of Financial Studies*. 5. pp. 531-552.
12. Leippold, M. & Wu, L. (2003) Design and Estimation of Quadratic Term Structure Models. *European Finance Review*. 7. pp. 47-73. DOI: 10.1093/rof/rfl002/.

УДК 519.872

DOI: 10.17223/19988605/37/6

А.А. Назаров, Е.А. Фёдорова

**МОДИФИКАЦИЯ МЕТОДА АСИМПТОТИЧЕСКОГО АНАЛИЗА
В УСЛОВИИ БОЛЬШОЙ ЗАГРУЗКИ НА ПРИМЕРЕ RQ-СИСТЕМЫ $M|M|1$** *Работа выполнена в рамках проекта Министерства образования и науки РФ №1.511.2014/К.*

В работе предлагаются три метода модификации асимптотического анализа в условии большой загрузки RQ-систем. Описание методов проведено на примере RQ-системы $M|M|1$. Предложены «промежуточная» асимптотика, имеющая вид гипергамма распределения, модификация «промежуточной» асимптотики в виде сдвига аргумента на параметр δ и полусумма распределений модифицированной и второй асимптотики. Представлены результаты численного анализа предложенных модификаций.

Ключевые слова: RQ-система; метод асимптотического анализа; большая загрузка; гипергамма-распределение.

Retrial Queueing System (системы массового обслуживания с повторными вызовами, или RQ-системы) – математические модели, широко применяемые для анализа и оптимизации различных телекоммуникационных систем, сетей мобильной связи, call-центров и других технических и экономических систем [1–6]. Характерной чертой таких моделей является наличие повторных обращений заявок к прибору после неудачной попытки обслуживания спустя некоторое случайное время. Такие ситуации могут быть вызваны не только отсутствием свободных серверов в моменты поступления заявок в систему, но техническими причинами.

Возникновение моделей RQ-систем прежде всего связывают с работами американских ученых R.I. Wilkinson [1] и J.W. Cohen [2] середины XX в., которые были посвящены практическим задачам, возникающим в телефонных сетях, и описанию влияния эффекта повторных вызовов на производительность технических систем. Первые подходы к математическому описанию RQ-систем были предприняты G. Gosztony [3] и A. Elldin [4].

Наиболее полное и детальное описание RQ-систем и их сравнение с классическими СМО было отражено в монографиях J.R. Artalejo, A. Gomez-Corral, G.I. Falin и J.G.C. Templeton [7, 8]. Ими получены аналитические результаты для RQ-систем $M|M|1$, $M|GI|1$, $M|M|C$ и других систем с пуассоновским входящим потоком, а также рассмотрены разнообразные методы для исследования таких систем и проведено детальное сравнение RQ-систем с их классическими аналогами [9].

На сегодняшний день исследованию RQ-систем посвящено большое количество работ [7], однако аналитических формул получено немного (лишь для систем с простейшим входящим потоком) [8]. В основном возникающие задачи решаются численными методами или с помощью имитационного моделирования [10–12], и как следствие, результаты таких исследований имеют узкое применение.

Ранее в работах [13, 14] был предложен метод асимптотического анализа (первого и второго порядков) для исследования RQ-систем при условии большой загрузки. Предложенным методом были исследованы системы с различными типами входящего потока и дисциплиной обслуживания (в том числе RQ-система $MMPP|GI|1$ [15]). Было показано, что асимптотика первого порядка имеет вид гамма-распределения вне зависимости от типа системы, но область ее применения достаточно узка. Асимптотика второго порядка позволила расширить область применения результатов в 4 раза, однако асимптотическая формула не обладает единообразием и требуется ее достаточно трудоемкий вывод для каждой системы.

В связи с этим в данной статье предлагаются способы модификации метода асимптотического анализа в условии большой загрузки RQ-систем, позволяющие увеличить точность асимптотического метода. Описание методов приведено на примере простейшей RQ-системы $M|M|1$ для возможности

сравнения асимптотических результатов с известным точным. В дальнейшем предложенные модификации могут быть применены для исследования различных RQ-систем, в том числе с непуассоновским входящим потоком.

1. Математическое описание RQ-системы M|M|1

Рассмотрим однолинейную RQ-систему (рис. 1), на вход которой поступает простейший поток заявок с параметром λ , время обслуживания каждой заявки распределено по экспоненциальному закону с параметром μ . Если поступившая заявка застаёт прибор свободным, то она занимает его для обслуживания. Если прибор занят, то заявка переходит в источник повторных вызовов (ИПВ), где осуществляет случайную задержку, продолжительность которой имеет экспоненциальное распределение с параметром σ . Из ИПВ после случайной задержки заявка вновь обращается к обслуживающему прибору с повторной попыткой его захвата. Если прибор свободен, то заявка из ИПВ занимает его для обслуживания, в противном случае она мгновенно возвращается в источник повторных вызовов для реализации следующей задержки.

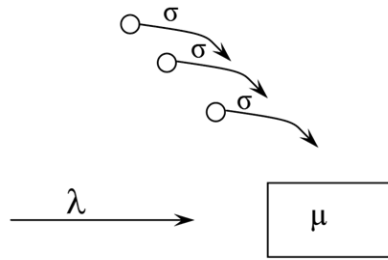


Рис. 1. RQ-система M|M|1

Пусть $i(t)$ – число заявок в ИПВ, а $k(t)$ определяет состояние прибора следующим образом:

$$k(t) = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если прибор занят.} \end{cases}$$

Обозначим $P\{k(t) = k, i(t) = i\} = P(k, i, t)$ вероятность того, что прибор в момент времени t находится в состоянии k и в источнике повторных вызовов находится i заявок.

Причем процесс $\{k(t), i(t)\}$ изменения состояний данной системы во времени является марковским. Тогда ставится задача нахождения распределения вероятностей числа заявок в источнике повторных вызовов такой системы.

Для распределения вероятностей $P(k, i, t)$ состояний рассматриваемой RQ-системы составим систему дифференциальных уравнений Колмогорова:

$$\begin{cases} \frac{\partial P(0, i, t)}{\partial t} = \mu P(1, i, t) - (\lambda + i\sigma)P(0, i, t), \\ \frac{\partial P(1, i, t)}{\partial t} = \lambda P(0, i, t) + (i+1)\sigma \cdot P(0, i+1, t) - (\lambda + \mu)P(1, i, t) + \lambda P(1, i-1, t). \end{cases} \quad (1)$$

Перейдем к стационарным частичным характеристическим функциям $H(k, u) = \sum_i e^{ju i} P(k, i)$, где

$P(k, i) = \lim_{t \rightarrow \infty} P(k, i, t)$, а j – мнимая единица.

Введем параметр $\rho = \frac{\lambda}{\mu}$, характеризующий загрузку системы. Тогда система (1) в стационарном режиме для характеристических функций переписывается в следующем виде:

$$\begin{cases} H(1, u) - \rho H(0, u) = -\frac{\sigma}{\mu} j \frac{\partial H(0, u)}{\partial u}, \\ \rho H(0, u) + (\rho(e^{ju} - 1) - 1)H(1, u) = j \frac{\sigma}{\mu} e^{-ju} \frac{\partial H(0, u)}{\partial u}. \end{cases} \quad (2)$$

2. Метод асимптотического анализа в условии большой загрузки

В работах [13–15] для решения системы уравнений (2) был предложен метод асимптотического анализа в условиях большой загрузки, т.е. при $\rho \uparrow 1$, который заключается в следующем.

Вводятся обозначения $\varepsilon = 1 - \rho$ (тогда в условии большой загрузки $\varepsilon \downarrow 0$), $u = \varepsilon w$, $H(0, u) = \varepsilon F_0(w, \varepsilon)$, $H(1, u) = F_0(w, \varepsilon)$.

Таким образом, система (2) принимает вид

$$\begin{cases} F_1(w, \varepsilon) - (1 - \varepsilon)\varepsilon F_0(w, \varepsilon) + j \frac{\sigma}{\mu} \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0, \\ (1 - \varepsilon)\varepsilon F_0(w, \varepsilon) + ((1 - \varepsilon)(e^{jw\varepsilon} - 1) - 1)F_1(w, \varepsilon) - j \frac{\sigma}{\mu} e^{-jw\varepsilon} \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0. \end{cases} \quad (3)$$

Далее используются следующие разложения:

$$F_k(w, \varepsilon) = F_k(w) + \varepsilon \cdot f_k(w) + O(\varepsilon^2), \quad (4)$$

где $O(\varepsilon^2)$ – бесконечно малая величина порядка ε^2 .

Допредельная характеристическая функция $H(u) = H(1, u) + H(0, u)$ в условиях большой загрузки может быть приближенно определена равенством

$$H(u) \approx F_1\left(\frac{u}{1 - \rho}\right) + O(\varepsilon^2).$$

Обозначим $h_1(u)$ асимптотическую характеристическую функцию первого порядка, которая равна

$$h_1(u) = F_1\left(\frac{u}{1 - \rho}\right).$$

В [13] было доказано, что $h_1(u)$ удовлетворяет следующей теореме.

Теорема 1. Асимптотическая характеристическая функция числа заявок в ИПВ в RQ-системе M|M|1 имеет вид характеристической функции гамма-распределения

$$h_1(u) = \left(1 - j \frac{u}{\beta}\right)^{-\alpha} \quad (5)$$

с параметрами формы $\alpha = \frac{\mu}{\sigma} + 1$ и масштаба $\beta = 1 - \rho$.

В качестве критерия допустимости асимптотических результатов использовалось условие: расстояние Колмогорова $\Delta \leq 0,05$.

С помощью численного анализа было показано, что метод асимптотического анализа первого порядка в условиях большой загрузки может быть применим при значениях загрузки $\rho \geq 0,95$, что является достаточно узкой областью применимости. Поэтому также была получена асимптотическая характеристическая функция второго порядка $h_2(u)$:

$$h_2(u) = F_1\left(\frac{u}{1 - \rho}\right) + (1 - \rho) \left[F_0\left(\frac{u}{1 - \rho}\right) + f_1\left(\frac{u}{1 - \rho}\right) \right].$$

Для получения этой асимптотики второго порядка вместо разложений функций (4) необходимо было рассмотреть разложения

$$F_k(w, \varepsilon) = F_k(w) + \varepsilon \cdot f_k(w) + \varepsilon^2 \cdot \varphi_k(w) + O(\varepsilon^3)$$

и, подставив их в систему (3), решить систему пяти дифференциальных уравнений.

Тогда допредельная характеристическая функция может быть определяется равенством

$$H(u) = h_2(u) + O(\varepsilon^3).$$

Была доказана следующая теорема о виде функции $h_2(u)$ [14].

Теорема 2. Асимптотическая характеристическая функция второго порядка распределения вероятностей числа заявок в ИПВ в RQ-системе M|M|1 имеет вид

$$h_2(u) = \left(1 - j \frac{u}{1-\rho}\right)^{-\frac{\mu+\sigma}{\sigma}} \left\{ 1 + (1-\rho) \left[-\frac{ju}{1-\rho} + \frac{\mu}{\sigma} \ln \left(1 - \frac{ju}{1-\rho}\right) - \left(\frac{\mu}{\sigma} + 1\right) \frac{\frac{ju}{1-\rho} - \frac{\left(\frac{ju}{1-\rho}\right)^2}{2}}{1 - \frac{ju}{1-\rho}} \right] \right\}.$$

Численный анализ показал, что область применения асимптотики второго порядка $\rho > 0,8$, т.е. увеличивается в 4 раза (по сравнению с (5)). Однако для более сложных RQ-систем, например MMPP|GI|1, приходилось решать систему из семи и более дифференциальных уравнений, что достаточно трудоемко.

3. Гипергамма-асимптотика

Заметив, что в ходе доказательства теоремы 1 была найдена и функция $F_0(w)$, было предложено провести численный анализ «промежуточной» асимптотики $h^*(u)$:

$$h^*(u) = F_1\left(\frac{u}{1-\rho}\right) + (1-\rho)F_0\left(\frac{u}{1-\rho}\right).$$

В работе [13] показано выполнение следующих равенств:

$$\begin{cases} F_0(w) = C(1-jw)^{-\frac{\mu}{\sigma}}, \\ F_1(w) = C(1-jw)^{-\frac{\mu}{\sigma}-1}. \end{cases}$$

Имеем условие нормировки $h(0) = H_0(0) + H_1(0) = 1$. Так как $H_1(u) \approx F_1(u)$, а $H_0(u) \approx (1-\rho)F_0(u)$, то получаем $h(0) = (1-\rho)F_0(0) + F_1(0) = (1-\rho)C + C \equiv 1$. Отсюда $C = \frac{1}{2-\rho}$.

Тогда имеем

$$h^*(u) = \frac{1}{2-\rho} \left(1 - j \frac{u}{1-\rho}\right)^{-\frac{\mu}{\sigma}-1} + \frac{1-\rho}{2-\rho} \left(1 - j \frac{u}{1-\rho}\right)^{-\frac{\mu}{\sigma}}.$$

Заметим, что слагаемые функции $h^*(u)$ являются характеристическими функциями гамма-распределения:

$$\Gamma_1(u) = \left(1 - \frac{ju}{\beta}\right)^{-\alpha}, \quad \Gamma_0(u) = \left(1 - \frac{ju}{\beta}\right)^{-\alpha+1}, \quad \text{где } \alpha = \frac{\mu}{\sigma} + 1 \text{ и } \beta = 1-\rho.$$

Таким образом, получили, что «промежуточная» асимптотическая характеристическая функция распределения числа заявок в ИПВ имеет вид взвешенной суммы двух характеристических функций гамма-распределения

$$h^*(u) = q\Gamma_1(u) + (1-q)\Gamma_0(u), \quad (6)$$

где $q = \frac{1}{2-\rho}$.

Будем называть такое распределение *гипергамма-распределением* (по аналогии с гиперэкспоненциальным).

В табл. 1 приведены Δ_1 , Δ_r , Δ_2 – расстояния Колмогорова между точным распределением и первой асимптотикой, гипергамма-распределением и второй асимптотикой соответственно.

Из табл. 1 видно, что гипергамма-асимптотика значительно точнее асимптотики первого порядка, но уступает асимптотике второго порядка.

Расстояние Колмогорова между точным и асимптотическими распределениями

Значения параметров системы		$\rho = 0,7$	$\rho = 0,8$	$\rho = 0,9$	$\rho = 0,95$
$\sigma = 10$	$\Delta 1$	0,251	0,170	0,086	0,043
	Δp	0,097	0,060	0,028	0,014
	$\Delta 2$	0,035	0,022	0,009	0,005
$\sigma = 2$	$\Delta 1$	0,266	0,179	0,091	0,046
	Δp	0,157	0,104	0,050	0,025
	$\Delta 2$	0,088	0,047	0,016	0,005
$\sigma = 1$	$\Delta 1$	0,304	0,208	0,106	0,053
	Δp	0,220	0,147	0,074	0,037
	$\Delta 2$	0,124	0,052	0,013	0,003

Однако вывод асимптотической формулы второго порядка для более сложных RQ-систем затруднителен, тогда как «промежуточная» асимптотика не требует дополнительных выкладок, так как ее форма одинакова для RQ-систем с различными типами входящего потока и дисциплины обслуживания, а параметры гипергамма-распределения совпадают с параметрами первой асимптотики. Поэтому целесообразность ее использования для большинства практических задач очевидна.

4. Модификация гипергамма-асимптотики

Известно, что $H_1(0) = R_1 = \rho$, $H_0(0) = R_0 = 1 - \rho$. Однако в полученной «промежуточной» асимптотике (6)

$$q\Gamma_1(0) = q \neq R_1 \text{ и } (1-q)\Gamma_0(0) = 1 - q \neq R_0.$$

Перейдем от характеристических функций к функциям распределения вероятностей, обозначив их соответственно $\tilde{F}_0(x)$ и $\tilde{F}_1(x)$. Тогда $q\tilde{F}_1(\infty) = q \neq R_1$ и $(1-q)\tilde{F}_0(\infty) = 1 - q \neq R_0$.

В связи с этим предлагается модифицировать асимптотику следующим образом:

$$\begin{cases} F_1(x) = q[\tilde{F}_1(x + \delta) - \tilde{F}_1(\delta)], \\ F_0(x) = \pi + (1 - q)[\tilde{F}_0(x + \delta) - \tilde{F}_0(\delta)], \end{cases}$$

где $F_1(x)$ и $F_0(x)$ – модифицированные частичные асимптотические функции распределения, а параметры δ и π могут быть найдены из условия нормировки:

$$\begin{cases} F_1(\infty) = q[1 - \tilde{F}_1(\delta)] = \rho, \\ F_0(\infty) = \pi + (1 - q)[1 - \tilde{F}_0(\delta)] = 1 - \rho. \end{cases} \quad (7)$$

Из первого уравнения (7) имеем

$$\tilde{F}_1(\delta) = 1 - \frac{\rho}{q} = \frac{q - \rho}{q}.$$

Подставляя введенные ранее обозначения, нетрудно получить

$$\tilde{F}_1(\delta) = (1 - \rho)^2. \quad (8)$$

Выразим величину π из второго уравнения (11):

$$\pi = \frac{(1 - \rho)^2 + (1 - \rho)\tilde{F}_0(\delta)}{2 - \rho}.$$

Подставляя (8), получим следующее выражение:

$$\pi = \frac{\tilde{F}_1(\delta) + (1 - \rho)\tilde{F}_0(\delta)}{2 - \rho} = q\tilde{F}_1(\delta) + (1 - q)\tilde{F}_0(\delta). \quad (9)$$

Модифицированная асимптотическая функция распределения записывается как сумма

$$F(x) = F_1(x) + F_0(x).$$

Учитывая полученные выражения (11)–(13), имеем следующее равенство:

$$F(x) = q\tilde{\Gamma}_1(x + \delta) + (1 - q)\tilde{\Gamma}_0(x + \delta).$$

Проведем численный анализ модификации асимптотики и сравним ее с результатами, полученными ранее асимптотики (табл. 2, где ΔM – расстояние Колмогорова между точным распределением и модифицированной асимптотикой).

Т а б л и ц а 2

Расстояние Колмогорова между точным, асимптотическими и модифицированными распределениями

Значения параметров системы		$\rho = 0,5$	$\rho = 0,6$	$\rho = 0,7$	$\rho = 0,8$	$\rho = 0,9$	$\rho = 0,95$
$\sigma = 10$	$\Delta 1$	0,378	0,321	0,251	0,170	0,086	0,043
	Δp	0,180	0,137	0,097	0,060	0,028	0,014
	ΔM	0,065	0,059	0,050	0,038	0,022	0,012
	$\Delta 2$	0,042	0,041	0,035	0,022	0,009	0,005
$\sigma = 2$	$\Delta 1$	0,432	0,342	0,266	0,179	0,091	0,046
	Δp	0,271	0,213	0,157	0,104	0,050	0,025
	ΔM	0,079	0,077	0,068	0,050	0,031	0,017
	$\Delta 2$	0,192	0,139	0,088	0,047	0,016	0,005
$\sigma = 1$	$\Delta 1$	0,487	0,394	0,304	0,208	0,106	0,053
	Δp	0,364	0,291	0,220	0,147	0,074	0,037
	ΔM	0,093	0,087	0,074	0,052	0,030	0,016
	$\Delta 2$	0,313	0,214	0,124	0,052	0,013	0,003

Данные табл. 2 показывают, что «модифицированное» распределение существенно улучшает результаты «промежуточной» асимптотики. При этом для построения модифицированного распределения необходимо лишь найти величину сдвига δ из условия (8).

Рассмотрим комбинацию «модифицированного» распределения и второй асимптотики

$$Pm^{(2)} = \frac{1}{2}(Pm + P^{(2)}).$$

Будем называть такое распределение «модификация-2». Проведем численный анализ модификации асимптотики и сравним ее с результатами, полученными ранее.

Т а б л и ц а 3

Расстояние Колмогорова между точным, асимптотическими и модифицированными распределениями

Значения параметров системы		$\rho = 0,5$	$\rho = 0,6$	$\rho = 0,7$	$\rho = 0,8$	$\rho = 0,9$	$\rho = 0,95$
$\sigma = 10$	$\Delta 1$	0,378	0,321	0,251	0,170	0,086	0,043
	Δp	0,180	0,137	0,097	0,060	0,028	0,014
	ΔM	0,065	0,059	0,050	0,038	0,022	0,012
	$\Delta 2$	0,042	0,041	0,035	0,022	0,009	0,005
	$\Delta M2$	0,034	0,024	0,018	0,013	0,009	0,005
$\sigma = 2$	$\Delta 1$	0,432	0,342	0,266	0,179	0,091	0,046
	Δp	0,271	0,213	0,157	0,104	0,051	0,025
	ΔM	0,079	0,077	0,068	0,050	0,031	0,017
	$\Delta 2$	0,192	0,139	0,088	0,047	0,016	0,005
	$\Delta M2$	0,111	0,073	0,039	0,019	0,013	0,008
$\sigma = 1$	$\Delta 1$	0,487	0,394	0,304	0,208	0,106	0,053
	Δp	0,364	0,291	0,220	0,147	0,074	0,037
	ΔM	0,093	0,087	0,074	0,052	0,030	0,016
	$\Delta 2$	0,313	0,214	0,124	0,052	0,013	0,003
	$\Delta M2$	0,150	0,087	0,046	0,022	0,012	0,007

В результате численного анализа был замечен интересный результат: «модификация-2» точнее второй асимптотики. При этом предложенная модификация применима для значений загрузки $\rho \geq 0,7$, а при больших значениях параметра задержки σ – для значений загрузки $\rho \geq 0,5$.

Заключение

Таким образом, в работе предложены три модификации метода асимптотического анализа в условии большой загрузки RQ-систем: «промежуточная» асимптотика, имеющая вид гипергамма-рас-

пределения; модификация «промежуточной» асимптотики в виде сдвига распределения на параметр δ ; полусумма распределений модифицированной и второй асимптотики. В статье также представлены результаты численного сравнения предложенных асимптотик с известным точным, позволяющие сделать заключения об области применения каждого метода.

Обобщая вышесказанное, можно сделать вывод, что для практических задач, не требующих высокой точности аппроксимации, целесообразно использование результатов «промежуточной» асимптотики в силу ее простоты построения (имеет единую форму вне зависимости от типа входящего потока) и достаточно широкой области применимости ($\rho \geq 0,8$), а для задач, требующих более высокой точности аппроксимации, целесообразно применять предлагаемые в работе модификации 1 и 2.

ЛИТЕРАТУРА

1. Wilkinson R.I. Theories for toll traffic engineering in the USA // The Bell System Technical Journal. 1956. V. 35, No. 2. P. 421–507.
2. Cohen J.W. Basic problems of telephone traffic and the influence of repeated calls // Philips Telecommunication Review. 1957. V. 18, No. 2. P. 49–100.
3. Gosztony G. Repeated call attempts and their effect on traffic engineering // Budavox Telecommunication Review. 1976. V. 2. P. 16–26.
4. Elldin A., Lind G. Elementary Telephone Traffic Theory. Ericsson Public Telecommunications, 1971.
5. Kuznetsov D.Yu., Nazarov A.A. Analysis of non-Markovian models of communication networks with adaptive protocols of multiple random access // Avtomatika i Telemekhanika. 2001. V. 5. P. 124–146.
6. Nazarov A.A., Tsoj S.A. Common approach to studies of Markov models for data transmission networks controlled by the static random multiple access protocols // Avtomatika i Vychislitel'naya Tekhnika. 2004. V. 4. P. 73–85.
7. Artalejo J.R., Gomez-Corral A. Retrial Queueing Systems. A Computational Approach. Springer, 2008.
8. Falin G.I., Templeton J.G.C. Retrial queues. London : Chapman & Hall, 1997.
9. Artalejo J.R., Falin G.I. Standard and retrial queueing systems: A comparative analysis // Revista Matematica Complutense. 2002. V. 15. P. 101–129.
10. Neuts M.F., Rao B.M. Numerical investigation of a multiserver retrial model // Queueing Systems. 2002. V. 7, No. 2. P. 169–189.
11. Ridder A. Fast simulation of retrial queues // Third Workshop on Rare Event Simulation and Related Combinatorial Optimization Problems. Pisa, 2000. P. 1–5.
12. Kim C.S., Mushko V.V., Dudin A. Computation of the steady state distribution for multi-server retrial queues with phase type service process // Annals of Operations Research. 2012. V. 201, No. 1. P. 307–323.
13. Moiseeva E., Nazarov A. Asymptotic Analysis of RQ-systems M/M/1 on Heavy Load Condition // Proceedings of the IV International Conference Problems of Cybernetics and Informatics. Baku, Azerbaijan, 2012. P. 164–166.
14. Назаров А.А., Фёдорова Е.А. Метод асимптотического анализа в условиях большой загрузки 2-го порядка на примере исследования RQ-системы M|M|1 // Информационные технологии и математическое моделирование (ИТММ-2013) : материалы XII всерос. науч.-практ. конф. с междунар. участием им. А.Ф. Терпугова : в 2 ч. Томск : Изд-во Том. ун-та, 2013. Ч. 2. С. 59–65.
15. Моисеева Е.А., Назаров А.А. Исследование RQ-системы MMPP|GI|1 методом асимптотического анализа // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 4 (25). С. 84–94.

Назаров Анатолий Андреевич, д-р техн. наук, профессор. E-mail: nazarov.tsu@gmail.com

Фёдорова Екатерина Александровна, канд. физ.-мат. наук. E-mail: moiskate@mail.ru

Томский государственный университет

Поступила в редакцию 14 августа 2016 г.

Nazarov Anatoly A., Fedorova Ekaterina A. (Tomsk State University, Russian Federation).

Modification of the asymptotic analysis method under heavy load condition on the example of the study of the retrial queueing system M|M|1.

Key words: retrial queueing system; asymptotic analysis method; heavy load; hypergamma distribution.

DOI: 10.17223/19988605/37/6

In the paper, three methods of the asymptotic analysis modification are offered for retrial queueing systems under heavy load condition as mathematical models of telecommunication systems. The characteristic feature of these models is existence of repeated attempts of calls to get service after a random time. There are a large number of papers devoted to retrial queues researching, but mainly problems are solved by numerical methods and simulation. Analytical results are obtained only for systems with Poisson input process.

Earlier we proposed the method of asymptotic analysis (first and second order) for retrial queueing systems researching under heavy load condition [13, 14]. It has been shown that the first order asymptotics has the form of gamma distributions for all types of the system, but its application range is quite narrow. The asymptotics of second order has allowed to extend the application range of the results by 4 times, but the asymptotic formula has not unified form and it needs time-taking rather laborious formula derivations for each system. Therefore, we suggest ways to modify the method of the asymptotic analysis under heavy load condition to increase the method accuracy.

The method is described on the example of the simplest retrial queue M|M|1 in order to compare the results with the known exact distribution. The system of differential Kolmogorov equations for Markov process $\{k(t), i(t)\}$ is composed, then it is rewritten for characteristic functions in steady state.

The parameter $\rho = \frac{\lambda}{\mu}$ is the load rate of the system. The principle of asymptotic analysis method under heavy load condition (where $\rho \uparrow 1$) is briefly described. Asymptotic characteristic functions of the first and second order are presented.

Note that during the derivation of the first order asymptotic formula the function $F_0(w)$ was obtained. We propose to study "intermediate" asymptotics $h^*(u)$, which has the form of a weighted sum of two gamma distribution characteristic functions:

$$h^*(u) = q\Gamma_1(u) + (1-q)\Gamma_0(u),$$

$$\text{where } \Gamma_1(u) = \left(1 - \frac{ju}{\beta}\right)^{-\alpha}, \quad \Gamma_0(u) = \left(1 - \frac{ju}{\beta}\right)^{-\alpha+1}, \quad q = \frac{1}{2-\rho}, \quad \alpha = \frac{\mu}{\sigma} + 1, \quad \text{и } \beta = 1 - \rho.$$

We suggest to call this distribution as *hypergamma distribution* (like a hyperexponential distribution).

The numerical analysis shows that hypergamma asymptotics is more accurately than the first order asymptotics, but it is worth than the second order asymptotics. However, the derivation of the second order asymptotic formula for more complex retrial queues is difficult, whereas "intermediate" asymptotics does not require additional calculations, because its shape is the same for all retrial queues and hypergamma distribution parameters are obtained in the first order asymptotic formula theorem.

Then a modification of hypergamma asymptotics of shifting the argument to the parameter δ is proposed:

$$F(x) = q\tilde{\Gamma}_1(x + \delta) + (1-q)\tilde{\Gamma}_0(x + \delta),$$

where δ is found from the condition $\tilde{\Gamma}_1(\delta) = (1-\rho)^2$.

The numerical analysis shows that "modified" distribution significantly improves results of hypergamma asymptotics.

In the paper conclusion a combination of "modified" distribution and the second asymptotics $Pm^{(2)} = \frac{1}{2}(Pm + P^{(2)})$ is considered.

As a result of numerical analysis, it was observed that this modification is more accurately than the second order asymptotics and can be applied for load rate $\rho \geq 0.7$ and even $\rho \geq 0.5$ for large values of delay parameter σ .

In summary, it can be concluded that for practical tasks which does not require high accuracy of approximation, it is useful to apply the "intermediate" asymptotics results because of its simplicity of construction and rather broad area of applicability ($\rho \geq 0.8$), and for applications requiring higher accuracy of approximation, it is advisable to apply the proposed modifications 1 and 2. In the future, the modifications can be applied to the study of various retrial queueing systems including with non Poisson incoming process.

REFERENCES

1. Wilkinson, R.I. (1956) Theories for toll traffic engineering in the USA. *The Bell System Technical Journal*. 35(2). pp. 421-507. DOI: 10.1002/j.1538-7305.1956.tb02388.x.
2. Cohen, J.W. (1957) Basic problems of telephone traffic and the influence of repeated calls. *Philips Telecommunication Review*. 18(2). pp. 49-100.
3. Gosztony, G. (1976) Repeated call attempts and their effect on traffic engineering. *Budavox Telecommunication Review*. 2. pp. 16-26.
4. Elldin, A. & Lind, G. (1971) *Elementary Telephone Traffic Theory*. Ericsson Public Telecommunications.
5. Kuznetsov, D.Yu., Nazarov, A.A. (2001) Analysis of non-Markovian models of communication networks with adaptive protocols of multiple random access. *Automation and Remote Control*. 62(5). pp. 124-146. DOI: 10.1023/A:1010231008994.
6. Nazarov, A.A. & Tsoj, S.A. (2004) Common approach to studies of Markov models for data transmission networks controlled by the static random multiple access protocols. *Avtomatika i Vychislitel'naya Tekhnika – Automatic Control And Computer Sciences*. 4. pp. 73-85. (In Russian).
7. Artalejo, J.R. & Gomez-Corral, A. (2008) *Retrial Queueing Systems. A Computational Approach*. Springer.
8. Falin, G.I. & Templeton, J.G.C. (1997) *Retrial queues*. London: Chapman & Hall.
9. Artalejo, J.R. & Falin, G.I. (2002) Standard and retrial queueing systems: A comparative analysis. *Revista Matematica Complutense*. 15. pp. 101-129. DOI: 10.5209/rev_REMA.2002.v15.n1.16950
10. Neuts, M.F. & Rao, B.M. (2002) Numerical investigation of a multiserver retrial model. *Queueing Systems*. 7(2). pp. 169-189. DOI: 10.1007/BF01158473
11. Ridder, A. (2000) Fast simulation of retrial queues. *Third Workshop on Rare Event Simulation and Related Combinatorial Optimization Problems*. Pisa. pp. 1-5.
12. Kim, C.S., Mushko, V.V. & Dudin, A. (2012) Computation of the steady state distribution for multi-server retrial queues with phase type service process. *Annals of Operations Research*. 201(1). pp. 307-323. DOI: 10.1007/s10479-012-1254-7
13. Moiseeva, E. & Nazarov, A. (2012) Asymptotic Analysis of RQ-systems M/M/1 on Heavy Load Condition. *Problems of Cybernetics and Informatics*. Proc. of the Fourth International Conference. Baku, Azerbaijan. pp. 164-166.
14. Nazarov, A. & Fedorova, E. (2013) [The second order asymptotic analysis method under heavy load condition on example of the retrial queue M|M|1 research]. *Informatsionnye tekhnologii i matematicheskoe modelirovanie (ITMM-2013)* [Information technology and mathematical modeling (ITMM 2013)]. Proc. of the 12th All-Russian Conference with International Participation. Vol. 2. Tomsk. pp. 59-65. (In Russian).
15. Moiseeva, E. & Nazarov, A. (2013) Researching of Retrial Queueing system MMPP[GI]1 by using asymptotic analysis method on heavy load condition. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(25). pp. 84-94. (In Russian).

УДК 519.254

DOI: 10.17223/19988605/37/7

О.В. Ниссенбаум, А.М. Харченко

АДАПТИВНАЯ ОЦЕНКА ВЕСОВОГО КОЭФФИЦИЕНТА В АЛГОРИТМЕ ДИНАМИЧЕСКОЙ КЛАСТЕРИЗАЦИИ ДАННЫХ

Рассматривается адаптивный алгоритм определения коэффициентов затухающей оконной модели динамического ЕМ-алгоритма кластеризации потока данных. Алгоритм предназначен для кластеризации данных с нормальным распределением в R^n , параметры которого изменяются во времени, что соответствует ситуации в реальных динамических системах, таких как компьютерные системы, сети связи и т.п. Для расчета весовых коэффициентов требуется хранение ограниченного объема данных, алгоритм эффективно вычислим, может применяться в системах реального времени. Приведены данные вычислительного эксперимента (на имитационной модели потока).

Ключевые слова: алгоритм кластеризации; поток данных; системы реального времени; затухающая оконная модель; адаптивный алгоритм.

Интеллектуальный анализ потоков данных на протяжении последних десятилетий вызывает живой интерес исследователей по всему миру. Кластерный анализ – эффективный инструмент анализа данных, закономерными являются многочисленные попытки его применения к потокам данных. В последние годы был предложен ряд алгоритмов динамической кластеризации, учитывающих специфику обработки потоков: высокую размерность, большой объем, постоянно поступающие измерения, которые требуют эффективных, работающих в режиме реального времени алгоритмов [1].

Впервые проблема кластеризации потоков данных, по-видимому, была затронута в 1980-е гг. в [2], а соответствующая модель формализована в 1998 г. [3]. В настоящее время публикуется множество работ, посвященных созданию и применению алгоритмов кластеризации к потокам данных, например [4–16].

Область применения динамической кластеризации данных достаточно широка. Такие алгоритмы востребованы при наблюдении за сложными системами, такими как информационные системы, компьютерные сети, в долговременных популяционных, когнитивных, социологических исследованиях в режиме реального времени [17–21]. Обработка больших объемов информации в режиме реального времени требует эффективных, быстродействующих алгоритмов, а также сжатия данных – необходимо отказаться от хранения всего массива данных (измерений). При этом информация, полученная недавно, зачастую имеет большую значимость, нежели старые данные.

Еще одной важной задачей при разработке алгоритмов машинного обучения является сокращение количества параметров алгоритма (user-defined parameter) путем разработки адаптивных механизмов [1]. Такой механизм для определения параметра затухающего временного окна предложен в данной работе и может быть применен не только для описанного здесь алгоритма кластеризации [22], но и других, таких как DenStream [23].

1. Динамический ЕМ-алгоритм для потока данных, взвешенных по времени

В [22] на основе [24] путем применения затухающей оконной модели построен алгоритм, кластеризующий потоки данных, причем каждой точке соответствует весовой коэффициент $w = w(\Delta t)$, являющийся функцией от времени, прошедшего с момента поступления измерения (точки), а вес кластера $W(t)$ является суммой весов всех входящих в него точек. В качестве весовой функции выбрана $w(\Delta t) = e^{-a\Delta t}$ ($a \geq 0$), обладающая следующими свойствами:

- 1) в момент возникновения точка имеет вес, равный 1: $w(0) = 1$;

2) с течением времени вес точки монотонно убывает к 0: $\lim_{\Delta t \rightarrow \infty} w(\Delta t) = 0$;

3) если известен вес точки через Δt_1 ед. вр. после ее возникновения, то вес этой же точки еще через Δt_2 ед. вр. рассчитывается по формуле $w(\Delta t_1 + \Delta t_2) = w(\Delta t_1)w(\Delta t_2)$;

4) если известен вес кластера в момент времени t и за последующий период времени длительности Δt в кластер не попадало новых точек, то его вес пересчитывается по формуле $W(t + \Delta t) = W(t)w(\Delta t)$.

Такой выбор весовой функции, во-первых, позволяет обрабатывать динамические данные в режиме реального времени; во-вторых, не требует хранения обработанных данных; в-третьих, учитывает «новые» данные с большими весами, чем «старые». Следует отметить, что подобный выбор весовой функции, обеспечивающей затухающее временное окно, является достаточно распространенной практикой и применен независимо, например в [23] (с точностью до основания).

Пусть в некоторый момент времени набор измерений (назовем их *исходными данными*) разбит классическим ЕМ-алгоритмом на множество кластеров. Каждый кластер C_k представлен функцией плотности нормального распределения в пространстве характеристик:

$$\varphi(x|\mu, \Sigma) = \{(2\pi)^d |\Sigma| \exp[(x-\mu)^T \Sigma^{-1} (x-\mu)]\}^{-1/2}, \quad (1)$$

где μ и Σ – координаты центра и ковариационная матрица кластера, рассчитанные согласно соответствующим формулам для нормального распределения, $x \in \mathbf{R}^d$ – координаты новой точки, индекс k – номер кластера, для простоты опущен. Весовые коэффициенты кластеров в начальный момент времени устанавливаются равными количеству точек в них.

Пусть теперь $\{x_1, x_2, \dots\} \in \mathbf{R}^d$ – *поток данных*, т.е. точки, поступающие в последовательные моменты времени $t_1, t_2, \dots$. В момент поступления очередной точки x она относится к одному из кластеров методом максимального правдоподобия. Вероятность π_k принадлежности точки x к кластеру C_k с весом W_k определяется формулой

$$\pi_k = \frac{W_k \varphi_k(x|\mu_k, \Sigma_k)}{\sum_{i=1}^K W_i \varphi_i(x|\mu_i, \Sigma_i)}. \quad (2)$$

При появлении в потоке очередной точки x выполняется следующий алгоритм.

Вход: x – новая точка, μ_k, Σ_k, W_k – характеристики и веса кластеров ($k = 1, 2, \dots, K$), Δt – время, прошедшее с момента получения предыдущей точки.

1. Пересчитать веса кластеров $W_k = e^{-a\Delta t} W_k$.

2. Рассчитать вероятности попадания точки x в кластеры согласно (2). Отнести x к одному из кластеров методом максимального правдоподобия.

3. Для этого кластера произвести пересчет центра μ , ковариационной матрицы Σ и веса W :

$$\mu_+ = \frac{W\mu_- + x}{W+1}; \Sigma_+ = \frac{W}{W+1} \left(\Sigma_- + \frac{(\mu_- - x)^2}{W+1} \right); W = W+1,$$

где индексы «-» и «+» соответствуют значениям до и после пересчета, индекс k опущен.

Заметим, что весовая функция содержит параметр $a \geq 0$, причем при $a = 0$ разработанный алгоритм совпадает с [24]. Данный параметр влияет на скорость устаревания точки: при его возрастании точка устаревает скорее. Очевидно, он может быть определен отдельно для каждого кластера и зависит от множества факторов: скорости изменения параметров кластера, частоты поступления точек и т.п.

2. Адаптивный алгоритм расчета параметра весовой функции

Назовем точку, полученную в момент времени t_0 , незначимой в момент времени $t = t_0 + \Delta t$, если ее вес ниже наперед заданного малого значения ε ($0 < \varepsilon \ll 1$): $w(\Delta t) = e^{-a\Delta t} \leq \varepsilon$. Откуда получаем $a = -\ln \varepsilon / \Delta t$, где Δt – время, за которое точка должна потерять значимость. Определим Δt как время, за которое математическое ожидание точек кластера выйдет за пределы доверительного эллипсоида с доверительной вероятностью p , определенного в момент t_0 . Положим, что за время Δt скорость движения центра кластера $\vec{v}$ изменится незначительно, так что можно принять ее постоянной. Тогда, зная направление скорости, рассчитаем r – расстояние от центра кластера до границы доверительного эллипсоида с заданной дове-

рительной вероятностью в этом направлении, и тогда $\Delta t = r/|\vec{v}|$, а оценка для параметра весовой функции будет вычислена как

$$a = -|\vec{v}| \ln \varepsilon / r. \quad (3)$$

Заметим, что при использовании (3) весовая функция W не зависит от единиц измерения времени и расстояния.

Построим оценку для $\vec{v}$, используя $x_1, x_2, \dots, x_N$ – последние N точек, попавшие в данный кластер, и моменты их поступления $t_1, t_2, \dots, t_N$. Пусть t_0 – момент времени, когда в кластере была получена точка, предшествующая известным, а μ_0 и Σ_0 – центр кластера и его ковариационная матрица в этот момент. Тогда, поскольку каждая из точек x_i является реализацией случайной величины, которая, в свою очередь, есть сумма нормально распределенной многомерной величины и неслучайного смещения, зависящего от скорости центра кластера и времени, можем заключить, что $x_i = \xi_i + \vec{v}(t_i - t_0)$, где ξ_i – реализация нормально распределенной случайной величины со средним μ_0 и ковариационной матрицей Σ_0 , $i = 1, 2, \dots, N$. Тогда, обозначив среднее для N последних точек как μ , имеем

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{\sum_{i=1}^N \xi_i}{N} + \vec{v} \frac{\sum_{i=1}^N (t_i - t_0)}{N}.$$

Среднее всех ξ_i есть оценка для μ_0 , поэтому для $\vec{v}$ можно построить оценку

$$\vec{v} = \frac{N(\mu - \mu_0)}{\sum_{i=1}^N (t_i - t_0)}, \quad (4)$$

приняв в качестве μ_0 оценку центра кластера в момент времени t_0 .

Для пересчета a предложим следующий подход. В начальный момент времени можно принять $a = 0$, т.е. предположить, что центр кластера не смещается. При попадании в кластер каждой точки накапливать значения μ и знаменателя в (4), а при попадании каждой N -й точки – пересчитывать a согласно (4) и (3). Такой подход позволяет не хранить обработанные точки для алгоритма пересчета a так же, как и для алгоритма кластеризации.

Относительно выбора параметра N можно сделать следующие предположения: N должно быть достаточным для репрезентативности выборки; с другой стороны, при увеличении N снижается частота пересчета коэффициента a . Таким образом, выбор N определяется желаемой частотой пересчета параметра затухания a и ограничен снизу минимально необходимым объемом выборки. Результаты экспериментов, иллюстрирующие вопрос качества оценивания скорости в зависимости от выбора N , приведены в третьем пункте.

В случае возможности выделить для каждого кластера хранилище точек объема N ; предложим другой подход к пересчету параметра a : при поступлении в кластер новой точки она добавляется в хранилище вместо самой старой, значение a пересчитывается согласно (4) и (3). В этом случае пересчет параметра затухания происходит при поступлении в кластер каждой точки, и N ограничено лишь выделенным объемом хранилища точек, с одной стороны, и минимальным значением для репрезентативности выборки – с другой.

3. Вычислительный эксперимент

Разработанный алгоритм был реализован на C#. Было проведено несколько серий экспериментов, в каждом из которых моделировалось 9 000 точек в одном кластере, центр которого смещался с постоянной скоростью, а ковариационная матрица оставалась неизменной. Рассчитывались оценки скорости согласно (4) при различных N . В момент поступления каждой точки производилась оценка скорости $\hat{v}$, для скаляра которой рассчитывалась относительная погрешность $\delta = |v - \hat{v}|/v$, были получены выборочные средние $\bar{\hat{v}}$, $\bar{\delta}$ среднеквадратические отклонения σ_v , σ_δ . Перечисленные выборочные характеристики при $|\vec{v}| = 0,04$ ед.расст./ед.вр. и постоянной ковариационной матрице $\Sigma = \begin{pmatrix} 400 & 0 \\ 0 & 400 \end{pmatrix}$ приведены в табл. 1.

С увеличением N погрешность, в среднем, уменьшается, как и ее среднеквадратическое отклонение.

Результаты эксперимента при $\nu = 0,04$

N	100	250	500
$\bar{\nu}$	0,0724	0,0394	0,0346
σ_{ν}	0,0196	0,0079	0,0044
$\bar{\delta}$	0,8281	0,1682	0,1544
σ_{δ}	0,4574	0,1041	0,0820

Траектория скаляра оценки скорости в сравнении с траекторией истинного его значения, для упомянутого эксперимента при $N = 500$ приведен на рис. 1.



Рис. 1. Траектории скаляров оценки скорости (черная линия) и его истинного значения (серая линия) при $N = 500$

Был проведен вычислительный эксперимент, в котором для одних и тех же исходных параметров было смоделировано 30 комплектов данных, полученных в трех кластерах с нормальным распределением в $\mathbf{R}^2$, причем местоположение центров кластеров изменялось со временем.

Начальное положение центра первого кластера $\mu_1 = (650, 700)$, ковариационная матрица диагональная, $\sigma_{11}^1 = \sigma_{22}^1 = 700$ (здесь верхний индекс – номер кластера), скорость движения центра $v_1 = 7$ ед.расст./ед.вр. Для второго кластера $\mu_2 = (100, 300)$, ковариационная матрица также диагональная, $\sigma_{11}^2 = \sigma_{22}^2 = 400$, $v_2 = 5$ ед.расст./ед.вр, для третьего аналогично $\mu_3 = (550, 500)$, $\sigma_{11}^3 = \sigma_{22}^3 = 200$, $v_3 = 4$ ед.расст./ед.вр, Направления скоростей задавались случайным образом. Точки моделировались с интервалом в $1/3$ ед.вр. с равными вероятностями попадания в каждый из кластеров. Обработка данных производилась предложенным алгоритмом, пересчет параметра a производился согласно (3) и (4). В момент окончания расчетов для каждого эксперимента было вычислено отношение отклонения оцененного центра кластера от истинного к истинному среднеквадратическому отклонению $\Delta\mu_k/\sigma_{11}^k$. Гистограммы их частот по всем 30 экспериментам приведены на рис. 2.

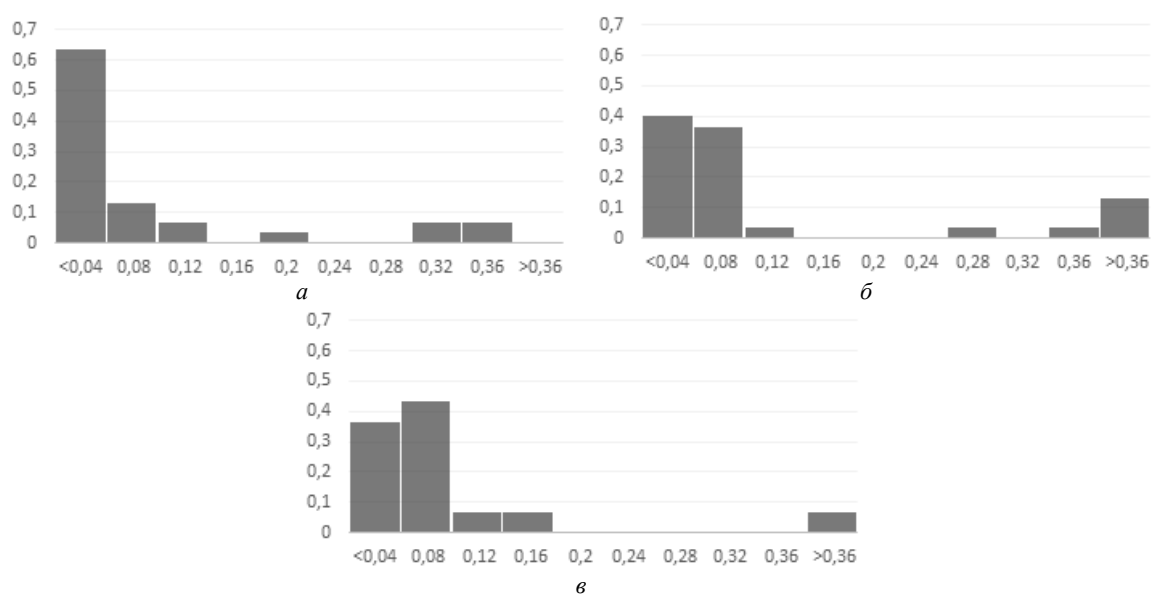


Рис. 2. Гистограммы частот относительных отклонений центров кластеров:
а – первого; б – второго; в – третьего в момент окончания эксперимента

На рис. 3 приведена визуализация одного из 30 результатов. Эллипсами обозначены доверительные области кластеров с вероятностью 0,9 (пунктирная линия – истинные, сплошная – полученные в предложенном алгоритме) в момент окончания эксперимента.

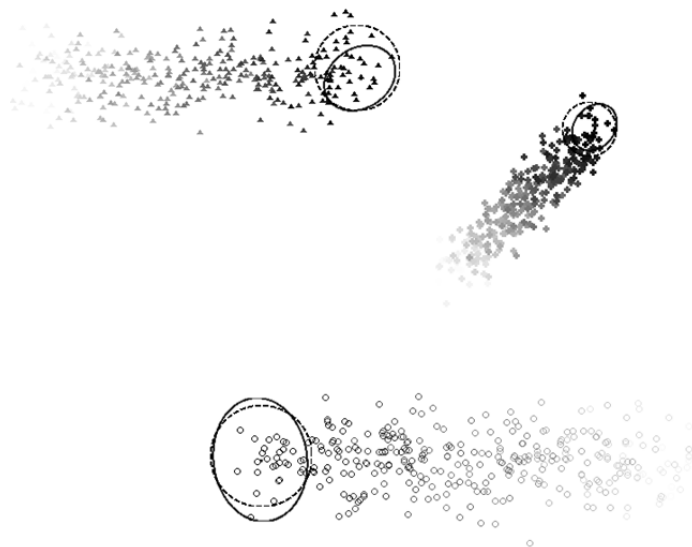


Рис. 3. Точки 1-го кластера обозначены кружками, 2-го – треугольниками, 3-го – крестиками; интенсивность цвета отражает новизну измерения (темный – новые точки, светлый – старые)

Заключение

Разработанный алгоритм нетребователен к ресурсам (время, память) и пригоден для оперативного мониторинга в больших динамических системах, таких как компьютерные системы и сети. Вычислительный эксперимент показал приемлемое качество его работы на имитационной модели.

ЛИТЕРАТУРА

1. Ding S., Wu F., Qian J., Jia H., Jin F. Research on data stream clustering algorithms // *Artif Intell Rev.* 2015. V. 43. P. 593–600.
2. Munro J., Paterson M. Selection and Sorting with Limited Storage // *Theoretical Computer Science.* 1980. P. 315–323.
3. Henzinger M., Raghavan P., Rajagopalan S. Computing on Data Streams // *Digital Equipment Corporation. SRC TN-1998-011.* August 1998.
4. Ackerman M., Martens R., Raupach C., Swierkot K., Lammersen C., Sohler C. StreamKM++: A clustering algorithm for data streams // *ACM J. Exper. Algor.* 2012. V. 17, No. 1.
5. Aggarwal C.C. A segment-based framework for modeling and mining data streams // *Knowl. Inf. Syst.* 2010. V. 30, No. 1. P. 1–29.
6. Csernel B., Clerot F., Hebrail G. Data stream clustering over tilted windows through sampling // *Proceedings of the Knowledge Discovery from Data Streams Workshop (ECML/PKDD).* 2006.
7. Dang X.H., Lee V.C.S., Ng W.K., Ciptadi A., Ong K.-L. An EM-based algorithm for clustering data streams in sliding windows // *Proceedings of the 14th International Conference on Database Systems for Advanced Applications. Lecture Notes in Computer Science.* Springer, 2009. V. 5463. P. 230–235.
8. Khalilian M., Mustapha N. Data stream clustering: challenges and issues // *Proceedings of International Multi Conference of Engineers and Computer Scientists.* 2010. P. 566–569.
9. Li Y., Tan B.H. Data stream clustering algorithm based on affinity propagation and density // *Advanced Materials Res.* 2011. V. 267. P. 444–449.
10. Ong K.L., Li W., Ng W.-K., Lim E.-P. SCLOPE: An algorithm for clustering data streams of categorical attributes // *Proceedings of the 6th International Conference on Data Warehousing and Knowledge Discovery (KDD'04).* 2004. P. 209–218.
11. Silva J.A., Hruschka E.R. Extending k-means-based algorithms for evolving data streams with variable number of clusters // *Proceedings of the 4th International Conference on Machine Learning and Applications (ICMLA'11).* 2011. V. 2. P. 14–19.
12. Cao F., Zhou A.Y. Fast clustering of data streams using graphics processors // *Journal of Software.* 2007. V. 18, No. 2. P. 291–302.
13. Zhu W.H., Yin J., Xie Y.H. Arbitrary shape cluster algorithm for clustering data stream // *Journal of Software.* 2006. V. 17, No. 3. P. 379–387.
14. Chandrika J., Ananda Kumar K.R. Dynamic Clustering Of High Speed Data Streams // *International Journal of Computer Science Issues.* 2012. V. 9, Issue 2, No. 1. P. 224–228.

15. Vorontsov K., Frei O., Apishev M., Romov P., Suvorova M., Yanina A. Non-bayesian additive regularization for multimodal topic modeling of large collections // Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications. 2015. P. 29–37.
16. Spirin N., Vorontsov K. Learning to rank with nonlinear monotonic ensemble // International Workshop on Multiple Classifier Systems: Springer Berlin Heidelberg, 2011. P. 16–25.
17. Qian Quan, Chao-Jie Xiao, Rui Zhang. Grid-based Data Stream Clustering for Intrusion Detection // International Journal of Network Security. 2013. V. 15, No. 1. P. 1–8.
18. Daniel A., Sternberg, Kacey Ballard, Joseph L. Hardey et al. The Largest Human Cognitive Performance Dataset Reveals Insights into the Effects of Lifestyle Factors and Aging // Frontiers in Human Neuroscience. 2013. V. 7. Article 292.
19. Zhang X., Sebag M., Germain-Renaud C. Multi-scale real-time grid monitoring with job stream mining // Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'09). 2009. P. 420–427.
20. Shurygin A.M., Strigunova M.S. Prediction of strong earthquakes with the use of scalar product of matrices with a sliding window // Pattern Recognition and Image Analysis (Advances in Mathematiccal Theory and Applications). 2007. V. 17, No. 1. P. 153–162.
21. Ниссенбаум О.В., Присяжнюк А.С. Адаптивный алгоритм отслеживания аномальной активности в компьютерной сети на основании характерных изменений оценок альтернирующего потока // Прикладная дискретная математика. Приложение. 2010. № 3. С. 55–58.
22. Ниссенбаум О.В. Алгоритм кластеризации потока данных с изменяющимися параметрами распределения // Вестник Тюменского государственного университета. 2013. № 7. С. 180–186.
23. Cao F., Ester M., Qian W., Zhou A. Density-based clustering over an evolving data stream with noise // Proceedings of the 6th SIAM International Conference on Data Mining. 2006. P. 328–339.
24. Mingzhou Song, Hongbin Wang. Highly efficient incremental estimation of Gaussian mixture models for online data stream clustering // Proceedings of SPIE 5803. 2005. P. 174–183.

Ниссенбаум Ольга Владимировна, канд. физ.-мат. наук. E-mail: onissenbaum@rambler.ru

Харченко Анастасия Михайловна. E-mail: kharchenkoam.86@gmail.com

Тюменский государственный университет

Поступила в редакцию 4 июня 2016 г.

Nissenbaum Olga V., Kharchenko Anastasia M. (Tyumen State University, Russian Federation).

Adaptive Estimation of Weight Coefficient in a Time-weighted Incremental EM-algorithm for Data Streams.

Keywords: clustering; time-weight; data stream; Gaussian mixture model; damped window model.

DOI: 10.17223/19988605/37/7

Application of data stream clustering is quite general. Applications of data streams include mining data generated by sensor networks, meteorological analysis, stock market analysis, computer network traffic monitoring, long-term population-based, cognitive, real-time sociological studies etc. These applications involve datasets that are far too large to fit in main memory and are typically stored in a secondary storage device. Real-time processing of large volumes of data requires efficient, fast algorithms and data compression. Moreover, recently obtained objects are often more important than the old ones, so several clustering algorithms implement a damped window model with a user-defined decay coefficient. This coefficient is not obvious. An important task in the development of machine learning algorithms is to reduce the number of algorithm parameters (user-defined parameter) through the development of adaptive mechanisms. Such a mechanism for determining the decay coefficient proposed in this paper.

A Gaussian mixture data stream algorithm is constructed with damped window model. Decay coefficient for the object according damped window is $w=w(\Delta t)$ where Δt is time since object arrived from the stream into cluster, weight of the cluster $W(t)$ is sum of all its objects. Recent objects receive higher weight than older ones, and the weights of the objects decrease with time. It is usually to assume $w(\Delta t)=e^{-a\Delta t}$ ($a \geq 0$), with following properties:

- 1) when arrives, the object weight is equal to one: $w(0)=1$;
- 2) over time object weight monotonic decreases to zero: $\lim_{\Delta t \rightarrow \infty} w(\Delta t) = 0$;

3) if weight of object $w(\Delta t_1)$ after Δt_1 units of time is known, then after more Δt_2 it can be easily recalculates as $w(\Delta t_1 + \Delta t_2) = w(\Delta t_1)w(\Delta t_2)$;

4) if weight of cluster $W(t)$ for the instant in time t is known, and during the next time period Δt there was no new objects arrived in this cluster, then weight of this cluster can be recalculates as $W(t + \Delta t) = W(t)w(\Delta t)$.

Let at initial instant in time we have a set of K Gaussian clusters C_k that described by their means μ_k and correlation matrixes Σ_k . Weight W_k at initial time defined as a number of objects in cluster C_k . Let $\{x_1, x_2, \dots\} \in \mathbf{R}^d$ is *data stream*, i.e. set of objects with timestamps $t_1, t_2, \dots$. When a new object x with timestamp t arrives to the cluster C_k , the following algorithm executes.

Input: new object x , cluster parameters: mean μ_k , correlation matrix Σ_k , weight W_k ($k=1, 2, \dots, K$), time period since last object arrives Δt .

1. Recalculate weights $W_k = e^{-a\Delta t} W_k$ ($k=1, 2, \dots, K$).

2. Calculate probabilities of belonging x to the k -th cluster $\pi_k = \frac{W_k \varphi_k(x | \mu_k, \Sigma_k)}{\sum_{i=1}^K W_i \varphi_i(x | \mu_i, \Sigma_i)}$. Put x into the most probable cluster.

3. For this cluster (in with x is, index of cluster is dropped) recalculate mean μ , correlation matrix Σ and weight W :

$$\mu_+ = \frac{W\mu_- + x}{W+1}; \Sigma_+ = \frac{W}{W+1} \left(\Sigma_- + \frac{(\mu_- - x)^2}{W+1} \right); W=W+1,$$

where indexes - and + corresponds to the values before and after recalculation.

Note that weight function depends on $a \geq 0$, that may be defined individually for each C_k . We propose the following algorithm to recalculate a separately for each cluster at the time the object arrives into cluster.

Input: $x_1, x_2, \dots, x_N$ – last N objects putted into cluster, $t_1, t_2, \dots, t_N$ – their timestamps, μ_0 – cluster mean Σ_0 – cluster correlation matrix.

1. At initial instance of time t_0 adopt $a=0$, i.e. we assume that cluster is not moving.

2. When object x_i with timestamp t_i ($i \leq N$) arrives into the cluster, accumulate variables $\mu = (x_1 + x_2 + \dots + x_N)/N$ and $\sum_{i=1}^N (t_i - t_0)$.

3. If $i=N$ calculate $\vec{v} = \frac{N(\mu - \mu_0)}{\sum_{i=1}^N (t_i - t_0)}$ using accumulated variables. Calculate $a = -|\vec{v}| \ln \varepsilon / r$, where r is Euclidean distance between

μ_0 and confidence ellipse boundary (confidence probability $p = 1 - \varepsilon$) in direction of $\vec{v}$ vector. Set $t_0 = t_N$, assign μ_0 and Σ_0 a corresponding values of cluster at the time t_N and return to step 1.

We perform an experiment using an imitation model of Gaussian mixture clusters with moving centroids. Some results are shown in this article. Proposed algorithm is undemanding to resources (time, memory) and therefore is suitable for real-time monitoring in large dynamic systems, such as computer systems and networks. Quality of decay coefficient adaptation is acceptable as follows from the experimental data.

REFERENCES

1. Ding, S., Wu, F., Qian, J., Jia, H. & Jin, F. (2015) Research on data stream clustering algorithms. *Artif Intell Rev.* 43. pp. 593-600. DOI: 10.1007/s10462-013-9398-7
2. Munro, J. & Paterson, M. (1980) Selection and Sorting with Limited Storage. *Theoretical Computer Science.* pp. 315-323. DOI: 10.1109/SFCS.1978.32
3. Henzinger, M., Raghavan, P. & Rajagopalan, S. (1998) Computing on Data Streams. *Digital Equipment Corporation. SRC TN-1998-011.*
4. Ackerman, M., Martens, R., Raupach, C., Swierkot, K., Lammersen, C. & Sohler, C. (2012) StreamKM++: A clustering algorithm for data streams. *ACM J. Exper. Algor.* 17(1). DOI: 10.1145/2133803.2184450
5. Aggarwal, C.C. (2010) A segment-based framework for modeling and mining data streams. *Knowl. Inf. Syst.* 30(1). pp. 1-29. DOI: 10.1007/s10115-010-0366-0
6. Csernel, B., Clerot, F. & Hebrail G. (2006) Data stream clustering over tilted windows through sampling. *Proceedings of the Knowledge Discovery from Data Streams Workshop (ECML/PKDD).*
7. Dang, X.H., Lee, V.C.S., Ng, W.K., Ciptadi, A. & Ong, K.-L. (2009) An EM-based algorithm for clustering data streams in sliding windows. *Proc. of the 14th International Conference on Database Systems for Advanced Applications.* Lecture Notes in Computer Science. Vol. 5463 Springer. pp. 230-235.
8. Khalilian, M. & Mustapha, N. (2010) Data stream clustering: challenges and issues. *Proc. of International Multi Conference of Engineers and Computer Scientists.* pp. 566-569.
9. Li, Y. & Tan, B.H. (2011) Data stream clustering algorithm based on affinity propagation and density. *Advanced Materials Res.* 267. pp. 444-449. DOI: 10.4028/www.scientific.net/AMR.267.444.
10. Ong, K.L., Li, W., Ng, W.-K. & Lim, E.-P. (2004) SCLOPE: An algorithm for clustering data streams of categorical attributes. *Proc. of the 6th International Conference on Data Warehousing and Knowledge Discovery (KDD'04).* pp. 209-218.
11. Silva, J.A. & Hruschka, E.R. (2011) Extending k-means-based algorithms for evolving data streams with variable number of clusters. *Proc. of the 4th International Conference on Machine Learning and Applications (ICMLA'11).* 2. pp. 14-19. DOI: 10.1109/ICMLA.2011.67.
12. Cao, F. & Zhou, A.Y. (2007) Fast clustering of data streams using graphics processors. *Journal of Software.* 18(2). pp. 291-302. DOI: 10.1360/jos180291
13. Zhu, W.H., Yin, J. & Xie, Y.H. (2006) Arbitrary shape cluster algorithm for clustering data stream. *Journal of Software.* 17(3). pp. 379-387.
14. Chandrika, J. & Ananda Kumar, K.R. (2012) Dynamic Clustering Of High Speed Data Streams. *International Journal of Computer Science Issues.* 9(2-1). pp.224-228. DOI: 10.1109/ICDE.2009.13.
15. Vorontsov, K., Frei, O., Apishev, M., Romov, P., Suvorova, M. & Yanina, A. (2015) Non-bayesian additive regularization for multimodal topic modeling of large collections. *Proc. of the 2015 Workshop on Topic Models: Post-Processing and Applications.* pp. 29-37. DOI: 10.1145/2809936.2809943
16. Spirin, N. & Vorontsov K. (2011) Learning to rank with nonlinear monotonic ensemble. *International Workshop on Multiple Classifier Systems.* Naples, Italy – June 15 - 17, 2011. Berlin, Heidelberg: Springer. pp.16-25.
17. Qian, Q., Chao-Jie, X. & Rui Zhang. (2013) Grid-based Data Stream Clustering for Intrusion Detection. *International Journal of Network Security.* 15(1). pp. 1-8.
18. Sternberg, D.A., Joseph, Ballard, K. & Hardey, J.L. et al. (2013) The Largest Human Cognitive Performance Dataset Reveals Insights into the Effects of Lifestyle Factors and Aging. *Frontiers in Human Neuroscience.* 7. DOI: 10.3389/fnhum.
19. Zhang, X., Sebag, M. & Germain-Renaud, C. (2009) Multi-scale real-time grid monitoring with job stream mining. *Proc. of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'09).* pp. 420-427. DOI: 10.1109/CCGRID.2009.20

20. Shurygin, A.M. & Strigunova, M.S. (2007) Prediction of strong earthquakes with the use of scalar product of matrices with a sliding window. *Pattern Recognition and Image Analysis (Advances in Mathematiccal Theory and Applications)*. 17(1). pp. 153-162. DOI: 10.1134/S1054661807010191
21. Nissenbaum, O.V. & Prisyagnyuk, A.C. (2010) Adaptive algorithm for anomalous network traffic indication based on alternating process. *Prikladnaya diskretnaya matematika – Applied Discrete Mathematics. Applications*. 3. pp. 55-58. (In Russian).
22. Nissenbaum, O.V. (2013) Clustering algorithm for data streams with changing distribution parameters. *Vestnik Tyumenskogo gosudarstvennogo universiteta*. 7. pp. 180-186. (In Russian).
23. Cao, F., Ester, M., Qian, W. & Zhou, A. (2006) Density-based clustering over an evolving data stream with noise. *Proc. of the 6th SIAM International Conference on Data Mining*. pp. 328-339.
24. Song, M. & Wang, H. (2005) Highly efficient incremental estimation of Gaussian mixture models for online data stream clustering. *Proceedings of SPIE* 5803. pp.174-183. DOI: 10.1117/12.601724

В.П. Шуленин

АСИМТОТИЧЕСКИЕ СВОЙСТВА МОДИФИЦИРОВАННЫХ СРЕДНИХ РАЗНОСТЕЙ ДЖИНИ

Изучаются свойства робастных оценок масштабного параметра. Показано, что модифицированная оценка средних разностей Джини имеет асимптотически нормальное распределение, является В-робастной оценкой и имеет ограниченную функцию влияния. Приводятся результаты сравнения оценок масштабного параметра в рамках гауссовской модели с засорением. Рассматривается адаптивный вариант предложенной оценки масштабного параметра.

Ключевые слова: масштабный параметр; робастные оценки; функция влияния; средняя разность Джини; U-статистики; адаптивные оценки.

В класс U -статистик Хёффдинга [1] входят многие конкретные оценки параметров, представляющие практический интерес. Обобщение класса U -статистик, описанное в работе [2], связанное с построением робастных оценок с ограниченными функциями влияния, приводит к рассмотрению U -статистик, основанных на урезанных выборках, что позволяет изучать многие известные в теории робастности оценки параметров с единых позиций и открывает широкие возможности для построения новых оценок. Например, выборочное α -урезанное среднее $\bar{X}_\alpha$, $0 \leq \alpha < 1/2$, обычно применяемое в качестве робастной оценки параметра положения случайной величины (с.в.) X с функцией распределения (ф.р.) F , традиционно рассматривалось и изучалось как представитель семейства L-оценок в виде линейных комбинаций порядковых статистик $X_{(1)}, \dots, X_{(n)}$ исходной выборки $X_1, \dots, X_n$. Доказанная ранее асимптотическая нормальность $\bar{X}_\alpha$ -оценок (см., например [3, 4]) также непосредственно следует и из результатов работы [2] (см. пример 8.6.15 в [6]). В качестве другого примера приведем выборочную оценку средних разностей Джини, которая используется в качестве оценки масштабного параметра, характеризующего степень разброса с.в. X , и записывается в виде $\hat{\Delta}_0 = [n(n-1)]^{-1} \sum |X_i - X_j|$. Эта оценка имеет асимптотически нормальное распределение [5, 6], её асимптотическая относительная эффективность по отношению к традиционно применяемой на практике оценке $\hat{S}_1(0) = \{n^{-1} \sum (X_i - \bar{X})^2\}^{1/2}$ стандартного отклонения очень высокая, в частности, при нормальном распределении Φ она равна $AOЭ_\Phi(\hat{\Delta}_0 : \hat{S}_1(0)) = 0,98$. И хотя $\hat{\Delta}_0$ -оценка подвержена меньшему влиянию выбросов в выборке, чем $\hat{S}_1(0)$ -оценка, обе они имеют неограниченные функции влияния Хампеля [4]. В качестве ещё одного примера оценок масштабного параметра приведём семейство интер- α -квантильных размахов $\hat{S}_3(\alpha) = [X_{(n-[an])} - X_{([an])}] / 2$, $0 < \alpha < 1/2$ [5, 6]. Эти оценки имеют ограниченные функции влияния, но их асимптотические эффективности по отношению к оценке стандартного отклонения при нормальном распределении Φ очень низкие. Например, для оценки интерквартильного размаха $\hat{S}_3(0,25)$ асимптотическая относительная эффективность равна $AOЭ_\Phi(\hat{S}_3(0,25) : \hat{S}_1(0)) = 0,37$. По этой причине в литературе (см., например, [5, 6]) рассматривают различные модификации оценок масштабного параметра с целью обеспечить и ограниченность функции влияния, и высокую эффективность по отношению к $\hat{S}_1(0)$ -оценке при нормальном распределении Φ . К таким оценкам относится рассматриваемая в данной работе модифицированная оценка средних разностей Джини $\hat{\Delta}_\alpha$, $0 \leq \alpha < 1/2$, определённая в (24). Эта оценка является

U -статистикой, основанной на урезанной выборке и её асимптотические свойства изучаются с использованием результатов работы [2]. В данном исследовании приводятся результаты сравнения $\hat{\Delta}_\alpha$ -оценок с другими оценками масштабного параметра в рамках гауссовской модели с засорением и предлагается адаптивный вариант $\hat{\Delta}_\alpha$ -оценок, для которых параметр α выбирается на основе информации, содержащейся в исходной выборке с использованием выборочной оценки функционала, характеризующего степень «затянутости хвостов» функции распределения F изучаемой случайной величины X .

1. U -статистики, основанные на урезанных выборках ($U_{\alpha\beta}$ -оценки)

Пусть $X_1, \dots, X_n$ – последовательность нормальных оценок распределения (н.о.р.) случайных величин с ф.р. $F(x)$ и плотностью $f(x)$, $x \in R^1$, $X_{(1)}, \dots, X_{(n)}$ – упорядоченная статистика исходной выборки $X_1, \dots, X_n$ и $X_{([\alpha n]+1)}, \dots, X_{(n-[\beta n])}$, обозначает $\alpha\beta$ -урезанную упорядоченную статистику выборки, где α и β – заданные пропорции урезания выборки, причем $0 \leq \alpha, \beta \leq 1/2$. Обозначим $n_{\alpha\beta} = n - [\alpha n] - [\beta n]$. Пусть задано «ядро» $h(X_1, \dots, X_m)$, $m < n$, которое является симметричной функцией своих аргументов. Множество m -наборов индексов $(i_1, \dots, i_m)$, удовлетворяющих условию $\{[\alpha n] + 1 \leq i_1 < \dots < i_m \leq n - [\beta n]\}$, обозначим через $C_{\alpha\beta}$, т.е.

$$C_{\alpha\beta} = \{ (i_1, \dots, i_m) : [\alpha n] + 1 \leq i_1 < \dots < i_m \leq n - [\beta n] \}. \quad (1)$$

Следуя работе [2], рассмотрим U -статистики, основанные на $\alpha\beta$ -урезанных выборках, которые определяются в виде

$$U_{n, \alpha\beta} = \left(\frac{n_{\alpha\beta}}{m} \right)^{-1} \sum_{C_{\alpha\beta}} h(X_{(i_1)}, \dots, X_{(i_m)}). \quad (2)$$

Ниже рассмотрим случай $\alpha = \beta$ и переобозначим множество индексов $C_{\alpha\beta}$ из (1) через

$$C_\alpha = \{ (i_1, \dots, i_m) : [\alpha n] + 1 \leq i_1 < \dots < i_m \leq n - [\alpha n] \}. \quad (3)$$

Далее введем следующие обозначения. Пусть $n_\alpha = n - 2[\alpha n]$, и определим ограниченную функцию $g_\alpha(x; F)$, $0 \leq \alpha < 1/2$, $x \in R^1$, в виде

$$g_\alpha(x; F) = \frac{I[F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha)]}{(1-2\alpha)^m} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \dots \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} h(x_1, \dots, x_{m-1}, x) \prod_{i=1}^{m-1} dF(x_i), \quad (4)$$

где $I[A]$ – индикатор события A . Далее пусть $\xi_\alpha = F^{-1}(\alpha)$, $\bar{\xi}_\alpha = F^{-1}(1-\alpha)$ и значения функции $g_\alpha(x; F)$ в точках ξ_α и $\bar{\xi}_\alpha$ обозначим соответственно в виде

$$A_\alpha = -g_\alpha(\xi_\alpha, F), \quad B_\alpha = g_\alpha(\bar{\xi}_\alpha, F). \quad (5)$$

Среднее значение и дисперсию $g_\alpha(X; F)$ соответственно обозначим в виде

$$U_\alpha(F) = M_F \{g_\alpha(X, F)\} = \frac{1}{(1-2\alpha)^m} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \dots \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} h(x_1, \dots, x_m) \prod_{i=1}^m dF(x_i) \quad (6)$$

и

$$\Delta_\alpha(F) = D_F \{g_\alpha(X, F)\} = \int_{\xi_\alpha}^{\bar{\xi}_\alpha} g_\alpha^2(x, F) dF(x) - U_\alpha^2(F). \quad (7)$$

Теорема 1. Предположим, что ф.р. F имеет плотность f , которая непрерывна и ограничена в точках ξ_α и $\bar{\xi}_\alpha$. Пусть функция $g_\alpha(x; F)$ непрерывна в точках ξ_α , $\bar{\xi}_\alpha$, и предположим, что для некоторых $a < \xi_\alpha$ и $b > \bar{\xi}_\alpha$ выполняется условие

$$\sup_{a \leq x_1, \dots, x_m \leq b} |h(x_1, \dots, x_m)| = M_0 < \infty. \quad (8)$$

Тогда $\sqrt{n}U_{n,\alpha}$ -статистика имеет асимптотически нормальное распределение, т.е. выполняется выражение, согласно которому закон ($L - Law$) распределения вероятностей случайной величины в фигурных скобках является стандартным нормальным

$$L\{\sqrt{n}[U_{n,\alpha} - U_\alpha(F)] / \sigma(U_\alpha, F)\} = N(0, 1) \text{ при } n \rightarrow \infty, \quad (9)$$

где $\sigma^2(U_\alpha, F)$ – асимптотическая дисперсия $\sqrt{n}U_{n,\alpha}$ -статистики, вычисляемая по формуле

$$\sigma^2(U_\alpha, F) = m^2 \{ \Delta_\alpha(F) + \alpha(1 - \alpha)(A_\alpha^2 + B_\alpha^2) + 2\alpha U_\alpha(F)(A_\alpha - B_\alpha) + 2\alpha^2 A_\alpha B_\alpha \}. \quad (10)$$

Доказательство. Схема доказательства этой теоремы основана на использовании результатов работы [2], в которой отмечается, что при выполнении предположений для плотности f , функции $g_\alpha(x; F)$ и ядра $h(x_1, \dots, x_m)$ условия (8) применимы леммы (2.1–2.4) этой работы и при этом выполняется выражение вида

$$U_{n,\alpha} = U_\alpha(F) + n^{-1} \sum_{i=1}^n IF(X_i; F, U_\alpha) + o_p(n^{-1/2}), \quad (11)$$

в котором $IF(x; F, U_\alpha)$ обозначает функцию влияния Хампеля [4] оценки $U_\alpha(F_n)$ функционала $U_\alpha(F)$, определенного формулой (6). Из выражения (11) следует асимптотическая нормальность $\sqrt{n}U_{n,\alpha}$ -статистик с использованием теоремы Слуцкого и центральной предельной теоремы с учетом, что элементы выборки $X_1, \dots, X_n$ являются последовательностью н.о.р. случайных величин [5, 6]. Справедливость формулы (10) проверяется непосредственно, путем вычисления функции влияния $IF(x; F, U_\alpha)$ с учетом, что

$$\int IF(x; F, U_\alpha) dF(x) = 0 \text{ и } \sigma^2(U_\alpha, F) = \int IF^2(x; F, U_\alpha) dF(x). \quad (12)$$

Стандартным способом убеждаемся (см.: [6. С. 195]), что функция влияния $IF(x; F, U_\alpha)$ имеет вид

$$IF(x; F, U_\alpha) = m \{ g_\alpha(x, F) - U_\alpha(F) + A_\alpha(\alpha - I[x \leq \xi_\alpha]) + B_\alpha(1 - \alpha - I[x \leq \bar{\xi}_\alpha]) \}. \quad (13)$$

Отметим важное обстоятельство. Учитывая, что функция $g_\alpha(x; F)$ ограничена, функция влияния $IF(x; F, U_\alpha)$ вида (13) также является ограниченной функцией и, следовательно, $U_{n,\alpha}$ -статистики являются В-робастными [4] и подвержены лишь ограниченному влиянию выбросов в выборке. Используя приведенное выражение (13) и вторую формулу в (12), после несложных преобразований получаем (10). Теорема доказана.

2. Средняя разность Джини и её модифицированный вариант

Обсуждение средней разности Джини как меры разброса случайных величин и её связь с кривой Лоренца приводятся в [7. С. 75]. Средняя разность Джини, как и медиана абсолютных разностей, относится ко второй группе функционалов, определяющих масштабный параметр (см. формулу (10.1.3) в [6]). В работах [5–7] функционал $T(F)$, определяющий среднюю разность Джини, записывают в разных вариантах. Обычно [7] его записывают в виде

$$\Delta_F = T(F) = \int \int |x - y| dF(x) dF(y). \quad (14)$$

Другая форма записи функционала $T(F)$ [5, 6] имеет вид

$$\Delta_F = T(F) = \int_0^1 F^{-1}(t)(4t - 2)dt = \int_0^1 F^{-1}(t)J(t)dt, \quad J(t) = 4t - 2, \quad 0 \leq t \leq 1. \quad (15)$$

Выборочная оценка $\tilde{\Delta}_0 = T(F_n)$ средней разности Джини Δ_F вида (14), построенная по выборке $X_1, \dots, X_n$ методом подстановки, записывается в виде

$$\tilde{\Delta}_0 = T(F_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|. \quad (16)$$

Если же при построении оценки не используются совпадающие индексы (i, j) , то оценку записывают в асимптотически эквивалентном варианте вида

$$\hat{\Delta}_0 = \frac{1}{n(n-1)} \sum_{i \neq j=1}^n |X_i - X_j|. \quad (17)$$

Отметим, что представление функционала $T(F)$ в виде (15) позволяет записать оценку в виде линейной комбинации порядковых статистик $X_{(1)}, \dots, X_{(n)}$ исходной выборки $X_1, \dots, X_n$, т.е. в виде

$$\hat{\Delta} = \sum_{i=1}^n a_{ni} X_{(i)}, \quad (18)$$

где весовые коэффициенты a_{ni} вычисляются по формуле

$$a_{ni} = \int_{(i-1)/n}^{i/n} J(t) dt = \int_{(i-1)/n}^{i/n} (4t-2) dt = \frac{2i-1-n}{n^2}, \quad (19)$$

и $J(t) = 4t - 2$, $0 \leq t \leq 1$, – функция, определяющая L -оценку [5, 6]. Отметим, что оценка $\hat{\Delta}_0 = T(F_n)$ средней разности Джини $\Delta_F = T(F)$, согласно теореме (7.1.25) в [6], асимптотически нормальна, т.е. выполняется выражение

$$L\{\sqrt{n}(\hat{\Delta}_0 - \Delta_F) / \sigma_F(\Delta_n)\} = N(0, 1) \text{ при } n \rightarrow \infty, \quad (20)$$

где асимптотическая дисперсия $\sqrt{n}\hat{\Delta}_0$ -оценки вычисляется по формуле

$$\sigma_F^2(\hat{\Delta}_0) = \int IF^2(x; F, \Delta_F) dF(x) = \int \varphi_F^2(x) dF(x) - \left(\int \varphi_F(x) dF(x) \right)^2, \quad (21)$$

в которой функция $\varphi_F(x)$ определяется в виде

$$\varphi_F(x) = 2x(2F(x) - 1) - 4\mu_F(x), \quad \mu_F(x) = \int_{-\infty}^x y dF(y). \quad (22)$$

Можно убедиться (см., например, [5, 6]), что функция влияния $IF(x; F, \Delta_F)$ оценки $\hat{\Delta}_0 = T(F_n)$ средней разности Джини $\Delta_F = T(F)$ записывается в виде

$$IF(x; F, \Delta_F) = \varphi_F(x) - \int \varphi_F(x) dF(x) = \varphi_F(x) - 2(\Delta_F - \mu_F), \quad x \in R^1. \quad (23)$$

Отметим, что функция влияния $IF(x; F, \Delta_F)$ не является ограниченной функцией и, следовательно, выборочные оценки средней разности Джини, так же как и оценка $\hat{S}_1(0)$ стандартного отклонения, подвержены сильному влиянию выбросов в выборке.

Рассмотрим теперь модифицированную оценку средней разности Джини, которая была предложена в [8]. Эта оценка является U -статистикой, основанной на урезанной выборке, и имеет вид

$$\hat{\Delta}_\alpha = \left\{ (n - 2[\alpha n])(n - 2[\alpha n] - 1) \right\}^{-1} \sum_{C_\alpha} |X_{(i)} - X_{(j)}|, \quad 0 \leq \alpha < 1/2. \quad (24)$$

Функционал $U_\alpha(F)$ из (6), соответствующий этой оценке при $m = 2$ и ядре $h(x_1, x_2) = |x_1 - x_2|$, записывается в виде

$$\Delta_\alpha(F) = \frac{1}{(1 - 2\alpha)^2} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \int |x - y| dF(x) dF(y), \quad 0 \leq \alpha < 1/2. \quad (25)$$

Далее функция $g_\alpha(x; F)$ из (4) для данного случая при $m = 2$ и $h(x_1, x_2) = |x_1 - x_2|$ запишется в виде

$$g_\alpha(x; F) = \frac{I[\xi_\alpha \leq x \leq \xi_{1-\alpha}]}{(1 - 2\alpha)^2} \{ \varphi_\alpha(x, F) / 2 + \mu_\alpha(F) \}, \quad (26)$$

где функция $\varphi_\alpha(x, F)$ и $\mu_\alpha(F)$ вычисляются по формулам

$$\varphi_\alpha(x, F) = 4xF(x) - 2x - \int_{F^{-1}(\alpha)}^x y dF(y), \quad \mu_\alpha(F) = \int_{\alpha}^{1-\alpha} F^{-1}(t) dt. \quad (27)$$

Величины A_α и B_α из (5) в данном случае, соответственно, равны

$$A_\alpha = [(1-2\alpha)\xi_\alpha - \mu_\alpha(F)] / (1-2\alpha)^2, \quad B_\alpha = [(1-2\alpha)\xi_{1-\alpha} - \mu_\alpha(F)] / (1-2\alpha)^2. \quad (28)$$

Приведем выражение для функции влияния $IF(x; F, \hat{\Delta}_\alpha)$ оценки $\hat{\Delta}_\alpha$ вида (24). Для простоты рассмотрим симметричные распределения, т.е. $F \in \mathfrak{S}_{S|0}$, тогда $\mu_\alpha(F) = 0$ и $\xi_\alpha = -\bar{\xi}_\alpha$. Введём дополнительные обозначения:

$$J_1 = J_1(F, \alpha) = \int_{\bar{\xi}_\alpha}^{\xi_\alpha} \varphi_\alpha(x, F) dF(x) = 2(1-2\alpha)^2 \Delta_\alpha(F), \quad J_2 = J_2(F, \alpha) = \int_{\bar{\xi}_\alpha}^{\xi_\alpha} \varphi_\alpha^2(x, F) dF(x). \quad (29)$$

С использованием формул (26)–(29), формула (13) для *ограниченной* функции влияния $IF(x; F, \hat{\Delta}_\alpha)$ оценки $\hat{\Delta}_\alpha$ переписывается в виде

$$IF(x; F, \hat{\Delta}_\alpha) = \frac{1}{(1-2\alpha)^2} \begin{cases} [\varphi_\alpha(x, F) - J_1(F, \alpha) - 4\alpha(1-2\alpha)\bar{\xi}_\alpha], & |x| \leq \bar{\xi}_\alpha, \\ [2(1-2\alpha)^2 \bar{\xi}_\alpha - J_1(F, \alpha)], & |x| > \bar{\xi}_\alpha. \end{cases} \quad (30)$$

Согласно (10) асимптотическая дисперсия $\sqrt{n}\hat{\Delta}_\alpha$ -оценок вычисляется по формуле

$$\begin{aligned} \sigma^2(F, \hat{\Delta}_\alpha) &= \int_{-\infty}^{\infty} IF^2(x; F, \hat{\Delta}_\alpha) dF(x) = \\ &= \frac{J_2(F, \alpha) - 2J_1^2(F, \alpha) + 8J_1(F, \alpha)\alpha(1-2\alpha)F^{-1}(\alpha) + 8\alpha(1-2\alpha)^3[F^{-1}(\alpha)]^2}{(1-2\alpha)^4}, \quad F \in \mathfrak{S}_{S|0}. \end{aligned} \quad (31)$$

Используя формулы (29) и (31), получаем выражение для асимптотической стандартизованной дисперсии $\hat{\Delta}_\alpha$ -оценок в виде

$$\tilde{\sigma}^2(F, \hat{\Delta}_\alpha) = \frac{\sigma^2(F, \hat{\Delta}_\alpha)}{\Delta_\alpha^2(F)} = 4\{J_2 - J_1^2 + 8\alpha(1-\alpha)J_1\xi_\alpha + 8\alpha(1-2\alpha)^3\xi_\alpha^2\} / J_1^2, \quad F \in \mathfrak{S}_{S|0}. \quad (32)$$

Пример 1. Приведём результаты вычислений асимптотической дисперсии оценки $\hat{\Delta}_0$ средней разности Джини при нормальном распределении, т.е. предполагаем, что $F(x) = \Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x \exp\{-x^2/2\} dx$. В данном случае функция $\varphi_F(x)$ из (22) запишется в виде $\varphi_\Phi(x) = 2x(2\Phi(x)-1) + (4/\sqrt{2\pi}) \exp\{-x^2/2\}$. Используя формулу (21), получаем (детали см. в примере (10.6.15) в [6]) значение асимптотической дисперсии $\sqrt{n}\hat{\Delta}_0$ -оценки средней разности Джини для нормального распределения, равное

$$\sigma^2(\Phi, \hat{\Delta}_0) = \int \varphi_\Phi^2(x) d\Phi(x) - \left(\int \varphi_\Phi(x) d\Phi(x) \right)^2 = \frac{4(\pi+6\sqrt{3})}{3\pi} - \frac{16}{\pi} = \frac{4(\pi+6\sqrt{3}-12)}{3\pi} \approx 0,651.$$

Стандартизованная асимптотическая дисперсия $\sqrt{n}\hat{\Delta}_0$ -оценки равна

$$\tilde{\sigma}^2(\Phi, \hat{\Delta}_0) = 4 \cdot \frac{\int \varphi_\Phi^2(x) d\Phi(x) - \left(\int \varphi_\Phi(x) d\Phi(x) \right)^2}{\left(\int \varphi_\Phi(x) d\Phi(x) \right)^2} = 4 \frac{4(\pi+6\sqrt{3}-12)}{3\pi \cdot (16/\pi)} = \frac{\pi+6\sqrt{3}-12}{3} \approx 0,511.$$

Пример 2. Рассмотрим супермодель с засорением $\mathfrak{S}_{\varepsilon, \tau}(\Phi) = \{F : F(x) = \Phi_{\varepsilon, \tau}(x)\}$, где

$$\Phi_{\varepsilon, \tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau), \quad 0 \leq \varepsilon \leq 1, \quad \tau \geq 1.$$

Численные значения асимптотической стандартизованной дисперсии $\tilde{\sigma}^2(F, \hat{\Delta}_0) = \sigma^2(F; \hat{\Delta}_0) / \Delta_F^2$ для $\sqrt{n}\hat{\Delta}_0$ -оценки средней разности Джини для супермодели $\mathfrak{S}_{\varepsilon, \tau}(\Phi)$, вычисленные по формуле (10.6.32) в [6], приведены в табл. 1.

Для сравнения оценки $\hat{\Delta}_0$ средней разности Джини с оценкой $\hat{S}_1(0)$ стандартного отклонения в рамках супермодели $\mathfrak{S}_{\varepsilon, \tau}(\Phi)$ в табл. 2 приведены значения их асимптотических относительных эффективностей $AOЭ_F(\hat{\Delta}_0 : \hat{S}_1(0)) = \tilde{\sigma}^2(F; \hat{S}_1(0)) / \tilde{\sigma}^2(F; \hat{\Delta}_0)$ для $F \in \mathfrak{S}_{\varepsilon, \tau}(\Phi)$.

Т а б л и ц а 1

Асимптотическая стандартизованная дисперсия $\tilde{\sigma}^2(F, \hat{\Delta}_0)$ для $F \in \mathfrak{Z}_{\varepsilon, \tau}(\Phi)$

$\tau \setminus \varepsilon$	0,00	0,001	0,005	0,01	0,05	0,10	0,20	0,30
$\tau = 3$	0,511	0,523	0,566	0,618	0,928	1,137	1,255	1,204
$\tau = 5$	0,511	0,558	0,735	0,933	1,887	2,256	2,159	1,831
$\tau = 10$	0,511	0,753	1,601	2,443	4,901	4,781	3,542	2,592

Т а б л и ц а 2

Асимптотическая относительная эффективность $AO\mathfrak{E}_F(\hat{\Delta}_0 : \hat{S}_1(0))$ для $F \in \mathfrak{Z}_{\varepsilon, \tau}(\Phi)$

$\tau \setminus \varepsilon$	0,00	0,001	0,005	0,01	0,05	0,10	0,20	0,30
$\tau = 3$	0,978	1,046	1,274	1,468	1,679	1,612	1,304	1,140
$\tau = 5$	0,978	1,634	3,011	3,518	2,512	1,712	1,183	1,010
$\tau = 10$	0,978	8,740	10,53	7,730	2,114	1,269	0,907	0,825

Из данных табл. 2 следует, что оценка $\hat{\Delta}_0$ средней разности Джини, проигрывая лишь 2% в эффективности оценке $\hat{S}_1(0)$ стандартного отклонения при нормальном распределении, становится предпочтительнее уже при небольших отклонениях от нормального распределения в рамках супермодели $\mathfrak{Z}_{\varepsilon, \tau}(\Phi)$. Отметим, что при увеличении параметра τ относительная эффективность $AO\mathfrak{E}_{\Phi, \tau}(\hat{\Delta}_0 : \hat{S}_1(0))$ принимает неожиданно большие значения. По всей вероятности, это является следствием того факта, что оценка $\hat{S}_1(0)$ стандартного отклонения имеет неограниченную (квадратично возрастающую) функцию влияния, и при увеличении τ и малых ε её дисперсия резко возрастает (см. также теорему (2) в [10]).

Пример 3. В табл. 3 приведены результаты сравнения различных оценок масштабного параметра, используя понятие дефекта оценки [12]. Напомним, что дефект $DE(F, \hat{\theta}_i)$ оценки $\hat{\theta}_i$, $i = 1, \dots, k$, среди сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ параметра θ при распределении F определяют в виде

$$DE(F, \hat{\theta}_i) = 1 - \min\{\sigma^2(F, \hat{\theta}_1), \dots, \sigma^2(F, \hat{\theta}_k)\} / \sigma^2(F, \hat{\theta}_i), \quad i = 1, \dots, k.$$

Отметим, что при сравнении оценок масштабного параметра используют вместо асимптотических дисперсий $\sigma^2(F, \hat{\theta}_i)$ их стандартизованные дисперсии $\tilde{\sigma}^2(F, \hat{\theta}_i)$, $i = 1, \dots, k$. Обычно дефекты оценок откладывают для наглядности на плоскости двух распределений [12]. В табл. 3 приведены дефекты оценок масштабного параметра с использованием их стандартизованных дисперсий в «плоскости двух распределений: $F_{(1)}$ – Гаусс, $F_{(3)}$ – Лаплас».

Т а б л и ц а 3

Дефекты оценок масштабного параметра для распределений $F_{(1)} = \Phi$ и $F_{(3)}$

Дефекты оценок	$\hat{S}_1(0)$	$\hat{S}_2(0)$	$\hat{\Delta}_0$	$\hat{S}_4$	$\hat{S}_3(0,10)$	$\hat{S}_3(0,25)$
$DE(F_{(1)}, \hat{S})$	0,00	0,12	0,02	0,14	0,38	0,63
$DE(F_{(3)}, \hat{S})$	0,20	0,00	0,04	0,21	0,35	0,52

Итак, среди сравниваемых оценок предпочтение следует отдать $\hat{\Delta}_0$ -оценке средней разности Джини. Отметим, что для симметричных распределений ограниченные функции влияния и асимптотические дисперсии $\hat{S}_3(0,25)$ -оценки интерквартильного размаха и оценки $\hat{S}_3^* = \text{med}\{|X_i - \text{med}(X)|, 1 \leq i \leq n\}$, совпадают. Отметим также, что оценка $\hat{S}_4 = \text{med}\{|X_i - X_j|, 1 \leq i < j \leq n\}$, в отличие от оценок $\hat{S}_1(0)$, $\hat{S}_2(0)$ и $\hat{\Delta}_0$, также имеет ограниченную функцию влияния.

Пример 4. Вычисления асимптотической стандартизованной дисперсии $\tilde{\sigma}^2(F, \hat{\Delta}_\alpha)$ для $\hat{\Delta}_\alpha$ -оценок вида (24) в рамках модели с засорением, т.е. для $F \in \mathfrak{F}_{\varepsilon, \tau}(\Phi)$, показывают, что она существенно зависит от параметра α . Например, для нормального распределения минимальное значение асимптотической стандартизованной дисперсии $\tilde{\sigma}^2(F, \hat{\Delta}_\alpha)$ для $\hat{\Delta}_\alpha$ -оценок достигается при $\alpha = 0$, а при $\tau = 3$ и $\varepsilon = 0,05$ минимум достигается при $\alpha = 0,05$. Для распределения Коши минимум достигается при $\alpha = 0,20$. Отмеченный факт приводит к необходимости адаптации параметра α к меняющемуся распределению в рамках заданной супермодели при практическом использовании $\hat{\Delta}_\alpha$ -оценок. Следуя работам [13, 14], определим адаптивный параметр $\hat{\alpha}(X_1, \dots, X_n)$ для $\hat{\Delta}_\alpha$ -оценок вида (24), вычисляемый по формуле

$$\hat{\alpha}(X_1, \dots, X_n) = \begin{cases} \alpha_1, & Q(F_n) \leq Q_1, \\ \frac{\alpha_{i+1} - \alpha_i}{Q_2 - Q_1} \{Q(F_n) - Q_i\} + \alpha_i, & Q_i < Q(F_n) < Q_{i+1}, i = 1, 2, \\ \alpha_3, & Q(F_n) \geq Q_3, \end{cases} \quad (33)$$

где параметры $\alpha_1, \alpha_2, \alpha_3, Q_1, Q_2$ и Q_3 задаются в соответствии с рассматриваемым типом супермодели. В формуле (33) $Q(F_n)$ – выборочная оценка, построенная по выборке $X_1, \dots, X_n$ методом подстановки, для функционала $Q(F; \nu, \mu)$, характеризующего степень затянутости хвостов распределения наблюдений [13]. Эта оценка записывается в виде

$$Q(F_n; \nu, \mu) = \frac{m}{k} \left(\sum_{i=n-k+1}^n X_{(i)} - \sum_{i=1}^k X_{(i)} \right) / \left(\sum_{i=n-m+1}^n X_{(i)} - \sum_{i=1}^m X_{(i)} \right), \quad k = [\nu n], \quad m = [\mu n], \quad (34)$$

где $0 < \nu < \mu \leq 0,5$ и $X_{(1)}, \dots, X_{(n)}$ – порядковые статистики выборки $X_1, \dots, X_n$. Следуя работе [13], ниже полагаем $\nu = 0,2$ и $\mu = 0,5$. В качестве примера рассмотрим достаточно широкую супермодель Тьюки в виде λ -аппроксимаций квантильных функций заданных распределений [11]. Эта супермодель определяется в виде

$$\mathfrak{F}_\lambda = \{F : F_\lambda^{-1}(t) = \lambda_1 + [t^{\lambda_3} - (1-t)^{\lambda_3}] / \lambda_2, \quad 0 \leq t \leq 1\}. \quad (35)$$

При определении адаптивного параметра $\hat{\alpha}(X_1, \dots, X_n)$ по формуле (33) в рамках супермодели Тьюки $\mathfrak{F}_\lambda$ примем следующие значения параметров: $\alpha_1 = 0$, $\alpha_2 = 0,2$, $\alpha_3 = 0,3$, $Q_1 = 1,76$, $Q_2 = 2,50$, $Q_3 = 4,30$. Результаты сравнения адаптивной $\hat{\Delta}_{\hat{\alpha}}$ -оценки средней разности Джини с семейством $\hat{\Delta}_\alpha$ -оценок, для которых параметр α , $0 \leq \alpha < 1/2$, *фиксирован*, приведены в табл. 4 в виде отношения асимптотической стандартизованной дисперсии оценки к минимальной стандартизованной дисперсии среди сравниваемых оценок при заданном распределении (т.е. при заданном значении параметра λ_3).

Т а б л и ц а 4

Отношения асимптотических стандартизованных дисперсий $\hat{\Delta}_\alpha$ -оценок для $F \in \mathfrak{F}_\lambda$

Ф.р.	λ_3	$Q(F_\lambda)$	$\hat{\Delta}_{0,0}$	$\hat{\Delta}_{0,01}$	$\hat{\Delta}_{0,05}$	$\hat{\Delta}_{0,10}$	$\hat{\Delta}_{0,20}$	$\hat{\Delta}_{0,30}$	$\hat{\Delta}_{\hat{\alpha}}$
Равномерная	1,000	1,60	1,00	1,14	1,67	2,50	5,00	10,0	1,00
	0,200	1,74	1,00	1,07	1,36	1,71	2,74	4,83	1,00
Гаусса–Лапласа	0,1349	1,76	1,00	1,06	1,29	1,59	2,47	4,28	1,00
	–0,0802	1,84	1,06	1,00	1,04	1,17	1,65	2,71	1,00
	–0,2450	1,91	1,51	1,11	1,00	1,05	1,35	2,12	1,00
	–0,5500	2,10	∞	1,67	1,11	1,00	1,10	1,57	1,00
Коши	–1,000	2,50	∞	3,87	1,59	1,16	1,00	1,22	1,00
	–2,000	4,38	∞	18,2	3,93	2,05	1,14	1,00	1,00

Приведенные в табл. 4 данные показывают, что адаптивная $\hat{\Delta}_{\hat{\alpha}}$ -оценка средней разности Джини обладает преимуществом перед урезанными $\hat{\Delta}_\alpha$ -оценками средней разности Джини, для которых параметр α , $0 \leq \alpha < 1/2$, *фиксирован*, при изменении распределения выборки в достаточно широком семей-

стве распределений $\mathfrak{Z}_\lambda$. Отметим, что приведенные результаты сравнения являются асимптотическими. При конечных объемах выборки требуются дополнительные исследования. Предварительные результаты моделирования показывают, что отмеченное преимущество адаптивных $\hat{\Delta}_\alpha$ -оценок начинает устойчиво проявляться уже при объемах выборки $n \geq 40$ и для других супермоделей.

Пример 5. Рассмотрим супермодель $\mathfrak{Z}\{\Phi(c)\}$ в виде семейства обобщенных гауссовских распределений с плотностью $f(x, c)$, которая зависит от параметра c и определяется в виде

$$f(x, c) = \frac{c}{2A(c)\Gamma(1/c)} \exp\{-[|x|/A(c)]^c\}, \quad x \in R^1, \quad 0,5 \leq c \leq 3,0, \quad A(c) = \sqrt{\Gamma(1/c)/\Gamma(3/c)}.$$

Отметим, что супермодель $\mathfrak{Z}\{\Phi(c)\}$ вида

$$\mathfrak{Z}\{\Phi(c)\} = \{F : F(x, c) = [c/\Gamma(1/c)A(c)] \int_{-\infty}^x \exp\{-[|x|/A(c)]^c\} dx\}, \quad x \in R^1,$$

включает при $c=1$ распределение Лапласа, и $c=2$ соответствует нормальному распределению.

Изменение эффективности для $0,5 < C < 1,75$ приведены в табл. 5.

Т а б л и ц а 5

Относительные эффективности $AO\mathfrak{E}_F(\hat{\Delta}_0 : \hat{S}_1(0))$ и $AO\mathfrak{E}_F(\hat{S}_2(0) : \hat{S}_1(0))$ для $F \in \mathfrak{Z}\{\Phi(c)\}$

c	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,50
$AO\mathfrak{E}_F(\hat{\Delta}_0 : \hat{S}_1(0))$	1,74	1,41	1,21	1,09	1,03	1,00	0,98	0,96
$AO\mathfrak{E}_F(\hat{S}_2(0) : \hat{S}_1(0))$	1,78	1,52	1,25	1,09	1,00	0,93	0,88	0,81

Итак, при изменении параметра c в интервале $0,5 \leq c < 1,75$ оценка $\hat{S}_2(0)$ среднего абсолютных отклонений эффективнее оценки $\hat{S}_1(0)$ стандартного отклонения. Кроме того, оценка $\hat{\Delta}_0$ средней разности Джини также предпочтительнее оценки $\hat{S}_1(0)$.

Заключение

Традиционно применяемые на практике оценки масштабного параметра, характеризующего степень разброса случайной величины, имеют неограниченные функции влияния Хампеля и обладают повышенной чувствительностью к наличию выбросов в выборке, что приводит к существенным искажениям статистических выводов. В данной работе предложена модифицированная оценка средних разностей Джини $\hat{\Delta}_\alpha$, $0 \leq \alpha < 1/2$, которая входит в класс U -статистик, основанных на урезанных выборках. Показано, что эта оценка асимптотически нормальна, имеет ограниченную функцию влияния, и, следовательно, защищена от влияния выбросов. При этом она обладает высокой эффективностью при нормальном распределении наблюдений. В работе приведены результаты сравнения предложенной оценки с другими оценками масштабного параметра в рамках гауссовской модели с засорением и предложен адаптивный вариант $\hat{\Delta}_\alpha$ -оценок, для которых параметр α , характеризующий пропорцию «урезания» исходной выборки, выбирается на основе информации, содержащейся в исходной выборке с использованием выборочной оценки функционала, характеризующего степень «затянутости хвостов» функции распределения F изучаемой случайной величины X .

ЛИТЕРАТУРА

1. Hoefding W. A class of statistics with asymptotically normal distribution // Ann. Math. Statist. 1948. V. 19. P. 292–325.
2. Janssen P., Serfling R., Veraverbeke M. Asymptotic normality of U-statistics based on trimmed samples // J. Statist. Planning and Inference. 1987. V. 16. P. 63–74.
3. Serfling R.J. Approximation Theorems of Mathematical Statistics. N. Y. : Wiley, 1980. 371 p.

4. Хампель Ф., Ронchetti Э., Рауссей П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М. : Мир, 1989. 512 с.
5. Шуленин В.П. Введение в робастную статистику. Томск : Изд-во Том. ун-та, 1993. 227 с.
6. Шуленин В.П. Математическая статистика. Ч. 3: Робастная статистика : учеб. Томск : Изд-во НТЛ, 2012. 520 с.
7. Кендэлл М., Стьюарт А. Теория распределений. М. : Наука, 1966. 587 с.
8. Шуленин В.П. Исследование устойчивости и асимптотических свойств урезанной средней разности Джини // Тр. IV Международной конференции по теории вероятности и математической статистике. Вильнюс, 1985. С. 330–332.
9. Шуленин В.П. Асимптотические свойства GL и U-статистик // Вестник Томского государственного университета. Приложение. 2004. № 9 (11). С. 184–190.
10. Bickel P.J., Lehmann E.L. Descriptive statistics for nonparametric models. III. Dispersion // Ann. Statist. 1976. V. 4, No. 6. P.1139–1158.
11. Ramberg J.S., Schmeiser B.W. An approximative method for generating symmetric random variables // Commun ACM. 1972. V. 15. P. 987–990.
12. Andrews D.F., Bickel P.J., Hampel F.R., Huber P. J., Rogers W.H., Tukey J.W. Robust estimation of location: survey and advances. N. Y. : Princeton Univ. Press, 1972. 375 p.
13. Hogg R.V. Adaptive robust procedures: A partial review and some suggestions for future applications and theory // J. Amer. Statist. Assoc. 1974. V. 69. P. 909–923.
14. Шуленин В.П. Адаптивная оценка урезанной средней разности Джини // Методы и программное обеспечение обработки информации и прикладного статистического анализа данных на ЭВМ. Минск, 1985. С. 113–114.

Шуленин Валерий Петрович, канд. техн. наук. E-mail: shvp@fpmk.tsu.ru
Томский государственный университет

Поступила в редакцию 21 апреля 2016 г.

Shulenin Valery P. (Tomsk State University, Russian Federation).

The asymptotic properties of the modified Gini's mean difference.

Keywords: scale parameter; robust estimation; influence function; asymptotic relative efficiency; adaptive estimators.

DOI: 10.17223/19988605/37/8

The paper proposes a modified estimator Gini's mean difference, which is part of a class U-statistics based on the trimmed samples. It is shown that this estimate is asymptotically normal, has a limited influence function, it has a high efficiency for a normal distribution of observations.

We assume that $X_1, \dots, X_n$ a random sample from a distribution function $F(x)$ and we assume that has a density $f(x)$, $x \in R^1$, $X_{(1)}, \dots, X_{(n)}$ - ordered statistics of the original sample $X_1, \dots, X_n$. Let $T(F)$, $F \in \mathfrak{F}$ common functional that characterizes the scale parameter, which describes the degree of dispersion of the study of a random variable X . We consider the functional which is defined as

$$\Delta_\alpha(F) = \frac{1}{(1-2\alpha)^2} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \int |x-y| dF(x) dF(y), \quad 0 \leq \alpha < 1/2.$$

Sample estimator of this functional, called a modified Gini's mean difference, written in the form

$$\hat{\Delta}_\alpha = \left\{ (n-2[\alpha n]) (n-2[\alpha n]-1) \right\}^{-1} \sum_{C_\alpha} |X_{(i)} - X_{(j)}|, \quad 0 \leq \alpha < 1/2, \quad C_\alpha = \{(i, j) : [\alpha n] + 1 \leq i < j \leq n - [\alpha n]\}.$$

The results of the comparison of the proposed $\hat{\Delta}_\alpha$ -estimators with other estimates of the scale parameter for Gaussian model with ε -fixed proportion of contamination, and proposed an adaptive version $\hat{\Delta}_\alpha$ -estimators for which the parameter α characterizing the proportion of "trimmed" of the original sample, selected on the basis of information contained in the original sample using a sample estimate functional, characterizing the degree of "heavy tails" of the distribution function of the random variable X under study.

REFERENCES

1. Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*. 19. pp. 292-325. DOI: 10.1214/aoms/1177730196
2. Janssen, P., Serfling, R. & Veraverbeke, M. (1987) Asymptotic normality of U-statistics based on trimmed samples. *J. Statist. Planning and Inference*. 16. pp. 63-74. DOI: 10.1016/0378-3758(87)90056-5
3. Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
4. Hampel, F., Ronchetti, E., Rousseeuw, P.J. & Stahel, W. (1989) *Robastnost' v statistike. Podkhod na osnove funktsiy vliyaniya* [Robust Statistics. The Approach Based on Influence Functions]. Translated from English. Moscow: Mir.
5. Shulenin, V.P. (1993) *Vvedenie v robastnuyu statistiku* [Introduction to robust statistics]. Tomsk: Tomsk State University.
6. Shulenin, V.P. (2012) *Matematicheskaya statistika. Ch. 3: Robastnaya statistika* [Math statistics. Part 3. Robust statistics]. Tomsk: NTL.
7. Kendall, M. & Stewart, A. (1966) *Theory of distributions* [Theory of distributions]. Translated from English. Moscow: Nauka.

8. Shulenin, V.P. (1985) [Investigation of the stability and asymptotic properties of the Gini truncated mean difference]. *Proc. of the Fouth International Conference on the Theory of Probability and Mathematical Statistics*. Vilnius. pp. 330–332. (In Russian).
9. Shulenin, V.P. (2004) The asymptotic properties of GL and U- statistics. *Vestnik Tomskogo gosudarstvennogo universiteta. Prilozhenie – Tomsk State University Journal. Appendix*. 9(11). pp. 184-190. (In Russian).
10. Bickel, P.J. & Lehmann, E.L. (1976) Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* 4(6). pp. 1139-1158. DOI: 10.1214/aos/1176343648
11. Ramberg, J.S. & Schmeiser, B.W. (1972) An approximative method for generating symmetric random variables. *Commun ACM*. 15. pp. 987-990. DOI: 10.1145/355606.361888
12. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P. J., Rogers, W.H. & Tukey, J.W. (1972) *Robust estimation of location: survey and advances*. Princeton, New York: Princeton Univ. Press.
13. Hogg, R.V. (1974) Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.* 69. pp. 909-923. DOI: 10.1080/01621459.1974.10480225
14. Shulenin, V.P. (1985) Adaptive estimation trimmed Gini mean difference. *Methods and software provision of information processing and application of statistical analysis of data on a computer*. Minsk. pp. 113-114. (In Russian).

УДК 519.218.72

DOI: 10.17223/19988605/37/9

Г.Ш. Цициашвили

ПУАССОНОВСКИЕ ПОТОКИ В СИСТЕМЕ С ПОВТОРНЫМ ОБСЛУЖИВАНИЕМ

Рассматриваются системы массового обслуживания с бесконечным числом каналов и повторным обслуживанием общего вида. Эти системы представляются в виде сетей с бесконечным числом каналов в узлах и имеют вид ориентированных деревьев. С помощью теоремы Бурке устанавливаются условия, когда потоки в сети являются пуассоновскими и независимыми.

Ключевые слова: многоканальная и бесконечно канальная системы массового обслуживания; винеровский процесс; число занятых каналов.

Системы массового обслуживания с повторным обслуживанием заявок вызывают большой интерес у специалистов по массовому обслуживанию [1, 2]. Особый интерес в последние годы вызывают системы с повторным обслуживанием и бесконечным числом каналов [3, 4]. В этих работах инициируется построение достаточно общих моделей обслуживания в сетях с бесконечным числом каналов в каждом узле. Особенностью таких сетей является их представление в виде ориентированного дерева. При анализе этих сетей с пуассоновским входным потоком появляется необходимость использовать теорему Бурке и применять теоретико-графовые построения для определения независимости отдельных потоков сети. В настоящей работе предпринимается попытка решения подобных вопросов в наиболее общем виде.

1. Потоки, выходящие из одного узла

Рассмотрим бесконечно канальную систему массового обслуживания с пуассоновским входным потоком и показательным распределением времени обслуживания заявок. Полагаем, что каждая заявка после окончания обслуживания может быть повторно обслужена (рис. 1). Возможны самые различные протоколы работы такой системы, например, время повторного обслуживания может иметь распределение, отличающееся от времени первоначального обслуживания, число повторных обслуживаний может быть отличным от единицы и т.д.

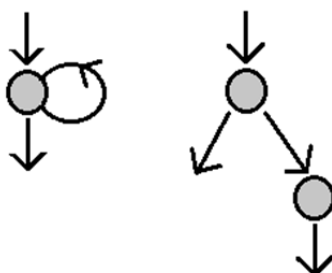


Рис. 1. Преобразование системы с однократным повторным обслуживанием в сеть

Самой общей может быть модель прохождения заявки через сеть, в каждом узле которой имеется бесконечное число каналов с показательным распределением времени обслуживания. Такая сеть описывается ориентированным деревом D , в котором каждое ребро направлено от корня дерева r (вершины, в которую не входит направленное ребро) к его листьям (вершинам, из которых не выходят направленные ребра). После окончания обслуживания в узле заявка может быть направлена в один из

узлов дерева, куда указывают его направленные ребра. Естественным обобщением этой модели является предположение, что время обслуживания заявки на приборе является гиперэкспоненциальным (вероятностной смесью экспоненциальных распределений). Данное обобщение никак не меняет вид сети, у которой вновь переходы заявок между узлами описываются ориентированным деревом.

Из известной теоремы Бурке [6] следует, что если сеть функционирует в стационарном режиме, то выходной поток из каждого узла, равно как и входной поток в каждый узел, является пуассоновским. Однако возникает вопрос о зависимости или независимости этих потоков.

Рассмотрим теперь пуассоновский поток точек $\Lambda = \{0 \leq t_1 \leq t_2 \leq \dots\}$ интенсивности λ . Предположим, что каждая точка потока Λ независимо с вероятностью $p_1, 0 < p_1 < 1$, становится точкой потока Λ_1 и с вероятностью $p_2 = 1 - p_1$ – точкой потока Λ_2 (рис. 2). Известно [5], что потоки Λ_1, Λ_2 также являются пуассоновскими с интенсивностями $\lambda p_1, \lambda p_2$ соответственно.

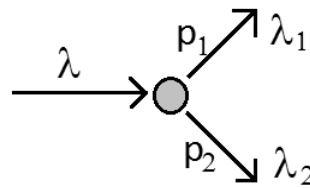


Рис. 2. Вероятностное раздвоение пуассоновского потока

Теорема 1. Потоки Λ_1, Λ_2 независимы.

Доказательство. Возьмем произвольный отрезок $[t, t+T]$, $t \geq 0, T > 0$, и обозначим n, n_1, n_2 количество точек потоков $\Lambda, \Lambda_1, \Lambda_2$ на этом отрезке. Вычислим вероятность

$$\begin{aligned} P(n_1 = k_1, n_2 = k_2) &= P(n = k_1 + k_2) C_{k_1+k_2}^{k_1} p_1^{k_1} p_2^{k_2} = \frac{e^{-\lambda T} (\lambda T)^{k_1+k_2}}{(k_1+k_2)!} \frac{(k_1+k_2)!}{k_1! k_2!} p_1^{k_1} p_2^{k_2} = \\ &= e^{-\lambda T p_1} \frac{(\lambda T p_1)^{k_1}}{k_1!} \cdot e^{-\lambda T p_2} \frac{(\lambda T p_2)^{k_2}}{k_2!} = P(n_1 = k_1) \cdot P(n_2 = k_2). \end{aligned}$$

Пусть теперь отрезки $[t^{(1)}, t^{(1)} + T^{(1)}], [t^{(2)}, t^{(2)} + T^{(2)}]$ не пересекаются. Обозначим $n^{(i)}, n_1^{(i)}, n_2^{(i)}$ количество точек потоков $\Lambda, \Lambda_1, \Lambda_2$ на отрезке $[t^{(i)}, t^{(i)} + T^{(i)}], i=1, 2$. Докажем независимость случайных величин $n_1^{(1)}, n_2^{(2)}$. Для этого, полагая $k_2^{(1)} = k^{(1)} - k_1^{(1)}, k_1^{(2)} = k^{(2)} - k_2^{(2)}$, вычислим вероятность

$$\begin{aligned} P(n_1^{(1)} = k_1^{(1)}, n_2^{(2)} = k_2^{(2)}) &= \sum_{k^{(1)} \geq k_1^{(1)}, k^{(2)} \geq k_2^{(2)}} P(n^{(1)} = k^{(1)}, n^{(2)} = k^{(2)}) C_{k^{(1)}}^{k_1^{(1)}} p_1^{k_1^{(1)}} p_2^{k_2^{(1)}} C_{k^{(2)}}^{k_1^{(2)}} p_1^{k_1^{(2)}} p_2^{k_2^{(2)}} = \\ &= \sum_{k^{(1)} \geq k_1^{(1)}, k^{(2)} \geq k_2^{(2)}} P(n^{(1)} = k^{(1)}) P(n^{(2)} = k^{(2)}) C_{k^{(1)}}^{k_1^{(1)}} p_1^{k_1^{(1)}} p_2^{k_2^{(1)}} C_{k^{(2)}}^{k_1^{(2)}} p_1^{k_1^{(2)}} p_2^{k_2^{(2)}} = \\ &= \sum_{k^{(1)} \geq k_1^{(1)}, k^{(2)} \geq k_2^{(2)}} e^{-\lambda T^{(1)}} \frac{(\lambda T^{(1)})^{k^{(1)}}}{k^{(1)}!} e^{-\lambda T^{(2)}} \frac{(\lambda T^{(2)})^{k^{(2)}}}{k^{(2)}!} \frac{k^{(1)}!}{k_1^{(1)}! k_2^{(1)}!} p_1^{k_1^{(1)}} p_2^{k_2^{(1)}} \frac{k^{(2)}!}{k_1^{(2)}! k_2^{(2)}!} p_1^{k_1^{(2)}} p_2^{k_2^{(2)}} = \\ &= e^{-\lambda T^{(1)} p_1} \frac{(\lambda T^{(1)} p_1)^{k_1^{(1)}}}{k_1^{(1)}!} e^{-\lambda T^{(2)} p_2} \frac{(\lambda T^{(2)} p_2)^{k_2^{(2)}}}{k_2^{(2)}!} \sum_{k_2^{(1)} \geq 0} e^{-\lambda T^{(1)} p_2} \frac{(\lambda T^{(1)} p_2)^{k_2^{(1)}}}{k_2^{(1)}!} \sum_{k_1^{(2)} \geq 0} e^{-\lambda T^{(2)} p_1} \frac{(\lambda T^{(2)} p_1)^{k_1^{(2)}}}{k_1^{(2)}!} = \\ &= e^{-\lambda T^{(1)} p_1} \frac{(\lambda T^{(1)} p_1)^{k_1^{(1)}}}{k_1^{(1)}!} e^{-\lambda T^{(2)} p_2} \frac{(\lambda T^{(2)} p_2)^{k_2^{(2)}}}{k_2^{(2)}!} = P(n_1^{(1)} = k_1^{(1)}) P(n_2^{(2)} = k_2^{(2)}). \end{aligned}$$

Таким образом, пары случайных величин $(n_1^{(1)}, n_2^{(1)}), (n_1^{(2)}, n_2^{(2)}), (n_1^{(1)}, n_2^{(2)}), (n_2^{(1)}, n_1^{(2)})$ состоят из независимых компонент. Следовательно, пуассоновские потоки Λ_1, Λ_2 независимы.

Замечание 1. Утверждение теоремы 1 распространяется на случай, когда пуассоновский поток Λ делится на $l > 1$ пуассоновских потоков с помощью вероятностного просеивания, при котором каждая точка потока Λ с вероятностью $p_j > 0$ попадает в поток $\Lambda_j, j = 1, \dots, l, \sum_{j=1}^l p_j = 1$.

2. Независимость конечного числа потоков сети

На множестве вершин ориентированного дерева D (в каждую вершину которого входит только одно направленное ребро) определим отношение частичного порядка $u_1 \succ u_2$, если в дереве D существует путь из вершины u_1 в вершину u_2 (здесь $u \succ u$). Сопоставим каждой вершине u дерева D множества вершин $P(u) = \{v : u \succ v\}, Q(u) = \{v : v \succ u\} \cup P(u)$.

Теорема 2. Предположим, что вершины $u_1, \dots, u_k$ дерева D таковы, что множества $P(u_1), \dots, P(u_k)$ не пересекаются. Тогда для любых вершин $v_1 \in P(u_1), \dots, v_k \in P(u_k)$ потоки $\Lambda(v_1), \dots, \Lambda(v_k)$, входящие в вершины $v_1, \dots, v_k$, соответственно, независимы.

Доказательство. Из определения дерева D следует, что в нем существует вершина u такая, что непересекающиеся множества $P(u_1), \dots, P(u_k)$ содержатся в множестве $P(u)$. В качестве u можно взять, например, корень дерева. Следовательно, в силу теоремы 1 и замечания 1 потоки, выходящие из вершины u , независимы. Отсюда автоматически следует независимость потоков $\Lambda(v_1), \dots, \Lambda(v_k)$.

Всюду далее без ограничения общности будем предполагать, что любая вершина дерева D за исключением листьев имеет отвлечение (из нее выходит несколько направленных ребер, рис. 3). Построим алгоритм выделения кластеров на множестве вершин дерева D . Пусть в дереве D зафиксирована вершина u_1 , не совпадающая с корневой вершиной, и определены соответствующие ей множества $P(u_1), Q(u_1), S_1 = Q(u_1)$. Если на шаге $i \geq 1$ заданы u_i, S_i , то возьмем вершину $u_{i+1} \notin S_i$ и положим $S_{i+1} = S_i \cup Q(u_i)$. В силу конечности числа вершин в дереве D эта процедура закончится на некотором шаге $l \geq 1$ построением вершин $u_1, \dots, u_l$ и соответствующих им множеств $S_1, \dots, S_l$.

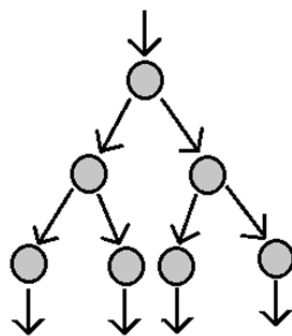


Рис. 3. Дерево с ответвлениями

Теорема 3. Множества вершин $P(u_1), \dots, P(u_l)$ в дереве D не пересекаются.

Доказательство. Пусть на шаге $i < l$ заданы вершины $u_1, \dots, u_i$ и соответствующие им множества $S_1, \dots, S_i; P(u_1), \dots, P(u_i)$. Тогда на шаге $i+1$ выполняется соотношение $u_{i+1} \notin S_i$ и, значит, множество $P(u_{i+1})$ не пересекается с множествами $P(u_1), \dots, P(u_i)$. В противном случае существуют

такие k , $1 \leq k \leq i$, $v \in P(u_k)$, что $v \in P(u_{i+1})$. Следовательно, $u_{i+1} \succ v, u_k \succ v, u_{i+1} \neq u_k$ и, значит, D не является ориентированным деревом.

Обозначим m число ребер, выходящих из корневой вершины дерева D и попадающих в вершины $v_1, \dots, v_m$, и положим M число листьев этого дерева.

Теорема 4 Справедливо неравенство

$$m \leq l \leq M. \quad (1)$$

Доказательство. Обозначим $V_1, \dots, V_m$ наборы вершин дерева D , удовлетворяющие соотношениям $V_j = P(v_j)$, $j = 1, \dots, m$. Из определения множеств $V_1, \dots, V_m$ следует, что они не пересекаются, а их объединение совпадает с множеством всех вершин дерева D за исключением корневой вершины. Поэтому для любого $j, 1 \leq j \leq m$, существует такое $i, 1 \leq i \leq l$, что $P(u_i) \subseteq P(v_j)$. В противном случае множество S_l не будет совпадать с множеством вершин дерева D , что противоречит определению S_l . Следовательно, справедливо неравенство $m \leq l$. По определению объединение $\bigcup_{i=1}^l P(u_i)$ непересекающихся в силу теоремы 3 множеств $P(u_1), \dots, P(u_l)$ содержит все листья ориентированного дерева D . Следовательно, $l \leq M$.

Замечание 2. Рекуррентная процедура, когда на каждом шаге выбирается вершина, в которую входит ребро-стрелка из корневой вершины, приводит к равенству $m = l$. Рекуррентная процедура, в которой на каждом шаге выбирается очередной лист дерева D , приводит к равенству $l = M$.

Замечание 3. Предположим, что вершины $u_1, \dots, u_k$ дерева D таковы, что множества $P(u_1), \dots, P(u_k)$ не пересекаются и объединение множеств $\bigcup_{i=1}^k Q(u_i)$ совпадает с множеством вершин дерева D . Тогда описанная выше рекуррентная процедура может быть реализована с использованием этого набора вершин (рис. 4).

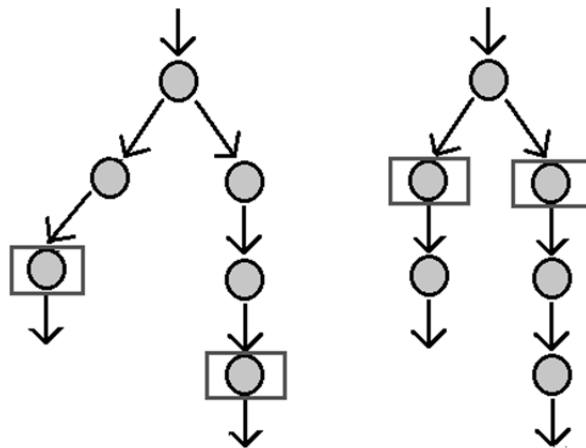


Рис. 4. Примеры совокупностей узлов с независимыми потоками

Заключение

Теоремы 2–4 справедливы для сетей, представляемых ориентированными деревьями общего вида без требования наличия ответвления в каждой вершине. Полученные в работе результаты можно распространить на сети, в каждом узле которых содержится конечное число каналов. От них возможен переход к сетям с бесконечным числом каналов путем выделения синергетических эффектов. В случае непуассоновского входного потока можно использовать результаты работы [7] для асимптотического анализа числа занятых каналов в каждом узле сети.

ЛИТЕРАТУРА

1. Nobel R.D., Tijms H.C. Optimal control for an MX/G/1 queue with two service modes // *European Journal of Operational Research*. 1999. Elsevier. V. 113(3). P. 610–619.
2. Nobel R.D., Tijms H.C. Waiting-time probabilities in the "M/G/1" retrial queue // *Statistica Neerlandica*. 2006. Netherlands Society for Statistics and Operations Research. V. 60 (1). P. 73–78.
3. Жидкова Л.А., Моисеева С.П. Исследование системы параллельного обслуживания кратных заявок простейшего потока // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2011. № 4 (17). С. 49–54.
4. Жидкова Л.А., Моисеева С.П. Математическая модель потоков покупателей двухпродуктовой торговой компании в виде системы массового обслуживания с повторными обращениями к блокам // *Известия Томского политехнического университета*. 2013. Т. 322, № 6. С. 5–9.
5. Хинчин А.Я. Работы по математической теории массового обслуживания. М. : Физматлит, 1963.
6. Burke P.J. The output of a queuing system // *Operations Research*. 1956. V. 4. P. 699–704.
7. Моисеев А.Н., Назаров А.А. Бесконечнолинейные системы и сети массового обслуживания. Томск : Изд-во НТЛ, 2015.

Цициашвили Гурами Шалвович, д-р физ.-мат. наук, профессор. E-mail: guram@iam.dvo.ru
Институт прикладной математики ДВО РАН,
Дальневосточный федеральный университет, г. Владивосток

Поступила в редакцию 28 апреля 2016 г.

Tsitsiashvili Gurami Sh. (Institute for Applied Mathematics of Far Eastern Branch of RAS, Far Eastern Federal University, Vladivostok).

Poisson flows in systems with retrial queues.

Keywords: systems with retrial queues; Poisson flows; independence of flows.

DOI: 10.17223/19988605/37/9

Queuing systems with retrial queues represent large interest in manifold applications of the queuing theory. Especially last years it is connected with infinite server queuing systems and retrial queues. These models may be considered in terms of queuing networks which have a structure of directed tree. An analysis of such networks may be based on the Burke theorem and the graph theory concepts in a definition of independence of random flows. In this paper an attempt to solve this problem in the most general terms is realized.

Consider infinite server queuing system with Poisson input flow and exponentially distributed service times. Assume that each customer after a service may be served iteratively. The different protocols of functioning for such system may be realized. For example, a distribution of repeated service time may differ from a distribution of initial service time, a number of repeated service may be different and so on.

The most general model of such system is a model of queuing network in which each node has infinite number of servers with exponential distribution of service times. Such network is described by a directed tree D in which any edge is directed from a tree root r (a node which have not incoming edges) to its leaves (nodes which have not outgoing edges). On service finishing a customer may be directed to some nodes where its outgoing edges show. Natural generalization of this model is based on an assumption that service times have hyperexponential distributions (probability mixtures of exponential distributions). This assumption does not change a view of this network in which transitions between nodes are described by oriented tree also.

From the Burke theorem we have that if the network is in a stationary regime then its outgoing and incoming flows of each node are Poisson. But a problem is on dependent or independent for these flows. This problem is solved as follows.

Consider now Poisson point flow $\Lambda = \{0 \leq t_1 \leq t_2 \leq \dots\}$ with the intensity λ . Assume that each point of the flow Λ independently with a probability $p_1, 0 < p_1 < 1$, becomes a point of a flow Λ_1 and with a probability $p_2 = 1 - p_1$ – a point of a flow Λ_2 . It is well known that the flows Λ_1, Λ_2 are Poisson flows also with intensities $\lambda p_1, \lambda p_2$, accordingly.

Theorem 1. *Flows Λ_1, Λ_2 are independent.*

On a set of the directed tree D (in which each node has a single incoming edge) define a relation of partial order $u_1 \succ u_2$, if in the tree D there is a way from the node to the node u_2 (here $u \succ u$). Contrast each node u of the tree D the nodes sets $P(u) = \{v: u \succ v\}$, $Q(u) = \{v: v \succ u\} \cup P(u)$.

Theorem 2. *Assume that nodes $u_1, \dots, u_k$ of the tree D satisfy the condition that the sets $P(u_1), \dots, P(u_k)$ do not intersect. Then for any nodes $v_1 \in P(u_1), \dots, v_k \in P(u_k)$ flows $\Lambda(v_1), \dots, \Lambda(v_k)$, incoming to the nodes $v_1, \dots, v_k$, appropriately are independent.*

Obtained results may be spread onto networks with final number of servers in their nodes and to transit from these networks to networks with infinite number of servers by analyzing synergetic effects. If input network flow is recurrent then it is possible to use results on asymptotic analysis of numbers of occupied servers in network nodes.

REFERENCES

1. Nobel, R.D. & Tijms, H.C. (1999) Optimal control for an MX/G/1 queue with two service modes. *European Journal of Operational Research*. 113(3). pp. 610-619. DOI: 10.1007/978-81-322-1680-3_6. 814

2. Nobel, R.D. & Tijms, H.C. (2006) Waiting-time probabilities in the "M/G/1" retrial queue. *Statistica Neerlandica. Netherlands Society for Statistics and Operations Research*. 60(1). pp. 73-78. DOI: 10.1111/j.1467-9574.2006.00312.x
3. Zhidkova, L.A. & Moiseeva, S.P. (2011) Investigation of parallel serving system with simplest flow of fold customers. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(17). pp. 49-54. (In Russian).
4. Zhidkova, L.A. & Moiseeva, S.P. (2013) Mathematical model of customer flows of two product trade company as queuing system with retrial queues. *Izvestiya Tomskogo politekhnicheskogo universiteta – Bulletin of the Tomsk Polytechnic University*. 322(6). pp. 5-9. (In Russian).
5. Khinchin, A.Ya. (1963) *Raboty po matematicheskoy teorii massovogo obsluzhivaniya* [Researchs on mathematical queuing theory]. Moscow: Fizmatgiz.
6. Burke, P.J. (1956) The output of a queuing system. *Operations Research*. 4. pp. 699-704. DOI: 10.1287/opre.4.6.699
7. Moiseev, A.N. & Nazarov, A.A. (2015) *Beskonechnolineynye sistemy i seti massovogo obsluzhivaniya* [Infinite server queuing systems and networks]. Tomsk. NTL.

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

УДК 004

DOI: 10.17223/19988605/37/10

И.Б. Бурдонов, А.С. Косачев

ОБОБЩЕННАЯ МОДЕЛЬ СИСТЕМЫ АВТОМАТОВ

Статья посвящена проблеме моделирования, композиции и детерминизма составных систем. Компоненты системы моделируются конечными автоматами с несколькими входами и выходами, а взаимодействие между ними – обменом сообщениями по симплексным каналам связи. Система описывается ориентированным графом связей, вершина которого соответствует автомату компонента, а дуга – каналу связи, соединяющему выход одного автомата с входом другого. Автоматы системы работают синхронно: на каждом такте каждый автомат выполняет один переход. Определяются ассоциативная композиция автоматов системы по графу связей и условия детерминированности системы.

Ключевые слова: ориентированные графы; взаимодействующие автоматы; композиция автоматов; детерминизм; сети.

Большинство сложных, особенно распределённых, систем представляет собой набор взаимодействующих компонентов. В данной статье компоненты моделируются конечными автоматами, а взаимодействие – обменом сообщениями между автоматами. Автомат может иметь несколько входов для приёма сообщений и может принимать сразу несколько сообщений; также автомат может иметь несколько выходов для отправки сообщений и посылать сразу несколько сообщений. Структура связей между компонентами моделируется ориентированным графом (будем называть его *графом связей*), вершинам которого соответствуют автоматы, а дуги называются *соединениями* и соответствуют симплексным каналам передачи сообщений. Соединение, ведущее из автомата A в автомат B , помечается выходом j автомата A и входом i автомата B . При этом каждый вход (выход) каждого автомата соединён не более чем с одним выходом (входом) другого (или того же самого) автомата. Входы (выходы) автоматов, которые не соединяются с выходами (входами) автоматов, являются *внешними*, через них осуществляется связь системы с её *окружением*. Автоматы системы работают синхронно: на каждом такте каждый автомат выполняет один переход. Также на каждом такте окружение системы посылает сообщения на некоторые внешние входы системы и принимает сообщения с некоторых внешних выходов системы.

Считается, что граф связей статический, т.е. не меняющийся в процессе работы системы. В этом случае система (так же как её компоненты) может моделироваться конечным автоматом, получающимся из автоматов-компонентов с помощью подходящего оператора композиции, учитывающего граф связей.

Такая модель системы является обобщением модели, предложенной в [1]. Там дуга графа связей представляла собой, по сути, очередь сообщений длины 1. Она моделируется автоматом с одним входом и одним выходом. Синхронно взаимодействуют автомат в вершине с автоматом дуги, инцидентной этой вершине. Если в графе связей из [1] каждую дугу разбить на два соединения, вставив в «середину» дуги новую вершину, в которую поместить автомат дуги, то получится система автоматов, являющаяся частным случаем системы, предлагаемой в данной статье. «Частность» этого случая заключается в том, что автомат дуги может теперь моделировать не только очередь сообщений длины 1, но и очередь любой другой (в том числе неограниченной) длины, стек и т.д. Более того, в данной статье мы, по сути, не делаем различия между автоматом вершины и автоматом дуги: автомат дуги тоже может иметь несколько входов и выходов.

Основная особенность предлагаемой модели автомата заключается в том, что переход автомата, предъявляя требования к состоянию всех входов и выходов автомата (указываются сообщения на них),

отдельно указывает ту часть входов и выходов, по которым сообщения принимаются или посылаются. Синхронность взаимодействия автоматов означает, что для каждого соединения требования автоматов, связанных этим соединением, должны быть согласованы. Это даёт возможность описывать более широкий спектр поведений автомата. Например, приоритетный приём сообщений: если на входах автомата имеется несколько сообщений, автомат может принять сообщения с наивысшим приоритетом, не принимая остальные сообщения. Также это позволяет автомату выполнять приём сообщений детерминированно, т.е. независимо от того, удастся или не удастся одновременно послать некоторое сообщение на некоторый выход. Если удастся послать сообщение, автомат выполняет один переход, а если не удастся – то другой. Обычно это приводит к недетерминизму: на один стимул выдаются разные реакции. Однако в нашей модели реакция – это требования к выходам, а не указание той части выходов, по которым сообщения реально передаются.

В статье определяется ассоциативная композиция переходов, автоматов и системы по известному графу связей. Ассоциативность важна для того, чтобы работа системы зависела только от множества её автоматов и соединений и не зависела от какого-либо упорядочивания этих множеств. Дополнительно определяется композиция автоматов, не связанных соединениями, что позволяет докомпоновать систему ровно до одного автомата, который может использоваться как компонент более сложных систем автоматов.

Исследуется проблема детерминизма: даётся определение детерминированного автомата и выясняются условия, при которых композиция детерминированных автоматов является детерминированной.

1. Модель

Пусть задано некоторое конечное множество сообщений M , которое мы будем называть *алфавитом* (сообщений). Пустое множество $\emptyset$ будем называть *пустым сообщением*, считать, что $\emptyset \notin M$, и обозначать $M_\emptyset \triangleq M \cup \{\emptyset\}$.

Автомат в алфавите M – это набор $A = (M, I, J, S, T, s_0)$, где:

- I – конечное множество *входов* автомата;
- J – конечное множество *выходов* автомата;
- S – конечное множество *состояний* автомата;
- $T \subseteq S \times X \times P \times Y \times Q \times S$ – множество *переходов* автомата, где:
- $X = \{x \mid x: I \rightarrow M_\emptyset\}$ – множество *стимулов*;
- $Y = \{y \mid y: I \rightarrow M_\emptyset\}$ – множество *реакций*;
- $P = 2^I$ – семейство подмножеств входов (по ним принимаются сообщения);
- $Q = 2^J$ – семейство подмножеств выходов (по ним посылаются сообщения);
- $s_0 \in S$ – *начальное состояние*;

причём выполнены следующие условия:

- 1) множества входов и выходов не пересекаются: $I \cap J = \emptyset$;
- 2) принимаются и посылаются только непустые сообщения:

$$\forall (s, x, p, y, q, t) \in T (\forall i \in p (x(i) \neq \emptyset) \ \& \ \forall j \in q (y(j) \neq \emptyset)),$$

что эквивалентно $\forall (s, x, p, y, q, t) \in T (p \subseteq x^{-1}(M) \ \& \ q \subseteq y^{-1}(M))$.

Очевидно, что из конечности множеств M, I, J и S следует конечность множеств X, Y, P, Q, T .

Для перехода $a = (s, x, p, y, q, t)$ состояние s будем называть *пресостоянием* перехода, а состояние t – *постсостоянием*, и будем обозначать $s_a = s, x_a = x, p_a = p, y_a = y, q_a = q, t_a = t$. Для автомата $A = (M, I, J, S, T, s_0)$ будем обозначать $I_A = I, J_A = J, S_A = S, T_A = T, s_{0A} = s_0$.

Работу автомата можно описать следующим образом. В текущем состоянии s проверяется состояние входов, т.е. какие сообщения можно принять с входов автомата. Это определяет стимул x . Далее рассматриваются переходы по этому стимулу, т.е. множество $T_{s,x} \subseteq T$ переходов из s по x . Каждый переход из множества $T_{s,x}$ определяет ту или иную реакцию y . Если для данных s и x имеются переходы с разными реакциями y , то недетерминированным образом выбирается реакция y как одна из реакций на переходах из множества $T_{s,x}$. После выбора реакции y проверяется состояние выходов, а именно какие

непустые сообщения, определяемые реакцией y , могут быть переданы по соответствующим выходам, а какие нет. Это определяет параметр q . Тем самым определяется множество $T_{s,x,y,q} \subseteq T_{s,x}$ переходов с данными s, x, y, q . Если это множество содержит более одного перехода, то недетерминированным образом выбирается один из них. Если выбран переход (s, x, p, y, q, t) , то параметр p определяет, какие непустые сообщения будут приняты с входов автомата. Таким образом, принимаются сообщения, определяемые стимулом x , по входам, определяемым параметром p ; посылаются сообщения, определяемые реакцией y , по выходам, определяемым параметром q , после чего автомат переходит в состояние t . Это описание работы автомата неформально. Формальное описание есть, по существу, формальное определение композиции автоматов, которое будет дано в следующем разделе.

Пусть V – конечное множество автоматов в алфавите M . Без ограничения общности можно считать, что входы, выходы и состояния автоматов разные: $\forall A, B \in V (A \neq B \Rightarrow I_A \cap I_B = \emptyset \ \& \ I_A \cap J_B = \emptyset \ \& \ J_A \cap J_B = \emptyset \ \& \ S_A \cap S_B = \emptyset)$. Это всегда можно сделать систематическим переименованием состояний, входов и выходов автоматов, в результате чего получатся автоматы, изоморфные исходным.

Система автоматов – это набор $R = (M, V, E)$, где $E : E_{Dom} \rightarrow E_{Im} \rightarrow$ биекция, определяющая *соединения*, где $E_{Dom} \subseteq J_R$, $E_{Im} \subseteq I_R$, $J_R = \cup \{J_A | A \in V\}$ – множество всех выходов всех автоматов, $I_R = \cup \{I_A | A \in V\}$ – множество всех входов всех автоматов. Для $R = (M, V, E)$ обозначим $V_R = V$, $E_R = E$.

Определим функцию $\phi : (I_R \cup J_R) \rightarrow V$, которая каждому входу или выходу системы R ставит в соответствие автомат, которому этот вход или выход принадлежит: $\forall A \in V \ \forall z \in I_A \cup J_A \ \phi(z) = A$. Будем говорить, что соединение (j, i) связывает выход автомата $\phi(j)$ и вход автомата $\phi(i)$. Вход $i \in I_R$ или выход $j \in J_R$, для которого не определено соединение, т.е. $i \notin E_{Im}$ или $j \notin E_{Dom}$, будем называть *внешним входом* или *внешним выходом* системы соответственно.

Системе $R = (M, V, E)$ соответствует ориентированный *граф связей*, вершинами которого являются автоматы из V , а помеченные дуги – это соединения, причём дуга $A \rightarrow B$ существует и помечена соединением (j, i) тогда и только тогда, когда $(j, i) \in E$, $\phi(j) = A$ и $\phi(i) = B$. Такой граф обозначим через $Og(R)$. Если снять ориентацию соединений, то получившийся неориентированный граф обозначим через $Ng(R)$. Будем говорить, что автоматы A и B *связаны* в системе, если в графе $Ng(R)$ существует (неориентированный) путь из A в B (возможно, пустой, если $A = B$). В противном случае разные автоматы A и B *не связаны*. Очевидно, что отношение связности автоматов является отношением эквивалентности и разбивает множество автоматов V на классы эквивалентности, которые будем называть *классами связности*.

Будем считать, что состояние автомата в системе $R = (M, V, E)$ является множеством и состояния разных автоматов не пересекаются: $\forall A, B \in V \ \forall s_A \in S_A \ \forall s_B \in S_B \ (A \neq B \Rightarrow s_A \cap s_B = \emptyset)$. Если уже задана система, в которой это условие не выполнено, то в каждом автомате каждое состояние s заменим на синглетон $\{s\}$, что даст возможность получить автоматы, изоморфные исходным, а требуемые условия будут выполнены.

Работу системы автоматов можно описать следующим образом. Прежде всего, окружение моделируется автоматом, число выходов (входов) которого равно числу внешних входов (выходов) системы, и добавляются внешние соединения, связывающие выходы (входы) окружения с внешними входами (выходами) системы. Получается замкнутая система без внешних входов и выходов. Все автоматы, включая окружение, срабатывают одновременно, и каждый автомат выполняет ровно один переход (явный или неявный, если используются правила умолчания). Если автоматы A и B связаны соединением (j, i) , где $j \in J_A$ и $i \in I_B$, то для выполняемых переходов $a \in T_A$ и $b \in T_B$ сообщение $y_a(j)$ и условие его передачи $j \in q_a$ на выходе j автомата A должны быть согласованы с сообщением $x_b(i)$ и условием его приёма $i \in p_b$ на входе i автомата B . Это означает, во-первых, что сообщение $y_a(j)$ на выходе j совпадает с сообщением $x_b(i)$ на входе i . И, во-вторых, сообщение либо передаётся из A по выходу j , т.е. $j \in q_a$, и принимается в B по входу i , т.е. $i \in p_b$, либо не передаётся из A по выходу j , т.е. $j \notin q_a$, и не принимается в B по входу i , т.е. $i \notin p_b$.

2. Композиция

Существует большое разнообразие как автоматоподобных моделей, так и операторов их (параллельной) композиции начиная с классической композиции в алгебрах процессов CSP – Communicating Sequential Processes [4] и CCS – Calculus of Communicating Systems [5]. Они определяют две разновидности композиции LTS (Labelled Transition System), в которых переходы помечены символами *действий* из заданного алфавита. Для систем ввода-вывода (IOTS – Input-Output Transition System), в которых действия разбиваются на стимулы (input) и реакции (output), предлагались различные модификации композиции [7–9], в частности для обнаружения тупиковых ситуаций (deadlock) с целью наблюдения *отказов* (refusal) [6–8]. Для систем с приоритетами и моделей *событий* (систем более общего вида) нами были также предложены соответствующие операторы композиции [10,11]. Вводимая в данном разделе композиция учитывает наличие у автомата нескольких входов и выходов и моделирует взаимодействие автоматов через соединения, каждое из которых связывает выход одного автомата с входом другого. Иными словами, это композиция автоматов по графу связей системы автоматов.

Введём способ сокращённой записи для преобразования функции при уменьшении её домена: для функции f и множества N обозначим $f|N \triangleq f \setminus \{(z, f(z)) \mid z \in N \cap \text{Dom}(f)\}$. Для любого множества N' имеем $f|(N \cup N') = (f|N)|N'$, и для любой функции g , такой, что $\text{Dom}(f) \cap \text{Dom}(g) = \emptyset$ и $N \cap \text{Dom}(g) = \emptyset$, имеем $(f \cup g)|N = (f|N) \cup g$. Будем считать, что операция “/” имеет тот же приоритет, что разность “\” и объединение “ $\cup$ ” множеств. В выражениях над множествами, где используются только операции “/”, “\” и “ $\cup$ ”, будем использовать бесскобочную запись и предполагать, что операции выполняются слева направо. Логическую эквивалентность обозначим символом « $\sim$ »: $a \sim b \triangleq (a \& b) \vee (\neg a \& \neg b)$.

Композиция определяется по соединению $(j, i) \in E$. Обозначим через A автомат, которому принадлежит выход j , т.е. $A = \varphi(j)$, а через B – автомат, которому принадлежит вход i , т.е. $B = \varphi(i)$.

Композиция переходов по соединению. Для соединения (j, i) определим условие $f(a, j, i, b)$ композиции двух переходов a и b и результат композиции $a[j, i]b$:

$$f(a, j, i, b) \triangleq a \in T_A \& b \in T_B \& (A = B \Rightarrow a = b) \& y_a(j) = x_b(i) \& (j \in q_a \sim i \in p_b),$$

$$f(a, j, i, b): a[j, i]b \triangleq (s_a \cup s_b, x_a \cup x_b \setminus \{i\}, p_a \cup p_b \setminus \{i\}, y_a \cup y_b \setminus \{j\}, q_a \cup q_b \setminus \{j\}, t_a \cup t_b).$$

Заметим, что если $A = B$, то

$$f(a, j, i, b) \triangleq a \in T_A \& y_a(j) = x_a(i) \& (j \in q_a \sim i \in p_a),$$

$$f(a, j, i, b): a[j, i]b \triangleq (s_a, x_a \setminus \{i\}, p_a \setminus \{i\}, y_a \setminus \{j\}, q_a \setminus \{j\}, t_a).$$

Композиция переходов естественным образом распространяется на композицию множеств переходов, определяемую как множество попарных композиций переходов. Для соединения (j, i) определим композицию множеств переходов G и H : $G[j, i]H \triangleq \{a[j, i]b \mid a \in G \& b \in H \& f(a, j, i, b)\}$.

Композиция автоматов по соединению. В композиции автоматов оставим только начальное состояние композиции и пре- и постсостояния переходов композиции. Это множество включает подмножество достижимых состояний. Для множества переходов T обозначим множество пре- и постсостояний этих переходов: $\text{States}(T) \triangleq \{s_a \mid a \in T\} \cup \{t_a \mid a \in T\}$. Для соединения (j, i) определим композицию автоматов A и B :

$$A[j, i]B \triangleq (M, I_A \cup I_B \setminus \{i\}, J_A \cup J_B \setminus \{j\}, \{s_{0A} \cup s_{0B}\} \cup \text{States}(T_A[j, i]T_B), T_A[j, i]T_B, s_{0A} \cup s_{0B}).$$

Заметим, что если в определении автомата используются правила умолчания, то композиция автоматов существенно использует композицию не только явных, но и неявных переходов.

Теорема 1. Композиция автоматов удовлетворяет условиям, налагаемым на автомат.

Доказательство см. в [2], теорема 1.

Композиция системы по соединению. Для системы R и соединения (j, i) определим

$$R[j, i] \triangleq (M, \bigvee \{A, B\} \cup \{A[j, i]B\}, E \setminus \{(j, i)\}).$$

Теорема 2. Композиция системы удовлетворяет условиям, налагаемым на систему.

Доказательство см. в [2], теорема 2.

Определённая таким образом композиция обладает свойством ассоциативности, которая важна для того, чтобы работа системы зависела только от множеств соединений, автоматов и их переходов и не зависела от какого-либо упорядочивания этих множеств. *Последовательностью всех соединений* для системы $R = (M, V, E)$ назовём последовательность Z соединений из E , в которой каждое соединение встречается ровно один раз, т.е. $Z : [1..|Z|] \rightarrow E$ является биекцией. Определим композицию системы R по последовательности всех соединений Z как $R[Z] \triangleq R[Z(1)][Z(2)]...[Z(|Z|)]$, если Z не пусто, и $R[Z] \triangleq R$, если Z пусто.

Теорема 3. Композиция системы R по всем её соединениям не зависит от порядка композиции: для любых двух последовательностей всех соединений Z и Z' имеет место $R[Z] = R[Z']$.

Доказательство см. в [2], теорема 3.

В системе $R[Z]$ нет соединений, т.е. $E_{R[Z]} = \emptyset$; автоматы такой системы не связаны друг с другом и соответствуют классам связности системы R , а все входы и выходы автоматов являются внешними. Если в $R[Z]$ больше одного автомата, то её нельзя использовать как компонент более сложной системы. Для того чтобы обойти это неудобство, определим композицию несвязанных автоматов. Её семантика достаточно прозрачна: если два автомата не связаны, то системы из этих автоматов эквивалентны одному автомату, в котором объединены входы и выходы обоих автоматов, а любые два перехода из разных автоматов объединены в один переход. Определим композицию без соединения формально.

Композиция переходов: $a[]b \triangleq (s_a \cup s_b, x_a \cup x_b, p_a \cup p_b, y_a \cup y_b, q_a \cup q_b, t_a \cup t_b)$.

В этом определении неявно предполагается, что как домены стимулов, так и домены реакций в компонуемых переходах не пересекаются: $Dom(x_a) \cap Dom(x_b) = Dom(y_a) \cap Dom(y_b) = \emptyset$. В противном случае стимул и / или реакция в композиционном переходе могли бы стать многозначными функциями. Мы не записали это требование как условие композиции двух переходов, поскольку такая композиция будет использоваться только для определения композиции несвязанных и, следовательно, разных автоматов, а в этом случае это требование автоматически выполнено.

Композиция множеств переходов: $G[]H \triangleq \{a[]b \mid a \in G \ \& \ b \in H\}$.

Композиция автоматов: $A[]B \triangleq (M, I_A \cup I_B, J_A \cup J_B, \{s_{0A} \cup s_{0B}\} \cup States(T_A[]T_B), T_A[]T_B, s_{0A} \cup s_{0B})$.

Композиция автоматов удовлетворяет условиям, налагаемым на автомат (аналог теоремы 1).

Композиция системы: $R[A, B] \triangleq (M, \bigvee \{A, B\} \cup \{A[]B\}, E)$.

Композиция системы удовлетворяет условиям, налагаемым на систему (аналог теоремы 2). Каждая композиция системы уменьшает на единицу число автоматов в ней. Поэтому через конечное число композиций система $R[Z]$ превратится в систему без соединений, состоящую из одного автомата, который обозначим через $R^\wedge$.

Нетрудно показать, что композицию несвязанных автоматов можно чередовать с композицией по соединениям. Можно только отметить, что композиция несвязанных автоматов уменьшает на 1 число компонентов связности в графе $Ng(R)$: автомат $A[]B$ оказывается связанным с каждым автоматом (отличным от A и B), с которым был связан автомат A или B .

3. Детерминизм

Автомат $A = (M, I, J, S, T, s_0)$ определен по всем стимулам, если выполнено условие, что в каждом состоянии есть переход по каждому стимулу: $\forall s \in S \ \forall x \in X \ \exists a \in T (s_a = s \ \& \ x_a = x)$.

Автомат $A = (M, I, J, S, T, s_0)$ вполне определен, если он определен по всем стимулам и существуют переходы по всем q : $\forall a \in T \ \forall q' \subseteq y_a^{-1}(M) \ \exists a' \in T (s_{a'} = s_a \ \& \ x_{a'} = x_a \ \& \ p_{a'} = p_a \ \& \ y_{a'} = y_a \ \& \ q_{a'} = q')$.

Вместо этих условий можно использовать *правила умолчания*, если в определении автомата указывать подмножество T^* явных переходов и дополнять его до множества T всех переходов согласно правилам умолчания:

1. Если в T^* нет явного перехода из некоторого состояния по некоторому стимулу, то это означает, что такой переход определён неявно без приёма и выдачи сообщений и без изменения состояния.

2. Если в T^* есть переход (s, x, p, y, q, t) , но для некоторого $q' \subseteq y^{-1}(M)$ нет перехода вида (s, x, p, y, q', t') , то это означает, что такой переход определён неявно без изменения состояния ($t' = t$).

Формально это означает

$$T \triangleq T^* \cup \{(s, x, \emptyset, J \times \{\Lambda\}, \emptyset, s) \mid s \in S \& x \in X \& \forall p, y, q, t (s, x, p, y, q, t) \notin T^*\} \cup \\ \cup \{(s, x, p, y, q', s) \mid \exists q, t (s, x, p, y, q, t) \in T^* \& q' \subseteq y^{-1}(M) \& \forall t' (s, x, p, y, q', t') \notin T^*\}.$$

Автомат A будем называть *детерминированным*, если выполнены следующие условия:

1. Состояние s и стимул x однозначно определяют приём сообщений p и реакцию y :

$$\forall a, b \in T_A (s_a = s_b \& x_a = x_b \Rightarrow p_a = p_b \& y_a = y_b).$$

2. Одинаково помеченные переходы из одного состояния совпадают, т.е. ведут в одно постсостояние:

$$\forall a, b \in T_A (s_a = s_b \& x_a = x_b \& p_a = p_b \& y_a = y_b \& q_a = q_b \Rightarrow t_a = t_b).$$

Проблема в том, что композиция детерминированных автоматов может оказаться недетерминированной. Это может произойти из-за цикла соединений, если в автоматах этого цикла есть две «параллельные», но разные цепочки компокуемых переходов, ведущих из одинаковых состояний и отличающихся в параметрах x и q только по входам и выходам соединений цикла. Для одной цепочки по циклу соединений передаются одни сообщения, а для другой – другие. Более строго, пусть в графе связей есть цикл как чередующаяся последовательность автоматов и неповторяющихся соединений $W = A_1, (j_1, i_2), A_2, (j_2, i_3), \dots, A_n, (j_n, i_{n+1})$, $A_1: \forall k, k'=1..n (j_k, i_{k+1}) \neq (j_{k'}, i_{k'+1})$. Обозначим $i_1 = i_{n+1}$. Пусть также в каждом из автоматов A_k имеются два перехода a_k и a'_k , такие, что если $A_k = A_{k'}$, то $a_k = a_{k'}$. Обозначим $a_{n+1} = a_1$ и $a'_{n+1} = a'_1$. Пусть переходы в соседних в цикле автоматов, т.е. переходы a_k и a_{k+1} , и переходы a'_k и a'_{k+1} компокуются по соединению (j_k, i_{k+1}) , т.е. $f(a_k, j_k, i_{k+1}, a_{k+1}) = f(a'_k, j_k, i_{k+1}, a'_{k+1}) = \text{true}$. Пусть для каждого $k = 1..n$ переходы a_k и a'_k начинаются в одном состоянии и отличаются в x только по входу i_k и в q только по выходу j_k , т.е. если $a_k = (s_k, x_k, p_k, y_k, q_k, t_k)$ и $a'_k = (s'_k, x'_k, p'_k, y'_k, q'_k, t'_k)$, то $s_k = s'_k, x_k \setminus \{i_k\} = x'_k \setminus \{i_k\}, q_k \setminus \{j_k\} = q'_k \setminus \{j_k\}$. Тогда композиция по всем соединениям цикла W содержит два перехода $a = a_1[j_1, i_2]a_2[j_2, i_3]...a_n[j_n, i_{n+1}]a_1$ и $a' = a'_1[j_1, i_2]a'_2[j_2, i_3]...a'_n[j_n, i_{n+1}]a'_1$, в которых $s_a = s_{a'}, x_a = x_{a'}$ и $q_a = q_{a'}$, но, вообще говоря, $p_k \neq p'_k, y_k \neq y'_k$ и $t \neq t'$.

В то же время можно заметить, что если хотя бы в одном автомате цикла, например в первом, по соединению (j_1, i_2) в разных переходах a_1 и a'_1 передаётся одно и то же сообщение $y_1(j_1) = y'_1(j_1)$, то недетерминизм не возникает. Действительно, из условия композиции переходов следует, что в следующем втором автомате $x_2(i_2) = y_1(j_1) = y'_1(j_1) = x'_2(i_2)$, а тогда, поскольку $x_2 \setminus \{i_2\} = x'_2 \setminus \{i_2\}$, имеем $x_2 = x'_2$. Отсюда, поскольку $s_2 = s'_2$, в силу первого условия детерминизма второго автомата имеем $p_2 = p'_2, y_2 = y'_2$, в частности $y_2(j_2) = y'_2(j_2)$. Продолжая аналогичные рассуждения по циклу соединений, получаем $x_k = x'_k, p_k = p'_k, y_k = y'_k$ для каждого k . Тогда для каждого k имеем $i_{k+1} \in p_{k+1} \sim i_{k+1} \in p'_{k+1}$, и по условию композиции $j_k \in q_k \sim j_k \in q'_k$, что вместе с $q_k \setminus \{j_k\} = q'_k \setminus \{j_k\}$ влечёт $q_k = q'_k$. Отсюда по второму условию детерминизма для каждого k имеем $t_k = t'_k$. Тем самым, $a_k = a'_k$ для каждого k , что влечёт $a = a'$. Таким образом, одинаковые сообщения $y_k(j_k) = y'_k(j_k)$ «склеивают» цепочки переходов $a_1 = a'_1, a_2 = a'_2, \dots, a_n = a'_n$, а композиция по такому циклу становится детерминированной.

В [1] композиция детерминированных автоматов детерминированная. В терминах данной статьи это означает, что система детерминированная, если выполнены следующие условия: 1) автоматы разбиты на два класса: класс «вершин» и класс «дуг», 2) каждое соединение связывает автоматы разных классов, 3) «вершина» определена по всем стимулам, а «дуга» вполне определена, 4) все автоматы детерминированные, 5) «дуга» имеет один вход и один выход, 6) «дуга» моделирует очередь сообщений длины 1. Класс «вершин» соответствует вершинам, а класс «дуг» – дугам графа связей из [1]. Условие 5 можно ослабить: достаточно, чтобы «дуга» имела один вход, но она может иметь несколько выходов. Вместо условия 6 достаточно, чтобы в «дуге» реакция зависела только от состояния (т.е. не зависела от стимула): $\forall A \in V_2 \forall a, a' \in T_A (s_a = s_{a'} \Rightarrow y_a = y_{a'})$.

Композиционный переход порождается множеством переходов в автоматах-операндах, которое содержит по одному переходу в каждой «дуге» и либо один переход в «вершине», либо «вершина» «стоит», если какие-то из сообщений, которые она могла бы послать на выходящие «дуги», ими не принимаются.

«Дуга» «склеивает» цепочки в каждом цикле, в который она входит, т.е. композиция по каждому такому циклу становится детерминированной. Для «дуги» в [1] это условие выполнено.

Теорема 4. Если система $R = (M, V, E)$ удовлетворяет указанным выше шести условиям, то после специальной композиции каждой «вершины» со всеми выходящими из неё «дугами» дальнейшая композиция по описанным в разделе 2 правилам дает детерминированный автомат.

Доказательство следует из теорем 1 и 2 в [3].

Заключение

Предложенная в данной статье модель системы автоматов может быть использована как база для тестирования составных систем. Особенно это полезно в том случае, когда в графе связей нет ошибок (*гипотеза о связях*), т.е. граф связей совпадает с заданным. В этом случае тестирование сводится к проверке правильности переходов каждого автомата. Однако проблема в том, что автомат может тестироваться только как часть системы, т.е. тестер не имеет непосредственного доступа к автомату, и вынужден осуществлять тестовые воздействия с помощью сообщений, посылаемых по внешним входным соединениям, которые ведут, быть может, в другие автоматы. Тестирование компонента такой системы похоже на тестирование в контексте [12–15], когда этот компонент рассматривается как тестируемая система, а остальные – как контекст. Существенное отличие, однако, в том, что в таком контексте тоже возможны ошибки, но если верна гипотеза о связях, то только в компонентах, а не в графе связей. С другой стороны, при тестировании проверяется работа сразу нескольких компонентов, через которые проходят сообщения. Поскольку автомат тестируется как часть системы, могут быть проверены не все его переходы, которые проверяются при автономном тестировании с прямым доступом к автомату. Речь идет только о переходах автоматов, *достижимых* при их работе в составе системы.

Тестирование композиционного автомата системы, получающегося композицией автоматов по заданному графу связей, обеспечивает проход по всем его достижимым переходам. При этом, конечно, проверяются все достижимые переходы автоматов-компонентов, но может делаться много «лишней работы». Гипотеза о связях позволяет существенно сократить время тестирования.

Предполагается, что нам известно, каким должен быть каждый автомат (задан граф переходов автомата с точностью до изоморфизма), и именно это проверяется при тестировании. Кроме того, тест может наблюдать как состояния автоматов в вершинах графа связей, так и сообщения на его дугах. Такие предположения могут быть оправданы, например, при имитационном тестировании аппаратуры (*simulation-based verification*) (см. например, [16]).

При этих условиях может быть предложен алгоритм построения набора тестов [3], основанный на фильтрации тестов, генерируемых для композиционного автомата системы: тест отбрасывается, если он покрывает только такие переходы в компонентах системы, которые покрываются уже отобранными тестами. Такой набор тестов будет полным (проверяет все достижимые переходы автоматов-компонентов) при выполнении двух условий: 1) верна гипотеза о связях, 2) система детерминирована. Дополнительно алгоритм может определить недостижимые переходы автоматов.

ЛИТЕРАТУРА

1. Бурдонов И.Б., Косачев А.С. Тестирование системы автоматов // Труды ИСП РАН. 2016. Т. 28 (1).
2. Бурдонов И.Б., Косачев А.С. Система автоматов: композиция по графу связей // Труды ИСП РАН. 2016. Т. 28 (1).
3. Бурдонов И.Б., Косачев А.С. Система автоматов: условия детерминизма и тестирование // Труды ИСП РАН. 2016. Т. 28 (1).
4. Hoare C.A.R. Communicating Sequential Processes. Englewood Cliffs, NJ : Prentice Hall International, 1985.
5. Milner R. Communication and Concurrency. Prentice-Hall, 1989.
6. Langerak R. A testing theory for LOTOS using deadlock detection // Protocol Specification, Testing, and Verification IX / eds. by E. Brinksma, G. Scollo, C.A. Vissers. North-Holland, 1990. P. 87–98.
7. Tretmans J. Test Generation with Inputs, Outputs and Repetitive Quiescence // Software-Concepts and Tools. 1996. Vol. 17, Issue 3.
8. van der Bijl M., Rensink A., Tretmans J. Compositional testing with ioco // Formal Approaches to Software Testing: Third International Workshop, FATES 2003, Montreal, Quebec, Canada, October 6th, 2003 / eds. by Alexandre Petrenko, Andreas Ulrich. LNCS Springer, 2003. V. 2931. P. 86–100.

9. Petrenko A., Yevtushenko N., Huo J.L. Testing Transition Systems with Input and Output Testers. Proc. IFIP TC6/WG6.1 15th Int. Conf. Testing of Communicating Systems, TestCom'2003. Sophia Antipolis, France, 2003. May 26–29. P. 129–145.
10. Бурдонов И.Б., Косачев А.С. Системы с приоритетами: конформность, тестирование, композиция // Программирование. 2009. № 4. С. 24–40.
11. Бурдонов И.Б., Косачев А.С. Согласование конформности и композиции // Программирование. 2013. № 6. С. 3–15.
12. Revised Working Draft on “Framework: Formal Methods in Conformance Testing”. JTC1/SC21/WG1/Project 54/1, ISO Interim Meeting, ITU-T on. Paris, 1995.
13. Bourdonov I.B., Kossatchev A.S., Kuliain V.V. Conformance theory of the systems with Refused Inputs and Forbidden Actions. M.: Nauka, 2008. 412 p. (in Russian)
14. Bourdonov I. Conformance theory (functional testing on formal model base). LAP LAMBERT Academic Publishing. Saarbrücken, Germany, 2011. 428 p. (in Russian)
15. Bourdonov I.B., Kossatchev A.S. Specification Completion for IOCO // Programming and Computer Software. 2011. V. 37 (1). P. 1–14.
16. Kamkin A.S., Chupilko M.M. Survey of modern technologies of simulation-based verification of hardware // Programming and Computer Software. 2011. V. 37 (3). P. 147–152.

Бурдонов Игорь Борисович, д-р физ.-мат. наук. E-mail: igor@ispras.ru

Косачев Александр Сергеевич, канд. физ.-мат. наук. E-mail: kos@ispras.ru

Институт системного программирования РАН (ИСП РАН), Москва

Поступила в редакцию 30 мая 2016 г.

Burdonov Igor B., Kossatchev Alexander S. (Institute for System Programming of the Russian Academy of Sciences, Moscow, Russian Federation).

Generalized model of automata system.

Keywords: directed graph; communicating automata; automata composition; determinism; networks.

DOI: 10.17223/19988605/37/10

The problem of modeling and composing of aggregate systems is considered. The system components are described by finite automata with multiple entries and exits. The communication between automata is described with message passing over simplex communication channels. The system is described with an oriented graph of links. Each node of the graph corresponds to the automaton of a component and an arc corresponds to a communication channel connecting an output of one automaton with input of another automaton. An automaton of the graph node in each state can accept multiple messages from its entries (at most one message from each input) and send multiple messages to its outputs (at most one message to each output). Inputs (outputs) of the automata not connected to outputs (inputs) of other automata are considered to be external and are used for communication between the system and its environment. The automata of the system operate synchronously: on each step each automaton performs one transition. A transition of an automaton imposes requirements on states of all its inputs and outputs (messages in them are specified) and explicitly specifies subset of inputs and outputs through which the messages are received or sent, respectively. Synchronous communication between automata means that for each link the requirements of the automata connected with this link must conform to each other. It makes possible to describe a wider spectrum of automata behavior. For example, a priority of message receiving: if there are multiple message at the automata inputs, it can receive messages with the highest priority and discard the rest of the messages. It also makes possible for the automaton to receive messages regardless of ability to simultaneously send some message to some output. A composition of the automata of the system according to the graph of links is defined and its associativity is proved. In conclusion, the problems of testing of aggregate systems are described.

REFERENCES

1. Burdonov, I.B. & Kossatchev, A.S. (2016) Testing of automata system. *Trudy ISP RAN – The proceeding of ISP RAS*. 28(1). (In Russian).
2. Burdonov, I.B. & Kossatchev, A.S. (2016) Automata system: composition on connection graph. *Trudy ISP RAN – The proceeding of ISP RAS*. 28(1). (In Russian).
3. Burdonov, I.B. & Kossatchev, A.S. (2016) Automata system: conditions of determinism and testing. *Trudy ISP RAN – The proceeding of ISP RAS*. 28(1). (In Russian).
4. Hoare, C.A.R. (1985) *Communicating Sequential Processes*. Englewood Cliffs, New Jersey: Prentice Hall International.
5. Milner, R. (1989) *Communication and Concurrency*. Prentice-Hall.
6. Langerak, R. (1990) A testing theory for LOTOS using deadlock detection. In: Brinksma, E., Scollo, G. & Vissers, C.A. (eds) *Protocol Specification, Testing, and Verification IX*. North-Holland. pp. 87–98.
7. Tretmans, J. (1996) Test Generation with Inputs, Outputs and Repetitive Quiescence. *Software-Concepts and Tools*. 17(3). DOI: 10.1007/3-540-61042-1_42
8. van der Bijl, M., Rensink, A. & Tretmans, J. (2003) Compositional testing with ioco. *Formal Approaches to Software Testing: Third International Workshop, FATES 2003*. Montreal, Quebec, Canada. October 6. 2931. pp. 86–100. DOI: 10.1007/978-3-540-24617-6_7
9. Petrenko, A., Yevtushenko, N. & Huo, J.L. (2003) Testing Transition Systems with Input and Output Testers. *Proc. IFIP TC6/WG6.1 15th Int. Conf. Testing of Communicating Systems, TestCom'2003*. Sophia Antipolis, France. May 26–29. pp. 129–145. DOI: 10.1016/S0140-3664(96)01157-7

10. Bourdonov, I.B. & Kossatchev, A.S. (2009) Systems with Priorities: Conformance, Testing, and Composition. *Programming and Computer Software*. 35(4). pp. 198-211. DOI: 10.1134/S0361768809040045
11. Bourdonov, I.B. & Kossatchev, A.S. (2013) Agreement between Conformance and Composition. *Programming and Computer Software*. 39(6). pp. 269-278. DOI: 10.1134/S0361768813060029
12. Revised Working Draft on “Framework: Formal Methods in Conformance Testing”. (1995) *JTC1/SC21/WG1/Project 54/1, ISO Interim Meeting, ITU-T on*. Paris.
13. Bourdonov, I.B., Kossatchev, A.S. & Kuliainin, V.V. (2008) *Conformance theory of the systems with Refused Inputs and Forbidden Actions*. Moscow: Nauka. (In Russian).
14. Bourdonov, I. (2011) *Conformance theory (functional testing on formal model base)*. Saarbrücken: LAP LAMBERT Academic Publishing.
15. Bourdonov, I.B. & Kossatchev, A.S. (2011) Specification Completion for IOCO. *Programming and Computer Software*. 37(1). pp. 1-14. DOI: 10.1134/S0361768811010014. 1
16. Kamkin, A.S. & Chupilko, M.M. (2011) Survey of modern technologies of simulation-based verification of hardware. *Programming and Computer Software*. 37(3). pp. 147-152. DOI: 10.1134/S0361768811030017

СВЕДЕНИЯ ОБ АВТОРАХ

БОГДАНОВ Александр Леонидович – доцент, кандидат технических наук, доцент кафедры математических методов и информационных технологий в экономике экономического факультета Томского государственного университета. E-mail: bogdanov.al@mail.ru

БОГДАНОВА Юлия Вячеславовна – кандидат физико-математических наук, доцент кафедры геологии и разработки нефтяных месторождений Института природных ресурсов Томского политехнического университета. E-mail: bogdanova-ju@tpu.ru

БУРДОНОВ Игорь Борисович – доктор физико-математических наук, ведущий научный сотрудник Института системного программирования РАН (г. Москва). E-mail: igor@ispras.ru

ЕФАНОВ Дмитрий Викторович – кандидат технических наук, доцент кафедры «Автоматика и телемеханика на железных дорогах» Петербургского государственного университета путей сообщения Императора Александра I (г. Санкт-Петербург). E-mail: TrES-4b@yandex.ru

ИГНАТЬЕВ Николай Александрович – доктор физико-математических наук, профессор кафедры программных и сетевых технологий механико-математического факультета Национального университета Узбекистана (г. Ташкент). E-mail: ignat-ev@rambler.ru

КОСАЧЕВ Александр Сергеевич – кандидат физико-математических наук, ведущий научный сотрудник Института системного программирования РАН (г. Москва). E-mail: kos@ispras.ru

ЛОХВИЦКИЙ Владимир Александрович – кандидат технических наук, докторант кафедры математического и программного обеспечения Военно-космической академии имени А.Ф. Можайского (г. Санкт-Петербург). E-mail: vovan296@mail.ru

МЕДВЕДЕВ Геннадий Алексеевич – профессор, доктор физико-математических наук, профессор факультета прикладной математики и информатики Белорусского государственного университета (г. Минск). E-mail: MedvedevGA@bsu.by

НАЗАРОВ Анатолий Андреевич – профессор, доктор технических наук, заведующий кафедрой теории вероятности и математической статистики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: nazarov.tsu@gmail.com.

НИССЕНБАУМ Ольга Владимировна – кандидат физико-математических наук, доцент кафедры информационной безопасности института математики и компьютерных наук Тюменского государственного университета. E-mail: onissenbaum@rambler.ru

ПАВЛИВ Дмитрий Александрович – аспирант факультета прикладной математики и информатики Белорусского государственного университета (г. Минск). E-mail: simoco@yandex.ru

УЛАНОВ Александр Викторович – кандидат технических наук, старший научный сотрудник военного института (научно-исследовательского) Военно-космической академии имени А.Ф. Можайского (г. Санкт-Петербург). E-mail: ulanov246@rambler.ru

ФЁДОРОВА Екатерина Александровна – кандидат физико-математических наук, научный сотрудник лаборатории математического моделирования Сибирского физико-технического института Томского государственного университета. E-mail: moiskate@mail.ru

ХАРЧЕНКО Анастасия Михайловна – инженер кафедры информационной безопасности Института математики и компьютерных наук Тюменского государственного университета. E-mail: kharchenkoam.86@gmail.com

ЦИЦИАШВИЛИ Гурам Шалвович – профессор, доктор физико-математических наук, профессор кафедры алгебры, геометрии и анализа ДВФУ, заведующий лабораторией вероятностных методов и системного анализа ИПМ ДВО РАН (г. Владивосток). E-mail: guram@iam.dvo.ru

ШУЛЕНИН Валерий Петрович – кандидат технических наук, доцент кафедры теоретической кибернетики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: shvp@fpmk.tsu.ru

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2016. № 4 (37)

Редактор Н.А. Афанасьева
Корректор Е.Г. Шумская
Оригинал-макет А.И. Лелоюр
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцовой

Подписано к печати 12.12.2016 г. Формат 60x84¹/₈.
Гарнитура Times. Печ. л. 12,5; усл. печ. л. 11,6.
Тираж 250 экз. Заказ № 2266. Цена свободная.
Дата выхода в свет 30.12.2016 г.

Журнал отпечатан на полиграфическом оборудовании
Издательского Дома Томского государственного университета
634050, г. Томск, Ленина, 36
Тел. 8(382-2)–52-98-49; 8(382-2)–53-15-28; 8(382-2)–52-96-75
Сайт: <http://publish.tsu.ru>; E-mail: rio.tsu@mail.ru