

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2014

№ 4 (29)

Свидетельство о регистрации: ПИ № ФС 77-29497 от 27 сентября 2007 г.



EDITORIAL BOARD

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

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

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

Sergey Vorobeychikov – Doctor of Science, Prof. of the Higher Mathematics and Mathematical Modeling Department

Gennady Koshkin – Doctor of Science, Prof. of the Theoretical Cybernetics Department

Yury Kostyuk – Doctor of Science, Prof. of the Theoretical Informatics Department

Anjela Matrosova – Doctor of Science, Prof., Head of the Programming Department

Anatoly Nazarov – Doctor of Science, Prof., Head of the Probability Theory and Mathematical Statistics Department

Sergey Sushchenko – Doctor of Science, Prof., Head of the Applied of Information Department, Dean of the Faculty of Informatics

EDITORIAL COUNCIL

Ana Rosa Cavalli
PhD, Prof.
University VII
Paris, France

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

Alexander Dudin
Doctor of Science, 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 Science, Prof.
University of Technology
Tallinn, Estonia

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

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

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

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

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

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

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

Коскин Геннадий Михайлович – проф., д-р физ.-мат. наук, проф. каф. теоретической кибернетики ТГУ

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

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

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

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

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
д-р техн. наук, проф.
Технологический университет,
Таллинн, Эстония

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

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

JORNAL 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).

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке, его подписной индекс 44031 в объединённом каталоге «Пресса России». 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 «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» включен в Перечень изданий для публикации основных результатов кандидатских и докторских диссертаций, а также входит в систему Российского Индекса Научного Цитирования (РИНЦ). Правила оформления статей приведены на сайте:

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

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Мухина О.О., Смагин В.И. Динамические локально-оптимальные системы управления по выходам для объектов с интервальными параметрами с запаздыванием по состоянию	4
---	---

ОБРАБОТКА ИНФОРМАЦИИ

Ашуров М. Ф. Сравнение потоковых методов классификации текстов художественной литературы на основе сжатия информации и подсчета подстрок	16
Гуменюк А.С., Поздниченко Н.Н., Шпынов С.Н. Формальный анализ строя локальной структуры нуклеотидных последовательностей	23
Довбыш А.С., Берест О.Б. Трёхальтернативная обучающаяся система поддержки принятия решений для автоматизации технологического процесса	31
Фёдорова Е.А. Вычисление моментов в RQ-системе MMPP M 1	41
Чимитова Е.В., Четвертакова Е.С. Построение деградиционной гамма-модели с учетом влияния объясняющих переменных	51
Шумилов Б.М., Ыманов У.С. Сплайн-вейвлеты, ортогональные многочленам, и оптимизация вычислений вейвлет-преобразования	61

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Бабанов А.М. Правила порождения ограничений в семантических моделях данных ORM и ERMM	68
--	----

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

Твардовский А.С., Евтушенко Н.В. К минимизации автоматов с временными ограничениями	77
Сведения об авторах	84

CONTENTS

CONTROL OF DYNAMICAL SYSTEMS

Mukhina O.O., Smagin V.I. Dynamic locally-optimal control systems for objects with interval parameters with state delay	4
--	---

DATA PROCESSING

Ashurov M.F. Comparison of stream-based fiction text classification methods based on data compression and counting substrings	16
Gumenyuk A.S., Pozdnichenko N.N., Shpynov S.N. Formal analysis of order in the local structure of the nucleotide sequences	23
Dovbysh A.S., Berest O.B. Three-alternative learning rating system for decision support system of process automation	31
Fedorova E.A. Calculation of moments in retrieval queueing system MMPP M 1	41
Chimitova E.V., Chetvertakova E.S. The construction of the gamma degradation model with covariates	51
Shumilov B.M., Ymanov U.S. Spline-wavelets, orthogonal to polynomials, and optimization of calculations of wavelet-transformation	61

INFORMATICS AND PROGRAMMING

Babanov A.M. Constraint specifications generating rules in semantic models ORM and ERMM	68
--	----

DISCRETE FUNCTION AND AUTOMATION

Tvardovskiy A.S., Yevtushenko N.V. Minimizing timed Finite State Machines	77
Brief information about the authors	84

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

УДК 681.5.01:62-50

О.О. Мухина, В.И. Смагин

ДИНАМИЧЕСКИЕ ЛОКАЛЬНО-ОПТИМАЛЬНЫЕ СИСТЕМЫ УПРАВЛЕНИЯ ПО ВЫХОДУ ДЛЯ ОБЪЕКТОВ С ИНТЕРВАЛЬНЫМИ ПАРАМЕТРАМИ С ЗАПАЗДЫВАНИЕМ ПО СОСТОЯНИЮ

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта (№ 13-08-01015 А).

Рассматривается задача динамического локально-оптимального управления по наблюдаемому выходу для дискретных объектов с интервальными параметрами и с запаздыванием по состоянию. Для ее решения предлагаются алгоритмы, в основе которых лежит оптимизация локального критерия без использования расширения пространства состояний. Управление определяется как функция измеряемых переменных с памятью, отслеживаемого сигнала и динамического звена. Исследуется асимптотическое поведение замкнутой системы.

Ключевые слова: локально-оптимальное слежение; запаздывание по состоянию; интервальные параметры; динамические системы управления; управление по выходу.

Локально-оптимальные дискретные системы управления являются частным случаем дискретного прогнозирующего управления (Model predictive control) с прогнозом на один такт. Основным достоинством метода локально-оптимального управления является существенное упрощение процедуры синтеза. Область применения метода МРС и, соответственно, метода локально-оптимального управления охватывает задачи управления техническими системами, производственными системами, управление запасами и финансовую математику [1–14].

В целях улучшения качества управления объектами применяется практика введения в закон управления наблюдателей Люенбергера [15] или динамической обратной связи пониженной размерности [16–18].

В настоящей работе предлагается осуществлять синтез следящих динамических систем управления по выходу на основе оптимизации локального критерия, при косвенных измерениях для дискретных объектов с интервальными параметрами на основе вероятностного метода с учетом запаздываний по состоянию. Управление определяется как функция измеряемых переменных, динамического звена и отслеживаемого сигнала. Исследуется асимптотическое поведение системы, строятся оценки для асимптотической точности слежения. Результаты работы являются развитием [9] на случай синтеза динамической системы управления по выходу для модели объекта с интервальными параметрами.

1. Постановка задачи

Пусть управляемый объект с запаздыванием по состоянию и канал наблюдений описываются уравнениями

$$\begin{aligned} x(k+1) &= (A + \sum_{i=1}^r A_i \theta_i) x(k) + (\tilde{A} + \sum_{i=1}^r \tilde{A}_i \theta_i) x(k-h) + (B + \sum_{i=1}^r B_i \theta_i) u(k) + q(k); \\ x(\tau) &= \varphi(\tau), \tau = -h, 1-h, 2-h, \dots, 0; k = 0, 1, 2, \dots; \\ y(k) &= Sx(k) + v(k). \end{aligned} \quad (1)$$

В (1), (2) $x(k) \in R^n$ – вектор состояний; $h > 0$ – величина временного запаздывания (целое число); $u(k) \in R^m$ – управление; $y(k) \in R^l$ – вектор измерений; $A, A_i, \tilde{A}, \tilde{A}_i, B, B_i, i = \overline{1, r}$ – матрицы соответств-

ющих размерностей; S – матрица канала наблюдения; матрицы B и S полного ранга; пары матриц (A, B) и $(\tilde{A}, B)$ управляемы, пары матриц (S, A) и $(S, \tilde{A})$ наблюдаемы; x_0 – начальные условия ($M\{x_0 x_0^T\} = P_{x_0}$); $q(k)$, $v(k)$ – гауссовские случайные последовательности входных возмущений и ошибок измерений с характеристиками: $M\{q(k)\} = 0$, $M\{v(k)\} = 0$, $M\{q(k)v^T(j)\} = 0$, $M\{q(k)q^T(j)\} = Q(k)\delta_{kj}$, $M\{v(k)v^T(j)\} = V(k)\delta_{kj}$ ($\delta_{i,j}$ – символ Кронекера, $Q(k) = Q^T(k) \geq 0$, $V(k) = V^T(k) \geq 0$ – неотрицательно определенные матрицы); θ_i – неопределенные параметры интервального типа ($-1 \leq \theta_i \leq 1, \forall i = \overline{1, n}$). Матрицы B и S полного ранга, пара матриц (A, B) управляема, пара матриц (S, A) наблюдаема.

Оптимизируемый локальный критерий имеет вид

$$I(k) = M\{(w(k+1) - z(k))^T C(w(k+1) - z(k)) + u^T(k)Du(k)\}, \quad (3)$$

где $w(k) = Hx(k)$ – управляемый выход системы (H – матрица выхода системы), $C = C^T \geq 0$, $D = D^T \geq 0$ – весовые матрицы, $z(k) \in R^n$ – отслеживаемый вектор, удовлетворяющий уравнению

$$z(k+1) = Fz(k) + q_z(k), \quad z(0) = z_0, \quad k = 0, 1, 2, \dots \quad (4)$$

В (4) $q_z(k)$ – гауссовская случайная последовательность с характеристиками: $M\{q_z(k)\} = 0$, $M\{q_z(k)q_z^T(j)\} = 0$, $M\{q_z(k)v^T(j)\} = 0$, $M\{q_z(k)q_z^T(j)\} = Q_z(k)\delta_{k,j}$, z_0 – начальные условия ($M\{z_0 z_0^T\} = P_{z_0}$, $M\{z_0 x_0^T\} = P_{z_0 x_0}$, $M\{x_0 z_0^T\} = P_{x_0 z_0}$), F – матрица динамики модели отслеживаемого сигнала.

Требуется найти управление объектом (1), используя наблюдения (2), минимизируя критерий (3).

Суть вероятностного подхода заключается в том, что неопределенные интервальные параметры θ_i заменяются независимыми случайными последовательностями $\bar{\theta}(k)$ с равномерным законом распределения на интервале $[-1, 1]$.

2. Оптимизация локального критерия

Динамический закон управления объектом (1) при измерениях (2) зададим в виде

$$u(k) = K_0(k)w(k) + K_1(k)y(k) + K_2(k)y(k-h) + K_3(k)z(k), \quad (5)$$

где коэффициенты передачи $K_0(k), K_1(k), K_2(k), K_3(k)$ подлежат определению, а переменная $w(k)$ определяется с помощью динамического звена заданной размерности [16, 17]:

$$w(k+1) = \bar{A}(k)w(k) + \bar{B}(k)y(k) + \bar{C}(k)z(k), \quad w(0) = 0. \quad (6)$$

В (6) $w(k) \in R^p$ ($1 \leq p \leq n$), $\bar{A}(k) = (L + MBK_0(k))$, $\bar{B}(k) = M(BK_1(k) + K)$, $\bar{C}(k) = MBK_3(k)$. Матрица M удовлетворяет уравнению

$$M(A - KS) - LM = 0,$$

где L – заданная устойчивая матрица (ее собственные числа лежат внутри единичного круга). Матрица K вычисляется так, чтобы $A - KS$ имела заданные собственные числа, часть из которых совпадала бы с собственными числами матрицы L . Так как пара матриц (S, A) наблюдаема, то такую матрицу K всегда можно построить, применив методы модального управления [19]. Собственные числа матрицы $A - KS$ разбиваются на две группы: $\lambda_i^{(1)}$ ($i = \overline{1, p}$) и $\lambda_j^{(2)}$ ($j = \overline{1, n-p}$), причем группа $\lambda_i^{(1)}$ состоит из чисел, которые совпадают с собственными числами матрицы L . Применяя метод [20], строится неособенная матри-

ца $\Lambda = [\Lambda_1, \Lambda_2]$, блоки которой Λ_1, Λ_2 формируются из ρ и $n - \rho$ линейно независимых столбцов матриц

$$\Delta_1 = \prod_{j=1}^{n-\rho} (A - KS - \lambda_j^{(2)} E), \quad \Delta_2 = \prod_{i=1}^{\rho} (A - KS - \lambda_i^{(1)} E).$$

Тогда матрицы M и K определяются следующими равенствами:

$$K = \Lambda_1, M = \Xi_1,$$

где Ξ_1 – соответствующий блок матрицы $\Lambda^{-1} = [\Xi_1^T, \Xi_2^T]^T$.

Решение задачи локально-оптимального слежения сформулируем в виде теоремы.

Теорема 1. Пусть $K = \Lambda_1$, $M = \Xi_1$. Если для объекта (1), канала измерений (2) и локального критерия (3) матрицы

$$\begin{aligned} \bar{C} &= (B^T H^T C H B + D + \frac{1}{3} \sum_{i=1}^r B_i^T H^T C H B_i) > 0, \\ \bar{P}(k) &= \begin{bmatrix} P_w(k) & SP_{xw}(k) & SP_{xw}(k-h, k) & P_{zw}(k) \\ P_{wx}(k)S^T & SP_x(k)S^T + V(k) & SP_x(k-h, k)S^T & P_{zx}(k)S^T \\ P_{wx}(k, k-h)S^T & SP_x(k, k-h)S^T & SP_x(k-h)S^T + V(k-h) & P_{zx}(k, k-h)S^T \\ P_{wz}(k) & SP_{xz}(k) & SP_{xz}(k-h, k) & P_z(k) \end{bmatrix} > 0 \end{aligned} \quad (7)$$

положительно определены для всех $k = 1, 2, \dots$, то оптимальные в смысле минимума критерия (3) коэффициенты передачи для управления (5) определяются матричными уравнениями

$$K_0^*(k) = aK_1^*(k) + bK_2^*(k) + cK_3^*(k) + d; \quad (8)$$

$$K_1^*(k) = eK_0^*(k) + fK_1^*(k) + gK_2^*(k) + h; \quad (9)$$

$$K_2^*(k) = mK_0^*(k) + nK_1^*(k) + pK_2^*(k) + \beta; \quad (10)$$

$$K_3^*(k) = sK_0^*(k) + tK_1^*(k) + lK_2^*(k) + k, \quad (11)$$

где

$$\begin{aligned} a &= -SP_{xw}(k)P_w^{-1}(k); \quad b = -SP_{xw}(k-h, k)P_w^{-1}(k); \quad c = -P_{zw}(k)P_w^{-1}(k); \\ d &= -\bar{C}^{-1}[\frac{1}{3} \sum_{i=1}^r B_i^T H^T C (H\tilde{A}_i P_{xw}(k-h, k) + H\tilde{A}_i P_{xw}(k-h, k)) + \\ &\quad + B^T H^T C (H\tilde{A} P_{xw}(k) + H\tilde{A} P_{xw}(k-h, k) - P_{zw}(k))]P_w^{-1}(k); \\ e &= -P_{wx}(k)S^T [SP_x(k)S^T + V(k)]^{-1}; \quad f = -SP_x(k-h, k)S^T [SP_x(k)S^T + V(k)]^{-1}; \\ g &= -P_{zx}(k)S^T [SP_x(k)S^T + V(k)]^{-1}; \\ h &= -\bar{C}^{-1}[\frac{1}{3} \sum_{i=1}^r B_i^T H^T C (H\tilde{A}_i P_x(k-h, k) + H\tilde{A}_i P_x(k))S^T + \\ &\quad + B^T H^T C (H\tilde{A} P_x(k) + H\tilde{A} P_x(k-h, k) - P_{zx}(k))S^T] [SP_x(k)S^T + V(k)]^{-1}; \\ m &= -P_{wx}(k, k-h)S^T [SP_x(k-h)S^T + V(k-h)]^{-1}; \quad n = -SP_x(k, k-h)S^T [SP_x(k-h)S^T + V(k-h)]^{-1}; \\ p &= -P_{zx}(k, k-h)S^T [SP_x(k-h)S^T + V(k-h)]^{-1}; \\ \beta &= -\bar{C}^{-1}[\frac{1}{3} \sum_{i=1}^r B_i^T H^T C (H\tilde{A}_i P_{xz}(k-h) + H\tilde{A}_i P_{xz}(k, k-h)) + B^T H^T C (H\tilde{A} P_{xz}(k, k-h) + \\ &\quad + H\tilde{A} P_{xz}(k-h) - P_{zx}(k, k-h))S^T] [SP_x(k-h)S^T + V(k-h)]^{-1}; \\ s &= -P_{wz}(k)P_z^{-1}(k); \quad t = -SP_{xz}(k)P_z^{-1}(k); \quad l = -SP_{xz}(k-h, k)P_z^{-1}(k); \end{aligned}$$

$$k = -\bar{C}^{-1} \left[\frac{1}{3} \sum_{i=1}^r B_i^T H^T C (H \tilde{A}_i P_{xz}(k-h, k) + H A_i P_{xz}(k)) + \right. \\ \left. + B^T H^T C (H A P_{xz}(k) + H \tilde{A} P_{xz}(k-h, k) - P_z(k)) S^T \right] P_z^{-1}(k). \quad (12)$$

В (12) введены обозначения:

$$P_z(k) = M \{ z(k) z^T(k) \}; \quad P_x(k) = M \{ x(k) x^T(k) \}; \\ P_{zx}(k, k-h) = P_{xz}^T(k-h, k) = M \{ z(k) x^T(k-h) \}; \quad P_{xz}(k, k-h) = P_{zx}^T(k-h, k) = M \{ x(k) z^T(k-h) \}; \\ P_x(k, k-h) = M \{ x(k) x^T(k-h) \}; \quad P_{zx}(k) = P_{xz}^T(k) = M \{ x(k) z^T(k) \}; \\ P_{zx}(k) = P_{xz}^T(k) = M \{ z(k) x^T(k) \}; \quad P_{zw}(k, k-h) = P_{wz}^T(k-h, k) = M \{ z(k) w^T(k-h) \}; \\ P_{wx}(k) = P_{xw}^T(k) = M \{ w(k) x^T(k) \}; \quad P_w(k) = M \{ w(k) w^T(k) \},$$

которые определяются системой разностных матричных уравнений с запаздываниями.

Доказательство. Для вычисления локального критерия получим уравнение состояния путем подстановки (5) в (1):

$$x(k+1) = (A + \sum_{i=1}^r A_i \theta_i) x(k) + (\tilde{A} + \sum_{i=1}^r \tilde{A}_i \theta_i) x(k-h) + (B + \sum_{i=1}^r B_i \theta_i) K_0(k) w(k) + \\ + (B + \sum_{i=1}^r B_i \theta_i) K_1(k) S x(k) + (B + \sum_{i=1}^r B_i \theta_i) K_1(k) v(k) + (B + \sum_{i=1}^r B_i \theta_i) K_2(k) S x(k-h) + \\ + (B + \sum_{i=1}^r B_i \theta_i) K_2(k) v(k-h) + (B + \sum_{i=1}^r B_i \theta_i) K_3(k) z(k) + q(k). \quad (13)$$

Учитывая (1), (2), (4), (5), (6), характеристики случайных последовательностей $q(k)$ и $v(k)$, вычислим значение локального критерия (3), а затем значения градиентов критерия по $K_0(k), K_1(k), K_2(k)$ и $K_3(k)$. Приравняв значения градиентов к нулю, получим уравнения, решения которых имеют вид

$$K_0(k) = -\bar{C}^{-1} [\bar{C} K_1(k) S P_{xw}(k) + \bar{C} K_2(k) S P_{xw}(k-h, k) + \bar{C} K_3(k) P_{zw}(k) + \\ + \frac{1}{3} \sum_{i=1}^r B_i^T H^T C H (A_i P_{xw}(k) + \tilde{A}_i P_{xw}(k-h, k)) + \\ + B^T H^T C (H A P_{xw}(k) + H \tilde{A} P_{xw}(k-h, k) - P_{zw}(k)) S^T] P_w^{-1}(k). \quad (14)$$

$$K_1(k) = -\bar{C}^{-1} [(\bar{C} K_0(k) P_{wx}(k) S + \bar{C} K_2(k) S P_x(k-h, k) + \bar{C} K_3(k) P_{zx}(k)) S^T + \\ + \frac{1}{3} \sum_{i=1}^r B_i^T H^T C H (A_i P_x(k) + \tilde{A}_i P_x(k-h, k)) S^T + B^T H^T C (H A P_x(k) + H \tilde{A} P_x(k-h, k) - \\ - P_{zx}(k)) S^T] [S P_x(k) S^T + V(k)]^{-1}. \quad (15)$$

$$K_2(k) = -\bar{C}^{-1} [\bar{C} (K_1(k) S P_x(k, k-h) + K_0(k) P_{wx}(k, k-h) + K_3(k) P_{zx}(k, k-h)) S^T + \\ + \frac{1}{3} \sum_{i=1}^r B_i^T H^T C H (A_i P_x(k, k-h) + \tilde{A}_i P_x(k-h)) + B^T H^T C (H A P_x(k, k-h) + \\ + H \tilde{A} P_x(k-h) - P_{zx}(k, k-h))] [S P_x(k-h) S^T + V(k-h)]^{-1}. \quad (16)$$

$$K_3(k) = -\bar{C}^{-1} [\bar{C} (K_0(k) P_{wz}(k) + K_1(k) S P_{xz}(k) + K_2(k) S P_{xz}(k-h, k)) + \\ + \frac{1}{3} \sum_{i=1}^r B_i^T H^T C H (A_i P_{xz}(k) + \tilde{A}_i P_{xz}(k-h, k)) + B^T H^T C (H A P_{xz}(k) + H \tilde{A} P_{xz}(k-h, k) - P_z(k))] P_z^{-1}(k). \quad (17)$$

Выполнив преобразование уравнений (14)–(17), получим

$$\bar{C} [K_0(k) P_w(k) + K_1(k) S P_{xw}(k) + K_2(k) S P_{xw}(k-h, k) + K_3(k) P_{zw}(k)] =$$

$$= -B^T H^T C[HAP_{xw}(k) + H\tilde{A}P_{xw}(k-h, k) - P_{zw}(k)] - \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_{xw}(k) + \tilde{A}_i P_{xw}(k-h, k)]; \quad (18)$$

$$\begin{aligned} & \bar{C}[K_0(k)P_{wx}(k)S^T + K_1(k)(SP_x(k)S^T + V(k)) + K_2(k)SP_x(k-h, k)S^T + K_3(k)P_{zx}(k)S^T] = \\ & = -B^T H^T C[HAP_x(k) + H\tilde{A}P_x(k-h, k) - P_{zx}(k)]S^T - \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k) + \tilde{A}_i P_x(k-h, k)]; \end{aligned} \quad (19)$$

$$\begin{aligned} & \bar{C}[K_0(k)P_{wx}(k, k-h)S^T + K_1(k)SP_x(k, k-h)S^T + K_2(k)(SP_x(k-h) + V(k-h)) + \\ & + K_3(k)P_{zx}(k, k-h)S^T] = -B^T H^T C[HAP_x(k, k-h) + H\tilde{A}P_x(k-h) - P_{zx}(k, k-h)]S^T - \\ & - \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k, k-h) + \tilde{A}_i P_x(k-h)]; \end{aligned} \quad (20)$$

$$\begin{aligned} & \bar{C}[K_0(k)P_{wz}(k) + K_1(k)SP_{xz}(k) + K_2(k)SP_{xz}(k-h, k) + K_3(k)P_z(k)] = \\ & = -B^T H^T C[HAP_{xz}(k) + H\tilde{A}P_{xz}(k-h, k) - P_z(k)] + \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k, k-h) + \tilde{A}_i P_x(k-h)]. \end{aligned} \quad (21)$$

Входящие в (18)–(21) моменты $P_x(k, j), P_z(k, j), P_{xz}(k, j), P_{zx}(k, j), P_{xw}(k, j), P_{wx}(k, j), P_{wz}(k, j), P_{zw}(k, j), P_w(k, j)$ определяются следующими формулами:

$$\begin{aligned} & P_x(k+1, j+1) = M\{x(k+1)x^T(j+1)\} = \\ & = \xi(i)P_x(k, j)\xi^T(j) + \xi(i)P_x(k, j-h)\xi_2^T(j) + \frac{1}{3} \sum_{i=1}^r \xi_i(k)P_x(k, j)\xi_i^T(j) + \frac{1}{3} \sum_{i=1}^r \xi_i(k)P_x(k, j-h)\xi_{2i}^T(j) + \\ & + \xi_2(k)P_x(k-h, j)\xi^T(j) + \xi_2(k)P_x(k-h, j-h)\xi_2^T(j) + \frac{1}{3} \sum_{i=1}^r \xi_{2i}(k)P_x(k-h, j)\xi_i^T(j) + \\ & + \frac{1}{3} \sum_{i=1}^r \xi_{2i}(k)P_x(k-h, j-h)\xi_{2i}^T(j) + Q_1(k, j); \quad P_x(0) = P_{x_0}. \end{aligned} \quad (22)$$

$$\begin{aligned} & P_{zx}(k+1, j+1) = M\{z(k+1)x^T(j+1)\} = \\ & = FP_{zx}(k, j)\xi^T(j) + FP_{zx}(k, j-h)\xi_2^T(j) + FP_{zw}(k, j)K_0^T(j)B^T + FP_z(k, j)K_3^T(j)B^T; \quad P_{zx}(0) = P_{z_0x_0}. \end{aligned} \quad (23)$$

$$\begin{aligned} & P_{xz}(k+1, j+1) = M\{x(k+1)z^T(j+1)\} = \xi(k)P_{xz}(k, j)F^T + \\ & + \xi_2(k)P_{xz}(k-h, j)F^T + BK_0(k)P_{wz}(k-h, j)F^T + BK_3(k)P_z(k, j)F^T; \quad P_{xz}(0) = P_{x_0z_0}. \end{aligned} \quad (24)$$

$$\begin{aligned} & P_{xw}(k+1, j+1) = M\{x(k+1)w^T(j+1)\} = \\ & = \xi(k)P_{xw}(k, j)\bar{A}^T(j) + \xi_2(k)P_{xw}(k-h, j)\bar{A}^T(j) + BK_0(k)P_{wx}(k, j)S^T\bar{B}^T(j) + \\ & + BK_0(k)P_w(k, j)\bar{A}^T(j) + \xi(k)P_x(k, j)S^T\bar{B}^T(j) + \xi_2(k)P_x(k-h, j)S^T\bar{B}^T(j) + \\ & + BK_3(k)P_{zw}(k, j)\bar{A}^T(j) + BK_3(k)P_{zx}(k, j)S^T\bar{B}^T(j) + \\ & + BK_1(k)V(k, j)\delta_{k,j}\bar{B}^T(j); \quad P_{xz}(0) = P_{x_0z_0}; \quad P_{xz}(0) = P_{x_0z_0}. \end{aligned} \quad (25)$$

$$\begin{aligned} & P_{wx}(k+1, j+1) = M\{w(k+1)x^T(j+1)\} = \\ & = \bar{A}(k)P_{wx}(k, j)\xi^T(j) + \bar{A}(k)P_{wx}(k, j-h)\xi_2^T(j-h) + \bar{A}(k)P_w(k, j)K_0^T(j)B^T + \\ & + \bar{A}(k)P_{wz}(k, j)K_3^T(j)B^T + \bar{B}(k)SP_x(k, j)\xi^T(j) + \bar{B}(k)SP_x(k, j-h)\xi_2^T(j) + \\ & + \bar{B}(k)SP_{xw}(k, j)K_0^T(j)B^T + \bar{B}(k)SP_{xz}(k, j)K_3^T(j)B^T + \bar{B}(k)V(k, j)\delta_{k,j}K_1^T(j)B^T; \quad P_{wx}(0) = P_{w_0x_0}. \end{aligned} \quad (26)$$

$$\begin{aligned} & P_w(k+1, j+1) = M\{w(k+1)w^T(j+1)\} = \bar{A}(k)P_w(k, j)\bar{A}^T(j) + \bar{A}(k)P_{wx}(k, j)S^T\bar{B}^T(j) + \\ & + \bar{B}(k)SP_{xw}(k, j)\bar{A}^T(j) + \bar{B}(k)SP_x(k, j)S^T\bar{B}^T(j) + \bar{B}(k)V(k, j)\delta_{k,j}\bar{B}^T(j); \quad P_w(0) = P_{w_0}. \end{aligned} \quad (27)$$

$$P_z(k+1, j+1) = M\{z(k+1)z^T(j+1)\} = FP_z(k, j)F^T + Q_z(i, j)\delta_{i,j}; \quad P_z(0) = P_{z_0}. \quad (28)$$

$$P_{zw}(k+1, j+1) = M\{z(k+1)w^T(j+1)\} = FP_{zx}(k, j)S^T\bar{B}^T(j) + FP_{zw}(k, j)\bar{A}^T(j); \quad P_{zw}(0) = P_{z_0w_0}. \quad (29)$$

$$P_{wz}(k+1, j+1) = M\{w(k+1)z^T(j+1)\} = \bar{A}(k)P_{wz}(k, j)F^T + \bar{B}(k)SP_{xz}(k, j)F^T; \quad P_{wz}(0) = P_{w_0z_0}. \quad (30)$$

Окончательно представим систему матричных уравнений (18)–(21) в следующем виде:

$$\begin{aligned}
& \bar{C}[K_0(k)|K_1(k)|K_2(k)|K_3(k)]\bar{P}(k) = \\
& = -[B^T H^T C[HAP_{xw}(k) + H\tilde{A}P_{xw}(k-h, k) - P_{zw}(k)] + \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_{xw}(k) + \tilde{A}_i P_{xw}(k-h, k)]] \\
& B^T H^T C[HAP_x(k) + H\tilde{A}P_x(k-h, k) - P_{zx}(k)]S^T + \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k) + \tilde{A}_i P_x(k-h, k)] \\
& B^T H^T C[HAP_x(k, k-h) + H\tilde{A}P_x(k-h) - P_{zx}(k, k-h)]S^T + \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k, k-h) + \tilde{A}_i P_x(k-h)] \\
& B^T H^T C[HAP_{xz}(k) + H\tilde{A}P_{xz}(k-h, k) - P_z(k)] - \frac{1}{3} \sum_{i=1}^r B_i^T H^T CH[A_i P_x(k, k-h) + \tilde{A}_i P_x(k-h)]. \quad (31)
\end{aligned}$$

В силу условия (7) матрицы $\bar{C}$ и $\bar{P}(k)$ невырождены для всех $k = 0, 1, 2, \dots$, следовательно, уравнение (31) разрешимо относительно блочной матрицы $[K_0(k)|K_1(k)|K_2(k)|K_3(k)]$ и имеет единственное решение (8)–(11), которое получается из (31) непосредственным вычислением.

3. Асимптотическое поведение

Теорема 2. Пусть в описании объекта (1), канала измерений (2), критерия (3) и модели отслеживаемого вектора (4) матрицы $A, A_i, \tilde{A}, \tilde{A}_i, B, B_i, Q, S, V, C, D, i = \overline{1, r}$, – постоянные; $F = E, q_z(k) = 0$. Тогда, если выполняется условие (7) теоремы 1, существует установившееся решение уравнений (22)–(26), матрицы $P_x = \lim_{k \rightarrow \infty} P_x(k) \geq 0, Q_1 = \lim_{k \rightarrow \infty} Q_1(k) \geq 0$, пара матриц $(A, \sqrt{Q_1})$ стабилизируема, тогда матрица динамики замкнутой системы $\xi = A + BK_1^* S$ асимптотически устойчива для $K_1^* = \lim_{k \rightarrow \infty} K_1^*(k)$.

В теореме 2 введены обозначения:

$$\begin{aligned}
& \xi(k) = A + BK_1^*(k)S; \quad \xi_2(k) = \tilde{A} + BK_2^*(k)S; \quad \xi_3(k) = BK_3^*(k) - E; \\
& \xi_i(k) = A_i + B_i K_1^*(k)S; \quad \xi_{2i}(k) = \tilde{A}_i + B_i K_2^*(k)S; \\
& Q_1(k, j) = \xi(i)P_{xw}(k, j)K_0^{*T}(j)B^T + \xi(i)P_{xz}(k, j)K_3^{*T}(j)B^T + \frac{1}{3} \sum_{i=1}^r \xi_i(k)P_{xw}(k, j)K_0^{*T}(j)B_i^T + \\
& + \frac{1}{3} \sum_{i=1}^r \xi_i(k)P_{xz}(k, j)K_3^{*T}(j)B_i^T + \xi_2(k)P_{xw}(k-h, j)K_0^{*T}(j)B^T + \xi_2(i)P_{xz}(k-h, j)K_3^{*T}(j)B^T + \\
& + \frac{1}{3} \sum_{i=1}^r \xi_{2i}(k)P_{xw}(k-h, j)K_0^{*T}(j)B_i^T + BK_0^*(k)P_{wx}(k, j)\xi^T(j) + BK_0^*(k)P_{wx}(k, j-h)\xi_{2i}^T(j) + \\
& + \frac{1}{3} \sum_{i=1}^r B_i K_0^*(k)P_{wx}(k, j)\xi_i^{*T}(j) + \frac{1}{3} \sum_{i=1}^r B_i K_0^*(k)P_{wx}(k, j-h)\xi_{2i}^{*T}(j) + BK_3^*(k)P_{zx}(k, j)\xi^T(j) + \\
& + BK_3^*(k)P_{zx}(k, j-h)\xi_{2i}^T(j) + \frac{1}{3} \sum_{i=1}^r B_i K_3^*(k)P_{zx}(k, j)\xi_i^{*T}(j) + \frac{1}{3} \sum_{i=1}^r B_i K_3^*(k)P_{zx}(k, j-h)\xi_{2i}^{*T}(j) + \\
& + BK_0^*(k)P_w(k, j)K_0^{*T}(j)B^T + BK_0^*(k)P_{wz}(k, j)K_3^{*T}(j)B^T + \frac{1}{3} \sum_{i=1}^r B_i K_0^*(k)P_w(k, j)K_0^{*T}(j)B_i^T + \\
& + \frac{1}{3} \sum_{i=1}^r B_i K_0^*(k)P_{wz}(k, j)K_3^{*T}(j)B_i^T + BK_1^*(k)V(k, j)\delta_{k,j}K_1^{*T}(j)B^T + \\
& + BK_2^*(k)V(k-h, j-h)\delta_{k-h,j-h}K_2^{*T}(j)B^T + BK_3^*(k)P_{zw}(k, j)K_0^{*T}(j)B^T + \\
& + BK_3^*(k)P_z(k, j)K_3^{*T}(j)B^T + \frac{1}{3} \sum_{i=1}^r B_i K_3^*(k)P_{zw}(k, j)K_0^{*T}(j)B_i^T +
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{3} \sum_{i=1}^r B_i K_3^*(k) P_z(k, j) K_3^{*T}(j) B_i^T + \frac{1}{3} \sum_{i=1}^r B_i K_1^*(k) V(k, j) \delta_{k,j} K_1^{*T}(j) B_i^T + \\
& + \frac{1}{3} \sum_{i=1}^r B_i K_2^*(k) V(k-h, j-h) \delta_{k-h, j-h} K_2^{*T} B_i^T + Q(k, j) \delta_{k,j}.
\end{aligned} \tag{32}$$

Доказательство. Если матрица $P_x \geq 0$, то из леммы 12.2 [21] при условии, что пара матриц $(A, \sqrt{Q_1})$ стабилизируема, следует, что матрица ξ асимптотически устойчива. Применяя теорему 3.6 [21], получаем, что если пара матриц $(A, \sqrt{Q_1})$ стабилизируема, то и пара матриц $(\xi, \sqrt{Q_1})$ – также стабилизируема. Этим доказывается справедливость теоремы.

Асимптотическую точность слежения определим, вычислив оценку критерия:

$$J = \lim_{k \rightarrow \infty} M\{\|x(k) - z\|^2\}, \tag{33}$$

где $\|\cdot\|$ – евклидова норма вектора, z – постоянный отслеживаемый вектор. Построим сначала оценку для критерия $J(k) = M\{\|x(k) - z\|^2\}$. Задавая далее условие, что $k \rightarrow \infty$, найдем оценку для критерия (33). При этом предположим, что условия теоремы 2 выполняются, а $\|\xi\|_s = \alpha_1$, $\|\xi_2\|_s = \alpha_2$, $\|\varphi\|_s = \Phi$, $\|\varphi_2\|_s = \Phi_2$ (здесь $\|\cdot\|_s$ – спектральная норма матрицы, $K_0^* = \lim_{k \rightarrow \infty} K_0^*(k)$, $K_1^* = \lim_{k \rightarrow \infty} K_1^*(k)$, $K_2^* = \lim_{k \rightarrow \infty} K_2^*(k)$, $K_3^* = \lim_{k \rightarrow \infty} K_3^*(k)$). Введем условие $\alpha_1^2 + \Phi^2 < 1$. Отметим, что выполнение этого условия обеспечивает асимптотическую устойчивость замкнутой системы с запаздываниями по состоянию. Учитывая (1), (2), (5) при коэффициентах передачи $K_0^*, K_1^*, K_2^*, K_3^*$, вычислим значение критерия (33) для $k+1$ такта:

$$\begin{aligned}
J(k+1) = & M\{x^T(k) \xi^T(k) \xi(k) x(k) + x^T(k) \xi^T(k) \xi_2(k) x(k-h) + x^T(k) \xi^T(k) B K_0^* w(k) + \\
& + x^T(k) \xi^T(k) \xi_3(k) z + x^T(k) \varphi^T(k) \varphi x(k) + x^T(k) \varphi^T(k) \varphi_2(k) x(k-h) + \\
& + x^T(k) \varphi^T(k) \sum_{i=1}^r B_i \theta_i K_0^* w(k) + x^T(k) \varphi^T(k) \sum_{i=1}^r B_i \theta_i K_3^* z + x^T(k-h) \xi_2^T(k) \xi(k) x(k) + \\
& + x^T(k-h) \xi_2^T(k) \xi_2(k) x(k-h) + \\
& + x^T(k-h) \xi_2^T(k) B K_0^* w(k) + x^T(k-h) \xi_2^T(k) \xi_3(k) z + x^T(k-h) \varphi_2^T(k) \varphi(k) x + \\
& + x^T(k-h) \varphi_2^T(k) \varphi_2(k) x(k-h) + x^T(k-h) \xi_2^T(k) \varphi_2^T(k) \sum_{i=1}^r B_i \theta_i K_0^* w(k) + \\
& + x^T(k-h) \varphi_2^T(k) \sum_{i=1}^r B_i \theta_i K_3^* z + w^T(k) K_0^{*T} B^T \xi(k) x(k) + w^T(k) K_0^{*T} B^T \xi_2(k) x(k-h) + \\
& + w^T(k) K_0^{*T} B^T B K_0^* w + w^T(k) K_0^{*T} B^T \xi_3(k) z(k) + z^T(k) \xi_3^T(k) \xi(k) x(k) + \\
& + z^T(k) \xi_3^T(k) \xi_2(k) x(k-h) + z^T(k) \xi_3^T(k) \xi_3(k) z(k) + z^T(k) \xi_3(k) B K_0^* w(k) + \\
& + w^T(k) \sum_{i=1}^r K_0^{*T} \theta_i^T B_i^T \varphi(k) x(k) + w^T(k) \sum_{i=1}^r K_0^{*T} \theta_i^T B_i^T \varphi_2(k) x(k-h) + \\
& + \sum_{i=1}^r w^T(k) K_0^{*T} \theta_i^T B_i^T \sum_{i=1}^r B_i \theta_i K_0^* w(k) + \sum_{i=1}^r w^T(k) K_0^{*T} \theta_i^T B_i^T \sum_{i=1}^r B_i \theta_i K_3^* z + \\
& + z^T \sum_{i=1}^r K_3^{*T} \theta_i^T B_i^T \varphi(k) x(k) + \sum_{i=1}^r z^T K_3^{*T} \theta_i^T B_i^T \varphi_2(k) x(k-h) + \\
& + \sum_{i=1}^r z^T K_3^{*T} \theta_i^T B_i^T \sum_{i=1}^r B_i \theta_i K_0^* w(k) + \sum_{i=1}^r z^T K_3^{*T} \theta_i^T B_i^T \sum_{i=1}^r B_i \theta_i K_3^* z\} + z^T \xi_3^T(k) \xi_3(k) z + tr \tilde{Q}(k),
\end{aligned} \tag{34}$$

где

$$\begin{aligned} \varphi = \sum_{i=1}^r (A_i \theta_i + B_i \theta_i K_1^* S); \quad \varphi_2 = \sum_{i=1}^r (\tilde{A}_i \theta_i + B_i \theta_i K_2^* S); \quad \tilde{Q}(k) = Q(k) \delta_{k,k} + BK_1^* V(k) K_1^{*\top} B^\top + \\ + \frac{1}{3} \sum_{i=1}^r B_i K_1^* V(k) K_1^{*\top} B_i^\top + BK_2^* V(k-h) K_1^{*\top} B^\top + \frac{1}{3} \sum_{i=1}^r B_i K_2^* V(k-h) K_1^{*\top} B^\top. \end{aligned}$$

Из (34) в силу неравенства Коши–Буняковского получим оценку

$$\begin{aligned} J(k+1) \leq (\alpha_1^2 + \Phi^2) J_1(k) + (\alpha_1 \alpha_2 + \Phi \Phi_2) J_2(k, k-h) + \alpha_1 (g + r_1) J_3(k) + \\ + \Phi(G+R) J_3(k) + (\alpha_1 \alpha_2 + \Phi \Phi_2) J_2(k-h, k) + (\alpha_2^2 + \Phi_2^2) J_1(k-h) + \alpha_2 (g + r_1) J_3(k-h) + \\ + \Phi_2(G+R) J_3(k-h) + \alpha_1 (g + r_1) J_3(k) + \alpha_2 (g + r_1) J_3(k-h) + \\ + (g + r_1)^2 + \Phi(G+R) J_3(k) + \Phi_2(G+R) J_3(k-h) + (G+R)^2 + \text{tr} \tilde{Q}, \end{aligned} \quad (35)$$

где

$$\begin{aligned} J_1(k) = M\{\|x(k)\|^2\}; \quad J_2(k, k-h) = M\{\|x(k)\| \times \|x(k-h)\|\}; \quad J_3(k) = M\{\|x(k)\|\}; \\ r_1 = \|\xi_3 z\|; \quad R = \left\| \sum_{i=1}^r B_i \theta_i K_3^* z \right\|; \quad G = \left\| \sum_{i=1}^r B_i \theta_i K_0^* w \right\|; \quad g = \|BK_0^* w\|; \quad \tilde{Q} = \lim_{k \rightarrow \infty} \tilde{Q}(k) = \lim_{k \rightarrow \infty} \tilde{Q}(k-1). \end{aligned}$$

Тогда, учитывая, что траектория замкнутой системы описывается уравнением

$$\begin{aligned} x(k) = (\xi + \varphi)x(k-1) + (\xi_2 + \varphi_2)x(k-h-1) + (B + \sum_{i=1}^r B_i \theta_i) K_0^* w(k-1) + \\ + (B + \sum_{i=1}^r B_i \theta_i) K_1^* v(k-1) + (B + \sum_{i=1}^r B_i \theta_i) K_2^* v(k-h-1) + (B + \sum_{i=1}^r B_i \theta_i) K_3^* z + q(k-1), \end{aligned}$$

вычислим рекуррентные соотношения для критериев $J_1(k)$, $J_2(k, k-h)$, $J_3(k)$, которые входят в состав (35):

$$\begin{aligned} J_1(k) \leq (\alpha_1^2 + \Phi^2)^k J_1(0) + (\alpha_1 \alpha_2 + \Phi \Phi_2) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_2(j-1, j-h-1) + \\ + 2\alpha_1 (r_2 + g) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-1) + 2\Phi(R+G) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-1) + \\ + (\alpha_1 \alpha_2 + \Phi \Phi_2) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_2(j-h-1, j-1) + \\ + (\alpha_2^2 + \Phi_2^2) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_1(j-h-1) + 2\alpha_2 (r_2 + g) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \\ + 2\Phi_2(G+R) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \frac{(\alpha_1^2 + \Phi^2)^k - 1}{(\alpha_1^2 + \Phi^2) - 1} ((g + r_2)^2 + (G+R)^2 + \text{tr} \tilde{Q}), \end{aligned} \quad (36)$$

где $r_2 = \|BK_3^* z\|$.

Рекуррентное соотношение для $J_2(k, k-h)$ примет вид

$$\begin{aligned} J_2(k, k-h) \leq (\alpha_1^2 + \Phi^2)^k J_2(0, -h) + (\alpha_1 \alpha_2 + \Phi \Phi_2) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_2(j-1, j-2h-1) + \\ + \alpha_1 (r_2 + g) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-1) + \Phi(G+R) \sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-1) + \end{aligned}$$

$$\begin{aligned}
& +(\alpha_1\alpha_2 + \Phi\Phi_2)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_1(j-h-1) + (\alpha_2^2 + \Phi_2^2)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_2(j-h-1, j-2h-1) + \\
& +\alpha_1(r_2 + g)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \\
& +\Phi_2(G + R)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \alpha_2(r_2 + g)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \\
& +\alpha_2(g + r_2)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-2h-1) + \Phi(R + G)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-h-1) + \\
& +\Phi_2(R + G)\sum_{j=1}^k (\alpha_1^2 + \Phi^2)^{k-j} J_3(j-2h-1) + \frac{(\alpha_1^2 + \Phi^2)^k - 1}{(\alpha_1^2 + \Phi^2) - 1}((g + r_2)^2 + (G + R)^2 + tr\tilde{Q}_1), \quad (37)
\end{aligned}$$

где $\tilde{Q}_1(k) = BK_1^*V(k-h-1)K_2^{*\top}B^\top + \frac{1}{3}\sum_{i=1}^r B_iK_1^*V(k-h-1)K_2^{*\top}B_i^\top$.

Рекуррентное соотношение для $J_3(k)$ имеет вид

$$J_3(k) \leq \alpha_1^k J_3(0) + \alpha_2 \sum_{j=1}^k \alpha_1^{k-j} J_3(j-h-1) + \frac{\alpha_1^k - 1}{\alpha_1 - 1} (r_2 + g). \quad (38)$$

Оценку критерия (34) построим, учитывая неравенства (36)–(38). Тогда при $k \rightarrow \infty$ из (34) получим

$$\begin{aligned}
J \leq & \frac{[(G + R)^2 + (g + r_2)^2 + tr\tilde{Q}][(\alpha_1^2 + \Phi^2) + (\alpha_2^2 + \Phi_2^2)]}{1 - (\alpha_1^2 + \Phi^2)} + \\
& + 2(\alpha_1\alpha_2 + \Phi\Phi_2) \frac{(G + R)^2 + (g + r_2)^2 + tr\tilde{Q}_1}{1 - (\alpha_1^2 + \Phi^2)} + \\
& + 2(g + r_2) \frac{(\alpha_1 + \alpha_2)(g + r_1) + (\Phi + \Phi_2)(G + R)}{1 - \alpha_1} + (g + r_1)^2 + (G + R)^2 + tr\tilde{Q}. \quad (39)
\end{aligned}$$

Из оценки (39) видно, что при практически естественных ограничениях на класс динамических систем метод локально-оптимального слежения при косвенных измерениях с ошибками обеспечивает асимптотическое слежение с точностью, определяемой интенсивностью аддитивных возмущений и ошибок в канале измерений, динамическими характеристиками замкнутой системы, значениями параметров объекта и коэффициентов передачи следящей системы управления.

4. Результаты моделирования

Пусть объект и локальный критерий описываются следующими матрицами и векторами:

$$\begin{aligned}
A &= \begin{bmatrix} 0,05 & 1 \\ -0,025 & 1 \end{bmatrix}; \tilde{A} = \begin{bmatrix} 0 & 0 \\ 0,03 & 0 \end{bmatrix}; B = \begin{bmatrix} 0,1 \\ 1 \end{bmatrix}; A_1 = \begin{bmatrix} 0,05 & 0 \\ 0 & 0 \end{bmatrix}; A_2 = \begin{bmatrix} 0 & 0 \\ 0,005 & 0 \end{bmatrix}; \\
A_3 &= \begin{bmatrix} 0 & 0,1 \\ 0 & 0 \end{bmatrix}; A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0,01 \end{bmatrix}; A_5 = A_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}; \tilde{A}_1 = \begin{bmatrix} 0,03 & 0 \\ 0 & 0 \end{bmatrix}; \tilde{A}_2 = \begin{bmatrix} 0 & 0 \\ 0,05 & 0 \end{bmatrix}; \\
\tilde{A}_3 &= \begin{bmatrix} 0 & 0,001 \\ 0 & 0 \end{bmatrix}; \tilde{A}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0,002 \end{bmatrix}; \tilde{A}_5 = \tilde{A}_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}; \\
B_1 &= B_2 = B_3 = B_4 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; B_5 = \begin{bmatrix} 0,005 \\ 0 \end{bmatrix}; B_6 = \begin{bmatrix} 0 \\ 0,025 \end{bmatrix}; \\
Q &= \begin{bmatrix} 0,02 & 0 \\ 0 & 0,02 \end{bmatrix}; S = \begin{bmatrix} 0 & 1 \end{bmatrix}; H = \begin{bmatrix} 1 & 0 \end{bmatrix}; C = 1; D = 0,2; F = 1; z = \begin{bmatrix} 10 \\ 10 \end{bmatrix}; h = 1.
\end{aligned}$$

В ходе моделирования сравнивалось качество двух систем управления. Первая система управления моделировалась с оптимальными коэффициентами передачи, вычисленными с учетом интервальных параметров; вторая – с коэффициентами, которые рассчитывались по номинальным значениям параметров.

В качестве критерия оценки качества сходимости вектора состояния $x(k)$ к желаемому значению $z(k)$ рассчитывается средняя ошибка оценивания:

$$e_i = \frac{\sum_{k=1}^N |x(k) - z(k)|}{N},$$

где $z(k)$ – желаемое значение вектора состояния.

В таблице приведены значения критерия качества сходимости для двух алгоритмов ($N = 100$) для 5 различных наборов интервальных параметров θ_i :

- алгоритм 1 – локально оптимальное управление для объекта с интервальными параметрами;
- алгоритм 2 – управление, вычисленное по номинальным значениям параметров [9].

Средние ошибки

Алгоритм	e_1	e_2	e_3	e_4	e_5	e_6
1	0,384	0,258	1,169	0,819	0,572	0,274
2	0,449	0,307	1,204	0,830	0,611	0,281

Из таблицы видно, что средняя ошибка отклонения вектора состояния $x(k)$ от отслеживаемого вектора $z(k)$ при локально-оптимальном управлении для объекта с интервальными параметрами меньше, чем при управлении, построенном по номинальным значениям параметров.

Заключение

Решена задача управления выходом для дискретного объекта с интервальными параметрами с запаздыванием по состоянию. Решение выполнено на основе синтеза локально-оптимальной следящей системы управления линейным динамическим объектом при косвенных измерениях с использованием вероятностного метода. Исследовано асимптотическое поведение системы. Показано, что оптимальная система управления с постоянными коэффициентами передачи обеспечивает более высокую точность слежения, чем система управления, синтезированная по номинальным значениям параметров.

ЛИТЕРАТУРА

1. Camacho E.F., Bordons C. Model Predictive Control. London : Springer-Verlag, 2004. 405 p.
2. Aggelogiannaki E., Doganis Ph., Sarimveis H. An Adaptive Model Predictive Control Configuration for Production-Inventory Systems // International Journal of Production Economics. 2008. V. 114. P. 165–178.
3. Wang W., Rivera D. A Novel Model Predictive Control Algorithm for Supply Chain Management in Semiconductor Manufacturing // 2005 American Control Conference, Portland, OR, 2005. P. 841–855.
4. Stoica C., Arahal M. Application of Robustified Model Predictive Control to a Production-Inventory System // 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference Shanghai, P.R. China, 2009. P. 3993–3998.
5. Henneta J.-C. A Globally Optimal Local Inventory Control Policy for Multistage Supply Chains // International Journal of Production Research. 2009. V. 47, is. 2. P. 435–453.
6. Dombrovskii V.V., Dombrovskii D.V. Predictive Control of Random-Parameter Systems with Multiplicative Noise. Application to Investment Portfolio Optimization // Automation and Remote Control. 2005. V. 66, is. 4. P. 583–595.
7. Dai L., Xia Y., Fu M., Mahmoud M. Discrete-Time Model Predictive Control. Advances in Discrete Time Systems. InTech, 2012. Chapter 4. P. 77–116.
8. Tang G., Sun H., Liu Y. Optimal Tracking Control for Discrete Time-Delay Systems with Persistent Disturbances // Asian Journal of Control. 2006. V. 8, No. 8. P. 135–140.
9. Мухина О.О., Смагин В.И. Локально-оптимальное управление по выходу для дискретных объектов с запаздыванием по состоянию // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 1 (26). С. 4–13.

10. *Patre B.M., Bandyopadhyay B.* Robust Control for Two-Time-Scale Discrete Interval Systems // *Reliable computing*. 2006. No. 12. P. 45–58.
11. *Lin T-S., Chan S-W.* Robust Adaptive Fuzzy Sliding Mode Control for a Class of Uncertain Discrete-Time Nonlinear Systems // *International Journal of Innovative Computing, Information and Control*. 2012. V. 8, No. 1(A). P. 347–359.
12. *Смагин В.И., Смагин С.В.* Адаптивное управление запасами с учетом ограничений и транспортных запаздываний // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2008. № 3(4). С. 19–26.
13. *Киселева М.Ю., Смагин В.И.* Управление производством, хранением и поставками товаров на основе прогнозирующей модели выхода системы // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2009. № 2(7). С. 24–31.
14. *Киселева М.Ю., Смагин В.И.* Управление с прогнозирующей моделью с учетом запаздывания по управлению // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2010. № 2(11). С. 5–12.
15. *Luenberger D.G.* An introduction to observers // *IEEE. Trans. Automatic Contr.* 1972. V. AC-16, No. 6. P. 596–602.
16. *Домбровский В.В.* Синтез динамических регуляторов пониженного порядка при H_∞ ограничениях // *Автоматика и телемеханика*. 1996. № 11. С. 10–17.
17. *Домбровский В.В.* Понижение порядка систем оценивания и управления. Томск : Изд-во Том. ун-та, 1994. 175 с.
18. *Луценко И.В., Садомцев Ю.В.* Синтез дискретных H2-оптимальных регуляторов пониженного порядка // *Автоматика и телемеханика*. 2009. № 10. С. 114–132.
19. *Андреев Ю.Н.* Управление конечномерными линейными объектами М. : Наука, 1976. 424 с.
20. *Абгарян К.А.* Матричные и асимптотические методы в теории линейных систем. М. : Наука, 1973. 432 с.
21. *Wonham W.M.* Linear Multivariable Control: A Geometric Approach. Springer-Verlag, 1979. 354 p.

Мухина Оксана Олеговна. E-mail: oksm7@sibmail.com

Смагин Валерий Иванович, д-р техн. наук, профессор. E-mail: vsm@mail.tsu.ru
Томский государственный университет

Поступила в редакцию 1 июня 2014 г.

Mukhina Oksana O., Smagin Valery I. (Tomsk State University, Russian Federation).

Dynamic locally-optimal control systems for objects with interval parameters and state delay.

Keywords: local-optimal control, state delay, dynamic control system, output control.

Consider the problem of dynamical locally-optimal control based on the observed output for discrete objects with interval parameters and delay in the state, described by the following difference equation:

$$x(k+1) = (A + \sum_{i=1}^r A_i \theta_i) x(k) + (\tilde{A} + \sum_{i=1}^r \tilde{A}_i \theta_i) x(k-h) + (B + \sum_{i=1}^r B_i \theta_i) u(k) + q(k);$$

$$x(\tau) = \varphi(\tau), \tau = -h, 1-h, 2-h, \dots, 0; k = 0, 1, 2, \dots,$$

where $x(k) \in R^n$ is a state vector, $h > 0$ is a positive integer time delay, $u(k) \in R^m$ is a control input, $A, A_i, \tilde{A}, \tilde{A}_i, B, B_i, i = \overline{1, r}$ are constant matrices of appropriate dimensions, $q(k)$ is the Gaussian random sequence of input disturbances, θ_i is uncertain parameters of interval type $(-1 \leq \theta_i \leq 1)$. The measurement channel is represented by equation

$$y(k) = Sx(k) + v(k),$$

where S is the matrix of measurement channel, $v(k)$ is the Gaussian random sequence of measurement errors.

To solve the problem, we propose an algorithm, which is based on the optimization of the local criteria

$$I(k) = M \left\{ (w(k+1) - z(k))^T C (w(k+1) - z(k)) + u^T(k) D u(k) \right\},$$

where $w(k) = Hx(k)$ is the controlled output of the system, $C = C^T \geq 0$ и $D = D^T \geq 0$ are weighting matrices, $z(k) \in R^n$ is the tracking vector, described by equation

$$z(k+1) = Fz(k) + q_z(k),$$

where $q_z(k)$ is the Gaussian random sequence, F is a matrix.

The control law of object is determined by the function of measured variables with time memory of the tracked signal and the dynamic element $w(k)$:

$$u(k) = K_0(k)w(k) + K_1(k)y(k) + K_2(k)y(k-h) + K_3(k)z(k).$$

The formulas for calculating the optimal transfer coefficients $K_0^*(k), K_1^*(k), K_2^*(k), K_3^*(k)$ are given.

In this paper, the proposed synthesis algorithms of output control do not use an extension method of the state space. The asymptotic properties of the closed-loop system are obtained. For the square criterion

$$J = \lim_{k \rightarrow \infty} M \{ \|x(k) - z\|^2 \},$$

which defines the asymptotic accuracy of tracking, it is shown that

$$J \leq \frac{[(G+R)^2 + (g+r_2)^2 + \text{tr} \tilde{Q}][(\alpha_1^2 + \Phi^2) + (\alpha_2^2 + \Phi_2^2)]}{1 - (\alpha_1^2 + \Phi^2)} +$$

$$+2(\alpha_1\alpha_2 + \Phi\Phi_2) \frac{(G+R)^2 + (g+r_2)^2 + tr\tilde{Q}_1}{1 - (\alpha_1^2 + \Phi^2)} +$$

$$+2(g+r_2) \frac{(\alpha_1 + \alpha_2)(g+r_1) + (\Phi + \Phi_2)(G+R)}{1 - \alpha_1} + (g+r_1)^2 + (G+R)^2 + tr\tilde{Q}.$$

So, we see that under natural restrictions on the class of dynamic systems, the method of locally optimal tracking on indirect measurements with errors provides asymptotic tracking with accuracy determined by the intensity of additive disturbances and errors in the observations, dynamic characteristics of a closed-loop system, values of the parameters of the object, and the transmission coefficients of the tracking control system.

The comparison of the simulation results of the two control systems is given with:

- the optimal transfer coefficients, calculated with using the interval parameters;
- the transfer coefficients, calculated with using the nominal values of the parameters.

The criterion of estimation of the quality of convergence of the state vector $x(k)$ to the desired value $z(k)$ shows that the average error of the deviation of the state vector $x(k)$ of the tracking vector $z(k)$ in the robust control is less than the control constructed on the nominal values of the parameters.

REFERENCES

1. Camacho E.F., Bordons C. *Model Predictive Control*. London: Springer-Verlag, 2004. 405 p.
2. Aggelogiannaki E., Doganis Ph., Sarimveis H. An Adaptive Model Predictive Control Configuration for Production-Inventory Systems. *International Journal of Production Economics*, 2008, vol. 114, pp. 165-178. DOI: 10.1016/j.ijpe.2008.01.003
3. Wang W., Rivera D. A Novel Model Predictive Control Algorithm for Supply Chain Management in Semiconductor Manufacturing. *American Control Conference*, 2005, Portland, OR, pp. 841-855. DOI: 10.1109/ACC.2005.1469933
4. Stoica C., Arahal M. Application of Robustified Model Predictive Control to a Production-Inventory System. *48th IEEE Conference on Decision and Control and 28th Chinese Control Conference Shanghai*, 2009, P.R. China, pp. 3993-3998. DOI: 10.1109/CDC.2009.5399740
5. Henneta J.-C. A Globally Optimal Local Inventory Control Policy for Multistage Supply Chains. *International Journal of Production Research*, 2009, vol. 47, Issue 2, pp. 435-453. DOI: 10.1080/00207540802426458
6. Dombrovskii V.V., Dombrovskii D.V. Predictive Control of Random-Parameter Systems with Multiplicative Noise. Application to Investment Portfolio Optimization. *Automation and Remote Control*, 2005, vol. 66, Issue 4, pp. 583-595.
7. Dai L., Xia Y., Fu M., Mahmoud M. Discrete-Time Model Predictive Control. *Advances in Discrete Time Systems. InTech*, 2012, Chapter 4, pp. 77-116. DOI: 10.5772/51122
8. Tang G., Sun H., Liu Y. Optimal Tracking Control for Discrete Time-Delay Systems with Persistent Disturbances. *Asian Journal of Control*, 2006, vol.8, no. 8, pp. 135-140. DOI: 10.1111/j.1934-6093.2006.tb00263.x
9. Mukhina O.O., Smagin V.I. Local-Optimal Output Control for Discrete Systems with State Delays. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2014, no. 1 (26), pp. 4-13. (In Russian).
10. Patre B.M., Bandyopadhyay B. Robust Control for Two-Time-Scale Discrete Interval Systems. *Reliable computing*, 2006, no. 12, pp. 45-58. DOI: 10.1007/s11155-006-2971-x
11. Lin T.-S., Chan S.-W. Robust Adaptive Fuzzy Sliding Mode Control for a Class of Uncertain Discrete-Time Nonlinear Systems. *International Journal of Innovative Computing, Information and Control*, 2012, vol. 8, no. 1(A), pp. 347-359.
12. Smagin V.I., Smagin S.V. Adaptive inventory control with restrictions and transport delays. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2008, no. 3(4), pp. 19-26. (In Russian).
13. Kiseleva M.Yu., Smagin V.I. Control of goods production, storage and delivery based on prediction model systems output. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2009, no. 2(7), pp. 24-31. (In Russian).
14. Kiseleva M.Yu., Smagin V.I. Model predictive control with time-delay in control input. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2010, no. 2 (11), pp. 5-12. (In Russian).
15. Luenberger D.G. An introduction to observers. *IEEE. Trans. Automatic Contr.*, 1972, V, AC-16, no. 6, pp. 596-602. DOI: 10.1109/TAC.1971.1099826
16. Dombrovskii V.V. Synthesis of the dynamic governor of reduced order under H_∞ restrictions. *Avtomatika i telemekhanika*, 1996, no. 11, pp. 10-17. (In Russian).
17. Dombrovskii V.V. *Ponizhenie poryadka sistem otsenivaniya i upravleniya* [Reduction of the order of evaluation and control systems]. Tomsk: Tomsk State University Publ., 1994. 175 p.
18. Lutsenko I.V., Sadomtsev Yu.V. Design of discrete H_2 -optimal reduced-order controllers. *Avtomatika i telemekhanika*, 2009, no. 10, pp. 114-132. (In Russian).
19. Andreev Yu.N. *Upravlenie konechnomernymi lineynymi ob"ektami* [Management of the finite-dimensional linear objects]. Moscow: Nauka Publ., 1976. 424 p.
20. Abgaryan K.A. *Matrichnye i asimptoticheskie metody v teorii lineynykh sistem* [Matrix and asymptotic methods in the theory of linear systems]. Moscow: Nauka Publ., 1973. 432 p.
21. Wonham W.M. *Linear Multivariable Control: A Geometric Approach*. Springer-Verlag, 1979. 354 p.

ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.237.8: 81'322.2

М.Ф. Ашуров

СРАВНЕНИЕ ПОТОКОВЫХ МЕТОДОВ КЛАССИФИКАЦИИ ТЕКСТОВ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ НА ОСНОВЕ СЖАТИЯ ИНФОРМАЦИИ И ПОДСЧЕТА ПОДСТРОК

Рассматривается задача сравнения качества методов классификации текстов на основе R -меры и PPM метода. Оценка качества классификаторов для каждой тестовой выборки проводилась с использованием F -меры. Проведена классификация по авторам на двух выбранных тестовых выборках с характерными особенностями, связанными с периодами написания художественных произведений. Для каждого классификатора выявлены факторы, которые влияют на снижение качества классификации.

Ключевые слова: классификация текстов; художественные произведения; R -мера; PPM метод; F -мера.

В современном мире возрастающий объём информации в электронном виде всё больше нуждается в классификации для лучшего хранения и дальнейшей обработки. Ручная классификация при таком объеме текстов будет слишком затратной по времени и человеческим усилиям. Данную проблему призвана решить компьютерная автоматизированная классификация, на основе которой компьютерные комплексы могут справляться с большими объемами информации.

1. Потокосые методы классификации

Многие авторы [1–4] среди большого количества разработанных на сегодняшний день методов решения этой задачи разделяют их на признаковые и потоковые. Признаговые методы (feature-based approaches) основаны на использовании численного представления текстов – векторами значений выбранных признаков. Основной недостаток применения признаковых методов заключён в сложности и трудоемкости определения необходимого и достаточного набора признаков, по которому будет проходить классификация.

В отличие от признаковых методов потоковые методы (stream-based approaches) не требуют каких-либо преобразований и изменений в тексте. Они непосредственно используют элементы текста, т.е. текст X рассматривается как последовательность (поток) из n элементов $x_1, x_2, \dots, x_n$ некоторого алфавита Q , при этом длина текстовой строки $n = |X|$. В качестве элемента текста x_n может быть выбран одиночный текстовый символ, слово, грамматический класс, любая группировка символов текста.

Среди потоковых методов О.Г. Шевелев [5] выделяет два основных направления:

- подсчет повторений строк (R -, C - и другие меры);
- сжатие информации (off-the-shelf, PPM).

Одной из отличительных особенностей первого направления является то, что результаты обработки могут быть представлены в естественном для человека виде и проанализированы позднее. Стоит отметить, что Д.В. Хмельёв в работе [6] приводит результаты сравнения для R -меры и метода PPM, однако данное сравнение происходит на текстах новостной ленты – публицистических статьях, в которых авторский стиль проявляется весьма ярко. Для более полной оценки этих направлений требуется провести их сравнение на текстах разного типа.

2. Усечённая R -мера

Методы, предложенные Хмелёвым и Тианом [2], строятся по изложенным выше принципам, но их отличительной особенностью в получении самой меры является результат подсчета определенных подстрок исследуемого текста, которые есть в супертексте (конкатенации всех текстов класса).

Усечённая R -мера близости [1] учитывает все возможные повторения всех подстрок длин от k_1 до k_2 испытуемого текста длины n в супертексте (в отличие от неусечённой, для которой $k_1 = 1$, $k_2 = n$):

$$\begin{aligned}R(X | S) &= r(X | S) / N, \\r(X | S) &= \sum_{k=k_1}^{k_2} c_k(X | S), \\N &= (2(n+1) - (k_1 + k_2))(k_2 - k_1 + 1) / 2, \\c_k(X | S) &= \sum_{i=k}^n c(x_{i-k+1} \dots x_n | S), \\c(x_{i-k+1} \dots x_n | S) &= \begin{cases} 1, & x_{i-k+1} \dots x_n \subset S, \\ 0, & x_{i-k+1} \dots x_n \not\subset S, \end{cases}\end{aligned}$$

где X – испытуемый текст; S – супертекст исследуемого класса; k – длина подстроки поиска; N – число подстрок.

Для ускорения вычисления R -меры можно использовать суффиксные массивы или суффиксные деревья.

3. PPM метод

Изначально PPM (prediction by partial matching) метод – метод контекстно-зависимого моделирования ограниченного порядка (finite-context modeling), позволяющий оценить вероятность символа в зависимости от предыдущих символов. Префикс строки, непосредственно предшествующий текущему символу, называется контекстом. Если для оценки вероятности используется контекст длины N , то данный случай является контекстно-ограниченной моделью степени N или порядка N (order- N , O- N). Чтобы получить хорошую оценку вероятности символа, необходимо учитывать контексты разных длин, т.е. PPM может быть представлена как вариант стратегии перемешивания: оценки вероятностей, сделанные на основании контекстов разных длин, объединяются в одну общую вероятность. Полученные оценки кодируются любым энтропийным кодером, например любым арифметическим кодером, за счет чего и происходит сжатие текста.

Алгоритм классификации первоначально требует получить сжатый PPM методом супертекст класса, после чего происходит конкатенация супертекста и испытуемого текста и дальнейшее их сжатие тем же способом. В простейшем наивном подходе мера близости исследуемого текста к классу вычисляется как разница длин сжатой конкатенации супертекста и испытуемого текста относительно сжатого супертекста. Выбор принадлежности текста к классу строится на выборе наименьшей разницы длин вышеописанного способа, т.е. используется идея наименьшей энтропии между текстами класса и испытуемого текста.

4. Постановка задачи

В 2004 г. в Томском государственном университете (ТГУ) научным коллективом под руководством В.В. Поддубного был основан проект программного комплекса «СтилеАнализатор» [7], используемый для выполнения, прежде всего, стилового анализа размеченных корпусов текстов. Первоначально в «СтилеАнализаторе» был реализован анализ текстов с помощью различных признаковых методов, но затем было принято решение реализовать в комплексе также потоковые методы и провести сравнительный анализ этих методов.

В данной работе анализ обоих классификаторов, основанных на усечённой R -мере и сжатия PPM, проводился нами на двух тестовых выборках. Первая выборка состоит из текстов авторов русской художественной прозы XIX в. Всего было использовано 9 авторов, представлявших авторские классы, около 100 произведений которых были выбраны в качестве «обучающей» выборки. Вторая выборка строится из текстов авторов русской художественной прозы 90-х гг. XX в. Использовалось 112 текстов 21 автора. Образованные этими произведениями супертексты каждого класса были уравнены по объёму. Тексты тестовой (контрольной) выборки каждого класса были составлены тоже равного объема, порядка 100 тыс. символов каждый, и в равном количестве, при этом 2 из них были получены из текстов обучающей выборки, а оставшаяся часть – из других произведений, не участвовавших в обучении.

Для проведения тестирования качества потоковых методов классификации был спроектирован и реализован на языке C# программный модуль, позволяющий проводить классификацию текстов на основе R -меры и сжатия PPM. Проблема выбора параметров k_1 и k_2 для усеченной R -меры решалась с учётом особенностей текстов на естественных языках. Минимальная длина подстроки k_1 равна 10 символам, так как более короткие подстроки начинают совпадать с большинством слов русской речи, обычно используемых всеми авторами, что не позволяет выделить стиливые особенности разных авторов. Максимальная длина подстроки k_2 равна 45 символам – такая длина может включать в себе даже словосочетание из 3 или 4 слов. Использование большей длины подстроки видится сомнительным, так как повтор такой длинной фразы у разных авторов представляется крайне маловероятным (проблема плагиата полностью исключалась из рассмотрения). Модуль классификатора на основе PPM метода использует алгоритм PPMD порядка 5, реализованный в RAR-архиваторе.

5. Оценка качества классификатора

Качество классификации по каждому классу оценивалось по текстам контрольной выборки F -мерой – средним гармоническим между полнотой (Recall) r (долей текстов, правильно приписываемых классу из всех текстов этого класса) и точностью (Precision) p (долей текстов, правильно приписываемых классу из всех текстов, приписываемых этому классу) [1]:

$$F = 2 \frac{p \times r}{p + r}.$$

Среднее значение F -меры по всем классам принималось за оценку качества потоковой классификации каждого классификатора в целом.

6. Результаты сравнения

Относительно первой тестовой выборки, состоящей из 9 классов, для усеченной R -меры были получены следующие характеристики, приведенные в табл. 1.

Т а б л и ц а 1

Точность, полнота и F -мера на текстах авторов XIX в. для R -меры

Автор	Precision	Recall	F-measure
Chehov Anton	0,60	0,75	0,67
Dostoevskii Fedor	1,00	0,42	0,59
Gogol Nikolay	0,92	1,00	0,96
Goncharov Ivan	0,67	1,00	0,80
Kuprin Alexandr	1,00	0,50	0,67
Leskov Nikolai	0,71	0,83	0,77
Saltikov-Shedrin Mihail	0,40	0,17	0,24
Tolstoi Lev	1,00	0,92	0,96
Turgenev Ivan	0,57	1,00	0,73

Результаты для классификатора, основанного на сжатии PPM, приведены в табл. 2.

Таблица 2

Точность, полнота и *F*-мера на текстах авторов XIX в. для PPM метода

Автор	Precision	Recall	F-measure
Chehov Anton	0,56	0,83	0,67
Dostoevskii Fedor	0,44	0,33	0,38
Gogol Nikolay	0,90	0,75	0,82
Goncharov Ivan	0,71	1,00	0,83
Kuprin Alexandr	0,40	0,33	0,36
Leskov Nikolai	0,71	0,83	0,77
Saltikov-Shedrin Mihail	0,67	0,17	0,27
Tolstoi Lev	1,00	0,92	0,96
Turgenev Ivan	0,69	0,92	0,79

Для лучшего сравнения общую оценку для каждого классификатора приведем в табл. 3.

Таблица 3

Значения микро- и макропоказателей *F*-меры классификаторов на текстах авторов XIX в.

Метод	F-мера (micro)	F-мера(macro)
R-мера	0,75	0,71
PPM	0,68	0,65

В целом классификатор на основе R-меры показывает лучшие результаты для этой выборки, чем классификатор с использованием сжатия PPM. Лишь на некоторых классах PPM-классификатор показывает немного лучшие результаты, что можно связать с присутствием в алгоритме возможности учесть количественное вхождение одинаковых подстрок, за счет чего происходит более плотное сжатие текста.

Относительно второй выборки, основанной на текстах 21 класса, для усеченной R-меры были получены следующие характеристики, приведенные в табл. 4.

Таблица 4

Точность, полнота и *F*-мера на текстах авторов XX в. для R-меры

Фамилия автора	Precision	Recall	F-measure
Agafonov	0,88	1,00	0,93
Aristov	1,00	1,00	1,00
Azarov	0,96	1,00	0,98
Baganov	0,79	1,00	0,88
Belkin	1,00	1,00	1,00
BelobrovPopov	1,00	1,00	1,00
Belomlinskaya	1,00	1,00	1,00
Belov	1,00	0,78	0,88
Bonch	1,00	0,27	0,43
Bronin	1,00	1,00	1,00
Burmistrov	0,50	1,00	0,67
Galkin	1,00	0,29	0,45
Gergenreder	1,00	1,00	1,00
Glushkin	1,00	1,00	1,00
Svetlana	1,00	1,00	1,00
Velboi	0,10	0,20	0,13
Vershovskii	1,00	0,19	0,32
Veter	0,57	1,00	0,73
Vitkovskii	1,00	0,33	0,50
Voronov	1,00	1,00	1,00
Vulf	1,00	1,00	1,00

Для классификатора, основанного на сжатии PPM, результаты также приведены в табл. 5.

Точность, полнота и F -мера на текстах авторов XX в. для PPM метода

Фамилия автора	Precision	Recall	F-measure
Agafonov	0,88	1,00	0,93
Aristov	0,96	1,00	0,98
Azarov	0,89	0,89	0,89
Baganov	0,96	0,88	0,92
Belkin	0,91	1,00	0,95
BelobrovPopov	1,00	1,00	1,00
Belomlinskaya	0,71	1,00	0,83
Belov	1,00	0,89	0,94
Bonch	0,91	0,91	0,91
Bronin	1,00	0,96	0,98
Burmistrov	0,48	0,94	0,64
Galkin	1,00	0,82	0,90
Gergenreder	0,95	1,00	0,98
Glushkin	1,00	1,00	1,00
Svetlana	1,00	1,00	1,00
Velboi	0,43	1,00	0,61
Vershovskii	1,00	0,19	0,32
Veter	0,95	1,00	0,98
Vitkovskii	1,00	0,33	0,50
Voronov	1,00	1,00	1,00
Vulf	1,00	1,00	1,00

Общая оценка для каждого классификатора на данной выборке представлена в табл. 6.

Таблица 6

Значения микро- и макропоказателей F -меры классификаторов на текстах авторов XIX в.

Метод	F -мера (micro)	F -мера(macro)
R-мера	0,85	0,80
PPM	0,90	0,87

На данной выборке классификация с помощью сжатия PPM показала себя несколько лучше – в выборке присутствует всего два класса, с которыми классификатор имеет проблемы, при этом только один класс вызывает у обоих классификаторов сбои. Успех классификации с помощью PPM метода на данной выборке достигнут во многом за счёт использования частотной характеристики символов в тексте для каждого конкретного класса.

Заключение

Оба классификатора на данных большинства рассмотренных примеров показали себя неплохо, но для каждого использованного метода стоит отметить факторы, которые снижают качество классификации. Метод на основе сжатия PPM крайне зависим от выбора модели и алгоритма PPM, использование новых его модификаций может увеличить процент успешного распознавания класса. Кроме того, данные тесты показали, что упор, сделанный на одну лишь частотную характеристику последовательностей символов в тексте в PPM методе, нередко увеличивает ошибку классификации при определении авторского стиля. В противоположность этому использование R -меры, как показало тестирование классификаторов на второй выборке, не столь эффективно, так как в ней используется информация лишь о присутствии подстроки в тексте и никак не учитывается частота данной подстроки в тексте. Использование частотной характеристики подстрок текста в развитии R -меры могло бы несколько улучшить качество классификации этим методом.

ЛИТЕРАТУРА

1. Christopher M.B. Pattern Recognition and Machine Learning // Springer Science. 2006.

2. *Khmelev D.V., Teahan W.J.* Verification of text collections for text categorization and natural language processing // Technical Report AIAA 03.1. School of Informatics, University of Wales. Bangor, 2003.
3. *Humnisett D., Teahan W.J.* Context-based methods for text categorization // Proceedings of the 27th Annual International ACM SIGIR Conference (SIGIR). The University of Sheffield, UK, 2004.
4. *Shevelyov O.G., Poddubny V.V.* Complex investigation of texts with the system "StyleAnalyzer" // Text and Language / ed. by P. Grzyber, E. Kelih, J. Macutek. Wien : Praesens Verlag, 2010. P. 207–212.
5. *Шевелев О.Г.* Методы автоматической классификации текстов на естественном языке : учеб. пособие. Томск : ТМЛ-Пресс, 2007. 144 с.
6. *Хмельев Д.В.* Классификация и разметка текстов с использованием методов сжатия данных. Краткое введение. 2003. URL: <http://compression.graphicon.ru/download/articles/classif/intro.html>
7. *Поддубный В.В., Шевелев О.Г., Кравцова А.С., Фатыхов А.А.* Словарно-аналитический блок системы «Стилеанализатор» // Научное творчество молодежи : материалы XIV Всерос. науч.-практ. конф. (15–16 апреля 2010 г.). Томск : Изд-во Том. ун-та, 2010. Ч. 1. С. 138–140.

Ашуров Михаил Фаритович. E-mail: therevenge.amf@gmail.com

Томский государственный университет

Поступила в редакцию 4 сентября 2014 г.

Ashurov Mihail F. (Tomsk State University, Russian Federation).

Comparison of stream-based fiction text classification methods based on data compression and counting substrings.

Keywords: text classification, fiction classification, *R*-measure, PPM method, *F*-measure.

We consider the problem of comparing the quality of the natural language text classification based on the stream methods using *R*-measure and PPM compression. The task of text classification requires a certain amount of texts for each class (genre, author, etc.) in the classifier training. The process of classifier training is reduced to a concatenation of all texts in supertext by each class. In stream-based text classification *X* is a sequence (stream) of *n* elements (characters, words, phrases, etc.) [a]. A matching of the test text to any class, in cases of *R*-measure, is performed through maximum measure of closeness to supertext of a class. In the PPM method, the matching is determined by the smallest difference in lengths of a compressed concatenation of supertext and the test text by the length of the compressed supertext.

The truncated *R*-measure [a] counts all the test text (length *n*) substrings with the lengths from *k*₁ to *k*₂ in the supertext (unlike the basic method with *k*₁ = 1 and *k*₂ = *n*):

$$R(X|S) = r(X|S)/N, \quad r(X|S) = \sum_{k=k_1}^{k_2} c_k(X|S), \quad N = (2(n+1) - (k_1 + k_2))(k_2 - k_1 + 1)/2,$$

$$c_k(X|S) = \sum_{i=k}^n c(x_{i-k+1} \dots x_n | S), \quad c(x_{i-k+1} \dots x_n | S) = \begin{cases} 1, & x_{i-k+1} \dots x_n \subset S, \\ 0, & x_{i-k+1} \dots x_n \not\subset S. \end{cases}$$

where *X* is the test text, *S* is the supertext of the examined class, *k* is the length of the search substring, *N* is the number of substrings.

The PPM method is a context-dependent modeling with limited order, which allows to estimate the probability of a symbol appearance depending on the previous symbols [b]. The probability estimate using a context of length *N* is a context-limited model of order *N* (order-*N*, O-*N*). A good probability estimation of symbol appearance requires to consider contexts with different lengths. The calculated estimates are encoded by any entropy encoder, allowing to compress a text. Initially, a classification algorithm obtains the class supertext compressed by PPM method, then the concatenation of supertext, and finally the test text should be compressed in the same manner.

Both classifiers were tested by two text specific samples: the Russian prose texts of the 19th century and the 90s of the 20th century. The first sample contains only 9 authors representing the classes, the second sample - 21 authors. Each sample contains about 100 texts that make a training sample. The supertexts formed by each class texts were normalized by volume. Test samples were obtained from texts that were out of training sample, except two texts for each class, and they were normalized by the number and volume (about 100 thousand characters for a test text).

To test the stream classification quality software modules that allow classify texts by *R*-measure and PPM method have been designed and implemented in C#. The problem of the *k*₁ and *k*₂ parameter selection in *R*-measure has been solved by natural language features. The minimum length of substrings (*k*₁) is equal to 10 characters because a shorter substrings can match with a lot of words in Russian language, which are common for all authors, and that can reduce stylistic features detection of different authors. The maximum length of the substrings (*k*₂) is 45 characters that allows to process even a phrase with 3 or 4 words. Using a greater length of substring seems not so useful because such long phrase cannot be repeated many times by various authors. Note that the problem of plagiarism is completely excluded from our consideration. Classifier Module based on the PPM method uses the PPMD algorithm of the order 5.

Classification quality for each class is estimated by *F*-measure, namely, the harmonic mean of precision and recall [a]. The mean of *F*-measure values by all classes is taken as the estimate of the stream-based classification quality in general.

The classifier quality characteristics for each class are obtained for the test samples. The means of *F*-measure for each classifier are calculated and the conclusions about the classification quality are made. The factors that reduce the classification quality were described for both methods.

[a] Shevelyov O.G. Automatic natural language text classification methods: Tutorial. Tomsk: TML-Press, 2007. 144 p.

[b] Khmelev D. V. Text classification and markup using compression methods. Introduction, 2003. URL: <http://compression.graphicon.ru/download/articles/classif/intro.html>

REFERENCES

1. Bishop C.B. *Pattern Recognition and Machine Learning*. Springer Science, 2006.
2. Khmelev D.V., Teahan W.J. Verification of text collections for text categorization and natural language processing. *Technical Report AILA 03.1*. School of Informatics, University of Wales. Bangor, 2003.
3. Humnisett D., Teahan W.J. Context-based methods for text categorization. *Proceedings of the 27th Annual International ACM SIGIR Conference (SIGIR)*, The University of Sheffield, UK, 2004. DOI: 10.1145/1008992.1009129
4. Shevelyov O.G., Poddubny V.V. *Complex investigation of texts with the system "StyleAnalyzer"*. In: Grzyber P., Kelih E., Macutek J. (eds.) *Text and Language*. Wien: Praesens Verlag, 2010, pp. 207-212.
5. Shevelev O.G. *Metody avtomaticheskoy klassifikatsii tekstov na estestvennom yazyke* [Automatic natural language text classification methods]. Tomsk: TML-Press Publ., 2007. 144 p. (in Russian).
6. Khmelev D.V. *Klassifikatsiya i razmetka tekstov s ispol'zovaniem metodov szhatiya dannykh* [Text classification and text markup using compression methods]. Available at: <http://compression.graphicon.ru/download/articles/classif/intro.html>.
7. Poddubny V.V., Shevelyov O.G., Kravtsova A.S., Fatihov A.A. [Dictionary-analysis unit of the system "StyleAnalyzer". *Nauchnoe tvorчество molodezhi : materialy XIV Vseros. nauch.-prakt. konf* [Scientific creativity of the youth. Proc. of the 14th All-Russian Scientific. Pract. conf.]. Tomsk: Tomsk State University Publ., 2010, pt. 1, pp. 138-140. (In Russian).

А.С. Гуменюк, Н.Н. Поздниченко, С.Н. Шпынов

ФОРМАЛЬНЫЙ АНАЛИЗ СТРОЯ ЛОКАЛЬНОЙ СТРУКТУРЫ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Рассматриваются средства для анализа локальной структуры нуклеотидных последовательностей, которые в предыдущих публикациях использовались для оценки порядка расположения компонентов нуклеотидной цепи в целом. Определены функции некоторых характеристик строя и представлены соответствующие формулы для их вычисления. Рассмотрены возможности использования этих функций для описания и исследования локальной структуры нуклеотидных цепей.

Ключевые слова: строй цепи; нуклеотидная последовательность; характеристики строя; функции характеристик строя; L-граммы; локальная структура нуклеотидной цепи.

В опубликованных ранее работах [1, 2] дано определение строя цепи [3] и представлены интегральные характеристики строя знаковых, в том числе нуклеотидных последовательностей, которые показали высокую чувствительность к расположению компонентов. На основе введенных формализмов были рассмотрены возможности сравнения, классификации, хеширования последовательностей с помощью характеристик строя.

Разрабатываемый авторами подход позволяет отображать локальную структуру знаковых цепей произвольной природы числовыми последовательностями, которые представляют расположение компонентов. Это, в свою очередь, открывает возможности для применения разнообразных математических методов анализа числовых массивов данных, которые без такого преобразования неприменимы непосредственно к символьным последовательностям.

1. Интегральные характеристики строя цепи

Прежние публикации [4, 5] представляли средства и числовые характеристики для анализа строя целостных и полноразмерных нуклеотидных цепей, а также составных частей таких цепей. Приведем некоторые из этих характеристик:

$$\Delta_{ij} = x_{i+1j} - x_{ij}; \quad x_{i+1j}, x_{ij} \in [1, n], \quad (1)$$

$$G = \sum_{j=1}^m \sum_{i=1}^{n_j} \log_2 \Delta_{ij}, \quad (2)$$

$$g = 1/n G = \log_2 \Delta_g = \sum_{j=1}^m n_j/n \log_2 \Delta_{gj}, \quad (3)$$

$$r = \Delta_g / D, \quad (4)$$

$$\Delta_g = \sqrt[n]{\prod_{j=1}^m \prod_{i=1}^{n_j} \Delta_{ij}}, \quad (5)$$

где x – номер места на позиции рассматриваемой полноразмерной цепи; x_{ij} – номер места i -го вхождения j -го элемента алфавита на позиции данной цепи; Δ_{ij} – интервал от i -го до $(i + 1)$ -го вхождения j -го символа; n_j – число вхождений j -го символа; n – длина цепи (число мест на её позиции); m – мощность алфавита рассматриваемой цепи (для нуклеотидной последовательности $m = 4$); G – глубина расположения компонентов в цепи; g – средняя удалённость компонентов в цепи; r – регулярность расположения компонентов в цепи ($0 < r < 1$); Δ_g – среднегеометрический интервал между любыми соседними одинаковыми символами; D – число описательных информаций, вычисляемое по формуле М. Мазура [6].

2. Функции характеристик строя цепи

Общепринятым методом исследования больших массивов данных измерений, лингвистических текстов, нуклеотидных последовательностей и длинных цепей другой природы является «просмотр окном» [7]. В настоящей работе представлены средства для анализа локальной структуры целостных полноразмерных последовательностей на основе характеристик строя отдельных, но связанных фрагментов (*L*-грамм) и возможности их использования.

Для формального определения функции характеристики строя введем ряд понятий, часть из которых дана в рамках «алгебры ментальных событий» [8].

Место – элементарная ячейка, предназначенная для хранения одного компонента цепи. **Позиция** – это упорядоченное множество мест. **Фрагмент** – участок полной цепи. **Окно** – позиция фрагмента (участок позиции полной цепи). **Размер окна** – количество мест на позиции окна. **Шаг** – это смещение окна на позиции полной цепи, позволяющее выделить следующий фрагмент цепи. **Размер шага** – размер смещения окна, измеряемый числом мест. **Функция характеристики строя цепи** – это упорядоченное множество значений характеристик строя, вычисленных для всех фрагментов, последовательно взятых на позиции полной цепи. **Отсчётное значение функции характеристики строя** – это значение функции характеристики строя, вычисленное для отдельного фрагмента, задаваемого его номером, длиной и размером шага.

Ниже представлены формулы для вычисления отсчётных значений некоторых функций характеристик строя.

$$\Delta_{ij} = x_{i+1j} - x_{ij}; \quad x_{i+1j}, x_{ij} \in [s * k, s * k + l], \quad (6)$$

$$f_G(k, l, s) = \sum_{j=1}^m \sum_{i=1}^{n_j} \log_2 \Delta_{ij}, \quad (7)$$

$$f_g(k, l, s) = f_G(k, l, s) / l, \quad (8)$$

$$f_{\Delta_g}(k, l, s) = \sqrt[l]{\prod_{j=1}^m \prod_{i=1}^{n_j} \Delta_{ij}}, \quad (9)$$

$$f_r(k, l, s) = f_{\Delta_g}(k, l, s) / f_D(k, l, s), \quad (10)$$

где x_{ij} – номер места i -го вхождения j -го элемента алфавита на позиции данного фрагмента; k – номер фрагмента; s – размер шага (при $s = 1$ фрагменты являются *L*-граммами); l – размер окна; $f_G(k, l, s)$ – функция глубины; $f_g(k, l, s)$ – функция средней удалённости; $f_{\Delta_g}(k, l, s)$ – функция среднего геометрического интервала; $f_r(k, l, s)$ – функция регулярности; $f_D(k, l, s)$ – функция числа описательных информаций.

Общее количество фрагментов при заданных длине цепи, шаге и размере фрагмента определяется в виде

$$k_{max} = \lceil n/s \rceil - l + s. \quad (11)$$

Заметим, что мощность алфавита данного фрагмента m может быть меньше мощности алфавита всей цепи (минимум 1, если фрагмент полностью заполнен одинаковыми компонентами). Аргументы функций (k, l, s) – натуральные числа. Таким образом, данные функции являются функциями дискретных аргументов. Зная все три параметра, можно посчитать отдельное значение такой функции. Также возможно вычислить многомерную функцию, изменяя не только номер фрагмента, но и два других параметра. Так как все функции характеристик строя не имеют обратных, задача восстановления значений аргументов по отдельному значению функции является поисковой и требует больших вычислительных ресурсов.

3. Свойства функций характеристик строя

На рис. 1 представлены графики функции удалённости 18S рибосомальной РНК организма *Cricetulus griseus* (Хомячок китайский) (GenBank id DQ235090.1?from=11629&to=13499) [9], вычисленные с размером окна 100 (рис. 1, а) и 25 (рис. 1, б) и шагом 2. Сравнение графиков показывает, что при

уменьшении размера окна график функции характеристики оказывается всё менее «сглажен» и в нем проявляются всё большие колебания.

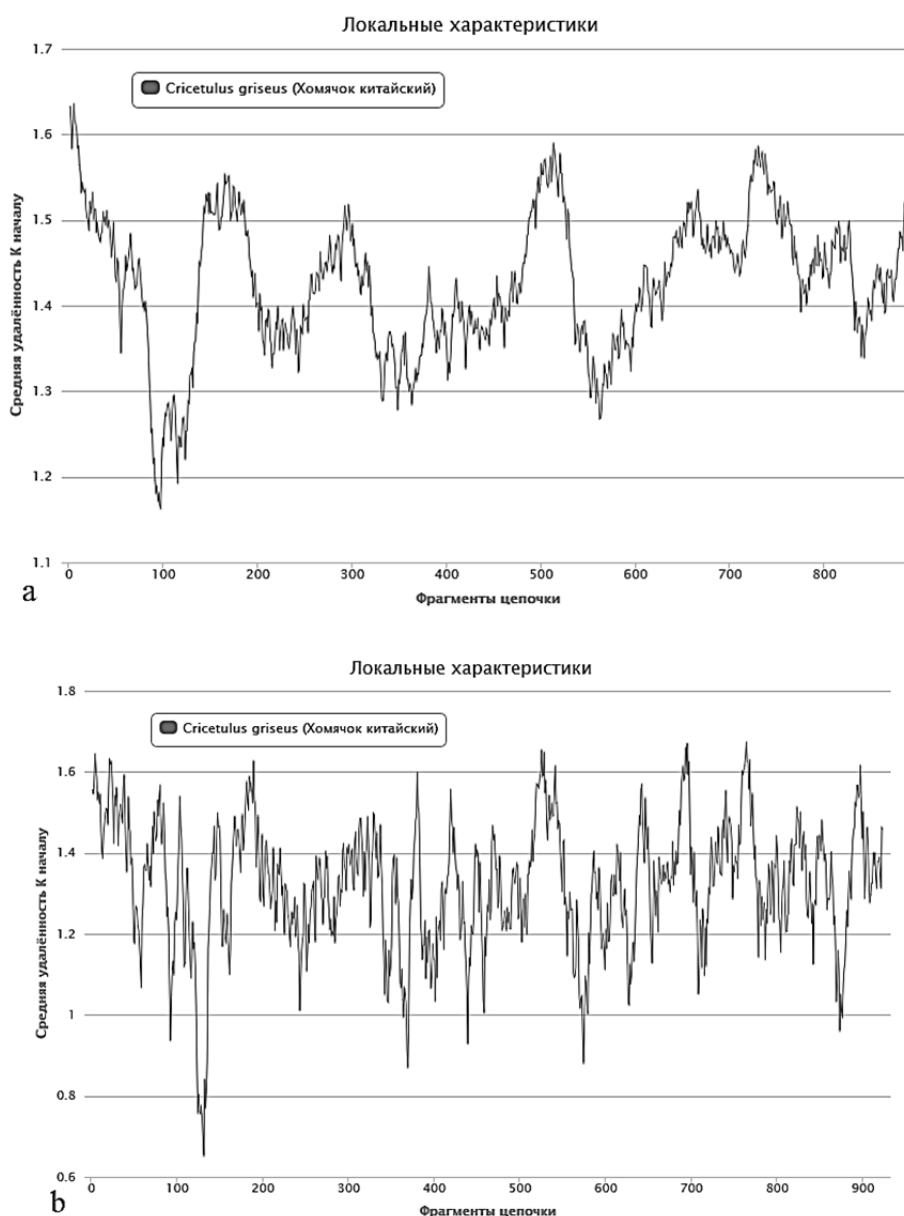


Рис. 1. Графики функции удалённости: *a* – при длине фрагментов 100 нуклеотидов; *b* – при длине фрагментов 25 нуклеотидов

Большой размер окна позволяет обнаруживать сходие по строю фрагменты большей длины. В пределе, при увеличении размера фрагментов, значения функции стремятся к значению соответствующей интегральной характеристики строя полноразмерной цепи. При уменьшении длины фрагментов отдельные значения функции позволяют выявлять всё более тонкие особенности расположения компонентов в пределах окна фрагмента. Предварительные исследования показали, что зависимость между размером окна и разбросом характеристики имеет гиперболический характер. Однако если размер окна уменьшается до мощности алфавита ($m = 4$), данная зависимость нарушается. Таким образом, как и предполагалось, неопределённость расположения фрагмента связана с неопределённостью значений функции, получаемых при заданном размере окна.

На рис. 2 представлены графики функции удалённости (рис. 2, *a*) и функции регулярности (рис. 2, *b*) той же нуклеотидной последовательности, что и на рис. 1, вычисленные с размером окна 50 и шагом 2. Из рисунка видно, что данные функции не являются функционально зависимыми и могут дополнять друг друга при комплексном описании локальной структуры нуклеотидных последовательностей.

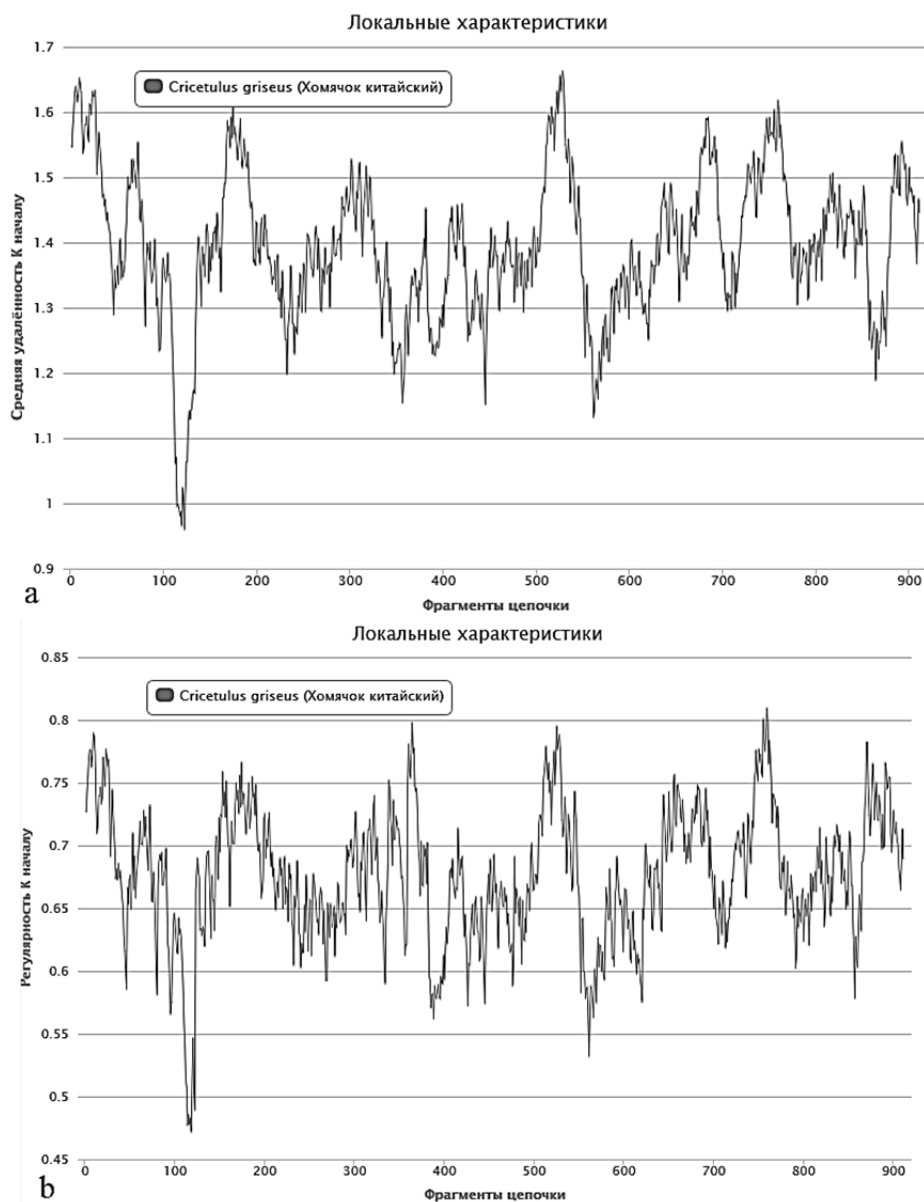


Рис. 2. Графики функций характеристик строя: *a* – функция удалённости; *b* – функция регулярности

Следует учитывать, что при увеличении длины шага уменьшается количество вычисляемых отсчётных значений, и для выявления совпадающих цепочек и фрагментов при таком разбиении требуется более сложная поисковая процедура, которая может потребовать дополнительных вычислительных ресурсов. Кроме того, при экспертном анализе сокращение количества отсчётных значений упрощает восприятие графического представления функций, вычисленных для длинных нуклеотидных последовательностей.

Дополнительных исследований требует выбор оптимального размера окна для решения различных задач, в том числе в зависимости от мощности алфавита исходной последовательности.

4. Применение функций характеристик строя для исследования нуклеотидных цепей

Ниже на представленных графиках видно, что особи одного вида (рис. 3) демонстрируют практически идентичную форму функции глубины f_G . Виды одного класса (рис. 4) имеют схожую форму этой функции, виды разных классов (рис. 5, 6) – сильно отличающуюся форму данной функции.



Рис. 3. Графики функции глубины 18S рибосомальных РНК двух особей одного вида (DQ235090.1, NR_045132.1)



Рис. 4. Графики функции глубины 18S рибосомальных РНК двух видов, принадлежащих к разным отрядам одного класса (DQ235090.1, AJ311675.1)



Рис. 5. Графики функции глубины 18S рибосомальных РНК двух видов, принадлежащих к разным классам одного типа (AJ311672.1, DQ235090.1)



Рис. 6. Графики функции глубины 18S рибосомальных РНК двух видов, принадлежащих к разным классам одного типа (AJ311672.1, EU637036.1)

Отметим, что значения интегральной характеристики этих же организмов G (таблица) соответствуют представленным графикам [4].

Значения интегральной характеристики исследуемых нуклеотидных последовательностей

№ пп	Название организма	G	№ пп	Название организма	G
1	<i>Cricetulus griseus</i>	2745,91	4	<i>Erinaceus europaeus</i>	2 671,10
2	<i>Cricetulus griseus</i> 2	2752,30	5	<i>Kareius bicoloratus</i>	2 722,66
3	<i>Crocodylus niloticus</i>	2625,63			

Перечислим другие применения функций характеристик строя: выделение повторяющихся одинаковых или схожих фрагментов; выделение разных фрагментов с одинаковым или схожим строем; поиск границ генов, интронов и экзонов, «слов»; установление функционального назначения некодирующих последовательностей; сравнение последовательностей; более надёжное хеширование, чем на основе интегральных характеристик.

Заключение

Сформулировано понятие функции характеристики строя цепи. Заданы формулы для вычисления значений некоторых функций характеристик строя и продемонстрированы возможности использования таких функций для анализа локальной структуры нуклеотидных последовательностей. Разработаны программные средства для вычисления и отображения функций характеристик строя [10]. Программные средства апробированы при сравнении рибосомальных РНК нескольких организмов. По результатам исследований выявлено влияние длины фрагментов (L -грамм) на форму функций характеристик строя.

Отмечены возможности использования функций характеристик строя для поиска схожих или совпадающих фрагментов в рамках одной или нескольких генетических цепочек, а также решения обратной задачи — поиск мест вхождения заданного фрагмента на позиции полноразмерной цепи.

Отображение строя нуклеотидных последовательностей функциями, кроме отмеченных средств, позволяет применять также классические методы математики, а именно математический анализ, спектральный анализ, корреляционный анализ и т.п., что было бы невозможно при непосредственном анализе самих знаковых последовательностей. Отмечено, что графическое представление функций характеристик строя делает возможным экспертный анализ длинных нуклеотидных цепей.

ЛИТЕРАТУРА

1. Гуменюк А.С., Морозенко Е.В., Родионов И.Н. Формализация анализа строя знаковых цепей // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 2(15). С. 15–23.
2. Gumenyuk A., Kostyshin A., Simonova S. An approach to the research of the structure of linguistic and musical texts // Glottometrics. 2002. No. 3. С. 61–69.
3. Гуменюк А.С. О средствах анализа взаимного расположения компонентов знаковой последовательности // Военная техника, вооружение и технологии двойного применения : материалы III Междунар. технолог. конгр. Омск : ОмГТУ, 2005. Ч. 2. С. 48–52.
4. Гуменюк А.С., Поздниченко Н.Н., Шпынов С.Н., Родионов И.Н. О средствах формального анализа строя нуклеотидных цепей // Математическая биология и биоинформатика. 2013. Т. 8, № 1. С. 373–397. URL: http://www.matbio.org/article.php?journ_id=15&id=158
5. Гуменюк А.С., Поздниченко Н.Н. Анализ строя нуклеотидных последовательностей // Материалы Всероссийской конференции с международным участием «Знания – онтологии – теории» (ЗОНТ-2013) 8–10 октября 2013 года. Новосибирск, 2013. Т. 2. С. 58–68.
6. Мазур М. Качественная теория информации. М. : Мир, 1974. 240 с.
7. Садовский М.Г. Информационно-статистический анализ нуклеотидных последовательностей : дис. ... д-ра физ.-мат. наук. Красноярск, 2004. 394 с.
8. Гуменюк А.С. О формализмах конструирования абстрактных объектов во внутреннем физическом пространстве информационной системы (Элементы алгебры ментальных событий) // Системный анализ в проектировании и управлении: труды X Междунар. науч.-практ. конф. СПб. : Изд-во Политех. ун.-та, 2006. Ч. 2. С. 172–181.
9. National Center for Biotechnology Information. URL: <http://www.ncbi.nlm.nih.gov/nuccore/>
10. Цымбал В.С., Поздниченко Н.Н. О разработке модуля для вычисления локальных характеристик строя нуклеотидных последовательностей // Материалы XII Всероссийской научно-практической конференции с международным участием «Информационные технологии и математическое моделирование (ИТММ-2013)». Томск : Изд-во Том. ун.-та, 2013. С. 61–65.

Гуменюк А.С., канд. техн. наук, доцент. E-mail: gumas45@mail.ru

Омский государственный технический университет

Поздниченко Н.Н. E-mail: nick670@yandex.ru

Омский государственный технический университет

Шпынов С.Н., д-р мед. наук. E-mail: stan63@inbox.ru

НИИЭМ им. Н.Ф. Гамалеи (г. Москва)

Поступила в редакцию 8 июня 2014 г.

Gumenyuk Alexander S., Pozdnichenko Nikolay N. (Omsk State Technical University, Russian Federation), *Shpynov Stanislav N.* (Gamaleya Institute of Epidemiology and Microbiology, Russian Federation).

Formal analysis of order in the local structure of the nucleotide sequences.

Keywords: chain's order, sequence, nucleotide sequence, order characteristics, functions of order characteristics, L-grams, local structure of nucleotide chain.

The definition of the chain order and integral characteristics of the order, in particular, for nucleotide sequences were presented in the previous papers. These characteristics showed the high sensitivity to the arrangement of components. The possibility of comparison, classification and hashing based on the introduced formalisms and using characteristics to order have been considered. The approach developed by the authors allows displaying the local structure of sign sequences of arbitrary nature by numerical sequences that represent the arrangement of their components.

The generally accepted method of studying large arrays of measurement data, linguistic texts, nucleotide sequences, and long sequences of another nature is the «window scan». This paper describes means for analysis of the local structure of complete full-length sequences based on the characteristics of the order of separate fragments (L-grams), named the functions of characteristics of order. The formulas for calculation of the values of these functions are of the following form:

$$\Delta_{ij} = x_{i+1j} - x_{ij}; \quad x_{i+1j}, x_{ij} \in [s * k, s * k + l],$$

$$f_G(k, l, s) = \sum_{j=1}^m \sum_{i=1}^{n_j} \log_2 \Delta_{ij},$$

$$f_g(k, l, s) = f_G(k, l, s) / l,$$

$$f_{\Delta_g}(k, l, s) = \sqrt[l]{\prod_{j=1}^m \prod_{i=1}^{n_j} \Delta_{ij}},$$

$$f_r(k, l, s) = f_{\Delta_g}(k, l, s) / f_D(k, l, s),$$

where x_{ij} is a number of position of i -th occurrence of j -th element of alphabet on position of current fragment; k is a number of fragment; s is a step size (when $s=1$ fragments become L -grams); l is a window length; $f_G(k, l, s)$ is a depth function; $f_g(k, l, s)$ is an average

remoteness function; $f_{\Delta_g}(k, l, s)$ is an average geometric interval function; $f_r(k, l, s)$ is a regularity function; $f_D(k, l, s)$ is a descriptive information function.

A larger window allows detecting fragments with similar order of greater length. Increasing the length of the fragments results in function values providing tending to a value of corresponding integral characteristic of the full length sequence. Reducing the length of the fragments allows using separate functions values for detection of more detailed features of the arrangement of components within the window. Preliminary studies showed that the relationship between the window length and the dispersion of the characteristic values is hyperbolic. However, if the window length is reduced to the cardinality of the alphabet ($m = 4$), this dependence is violated. Thus, as expected, the uncertainty of the location of the fragment is associated with the uncertainty of the function values obtained for a given window length. Selection of an optimal window length for various tasks, including, dependence on the cardinality of the alphabet of the original sequence, requires additional research.

Software for calculating and displaying the functions of characteristics of order is developed and tested on ribosomal RNA of several organisms.

Research revealed the influence of fragment (L -gram) length on the shape of functions of characteristics of order. The possibility of using the functions of the characteristics of order for finding similar or overlapping fragments in one or more sequences is considered, as well as – the inverse problem – finding of occurrences of the specified fragments in the complete genome sequence. Displaying the order of nucleotide sequences with functions, besides noted means also allows using the classical methods of mathematics, such as: mathematical analysis, spectral analysis, correlation analysis, etc., that would be impossible with the direct analysis of symbolic sequences themselves. It is noted that the graphical representation of functions of characteristics of order allows carrying expert analysis of long nucleotide sequences, including complete genome sequences.

REFERENCES

1. Gumenyuk A.S., Morozenko E.V., Rodionov I.N. Formalization of analysis of order of sign chains. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2011, no. 2(15), pp. 15-23. (In Russian).
2. Gumenyuk A., Kostyshin A., Simonova S. An approach to the research of the structure of linguistic and musical texts. *Glottometrics*, 2002, no. 3, pp. 61-69.
3. Gumenyuk A.S. [On the means of analysis of mutual arrangement of the components of the sign sequence]. *Voennaya tekhnika, voornuzhenie i tekhnologii dvoynogo primeneniya : materialy III Mezhdunar. tekhnolog. Kongressa [Military equipment, arms and double purpose technologies. Procs. of the 3rd International Congress of Technology]*. Omsk: OmSTU Publ., 2005, pt. 2, pp. 48-52. (In Russian).
4. Gumenyuk A.S., Pozdnichenko N.N., Shpynov S.N., Rodionov I.N. Formal Analysis of Structures of Nucleotide Chains. *Matematicheskaya biologiya i bioinformatika – Mathematical Biology and Bioinformatics*, 2013, vol. 8, no. 1, pp. 373-397. Available at: http://www.matbio.org/arti-cle.php?journ_id=15&id=158.
5. Gumenyuk A.S., Pozdnichenko N.N. [Analysis of order of the nucleotide sequences]. *Materialy Vserossiyskoy konferentsii s mezhdunarodnym uchastiem "Znaniya – ontologii – teorii" (ZONT-2013) [The 4th All-Russian Conference "Knowledge – Ontology – Theory" (KONT-13)]*. Novosibirsk, 2013, vol. 2, pp. 58-68. (In Russian).
6. Mazur M. *Kachestvennaya teoriya informatsii [Qualitative information theory]*. Moscow: Mir Publ., 1974. 240 p.
7. Sadovskiy M.G. *Informatsionno-statisticheskiy analiz nukleotidnykh posledovatel'nostey: dis. d-ra fiz.-mat. nauk [Information and statistical analysis of nucleotide sequences. Physics and Mathematics Doc. Thesis]*. Krasnoyarsk, 2004. 394 p.
8. Gumenyuk A.S. [On the formalism of constructing abstract objects in the inner physical space of information system (Elements of the algebra of mental events)]. *Sistemnyy analiz v proektirovanii i upravlenii: trudy X Mezhdunar. nauch.-prakt. konf. [The system analysis in the design and management. Procs. of the 10th International Scientific and Practical Conference]*. St. Petersburg: Polytechnic University Publ., 2006, pt. 2, pp. 172-181.
9. *National Center for Biotechnology Information*. Available at: <http://www.ncbi.nlm.nih.gov/nuccore/>
10. Tsybmal V.S., Pozdnichenko N.N. [About the development of the module for the calculation of local characteristics of order of the nucleotide sequences]. *Materialy XII Vserossiyskoy nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem "Informatsionnye tekhnologii i matematicheskoe modelirovanie (ITMM-2013)" [Proc. of 12th All-Russian research conference "Information technologies and mathematical modeling" (ITMM-2013)]*. Tomsk: Tomsk State University Publ., 2013, pp. 61-65.

А.С. Довбыш, О.Б. Берест

ТРЕХАЛЬТЕРНАТИВНАЯ ОБУЧАЮЩАЯСЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ АВТОМАТИЗАЦИИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Рассматривается задача информационного синтеза обучающейся системы поддержки принятия решений для управления технологическим процессом выращивания сцинтилляционных монокристаллов из расплава. В качестве критерия функциональной эффективности предложена модификация информационной меры Кульбака для трёхальтернативной системы оценок решений. Выполнен сравнительный анализ функциональной эффективности синтезированной системы с двухальтернативной.

Ключевые слова: система поддержки принятия решений; распознавание; обучение; трёхальтернативное решение; критерий функциональной эффективности; информационная мера Кульбака.

Современный этап развития информационных технологий анализа и синтеза систем поддержки принятия решений (СППР) для управления сложными технологическими процессами в условиях априорной неопределённости характеризуется наделением их интеллектуальной составляющей для моделирования когнитивных процессов принятия решений, свойственных человеку [1–3]. Основной нерешённой проблемой методов интеллектуального анализа данных в технологии Data Mining [4] является обеспечение инвариантности алгоритмов принятия решений от произвольных начальных условий. Одним из перспективных подходов к повышению функциональной эффективности СППР для управления технологическими процессами является использование идей и методов информационно-экстремальной интеллектуальной технологии (ИЭИ-технологии), основанной на максимизации информационной способности системы в процессе её обучения [5–7]. В работе [8, 9] рассматривалась задача информационно-экстремального синтеза обучающейся СППР при двухальтернативной системе оценок решений, основанная на методе «ближайшего соседа» [10]. Одним из путей повышения достоверности оценки функционального состояния технологического процесса является переход от двухальтернативной системы оценок принимаемых решений к трёхальтернативной в форме «Меньше нормы» – «Норма» – «Больше нормы».

В статье рассматривается задача оценки функциональной эффективности информационно-экстремального обучения СППР для управления выращиванием сцинтилляционных монокристаллов при использовании трёхальтернативной системы оценок управляющих решений.

1. Формализованная постановка задачи информационного синтеза обучающейся СППР

Пусть для заданного алфавита классов распознавания $\{X_m^0 | m = 1, \dots, M\}$, характеризующих M функциональных состояний СППР, получена многовекторная обучающая матрица $\|y^{(j)}_{m,i}\|$ типа «объект – свойство», в которой строка $\{y^{(j)}_{m,i} | i = 1, \dots, N\}$ является вектором-реализацией (далее реализация) класса X_m^0 ; N – количество признаков распознавания, а столбец – случайной обучающей выборкой $\{y^{(j)}_{m,i} | j = 1, \dots, n\}$, где n – объём выборки. Кроме того, дан структурированный вектор пространственно-временных параметров функционирования $g = \langle g_1, \dots, g_\xi, \dots, g_E \rangle$, влияющих на функциональную эффективность обучающейся СППР, с соответствующими для них ограничениями $R_\xi = (g_1, \dots, g_\xi, \dots, g_E) \leq 0$. Необходимо в процессе обучения найти оптимальные значения параметров функционирования $\{g^*_\xi\}$, обеспечивающих максимум информационного критерия функциональной эффективности (КФЭ) обучения системы в рабочей (допустимой) области определения его функции

$$E_m^* = \max_G E_m, \quad (1)$$

где E_m – КФЭ обучения СППР распознавать реализации класса X_m^0 ; G – рабочая (допустимая) область значений функции КФЭ.

При этом решением частной задачи информационного синтеза обучающейся системы является определение оптимального значения параметра g_{ξ}^* :

$$g_{\xi}^* = \arg_{G_{\xi}} E_m^*,$$

где G_{ξ} – область допустимых значений параметра g_{ξ} .

Таким образом, в рамках ИЭИ-технологии машинное обучение рассматривается как процесс оптимизации пространственно-временных параметров функционирования СППР по информационному КФЭ (1).

В режиме экзамена – непосредственного распознавания в рабочем режиме функционирования – СППР должна принять решение о принадлежности распознаваемой реализации одному из классов заданного алфавита.

2. Оценка функциональной эффективности обучения СППР на базе критерия Кульбака

Центральным вопросом информационного синтеза обучающейся СППР в рамках ИЭИ-технологии является конструирование общего КФЭ обучения системы. Для построения КФЭ в качестве функции, зависящей от эмпирических значений признаков распознавания, рассмотрим симметрический информационный критерий в виде дивергенции Кульбака–Лейблера [11]:

$$D_{KL}(p, q) = \sum_{x \in R} (p(x) - q(x)) \ln \frac{p(x)}{q(x)}, \quad (2)$$

где $p(x)$, $q(x)$ – функции вероятности двух наборов случайной дискретной величины x , принимающей значение во множестве рациональных чисел R ; D_{KL} – расстояние между ансамблями $\{p\}$ и $\{q\}$.

Практическое применение формула (2) нашла в методах ИЭИ-технологии, где используется для вычисления КФЭ информационная мера Кульбака в виде [6]:

$$E_m^{(k)} = [P_{t,m}^{(k)} - P_{f,m}^{(k)}] \cdot \log_2 \frac{P_{t,m}^{(k)}}{P_{f,m}^{(k)}}, \quad (3)$$

где $P_{t,m}^{(k)}$ – полная вероятность правильного распознавания реализаций класса X_m^0 на k -м шаге обучения; $P_{f,m}^{(k)}$ – полная вероятность неправильного распознавания реализаций класса X_m^0 на k -м шаге обучения.

Если за основную априорную гипотезу γ_1 примем нахождение значения признака распознавания в поле допусков δ , а за альтернативную гипотезу γ_2 – нахождение признака за пределами поля допусков, то гипотезы μ_1 и μ_2 будут соответственно апостериорными. Тогда полные вероятности правильного и неправильного принятия решений для двухальтернативной системы их оценок имеют вид

$$\begin{aligned} P_{t,m}^{(k)} &= p(\mu_1) p(\gamma_1 / \mu_1) + p(\mu_2) p(\gamma_2 / \mu_2); \\ P_{f,m}^{(k)} &= p(\mu_1) p(\gamma_2 / \mu_1) + p(\mu_2) p(\gamma_1 / \mu_2). \end{aligned} \quad (4)$$

Поскольку в формулах (4) условные вероятности $p(\gamma_2 / \mu_1)$ – ошибка первого рода α , $p(\gamma_1 / \mu_2)$ – ошибка второго рода β , $p(\gamma_1 / \mu_1)$ – первая достоверность D_1 , $p(\gamma_2 / \mu_2)$ – вторая достоверность D_2 , то после их преобразования по формуле Байеса при равновероятных гипотезах и подстановки в формулу (3) получим

$$E_{Km}^{(k)} = \log_2 \left(\frac{2 - (\alpha_m^{(k)}(d) + \beta_m^{(k)}(d))}{\alpha_m^{(k)}(d) + \beta_m^{(k)}(d)} \right) \cdot [1 - (\alpha_m^{(k)}(d) + \beta_m^{(k)}(d))]. \quad (5)$$

Рассмотрим классификатор с трёхальтернативной системой оценок, вычисляемых в процессе анализа реализаций трёх обучающих матриц, каждая из которых соответствует определенному состоянию технологического процесса.

В работах [5, 12] получено выражение КФЭ обучения СППР с унимодальным (вложенным) классификатором, графическое представление которого показано на рис. 1.

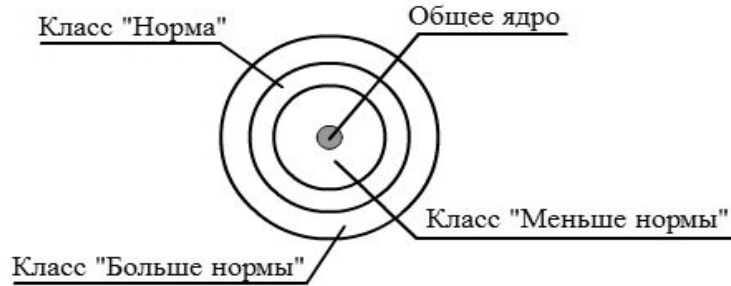


Рис. 1. Схема расположения классов для унимодального классификатора

В данном случае основная гипотеза γ_1 будет свидетельствовать о принадлежности признака показателю «Норма», гипотеза γ_2 – о принадлежности показателю «Меньше нормы» и гипотеза γ_3 – о принадлежности показателю «Больше нормы». Соответственно, их апостериорные гипотезы: μ_1 – значение признака действительно находится в поле допусков δ , μ_2 – левее поля допусков δ и μ_3 – правее поля допусков δ .

При этом возможные исходы для трёхальтернативного решения можно представить в виде девяти характеристик [5]:

- 1) первая достоверность $D_1^{(k)} = p(\gamma_1/\mu_1)$;
- 2) первая ошибка первого рода $\alpha_1^{(k)} = p(\gamma_2/\mu_1)$;
- 3) вторая ошибка первого рода $\alpha_2^{(k)} = p(\gamma_3/\mu_1)$;
- 4) вторая достоверность $D_2^{(k)} = p(\gamma_2/\mu_2)$;
- 5) первая ошибка второго рода $\beta_1^{(k)} = p(\gamma_1/\mu_2)$;
- 6) вторая ошибка второго рода $\beta_2^{(k)} = p(\gamma_3/\mu_2)$;
- 7) третья достоверность $D_3^{(k)} = p(\gamma_3/\mu_3)$;
- 8) первая ошибка третьего рода $\sigma_1^{(k)} = p(\gamma_1/\mu_3)$;
- 9) вторая ошибка третьего рода $\sigma_2^{(k)} = p(\gamma_2/\mu_3)$.

Согласно принципу Бернулли–Лапласа примем гипотезы равновероятными, и тогда полные вероятности правильного и неправильного принятия решений соответственно равны

$$\begin{aligned} P_{t,m}^{(k)} &= p(\mu_1)p(\gamma_1 / \mu_1) + p(\mu_2)p(\gamma_2 / \mu_2) + p(\mu_3)p(\gamma_3 / \mu_3); \\ P_{f,m}^{(k)} &= \sum_{i=2}^3 (p(\mu_1)p(\gamma_i / \mu_1) + p(\mu_i)p(\gamma_1 / \mu_i)) + p(\mu_2)p(\gamma_3 / \mu_2) + p(\mu_3)p(\gamma_2 / \mu_3). \end{aligned} \quad (6)$$

После подстановки выражений (5) в формулу (3) КФЭ СППР с трёхальтернативным унимодальным классификатором принимает вид [12]:

$$E_m^{(k)} = \frac{1}{3} \{D_{1,m}^{(k)} + 1 - 2[\beta_m^{(k)} + \sigma_m^{(k)}]\} \log_2 \frac{2D_{1,m}^{(k)} + 4 - 4[\beta_m^{(k)} + \sigma_m^{(k)}]}{1 - D_{1,m}^{(k)} + 2[\beta_m^{(k)} + \sigma_m^{(k)}]}.$$

Таким образом, информационная мера Кульбака является функционалом от точностных характеристик принимаемых решений и поэтому может использоваться в качестве общего КФЭ машинного обучения.

3. Трехальтернативный КФЭ обучения СППР с полимодальным классификатором

Унимодальный классификатор является частным случаем полимодального, имеющего не один, а несколько центров рассеивания реализаций образов.

Рассмотрим способ формирования КФЭ обучения СППР с трёхальтернативным полимодальным гиперсферическим классификатором для общего случая, когда классы распознавания пересекаются (рис. 2).



Рис. 2. Возможные варианты принадлежности распознаваемой реализации к областям классов распознавания

В табл. 1 показаны возможные варианты нахождения в пространстве признаков распознавания реализации базового класса X_1^0 , соответствующего функциональному состоянию СППР «Норма». При этом отнесение этой реализации к классу X_1^0 в случае, когда она принадлежит двум и более классам, будем считать неэффективным.

Таблица 1

Возможные варианты нахождения вектора-реализации класса «Норма» относительно всех классов распознавания

Правильное распознавание	Вектор-реализация принадлежит только к базовому классу
	Вектор-реализация принадлежит к базовому классу и классу «Меньше нормы». Неэффективное распознавание
	Вектор-реализация принадлежит к базовому классу и классу «Больше нормы». Неэффективное распознавание
	Вектор-реализация принадлежит ко всем классам распознавания. Неэффективное распознавание
Неправильное распознавание	Вектор-реализация принадлежит к классу «Меньше нормы»
	Вектор-реализация принадлежит к классу «Больше нормы»
	Вектор-реализация принадлежит к классам «Меньше нормы» и «Больше нормы»
	Вектор-реализация не принадлежит ни к одному из существующих классов

Таким образом, только в первом случае классификатор распознаёт реализацию класса X_1^0 с максимальной полной достоверностью правильного принятия решений. В случаях 2–4 (рис. 2), где имеет место пересечение классов, распознавание считается неэффективным.

Априорные гипотезы γ_1 , γ_2 и γ_3 являются такими же, как и в случае унимодального классификатора, а согласно табл. 1 апостериорные гипотезы имеют следующие значения:

- μ_1 – реализация находится в поле допусков δ класса «Норма» (область 1);
- μ_2 – реализация находится в поле допусков δ класса «Меньше нормы» (область 5);
- μ_3 – реализация находится в поле допусков δ класса «Больше нормы» (область 6);
- μ_4 – реализация находится в поле допусков δ класса «Меньше нормы» и класса «Норма» одновременно (область 2);

– μ_5 – реализация находится в поле допусков δ класса «Больше нормы» и класса «Норма» одновременно (см. рис. 2, область 3);

– μ_6 – реализация находится в поле допусков δ класса «Меньше нормы», класса «Норма» и класса «Больше нормы» одновременно (область 4);

– μ_7 – реализация находится в поле допусков δ класса «Больше нормы» и класса «Меньше нормы» одновременно (область 7).

Кроме того, введем дополнительную гипотезу μ_8 – реализация находится вне границ существующих классов, т.е. в области δ . Тогда возможные исходы для трёхальтернативных решений будут оцениваться следующими точностными характеристиками:

- 1) первая достоверность $D_1^{(k)} = p(\gamma_1/\mu_1)$;
- 2) первая ошибка второго рода $\alpha_2^{(k)} = p(\gamma_1/\mu_2)$;
- 3) первая ошибка третьего рода $\alpha_3^{(k)} = p(\gamma_1/\mu_3)$;
- 4) первая ошибка четвертого рода $\alpha_4^{(k)} = p(\gamma_1/\mu_4)$;
- 5) первая ошибка пятого рода $\alpha_5^{(k)} = p(\gamma_1/\mu_5)$;
- 6) первая ошибка шестого рода $\alpha_6^{(k)} = p(\gamma_1/\mu_6)$;
- 7) первая ошибка седьмого рода $\alpha_7^{(k)} = p(\gamma_1/\mu_7)$;
- 8) первая ошибка восьмого рода $\alpha_8^{(k)} = p(\gamma_1/\mu_8)$;
- 9) вторая достоверность $D_2^{(k)} = p(\gamma_2/\mu_2)$;
- 10) вторая ошибка первого рода $\beta_2^{(k)} = p(\gamma_2/\mu_1)$;
- 11) вторая ошибка третьего рода $\beta_3^{(k)} = p(\gamma_2/\mu_3)$;
- 12) вторая ошибка четвертого рода $\beta_4^{(k)} = p(\gamma_2/\mu_4)$;
- 13) вторая ошибка пятого рода $\beta_5^{(k)} = p(\gamma_2/\mu_5)$;
- 14) вторая ошибка шестого рода $\beta_6^{(k)} = p(\gamma_2/\mu_6)$;
- 15) вторая ошибка седьмого рода $\beta_7^{(k)} = p(\gamma_2/\mu_7)$;
- 16) вторая ошибка восьмого рода $\beta_8^{(k)} = p(\gamma_2/\mu_8)$;
- 17) третья достоверность $D_3^{(k)} = p(\gamma_3/\mu_3)$;
- 18) третья ошибка первого рода $\sigma_1^{(k)} = p(\gamma_3/\mu_1)$;
- 19) третья ошибка второго рода $\sigma_2^{(k)} = p(\gamma_3/\mu_2)$;
- 20) третья ошибка четвертого рода $\sigma_4^{(k)} = p(\gamma_3/\mu_4)$;
- 21) третья ошибка пятого рода $\sigma_5^{(k)} = p(\gamma_3/\mu_5)$;
- 22) третья ошибка шестого рода $\sigma_6^{(k)} = p(\gamma_3/\mu_6)$;
- 23) третья ошибка седьмого рода $\sigma_7^{(k)} = p(\gamma_3/\mu_7)$;
- 24) третья ошибка восьмого рода $\sigma_8^{(k)} = p(\gamma_3/\mu_8)$.

Таким образом, приведенные 24 точностные характеристики позволяют полностью оценить принадлежность реализаций к одному из трёх классов распознавания.

Поскольку для системы распознавания независимо от используемой системы оценок принятия решений базовым является класс X^0_1 – «Норма», то, воспользовавшись формулой (3), полные вероятности правильного и неправильного принятия решений представим в следующем виде:

$$\begin{aligned}
 P_t^{(k)} &= p(\mu_1)p(\gamma_1/\mu_1) + p(\mu_2)p(\gamma_2/\mu_2) + p(\mu_3)p(\gamma_3/\mu_3); \\
 P_f^{(k)} &= p(\mu_1)p(\gamma_2/\mu_1) + p(\mu_1)p(\gamma_3/\mu_1) + p(\mu_2)p(\gamma_1/\mu_2) + p(\mu_2)p(\gamma_3/\mu_2) + p(\mu_3)p(\gamma_1/\mu_3) + \\
 &+ p(\mu_3)p(\gamma_2/\mu_3) + \sum_{i=4}^8 (p(\mu_i)p(\gamma_1/\mu_i) + p(\mu_i)p(\gamma_2/\mu_i) + p(\mu_i)p(\gamma_3/\mu_i)).
 \end{aligned} \quad (7)$$

В соответствии с принципом Лапласа–Бернулли сделаем допущение о равных вероятностях попадания реализаций в каждую область, т.е., приняв $p(\mu_1) = p(\mu_2) = p(\mu_3) = p(\mu_4) = p(\mu_5) = p(\mu_6) = p(\mu_7) = p(\mu_8) = 1/8$, формулу (7) представим в виде

$$\begin{aligned}
P_t^{(k)} &= \frac{1}{8}(D_1^{(k)} + D_2^{(k)} + D_3^{(k)}); \\
P_f^{(k)} &= \frac{1}{8}(\alpha_2^{(k)} + \alpha_3^{(k)} + \beta_1^{(k)} + \beta_3^{(k)} + \sigma_1^{(k)} + \sigma_2^{(k)} + \sum_{i=4}^8 (\alpha_i^{(k)} + \beta_i^{(k)} + \sigma_i^{(k)})).
\end{aligned} \tag{8}$$

После соответствующей подстановки выражений (8) в формулу (3) получим

$$\begin{aligned}
E^{(k)} &= [P_t^{(k)} - P_f^{(k)}] \log_2 \frac{P_t^{(k)}}{P_f^{(k)}} = \frac{1}{8}(D_1^{(k)} + D_2^{(k)} + D_3^{(k)} - (\alpha_2^{(k)} + \alpha_3^{(k)} + \beta_1^{(k)} + \beta_3^{(k)} + \sigma_1^{(k)} + \sigma_2^{(k)} + \\
&+ \sum_{i=4}^8 (\alpha_i^{(k)} + \beta_i^{(k)} + \sigma_i^{(k)}))) \log_2 \frac{D_1^{(k)} + D_2^{(k)} + D_3^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)} + \beta_1^{(k)} + \beta_3^{(k)} + \sigma_1^{(k)} + \sigma_2^{(k)} + \sum_{i=4}^8 (\alpha_i^{(k)} + \beta_i^{(k)} + \sigma_i^{(k)})}.
\end{aligned} \tag{9}$$

С учётом того, что сумма полных вероятностей правильного и неправильного принятия решения равна единице для каждого m -го класса, т.е. $P_{t,m}^{(k)} + P_{f,m}^{(k)} = 1$, формула (9) примет вид

$$\begin{aligned}
E^{(k)} &= \frac{1}{8}(D_1^{(k)} + D_2^{(k)} + D_3^{(k)} - (3 - D_1^{(k)} - D_2^{(k)} - D_3^{(k)})) \log_2 \frac{D_1^{(k)} + D_2^{(k)} + D_3^{(k)}}{(3 - D_1^{(k)} - D_2^{(k)} - D_3^{(k)})} = \\
&= \frac{1}{8}(2(D_1^{(k)} + D_2^{(k)} + D_3^{(k)}) - 3) \log_2 \frac{D_1^{(k)} + D_2^{(k)} + D_3^{(k)}}{(3 - D_1^{(k)} - D_2^{(k)} - D_3^{(k)})}.
\end{aligned} \tag{10}$$

Таким образом, анализ выражения (10) показывает, что информационный КФЭ является функционалом от точностных характеристик, значения которых, в свою очередь, зависят от геометрических параметров контейнеров классов распознавания.

Рассмотрим вычислительный аспект оценки функциональной эффективности обучающейся СППР. При ограниченном объеме обучающих матриц введём следующие оценки точностных характеристик:

$$D_1^{(k)} = \frac{K_1^{(k)}}{n_{\min}}; \quad D_2^{(k)} = \frac{K_2^{(k)}}{n_{\min}}; \quad D_3^{(k)} = \frac{K_3^{(k)}}{n_{\min}}, \tag{11}$$

где $K_1^{(k)}$ – количество событий, состоящих в принадлежности распознаваемых реализаций к классу X_m^0 , характеризующему состояние процесса «Норма», если они действительно принадлежат обучающей матрице данного класса на k -м шаге обучения; $K_2^{(k)}$ – количество событий, состоящих в принадлежности реализаций к классу «Меньше нормы», если они действительно принадлежат обучающей матрице данного класса; $K_3^{(k)}$ – количество событий, состоящих в принадлежности реализаций к классу «Больше нормы», если они действительно принадлежат обучающей матрице данного класса; $n_{\min}$ – минимальный объём репрезентативной обучающей выборки.

После соответствующей подстановки оценок точностных характеристик (11) в выражение (10) рабочая формула для вычисления трёхальтернативного КФЭ примет вид

$$E^{(k)} = \frac{1}{8}(2(K_1^{(k)} + K_2^{(k)} + K_3^{(k)}) / n_{\min} - 3) \log_2 \frac{K_1^{(k)} + K_2^{(k)} + K_3^{(k)}}{(3n_{\min} - K_1^{(k)} - K_2^{(k)} - K_3^{(k)})}. \tag{12}$$

В формуле (12) подсчет количества событий для каждого класса распознавания происходит следующим образом:

1) одновременное изменение в процессе обучения радиусов контейнеров трех классов в интервале $[1; d_{\min}]$, где $d_{\min}$ – минимальное межклассовое кодовое расстояние (рис. 3);

2) вычисление в бинарном пространстве Хэмминга кодового расстояния $d(x_m \oplus x^j)$ между центрами гиперсфер и реализациями обучающих матриц, где x_m – эталонный вектор-реализация, вершина которого определяет центр контейнера, и x^j – j -я реализация класса X_m^0 ;

3) отнесение реализации к классу распознавания X_m^0 , для которого максимальна функция принадлежности

$$\mu_m = 1 - \frac{d(x_m \oplus x^j)}{r_m},$$

где r_m – текущий радиус контейнера класса X_m^0 .

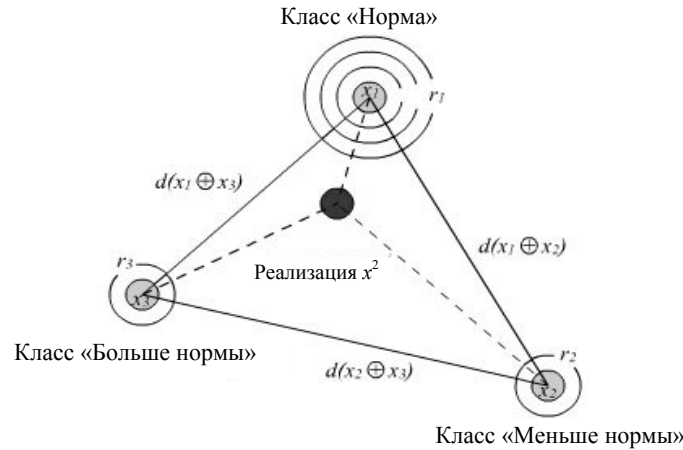


Рис. 3. Расположение классов в бинарном пространстве Хэмминга в общем случае

Для качественного сравнения функций двухальтернативного (5) и трёхальтернативного (12) критериев преобразуем функцию (12) к виду

$$E^{(k)} = \frac{1}{8} (2t / n_{\min} - 3) \log_2 \frac{t}{(3n_{\min} - t)}, \quad (13)$$

где t – количество правильно распознанных реализаций в процессе обучения СППР. При этом на них накладываются ограничения

$$\begin{aligned} t &= (K_1^{(k)} + K_2^{(k)} + K_3^{(k)}), \\ t &\in (0; 3n_{\min}). \end{aligned} \quad (14)$$

На рис. 4 показаны графики зависимости модифицированных критериев Кульбака (5) и (12) от числа правильно распознанных реализаций, вычисленные при двухальтернативной и трёхальтернативной системах оценок решений и одинаковом количестве реализаций каждого класса распознавания в обучающей матрице ($n_{\min} = 61$).

Анализ рис. 4 показывает, что функции информационных критериев не являются взаимоднозначными. При этом рабочими (допустимыми) областями определения функций критериев являются светлые участки графиков, в которых значения критериев возрастают с увеличением количества правильно распознанных реализаций. Например, для двухальтернативных критериев (рис. 4, а) рабочая область, где $D_1^k > 0,5$ и $D_2^k > 0,5$, начинается при $t > 61$, а для трёхальтернативных критериев (см. рис. 4, б) эта область, где $D_1^k > 0,5$, $D_2^k > 0,5$ и $D_3^k > 0,5$, начинается при $t > 91$.

Рассмотрим применение трёхальтернативного критерия (12) на примере оценки функциональной эффективности обучения СППР для управления выращиванием скнтилляционных монокристаллов с расплава [13].

Известно, что задача управления ростом монокристалла из расплава формулируется как задача стабилизации его диаметра при программно-логическом управлении остальными параметрами ростовой установки. Из архивных данных была сформирована обучающая матрица для трёх классов распознавания, характеризующих три функциональных состояния технологического процесса, соответствующих как технологическому режиму, так и отклонениям от него. Обучающая матрица для

каждого класса содержала 61 реализацию, состоящую из 45 структурированных первичных и вторичных признаков.

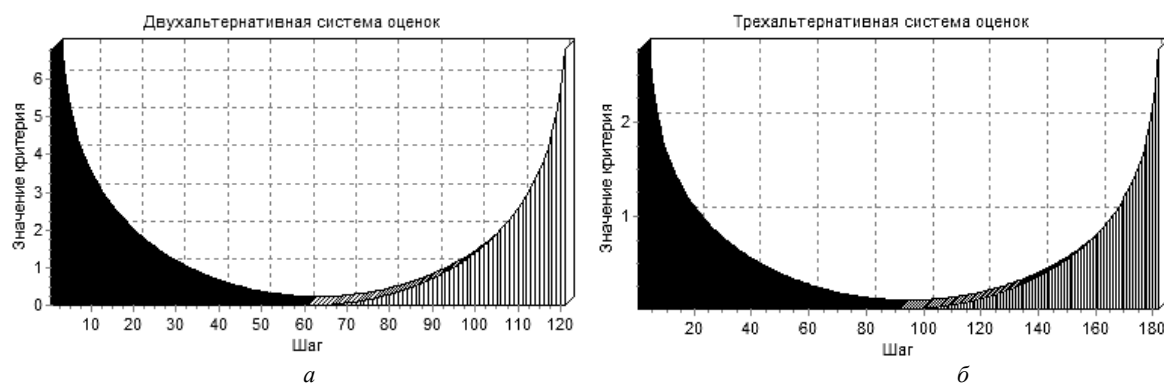


Рис. 4. Графики зависимости информационных критериев от правильно распознанных реализаций: a – для двухальтернативных решений; b – для трёхальтернативных решений

На рис. 5 показан график КФЭ (12), определяемый на каждом шаге обучения СППР при заданном параметре поля контрольных допусков на признаки распознавания $\delta = \pm 40$ (в относительных единицах). При этом на каждом шаге обучения СППР изменялся на заданную величину радиус одного контейнера, в то время как радиусы контейнеров остальных двух классов оставались фиксированными. Начальные значения радиусов контейнеров равнялись одной кодовой единице относительно их эталонных векторов-реализаций, а в процессе обучения были перебраны все возможные допустимые их значения.

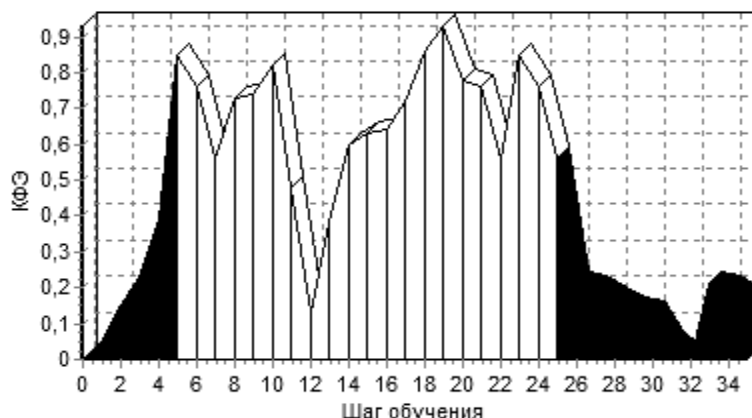


Рис. 5. График зависимости КФЭ для трех классов распознавания от радиуса контейнеров классов распознавания

Анализ рис. 5 показывает, что максимальное значение КФЭ $E^* = 0,92$ было получено на четвёртом шаге обучения при значениях радиусов в кодовых единицах $r_1 = 4$, $r_2 = 4$ и $r_3 = 3$ соответственно.

Для сравнения был реализован алгоритм обучения СППР для управления выращиванием скитилляционных монокристаллов с использованием двухальтернативного КФЭ (5) при таких же входных обучающих матрицах и параметре δ поля контрольных допусков. В результате при аналогичных радиусах контейнеров было получено меньшее по сравнению с трёхальтернативным критерием значение усреднённого для трёх классов распознавания КФЭ ($E^* = 0,72$).

Заключение

Предложенный КФЭ на базе модификации информационной меры Кульбака для трёхальтернативной системы оценок решений позволяет оценивать функциональную эффективность обучения СППР

для распределений реализаций классов распознавания в пространстве признаков произвольной конфигурации и, таким образом, обладает универсальностью по сравнению с аналогичным критерием оптимизации параметров обучения для унимодального классификатора. Кроме того, использование предложенного КФЭ для построения в процессе обучения СППР решающих правил характеризуется более высокими достоверностью и оперативностью по сравнению с двухальтернативным критерием. При этом повышение оперативности алгоритма обучения достигается за счёт отсутствия в процессе его реализации усреднения значений КФЭ и одновременного построения решающих правил для распознавания трёх классов.

ЛИТЕРАТУРА

1. Hastie T., Tibshirani R., Friedman J., Hastie T. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd ed. Springer Series in Statistics, 2009. 745 p.
2. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкина Л.Д. Прикладная статистика: Классификация и снижение размерности. М. : Финансы и статистика, 1989. 607 с.
3. Бендат Дж. Прикладной анализ случайных данных : пер. с англ. М. : Мир, 1989. 540 с.
4. Han J., Kamber M., Pei J. Data mining: concepts and techniques. 3rd ed. Morgan Kaufmann, Elsevier, 2012. 744 p.
5. Красноясовский А.С. Информационный синтез интеллектуальных систем управления: Подход, основанный на методе функционально-статистических испытаний. Сумы : Изд-во СумГУ, 2004. 261 с.
6. Довбыш А.С. Основы проектирования интеллектуальных систем : учеб. пособие. Сумы : Изд-во СумГУ, 2009. 171 с.
7. Довбыш А.С., Васильев А.В., Любчук В.А. Интеллектуальные информационные технологии в электронном обучении. Сумы : Изд-во СумГУ, 2013. 177 с.
8. Берест О.Б., Довбыш А.С., Козьмин Ю.С. Информационно-экстремальный алгоритм обучения системы управления выращиванием сцинтилляционных монокристаллов // Вестник «ХПИ»: Системный анализ, управление и информационные технологии. 2012. № 30. С. 54–60.
9. Берест О.Б., Довбыш А.С. Оценка функциональной эффективности обучения автоматизированной системы управления технологическим процессом // Вестник СумГУ. Сер. Технические науки. 2012. № 2. С. 39–47.
10. Duda R.O., Hart P.E., Stork D.G. Pattern classification. Second ed. N.Y. : John Wiley&Sons, 2001. 738 p.
11. Довбыш А.С., Джулгам С.А.С.М., Стадник А.А. Информационно-экстремальный алгоритм обучения системы диагностирования патологических процессов // Инновации в науке : сб. статей. Новосибирск : НПИ «СибАК», 2013. № 23. С. 45–54.
12. Kullback S., Leibler R.A. On information and sufficiency // The Annals of Mathematical Statistics. 1951. V. 22, No 1. P. 79–86.
13. Суздаль В.С., Стадник П.Е., Герасимчук Л.И., Епифанов Ю.М. Сцинтилляционные монокристаллы: автоматизированное выращивание. Харьков : ИСМА, 2009. 259 с.

Довбыш Анатолий Степанович, д-р техн. наук, профессор. E-mail: kras@id.sumdu.edu.ua

Берест Олег Борисович. E-mail: Berest_Oleg@mail.ru

Сумский государственный университет

Поступила в редакцию 10 мая 2014 г.

Dovbysh Anatoly S., Berest Oleg B. (Sumy State University, Ukraine).

Three-alternative learning rating system for decision support system of process automation.

Keywords: decision support system, recognition, learning, three-alternative decision, criterion for functional efficiency, Kullback information measure.

The problem of information synthesis of learning decision support system for scintillator crystals growth from the melt is under consideration. As a criterion of functional efficiency modification of Kullback - Leibler information measure is proposed:

$$D_{KL}(p, q) = \sum_{x \in \mathcal{X}} (p(x) - q(x)) \ln \frac{p(x)}{q(x)},$$

where $p(x)$ and $q(x)$ are the probability functions of two sets of discrete random value $x \in R$, and D_{KL} is the distance between the ensembles $\{p\}$ and $\{q\}$.

To improve the functional efficiency of decision support systems (DSS), the ideas and methods of information and intellectual extreme technology were used [5-9], based on maximizing the information ability of the system in the learning process. The multimodal three-alternative learning rating system is proposed as more appropriate for this process. The formula of the criteria is

$$E^{(k)} = \frac{1}{8} (2(K_1^{(k)} + K_2^{(k)} + K_3^{(k)}) / n_{\min} - 3) \log_2 \frac{K_1^{(k)} + K_2^{(k)} + K_3^{(k)}}{(3n_{\min} - K_1^{(k)} - K_2^{(k)} - K_3^{(k)})},$$

where $K_1^{(k)}$, $K_2^{(k)}$, and $K_3^{(k)}$ are respectively the number of events, which define belonging recognizable implementations to the classes «LESS THAN NORMAL», «NORMAL», «MORE THAN NORMAL», if they are really belong to the training data matrixes classes on the k^{th} step of learning; $n_{\min}$ is the minimum amount of representative training sample.

This approach provides the opportunity to work with three classes in the same recognition feature space, monitoring the radiuses of corresponding containers. This improves the quality of the recognition system. Proposed criteria for three-alternative system allows to evaluate the functional efficiency of the DSS with random distributions of recognition implementations in the feature space, and thus,

has the versatility compared to the unimodal classifier with the same learning parameters [12]. In addition, the use of the MFE to build learning rules characterized by higher reliability and efficiency in comparison with two-alternative criterion. This increase in efficiency of learning algorithm is achieved due to the absence in the process of its implementation averaging the criteria and the simultaneous construction of decision rules for three classes.

REFERENCES

1. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2009. 745 p.
2. Aivazyan S.A., Buchshtaber V.M., Enukov I.S., Meshalkina L.D. *Prikladnaya statistika: Klassifikatsiya i snizhenie razmernosti* [Applied Statistics: Classification and Dimension Reduction]. Moscow: Finance and Statistics Publ., 1989. 607 p. (in Russian).
3. Bendat J.S., Piersol A.G. *Prikladnoy analiz sluchaynykh dannykh* [Random Data: Analysis and Measurement Procedures]. Translated from English. Moscow: Mir Publ., 2010. 640 p.
4. Han J., Kamber M., Pei J. *Data mining: concepts and techniques*. Morgan Kaufmann, Elsevier, 2012. 744 p.
5. Krasnopoyasovsky A.S. *Informatsionnyy sintez intellektual'nykh sistem upravleniya: Podkhod, osnovanny na metode funktsional'no-statisticheskikh ispytaniy* [Information Synthesis of Intelligent Control Systems: An Approach Based on the Method of Functional-Statistical Tests]. Sumy: Sumy State University Publ., 2004. 261 p. (In Ukrainian).
6. Dovbysh A.S. *Osnovy proektirovaniya intellektual'nykh sistem* [The fundamentals of Intelligent Systems Design]. Sumy: Sumy State University Publ., 2009. 171 p. (In Ukrainian).
7. Dovbysh A.S., Vassiliev A.V., Lyubchak V.A. *Intellektual'nye informatsionnye tekhnologii v elektronnom obuchenii* [Intelligent information technology in e-learning]. Sumy: Sumy State University Publ., 2013. 177 p. (In Ukrainian).
8. Berest O.B., Dovbysh A.S., Kozmin U.S. Informatsionno-ekstremal'nyy algoritm obucheniya sistemy upravleniya vyrashchivaniem stsintillyatsionnykh monokristallov [The information - extreme learning algorithm of intellectual control scintillation monocrystal growth system]. *Vestnik KhPI: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii*, 2012, no. 30, pp. 54-60. (In Ukrainian).
9. Berest O.B., Dovbysh A.S. Otsenka funktsional'noy effektivnosti obucheniya avtomatizirovannoy sistemy upravleniya tekhnologicheskimm protsessom [Influence of vector – realizations number on functional efficiency of recognition system learning by Kullback's measure]. *Visnyk of the SSU*, 2012, no 2, pp. 39-47. (In Ukrainian).
10. Duda R.O., Hart P.E., Stork D.G. *Pattern Classification*. New York: John Wiley&Sons, 2001. 738 p.
11. Dovbysh A.S., Dzhulgam S.A.S.M., Stadnik A.A. Informatsionno-ekstremal'nyy algoritm obucheniya sistemy diagnostirovaniya patologicheskikh protsessov [The information - extreme learning algorithm of pathological processes diagnostic system]. *Innovatsii v nauke*, 2013, no 23, pp. 45-54.
12. Kullback S., Leibler R.A. On information and sufficiency. *The Annals of Mathematical Statistics*, 1951, vol. 22, no. 1, pp. 79-86. DOI: 10.1214/aoms/1177729694
13. Suzdal V.S., Stadnik P.E., Gerasimchuk L.I., Epifanov Y.M. *Stsintillyatsionnye monokristally: avtomatizirovannoe vyrashchivanie* [Scintillation single crystals: Automated growth]. Kharkiv: ISMA Publ., 2009. 259 p.

Е.А. Фёдорова

ВЫЧИСЛЕНИЕ МОМЕНТОВ В RQ-СИСТЕМЕ MMPP|M|1

Работа выполнена в рамках проекта Министерства образования и науки РФ № 1.511.2614/К.

Исследуется RQ-система MMPP|M|1 методом моментов. Найдены «квазитошные» формулы для вычисления первого и второго начальных моментов распределения вероятностей числа заявок в ИПВ. Проведен численный анализ полученных результатов.

Ключевые слова: RQ-система; источник повторных вызовов; метод моментов.

Исследования реальных систем передачи данных и сетей сотовой связи привели к тому, что описанные системы требуют рассмотрения математических моделей, выходящих за рамки множества классических систем массового обслуживания – систем с ожиданием и систем с потерями. Таким образом, стали выделять новый класс систем, который получил название RQ-системы (Retrial Queueing Systems [1–3]) или системы с повторными вызовами.

Принципиальное отличие RQ-систем от классических систем массового обслуживания состоит в том, что заявки, пришедшие в систему и обнаружившие прибор занятым, не покидают систему, а идут в источник повторных вызовов, где после некоторой задержки пытаются снова занять прибор для обслуживания.

Первые работы, посвященные исследованию систем с повторными вызовами, были опубликованы в середине XX в. R.I. Wilkinson [4] и J.W. Cohen [5]. Основные подходы к описанию систем с ИПВ были рассмотрены G. Gosztony [6], A. Elldin [7] и др. Большинство первых работ посвящено описанию практических задач и влиянию эффекта повторных вызовов в телекоммуникационных системах.

Наиболее полное и глубокое исследование различных процессов в системах с повторными вызовами проведено в работах Г.И. Фалина, J.R. Artalejo, A. Gomez-Corral и J.G.C. Templeton [1–3]. Ими получены допредельные характеристические функции для RQ-систем M|M|1, M|GI|1, M|M|с и т.д., а также рассмотрены разнообразные методы для исследования RQ-систем.

Большинство исследований Retrial Queueing System реализуются численно или с помощью имитационного моделирования [08–11]. Аналитические методы получены только в тех случаях, когда модели потока и дисциплина обслуживания относительно просты (например, пуассоновский поток и экспоненциальное распределение закона обслуживания) [1]. RQ-системы с входящими MAP- и BMAP-потоками исследуются в работах В.И. Клименок, А.Н. Дудина [12], в которых используются преимущественно матричные методы исследования. Кроме того, матричные методы исследования RQ-систем используются также в работах M.F. Neuts, J.R. Artalejo, A.G. Gomez-Corral [13], J.E. Diamond, A.S. Alfa [14] и др. В ТГУ под руководством А.А. Назарова активно развиваются различные асимптотические методы исследования моделей СМО, в том числе и RQ-систем [15, 16]. Асимптотические и приближенные методы исследования RQ-систем развивались также Г.И. Фалиным [17], В.В. Анисимовым [18] и др.

Часто для построения аппроксимирующих распределений необходимо, чтобы были известны точные моменты искомых распределений. Однако в RQ-системах известны аналитические формулы вычисления моментов распределения вероятностей числа заявок в источнике повторных вызовов лишь для систем с простейшим входящим потоком. В данной работе предложено вычислить моменты для RQ-систем с входящим марковским модулированным потоком (MMPP).

1. Математическая модель

Рассмотрим однолинейную RQ-систему с источником повторных вызовов (ИПВ), на вход которой поступает MMPP-поток заявок с матрицей условных интенсивностей $r\Lambda$ и матрицей инфинитези-

мальных характеристик Q (рис. 1), 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 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 заявка мгновенно возвращается в источник повторных вызовов для реализации следующей задержки.

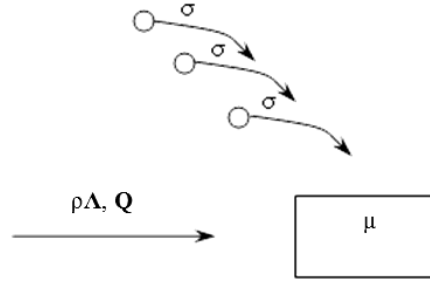


Рис. 1. RQ-система MMPP|M|1

Обозначим вектор-столбец $\mathbf{R}$ – стационарное распределение вероятностей значения цепи Маркова, управляющей входящим MMPP-поток, которое определяется из следующей системы:

$$\begin{cases} \mathbf{R}\mathbf{Q} = \mathbf{0}, \\ \mathbf{R}\mathbf{E} = \mathbf{1}, \end{cases} \quad (1)$$

где $\mathbf{E}$ – единичный вектор-столбец, $\mathbf{0}$ – вектор-столбец с нулевыми элементами.

Очевидно, что интенсивность входящего потока равна $\lambda = \mathbf{R} \cdot \rho\mathbf{\Lambda} \cdot \mathbf{E}$.

Пусть параметры системы таковы, что выполняется

$$\mathbf{R} \cdot \mathbf{\Lambda} \cdot \mathbf{E} = \mu. \quad (2)$$

Тогда загрузка системы определяется как $\rho = \lambda / \mu = \lambda / (\mathbf{R} \cdot \mathbf{\Lambda} \cdot \mathbf{E})$.

Для данной системы ставится задача найти математическое ожидание и дисперсию распределения вероятностей числа заявок в источнике повторных вызовов такой системы.

Пусть $i(t)$ – процесс, характеризующий число заявок в ИПВ. Процесс $i(t)$ не является марковским. Однако его можно марковизировать путем введения дополнительных компонент: $n(t)$ – цепь Маркова, управляющая MMPP-поток, а процесс $k(t)$ определяет состояние прибора следующим образом:

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

Обозначим $P\{k(t) = k, i(t) = i, n(t) = n\} = P(k, i, n, t)$ – вероятность того, что прибор в момент времени t находится в состоянии k , управляющая MMPP-поток цепь Маркова – в состоянии n , и в источнике повторных вызовов находится i заявок.

Очевидно, что процесс $\{k(t), i(t), n(t)\}$ изменения состояний данной системы во времени является марковским. Для получения распределения вероятностей $P(k, i, n, t)$ состояний рассматриваемой RQ-системы составим систему уравнений Колмогорова:

$$\begin{cases} \frac{\partial P(0, n, i, t)}{\partial t} = \mu P(1, n, i, t) - (\rho\lambda_n + i\sigma + q_{nn})P(0, n, i, t) + \sum_{v \neq n} P(0, v, i, t) \cdot q_{vn}, \\ \frac{\partial P(1, n, i, t)}{\partial t} = -(\rho\lambda_n + \mu + q_{nn})P(1, n, i, t) + \rho\lambda_n P(0, n, i, t) + \\ + \rho\lambda_n P(1, n, i-1, t) + \sigma(i+1) \cdot P(0, n, i+1, t) + \sum_{v \neq n} P(1, v, i, t) q_{vn}. \end{cases} \quad (3)$$

Обозначим векторы-строки $\mathbf{P}(k,i) = \{P(k,1,i) \ P(k,2,i) \ \dots \ P(k,N,i)\}$, где в стационарном режиме $\lim_{t \rightarrow \infty} P(k,i,n,t) = P(k,i,n)$. Тогда в матричном виде система (3) примет вид

$$\begin{cases} \mathbf{P}(0,i)(\mathbf{Q} - \rho\mathbf{A} - i\sigma \cdot \mathbf{I}) + \mu\mathbf{P}(1,i) = \mathbf{0}, \\ \mathbf{P}(1,i)(\mathbf{Q} - \rho\mathbf{A} - \mu\mathbf{I}) + \mathbf{P}(0,i)\rho\mathbf{A} + \mathbf{P}(1,i-1)\rho\mathbf{A} + \sigma(i+1) \cdot \mathbf{P}(0,i+1) = \mathbf{0}, \end{cases} \quad (4)$$

где $\mathbf{I}$ – единичная матрица.

Перейдем в системе (4) к частичным характеристическим функциям $\mathbf{H}(k,u) = \sum_i e^{ju_i} \mathbf{P}(k,i)$, где $j = \sqrt{-1}$ – мнимая единица. Тогда система уравнений (4) для характеристических функций переписывается в виде

$$\begin{cases} \mathbf{H}(0,u)(\mathbf{Q} - \rho\mathbf{A}) + j\sigma \frac{\partial \mathbf{H}(0,u)}{\partial u} + \mu\mathbf{H}(1,u) = \mathbf{0}, \\ \mathbf{H}(1,u)(\mathbf{Q} - \rho\mathbf{A} - \mu\mathbf{I}) + \mathbf{H}(0,u)\rho\mathbf{A} + \mathbf{H}(1,u)\rho\mathbf{A}e^{ju} - j\sigma e^{-ju} \frac{\partial \mathbf{H}(0,u)}{\partial u} = \mathbf{0}. \end{cases} \quad (5)$$

Из системы (5) найдем первый и второй моменты распределения вероятностей числа заявок в ИПВ, для этого будем использовать метод моментов.

2. Метод моментов

Обозначим $\mathbf{m}_k = -j \frac{\partial \mathbf{H}(k,u)}{\partial u} \Big|_{u=0}$, где $k=1,2$, тогда математическое ожидание числа заявок в

ИПВ вычисляется как $m = (\mathbf{m}_0 + \mathbf{m}_1)\mathbf{E} = \mathbf{m}\mathbf{E}$.

Кроме того, введем обозначения $\{\mathbf{R}_0, \mathbf{R}_1\}$ – двумерное совместное стационарное распределение вероятностей состояний ММРР-потока и состояний прибора. Очевидно, что для векторов $\mathbf{R}_0$ и $\mathbf{R}_1$ выполняются равенства $(\mathbf{R}_0 + \mathbf{R}_1)\mathbf{Q} = \mathbf{0}$ и $\mathbf{R}_0\mathbf{E} + \mathbf{R}_1\mathbf{E} = 1$.

Метод моментов состоит из нескольких этапов

Этап 1. Примем $u=0$ в системе (5). Получим следующую систему:

$$\begin{cases} \mathbf{R}_0(\mathbf{Q} - \rho\mathbf{A}) - \sigma\mathbf{m}_0 + \mu\mathbf{R}_1 = \mathbf{0}, \\ \mathbf{R}_1(\mathbf{Q} - \mu\mathbf{I}) + \mathbf{R}_0\rho\mathbf{A} + \sigma\mathbf{m}_0 = \mathbf{0}. \end{cases} \quad (6)$$

Умножим уравнения системы (6) на единичный вектор-столбец:

$$\begin{cases} \mathbf{R}_0(\mathbf{Q} - \rho\mathbf{A})\mathbf{E} - \sigma\mathbf{m}_0\mathbf{E} + \mu\mathbf{R}_1\mathbf{E} = 0, \\ \mathbf{R}_1(\mathbf{Q} - \mu\mathbf{I})\mathbf{E} + \mathbf{R}_0\rho\mathbf{A}\mathbf{E} + \sigma\mathbf{m}_0\mathbf{E} = 0. \end{cases}$$

В результате упрощения получим два одинаковых уравнения:

$$\begin{cases} -\mathbf{R}_0\rho\mathbf{A}\mathbf{E} - \sigma\mathbf{m}_0\mathbf{E} + \mu\mathbf{R}_1\mathbf{E} = 0, \\ -\mathbf{R}_1\mu\mathbf{E} + \mathbf{R}_0\rho\mathbf{A}\mathbf{E} + \sigma\mathbf{m}_0\mathbf{E} = 0. \end{cases}$$

Из этих уравнений можно выразить $\mathbf{m}_0\mathbf{E}$:

$$\sigma\mathbf{m}_0\mathbf{E} = \mu - \mathbf{R}_0(\mu\mathbf{E} + \rho\mathbf{A}\mathbf{E}). \quad (7)$$

Объединив первое уравнение системы (6) и уравнение (7), получаем систему для определения вектора $\mathbf{m}_0$ (при условии, что будут найдены векторы $\mathbf{R}_0$ и $\mathbf{R}_1$):

$$\begin{cases} \sigma\mathbf{m}_0 = \mathbf{R}_0(\mathbf{Q} - \rho\mathbf{A}) + \mu\mathbf{R}_1, \\ \sigma\mathbf{m}_0\mathbf{E} = \mu - \mathbf{R}_0(\mu\mathbf{E} + \rho\mathbf{A}\mathbf{E}). \end{cases} \quad (8)$$

Этап 2. Продифференцируем систему (5) по переменной u :

$$\begin{cases} \frac{\partial \mathbf{H}(0,u)}{\partial u}(\mathbf{Q} - \rho\mathbf{\Lambda}) + j\sigma \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2} + \mu \frac{\partial \mathbf{H}(1,u)}{\partial u} = \mathbf{0}, \\ \frac{\partial \mathbf{H}(1,u)}{\partial u}(\mathbf{Q} - \rho\mathbf{\Lambda} - \mu\mathbf{I}) + \frac{\partial \mathbf{H}(0,u)}{\partial u}\rho\mathbf{\Lambda} + \frac{\partial \mathbf{H}(1,u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} + j\mathbf{H}(1,u)\rho\mathbf{\Lambda}e^{ju} + \\ + j^2\sigma e^{-ju} \frac{\partial \mathbf{H}(0,u)}{\partial u} - j\sigma e^{-ju} \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2} = \mathbf{0}. \end{cases} \quad (9)$$

Обозначим $\mathbf{d}_k = j^2 \frac{\partial^2 \mathbf{H}(k,u)}{\partial u^2} \Big|_{u=0}$, где $k=1,2$, тогда второй момент распределения вероятностей

числа заявок в ИПВ вычисляется как $d = (\mathbf{d}_0 + \mathbf{d}_1)\mathbf{E} = \mathbf{dE}$.

Примем $u = 0$ в системе (9):

$$\begin{cases} \mathbf{m}_0(\mathbf{Q} - \rho\mathbf{\Lambda}) - \sigma\mathbf{d}_0 + \mu\mathbf{m}_1 = \mathbf{0}, \\ \mathbf{m}_1(\mathbf{Q} - \mu\mathbf{I}) + \mathbf{m}_0\rho\mathbf{\Lambda} + \mathbf{R}_1\rho\mathbf{\Lambda} - \sigma\mathbf{m}_0 + \sigma\mathbf{d}_0 = \mathbf{0}. \end{cases} \quad (10)$$

Сложив уравнения системы (10), получим

$$(\mathbf{m}_0 + \mathbf{m}_1)\mathbf{Q} + \mathbf{R}_1\rho\mathbf{\Lambda} - \sigma\mathbf{m}_0 = \mathbf{0}. \quad (11)$$

Умножим уравнение (11) на единичный вектор-столбец:

$$\sigma\mathbf{m}_0\mathbf{E} = \mathbf{R}\rho\mathbf{\Lambda}\mathbf{E} - \mathbf{R}_0\rho\mathbf{\Lambda}\mathbf{E}.$$

Учитывая условие (2), имеем следующее выражение:

$$\sigma\mathbf{m}_0\mathbf{E} = \rho\mu - \rho\mathbf{R}_0\mathbf{\Lambda}\mathbf{E}. \quad (12)$$

Приравняем (12) к (7):

$$\rho\mu - \rho\mathbf{R}_0\mathbf{\Lambda}\mathbf{E} = \mu - \mathbf{R}_0(\mu\mathbf{E} + \rho\mathbf{\Lambda}\mathbf{E}).$$

Отсюда несложно получить, что выполняются следующие равенства:

$$\begin{cases} \mathbf{R}_0\mathbf{E} = 1 - \rho, \\ \mathbf{R}_1\mathbf{E} = \rho. \end{cases} \quad (13)$$

Умножим уравнения системы (10) на единичный вектор-столбец:

$$\begin{cases} -\rho\mathbf{m}_0\mathbf{\Lambda}\mathbf{E} - \sigma\mathbf{d}_0\mathbf{E} + \mu\mathbf{m}_1\mathbf{E} = 0, \\ -\mu\mathbf{m}_1\mathbf{E} + \rho\mathbf{m}_0\mathbf{\Lambda}\mathbf{E} + \rho\mathbf{R}_1\mathbf{\Lambda}\mathbf{E} - \sigma\mathbf{m}_0\mathbf{E} + \sigma\mathbf{d}_0\mathbf{E} = 0. \end{cases}$$

В итоге некоторых преобразований получим дополнительное уравнение

$$\sigma\mathbf{d}_0\mathbf{E} = \mu\mathbf{m}_1\mathbf{E} - \rho\mathbf{m}_0\mathbf{\Lambda}\mathbf{E}. \quad (14)$$

Этап 3. Продифференцируем систему (9):

$$\begin{cases} \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2}(\mathbf{Q} - \rho\mathbf{\Lambda}) + j\sigma \frac{\partial^3 \mathbf{H}(0,u)}{\partial u^3} + \mu \frac{\partial^2 \mathbf{H}(1,u)}{\partial u^2} = \mathbf{0}, \\ \frac{\partial^2 \mathbf{H}(1,u)}{\partial u^2}(\mathbf{Q} - \rho\mathbf{\Lambda} - \mu\mathbf{I}) + \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2}\rho\mathbf{\Lambda} + j \frac{\partial \mathbf{H}(1,u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} + \frac{\partial^2 \mathbf{H}(1,u)}{\partial u^2}\rho\mathbf{\Lambda}e^{ju} + j^2\mathbf{H}(1,u)\rho\mathbf{\Lambda}e^{ju} + \\ + j \frac{\partial \mathbf{H}(1,u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} - j^3\sigma e^{-ju} \frac{\partial \mathbf{H}(0,u)}{\partial u} + j^2\sigma e^{-ju} \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2} + j^2\sigma e^{-ju} \frac{\partial^2 \mathbf{H}(0,u)}{\partial u^2} - j\sigma e^{-ju} \frac{\partial^3 \mathbf{H}(0,u)}{\partial u^3} = \mathbf{0}. \end{cases} \quad (15)$$

Обозначим вектор-строки $\mathbf{e}_k = -j^3 \frac{\partial^3 \mathbf{H}(k, u)}{\partial u^3} \Big|_{u=0}$, где $k=1, 2$. Аналогично предыдущему этапу,

примем $u = 0$ в системе (15):

$$\begin{cases} \mathbf{d}_0(\mathbf{Q} - \rho\mathbf{\Lambda}) - \sigma\mathbf{e}_0 + \mu\mathbf{d}_1 = \mathbf{0}, \\ \mathbf{d}_1(\mathbf{Q} - \mu\mathbf{I}) + \mathbf{d}_0\rho\mathbf{\Lambda} + 2\mathbf{m}_1\rho\mathbf{\Lambda} + \mathbf{R}_1\rho\mathbf{\Lambda} + \sigma\mathbf{m}_0 - 2\sigma\mathbf{d}_0 + \sigma\mathbf{e}_0 = \mathbf{0}. \end{cases} \quad (16)$$

Сложив уравнения системы (16), получим следующее равенство:

$$(\mathbf{d}_0 + \mathbf{d}_1)\mathbf{Q} - 2\sigma\mathbf{d}_0 + 2\mathbf{m}_1\rho\mathbf{\Lambda} + \mathbf{R}_1\rho\mathbf{\Lambda} + \sigma\mathbf{m}_0 = \mathbf{0}.$$

Умножим полученное уравнение на единичный вектор-столбец.

$$2\sigma\mathbf{d}_0\mathbf{E} = 2\mathbf{m}_1\rho\mathbf{\Lambda}\mathbf{E} + \mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} + \sigma\mathbf{m}_0\mathbf{E}.$$

Подставляя в последнее уравнение выражение (14), можно получить следующее скалярное уравнение:

$$\mathbf{m}_1(\mu\mathbf{E} - \rho\mathbf{\Lambda}\mathbf{E}) = \frac{1}{2}\mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} + \frac{1}{2}\mathbf{m}_0(2\rho\mathbf{\Lambda}\mathbf{E} + \sigma\mathbf{E}). \quad (17)$$

Из матричного уравнения (11) и скалярного уравнения (17) получаем систему для определения вектора $\mathbf{m}_1$:

$$\begin{cases} \mathbf{m}_1\mathbf{Q} = \sigma\mathbf{m}_0 - \mathbf{m}_0\mathbf{Q} - \mathbf{R}_1\rho\mathbf{\Lambda}, \\ \mathbf{m}_1(\mu\mathbf{E} - \rho\mathbf{\Lambda}\mathbf{E}) = \frac{1}{2}\mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} + \frac{1}{2}\mathbf{m}_0(2\rho\mathbf{\Lambda}\mathbf{E} + \sigma\mathbf{E}). \end{cases} \quad (18)$$

Так как векторы $\mathbf{m}_1$ и $\mathbf{d}_0$ связаны соотношением (14), то из (10) для определения вектора $\mathbf{d}_0$ получаем систему

$$\begin{cases} \sigma\mathbf{d}_0 = \mathbf{m}_0(\mathbf{Q} - \rho\mathbf{\Lambda}) + \mu\mathbf{m}_1, \\ \sigma\mathbf{d}_0\mathbf{E} = \mu\mathbf{m}_1\mathbf{E} - \rho\mathbf{m}_0\mathbf{\Lambda}\mathbf{E}. \end{cases} \quad (19)$$

Этап 4. Немного преобразуем систему (15):

$$\begin{cases} \frac{\partial^2 \mathbf{H}(0, u)}{\partial u^2}(\mathbf{Q} - \rho\mathbf{\Lambda}) + j\sigma \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3} + \mu \frac{\partial^2 \mathbf{H}(1, u)}{\partial u^2} = \mathbf{0}, \\ \frac{\partial^2 \mathbf{H}(1, u)}{\partial u^2}(\mathbf{Q} - \rho\mathbf{\Lambda} - \mu\mathbf{I}) + \frac{\partial^2 \mathbf{H}(0, u)}{\partial u^2}\rho\mathbf{\Lambda} + 2j \frac{\partial \mathbf{H}(1, u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} + \frac{\partial^2 \mathbf{H}(1, u)}{\partial u^2}\rho\mathbf{\Lambda}e^{ju} + \\ + j^2 \mathbf{H}(1, u)\rho\mathbf{\Lambda}e^{ju} - j^3 \sigma e^{-ju} \cdot \frac{\partial \mathbf{H}(0, u)}{\partial u} + 2j^2 \sigma e^{-ju} \cdot \frac{\partial^2 \mathbf{H}(0, u)}{\partial u^2} - j\sigma e^{-ju} \cdot \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3} = \mathbf{0}. \end{cases}$$

Продифференцируем полученную систему:

$$\begin{cases} \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3}(\mathbf{Q} - \rho\mathbf{\Lambda}) + j\sigma \frac{\partial^4 \mathbf{H}(0, u)}{\partial u^4} + \mu \frac{\partial^3 \mathbf{H}(1, u)}{\partial u^3} = \mathbf{0}, \\ \frac{\partial^3 \mathbf{H}(1, u)}{\partial u^3}(\mathbf{Q} - \rho\mathbf{\Lambda} - \mu\mathbf{I}) + \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3}\rho\mathbf{\Lambda} + 2j \frac{\partial^2 \mathbf{H}(1, u)}{\partial u^2}\rho\mathbf{\Lambda}e^{ju} + 2j^2 \frac{\partial \mathbf{H}(1, u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} + \frac{\partial^3 \mathbf{H}(1, u)}{\partial u^3}\rho\mathbf{\Lambda}e^{ju} + \\ + j \frac{\partial^2 \mathbf{H}(1, u)}{\partial u^2}\rho\mathbf{\Lambda}e^{ju} + j^3 \mathbf{H}(1, u)\rho\mathbf{\Lambda}e^{ju} + j^2 \frac{\partial \mathbf{H}(1, u)}{\partial u}\rho\mathbf{\Lambda}e^{ju} + j^4 \sigma e^{-ju} \frac{\partial \mathbf{H}(0, u)}{\partial u} - j^3 \sigma e^{-ju} \frac{\partial^2 \mathbf{H}(0, u)}{\partial u^2} - \\ - 2j^3 \sigma e^{-ju} \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3} + 2j^2 \sigma e^{-ju} \frac{\partial^4 \mathbf{H}(0, u)}{\partial u^4} + j^2 \sigma e^{-ju} \frac{\partial^3 \mathbf{H}(0, u)}{\partial u^3} - j\sigma e^{-ju} \frac{\partial^4 \mathbf{H}(0, u)}{\partial u^4} = \mathbf{0}. \end{cases} \quad (20)$$

Обозначим моменты 4-го порядка следующим образом: $\mathbf{g}_k = j^4 \frac{\partial^4 \mathbf{H}(k, u)}{\partial u^4} \Big|_{u=0}$, где $k = 1, 2$. Примем

$u = 0$ в системе (20):

$$\begin{cases} \mathbf{e}_0(\mathbf{Q} - \rho\mathbf{\Lambda}) - \sigma\mathbf{g}_0 + \mu\mathbf{e}_1 = \mathbf{0}, \\ \mathbf{e}_1(\mathbf{Q} - \mu\mathbf{I}) + \mathbf{e}_0\rho\mathbf{\Lambda} + 3\mathbf{m}_1\rho\mathbf{\Lambda} + 3\mathbf{d}_1\rho\mathbf{\Lambda} + \mathbf{R}_1\rho\mathbf{\Lambda} - \sigma\mathbf{m}_0 + 3\sigma\mathbf{d}_0 - 3\sigma\mathbf{e}_0 + \sigma\mathbf{g}_0 = \mathbf{0}. \end{cases} \quad (21)$$

Сложив уравнения системы (21), получим

$$(\mathbf{e}_0 + \mathbf{e}_1)\mathbf{Q} + 3\mathbf{m}_1\rho\mathbf{\Lambda} + 3\mathbf{d}_1\rho\mathbf{\Lambda} + \mathbf{R}_1\rho\mathbf{\Lambda} - \sigma\mathbf{m}_0 + 3\sigma\mathbf{d}_0 - 3\sigma\mathbf{e}_0 = \mathbf{0}. \quad (22)$$

Умножим уравнение (22) на единичный вектор-столбец:

$$3\sigma\mathbf{e}_0\mathbf{E} = 3\mathbf{m}_1\rho\mathbf{\Lambda}\mathbf{E} + 3\mathbf{d}_1\rho\mathbf{\Lambda}\mathbf{E} + \mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} - \sigma\mathbf{m}_0\mathbf{E} + 3\sigma\mathbf{d}_0\mathbf{E}. \quad (23)$$

Подставим первое уравнение системы (15) в выражение (23):

$$3\mu\mathbf{d}_1\mathbf{E} - 3\mathbf{d}_1\rho\mathbf{\Lambda}\mathbf{E} = 3\mathbf{d}_0\rho\mathbf{\Lambda}\mathbf{E} + 3\mathbf{m}_1\rho\mathbf{\Lambda}\mathbf{E} + \mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} - \sigma\mathbf{m}_0\mathbf{E} + 3\sigma\mathbf{d}_0\mathbf{E}.$$

Отсюда получим следующее уравнение:

$$\mathbf{d}_1(\mu\mathbf{E} - \rho\mathbf{\Lambda}\mathbf{E}) = \mathbf{d}_0(\rho\mathbf{\Lambda}\mathbf{E} + \sigma\mathbf{E}) + \mathbf{m}_1\rho\mathbf{\Lambda}\mathbf{E} + \frac{1}{3}\mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} - \frac{1}{3}\sigma\mathbf{m}_0\mathbf{E}. \quad (24)$$

Объединяя уравнения (16) и (24), составим систему для вычисления вектора $\mathbf{d}_1$:

$$\begin{cases} \mathbf{d}_1\mathbf{Q} = -\mathbf{d}_0\mathbf{Q} + 2\sigma\mathbf{d}_0 - 2\mathbf{m}_1\rho\mathbf{\Lambda} - \mathbf{R}_1\rho\mathbf{\Lambda} - \sigma\mathbf{m}_0, \\ \mathbf{d}_1(\rho\mathbf{\Lambda}\mathbf{E} + \mu\mathbf{E}) = \mathbf{d}_0(\rho\mathbf{\Lambda}\mathbf{E} + \sigma\mathbf{E}) - \mathbf{m}_1\rho\mathbf{\Lambda}\mathbf{E} - \frac{1}{3}\mathbf{R}_1\rho\mathbf{\Lambda}\mathbf{E} - \frac{1}{3}\sigma\mathbf{m}_0\mathbf{E}. \end{cases} \quad (25)$$

Таким образом, мы получили формулы (8), (18), (19) и (25) для вычисления компонент математического ожидания и второго начального момента распределения вероятностей числа заявок в ИПВ.

3. Квазиточные моменты

Для использования формул (8), (18), (19) и (25) необходимо знать векторы $\mathbf{R}_0$ и $\mathbf{R}_1$ в явном виде. В ходе вычислений мы получили лишь соотношение (13):

$$\begin{cases} \mathbf{R}_0\mathbf{E} = 1 - \rho, \\ \mathbf{R}_1\mathbf{E} = \rho. \end{cases}$$

Кроме того, из начального описания системы и введенных обозначений известно

$$\begin{cases} (\mathbf{R}_0 + \mathbf{R}_1)\mathbf{Q} = \mathbf{0}, \\ \mathbf{R}_0\mathbf{E} + \mathbf{R}_1\mathbf{E} = 1. \end{cases}$$

Очевидно, что при дальнейшем использовании метода моментов к системе (5), т.е. при получении формул для моментов высших порядков, дополнительной информации о векторах $\mathbf{R}_0$ и $\mathbf{R}_1$ получить невозможно. А с помощью указанных систем уравнений однозначно векторы $\mathbf{R}_0$ и $\mathbf{R}_1$ определить нельзя.

В связи с этим предлагается аппроксимировать векторы $\{\mathbf{R}_0, \mathbf{R}_1\}$ векторами, пропорциональными вектору $\mathbf{R}$. Тогда несложно показать, что они определяются следующим образом:

$$\begin{cases} \mathbf{R}_0 = (1 - \rho) \cdot \mathbf{R}, \\ \mathbf{R}_1 = \rho \cdot \mathbf{R}. \end{cases} \quad (26)$$

По определению, векторы $\{\mathbf{R}_0, \mathbf{R}_1\}$ – двумерное совместное стационарное распределение вероятностей состояний ММРР-потока и состояний прибора. Поэтому такой вид векторов $\mathbf{R}_0$ и $\mathbf{R}_1$ возможен в

предположении о независимости распределений вероятностей состояний прибора и состояний цепи Маркова, управляющей ММРР-поток.

4. Численный анализ

Для определения точности предлагаемой аппроксимации проведем численное сравнение полученных моментов с их точными значениями.

Возьмем параметры системы следующими: $\mu = 1$,

$$Q = \begin{pmatrix} -0,5 & 0,4 & 0,1 \\ 0,2 & -0,5 & 0,3 \\ 0,1 & 0,2 & -0,3 \end{pmatrix}, \Lambda = \begin{pmatrix} 0,780 & 0 & 0 \\ 0 & 0,936 & 0 \\ 0 & 0 & 1,170 \end{pmatrix}.$$

Будем исследовать зависимость результатов вычислений от значений параметра загрузки ρ и задержки σ .

В табл. 1 представим относительную погрешность численного сравнения значений математического ожидания и дисперсии, вычисленных с помощью описанных выше формул и полученных с помощью численного алгоритма.

Замечание. Численный алгоритм имеет естественные ограничения для числа заявок в ИПВ $i \geq 150$, так как это соответствует необходимости решения системы линейных алгебраических уравнений (4) размерностью до 1 000. В таких случаях провести сравнение не представляется возможным, поэтому в таблице присутствуют пустые ячейки.

Т а б л и ц а 1

Относительная погрешность вычислений

ρ	$\sigma = 0,1$		$\sigma = 0,5$		$\sigma = 1$		$\sigma = 2$		$\sigma = 10$	
	M	D	M	D	M	D	M	D	M	D
$\rho = 0,1$	0,010	0,122	<0,001	0,046	0,007	0,073	0,015	0,089	0,026	0,106
$\rho = 0,3$	0,007	0,079	<0,001	0,020	0,006	0,033	0,012	0,041	0,022	0,046
$\rho = 0,5$	0,004	0,044	<0,001	0,006	0,004	0,013	0,009	0,015	0,017	0,016
$\rho = 0,7$	0,002	0,018	<0,001	0,001	0,002	0,003	0,005	0,004	0,010	0,003
$\rho = 0,8$	<0,001	0,015	<0,001	<0,001	0,001	<0,001	0,003	0,002	0,007	0,001
$\rho = 0,9$			0,001	0,010	0,001	0,002	0,001	0,001	0,004	0,004

Представим сравнение (табл. 2) значений математического ожидания и дисперсии для следующих параметров системы: $\mu = 1$,

$$Q = \begin{pmatrix} -0,5 & 0,45 & 0,05 \\ 0,28 & -0,3 & 0,02 \\ 0,5 & 0 & -0,5 \end{pmatrix}, \Lambda = \begin{pmatrix} 0,594 & 0 & 0 \\ 0 & 1,187 & 0 \\ 0 & 0 & 1,781 \end{pmatrix}.$$

Т а б л и ц а 2

Относительная погрешность вычислений

ρ	$\sigma = 0,1$		$\sigma = 0,5$		$\sigma = 1$		$\sigma = 2$		$\sigma = 10$	
	M	D	M	D	M	D	M	D	M	D
$\rho = 0,1$	0,042	0,653	0,013	0,049	0,008	0,127	0,029	0,160	0,061	0,178
$\rho = 0,3$	0,029	0,401	0,009	0,010	0,006	0,051	0,022	0,066	0,048	0,067
$\rho = 0,5$	0,017	0,210	0,006	0,004	0,004	0,017	0,015	0,022	0,035	0,019
$\rho = 0,7$	0,007	0,080	0,003	0,004	0,002	0,004	0,008	0,005	0,020	0,002
$\rho = 0,8$	0,003	0,049	0,001	0,009	0,001	0,002	0,005	0,002	0,013	<0,001
$\rho = 0,9$			<0,001	0,006	0,003	0,023	0,002	0,002	0,007	0,004

Для случая входящего простейшего потока известны точные аналитические формулы вычисления первого и второго начальных моментов:

$$m = \frac{\rho^2}{1-\rho} \left(\frac{\mu}{\sigma} + 1 \right), \quad (27)$$

$$d = \frac{\rho^2}{(1-\rho)^2} \left(1 + \rho + \frac{\mu}{\sigma} + 2\rho^2 \frac{\mu}{\sigma} + \rho^2 \left(\frac{\mu}{\sigma} \right)^2 \right). \quad (28)$$

Для получения простейшего входящего потока в рассматриваемой системе положим элементы матрицы условных интенсивностей Λ одинаковыми или размерность матриц Λ и Q примем равной 1. Однако в этом случае несложно показать, что формулы (8), (18), (19) и (25) не зависят от R_1 и R_0 и совпадут с выражениями (27), (28).

Анализ табл. 1 и 2 позволяет сделать вывод, что предложенные формулы достаточно близки к точным для большей области изменений значений параметров системы. В частности, для значений интенсивности задержки заявки в ИПВ, близких к интенсивности обслуживания, абсолютная погрешность вычислений математического ожидания не превышает 3%. Таким образом, предложенные формулы вычисления первого и второго моментов распределения вероятностей числа заявок в источнике повторных вызовов можно назвать почти точными или «квазиточными», так же как и сами значения моментов.

Заключение

Системы массового обслуживания с повторными вызовами, рассматриваемые в данной статье, являются математическими моделями реальных процессов, возникающих в телекоммуникационных системах. В связи с этим научные результаты, полученные в этой области, имеют как теоретическую, так и практическую ценность для развития информационно-коммуникационных технологий и, как следствие, имеют большую значимость для экономики и социальной сферы.

В работе проведено исследование RQ-системы ММРР[М]1 методом моментов. Найдены приближенные формулы для вычисления первого и второго моментов распределения вероятностей числа заявок в источнике повторных вызовов. Сделано предположение о независимости распределений состояний цепи Маркова, управляющего ММРР-потоком и состояний прибора. Проведенный численный анализ показал, что в таком случае моменты достаточно близки к точным для широкого спектра значений параметров системы.

Таким образом, вычисленные предлагаемым способом моменты можно назвать «квазиточными». В дальнейшем исследовании они могут быть применимы для построения аппроксимирующих распределений (например, гауссовского, гамма-распределения или геометрического).

ЛИТЕРАТУРА

1. *Falin G.L., Templeton J.G.C.* Retrial queues. London : Chapman & Hall, 1997. 328 p.
2. *Artalejo J.R., Gomez-Corral A.* Retrial Queueing Systems: A Computational Approach. Berlin : Springer, 2008. 267 p.
3. *Artalejo J.R., Falin G.I.* Standard and retrial queueing systems: A comparative analysis // *Revista Matemática Complutense*. 2002. V. 15. P. 101–129.
4. *Wilkinson R.I.* Theories for toll traffic engineering in the USA // *The Bell System Technical Journal*. 1956. V. 35, No. 2. P. 421–507.
5. *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.
6. *Gosztory G.* Repeated call attempts and their effect on traffic engineering // *Budavox Telecommunication Review*. 1976. No. 2. P. 16–26.
7. *Elldin A., Lind G.* Elementary Telephone Traffic Theory. Ericsson Public Telecommunications, 1971.
8. *Jonin G.L., Sedol J.J.* Telephone systems with repeated calls // *Proc. of the 6th International Teletraffic Congress*. Munich, 1970. P. 435/1–5.
9. *Степанов С.Н.* Численные методы расчета систем с повторными вызовами. М. : Наука, 1983. 230 с.
10. *Neuts M.F., Rao B.M.* Numerical investigation of a multiserver retrial model // *Queueing Systems*. 1990. V. 7. P. 169–190.
11. *Ridder F.* Fast simulation of retrial queues // *Third Workshop on Rare Event Simulation and Related Combinatorial Optimization Problems*. Pisa, 2000. P. 1–5.
12. *Dudin A.N., Klimenok V.I.* Queueing System BMAP/G/1 with repeated calls // *Mathematical and Computer Modelling*. 1999. V. 30, No. 3–4. P. 115–128.

13. Artalejo J.R., Gomez-Corral A., Neuts M.F. Analysis of multiserver queues with constant retrial rate // European Journal of Operational Research. 2001. V. 135. P. 569–581.
14. Diamond J.E., Alfa A.S. Matrix analytical methods for M/PH/1 retrial queues // Stochastic Models. 1995. V. 11. P. 447–470.
15. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск : Изд-во НТЛ, 2006. 112 с.
16. Моисеева Е.А., Назаров А.А. Исследование RQ-системы MMPP|GI|1 методом асимптотического анализа // Вестник ТГУ. УВТИ. 2013. № 4. С. 84–94.
17. Фалин Г.И. Асимптотические свойства распределения числа требований в системе типа m/g/1/∞ с повторными вызовами // ВИНТИ. 1983. № 5418-83.
18. Anisimov V.V. Asymptotic analysis of highly reliable retrial systems with finite capacity // Queues, Flows, Systems, Networks. Proc. of the International Conference «Modern Mathematical Methods of Investigating the Telecommunicational Networks». Minsk, 1999. P. 7–12.

Фёдорова Екатерина Александровна. E-mail: moiskate@mail.ru
Томский государственный университет

Поступила в редакцию 2 сентября 2014 г.

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

Calculation of moments in retrial queueing system MMPP|M|1.

Key words: retrial queueing system, orbit, method of moments

In the paper, we investigate a retrial queueing system with the input Markovian Modulated Poisson Process (MMPP) which is defined by matrixes $\rho\lambda$ and Q , and the service time of each call is distributed by the exponential law with rate μ .

We introduce the following denotations: $i(t)$ is the process defining the call number in the orbit, $n(t)$ is the underlying Markov chain of the input MMP-process, and $k(t)$ describes states of the server.

Stochastic multidimensional process $\{k(t), i(t), n(t)\}$ is a Markov process. To obtain the probability distribution $P(k, i, n, t)$, the system of the Kolmogorov differentiate equations is composed. This system is written in matrix form at stationary regime. Then, it is rewritten with making use of characteristic functions and the obtained system is studied by the method of moments.

Using mathematical transforms, the following systems for components of the mean and the second order initial moment of probability distribution of calls number in the orbit are obtained:

$$\begin{cases} \sigma m_0 = R_0(Q - \rho\Lambda) + \mu R_1, \\ \sigma m_0 E = \mu - R_0(\mu E + \rho\Lambda E), \end{cases}$$

$$\begin{cases} m_1 Q = \sigma m_0 - m_0 Q - R_1 \rho\Lambda, \\ m_1(\mu E - \rho\Lambda E) = \frac{1}{2} R_1 \rho\Lambda E + \frac{1}{2} m_0(2\rho\Lambda E + \sigma E), \end{cases}$$

$$\begin{cases} \sigma d_0 = m_0(Q - \rho\Lambda) + \mu m_1, \\ \sigma d_0 E = \mu m_1 E - \rho m_0 \Lambda E, \end{cases}$$

$$\begin{cases} d_1 Q = -d_0 Q + 2\sigma d_0 - 2m_1 \rho\Lambda - R_1 \rho\lambda - \sigma m_0, \\ d_1(\rho\Lambda E + \mu E) = d_0(\rho\Lambda E + \sigma E) - m_1 \rho\Lambda E - \frac{1}{3} R_1 \rho\Lambda E - \frac{1}{3} \sigma m_0 E. \end{cases}$$

However, the form of vectors $\{R_0, R_1\}$, which is two-dimensional joint stationary probability distribution of the MMP-process and the server states, is unknown. So, we assume the independence of the MMP-process and the server states distributions, and the vectors $\{R_0, R_1\}$ have the following form:

$$\begin{cases} R_0 = (1 - \rho) \cdot R, \\ R_1 = \rho \cdot R. \end{cases}$$

The analysis of the results shows that moments calculated through obtained formulas are sufficiently close to exact ones obtained by numerical methods. Thus, the moments calculated by proposed way can be called «quasiexact». The formulas for moments calculation can be used in future researching by composing approximate distributions and also for other practical problems.

REFERENCES

1. Falin G.L., Templeton J.G.C. *Retrial queues*. London: Chapman & Hall, 1997. 328 p.
2. Artalejo J.R., Gomez-Corral A. *Retrial Queueing Systems: A Computational Approach*. Berlin: Springer, 2008. 267 p.
3. Artalejo J.R., Falin G.I. Standard and retrial queueing systems: A comparative analysis. *Revista Matemática Complutense*, 2002, vol. 15, pp. 101-129.
4. Wilkinson R.I. Theories for toll traffic engineering in the USA. *The Bell System Technical Journal*, 1956, vol. 35, no. 2, pp. 421-507. DOI: 10.1002/j.1538-7305.1956.tb02388.x

5. Cohen J.W. Basic problems of telephone traffic and the influence of repeated calls. *Philips Telecommunication Review*, 1957, vol. 18, no. 2, pp. 49-100.
6. Gosztonyi G. Repeated call attempts and their effect on traffic engineering. *Budavox Telecommunication Review*, 1976, no. 2, pp. 16-26. DOI: 10.1109/MILCOM.2006.302071
7. Elldin A., Lind G. *Elementary Telephone Traffic Theory*. Ericsson Public Telecommunications, 1971.
8. Jonin G.L., Sedol J.J. Telephone systems with repeated calls. *Proc. of the 6th International Teletraffic Congress*. Munich, 1970, pp. 435/1-5.
9. Stepanov S.N. *Chislennyye metody rascheta sistem s povtornymi vyzovami* [Numerical methods of calculation of retrial queues]. Moscow: Nauka Publ., 1983. 230 p.
10. Neuts M.F., Rao B.M. Numerical investigation of a multiserver retrial model. *Queueing Systems*, 1990, vol.7, pp. 169-190. DOI: 10.1007/BF01158473
11. Ridder F. Fast simulation of retrial queues. *Third Workshop on Rare Event Simulation and Related Combinatorial Optimization Problems*. Pisa, 2000, pp. 1-5.
12. Dudin A.N., Klimenok V.I. Queueing System BMAP/G/1 with repeated calls. *Mathematical and Computer Modelling*, 1999, vol. 30, no. 3-4, pp. 115-128. DOI: 10.1016/S0895-7177(99)00136-3
13. Artalejo J.R., Gomez-Corral A., M.F. Neuts. Analysis of multiserver queues with constant retrial rate. *European Journal of Operational Research*, 2001, vol.135, pp. 569-581. DOI: 10.1016/S0377-2217(00)00330-1
14. Diamond J.E., Alfa A.S. Matrix analytical methods for M/PH/1 retrial queues. *Stochastic Models*, 1995, vol. 11, pp. 447-470. DOI: 10.1080/15326349508807355
15. Nazarov A.A., Moiseeva S.P. *Metod asimptoticheskogo analiza v teorii massovogo obsluzhivaniya* [Asymptotic analysis method on queueing theory]. Tomsk: NTL Publ., 2006, 112 p.
16. Moiseeva E.A., Nazarov A.A. 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*, 2013, no. 4(25), pp. 84-94. (In Russian).
17. Falin G.I. Asimptoticheskie svoystva raspredeleniya chisla trebovaniy v sisteme tipa $m/g/1/\infty$ s povtornymi vyzovami [Asymptotic properties of probability distribution of the number of request in system $m/g/1/\infty$ with repeated calls]. *VINITI*, 1983, no. 5418-83. (In Russian).
18. Anisimov V.V. Asymptotic analysis of highly reliable retrial systems with finite capacity. *Queues, Flows, Systems, Networks. Proc. of the International Conference "Modern Mathematical Methods of Investigating the Telecommunicational Networks"*. Minsk, 1999, pp. 7-12.

Е.В. Чимитова, Е.С. Четвертакова

**ПОСТРОЕНИЕ ДЕГРАДАЦИОННОЙ ГАММА-МОДЕЛИ
С УЧЕТОМ ВЛИЯНИЯ ОБЪЯСНЯЮЩИХ ПЕРЕМЕННЫХ**

*Исследование выполнено при поддержке Министерства образования и науки Российской Федерации
в рамках государственного задания №2014/138 (проект № 1689).*

Рассматривается деградационная модель надежности с учетом влияния объясняющих переменных, в основе которой лежит предположение о принадлежности независимых приращений случайного процесса изменения показателя деградации гамма-распределению. Предлагается подход к проверке статистической гипотезы о согласии с деградационной гамма-моделью с использованием непараметрических критериев типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга. С использованием методов имитационного моделирования проводится исследование распределений статистик и мощности рассматриваемых критериев относительно различных видов конкурирующих гипотез.

Ключевые слова: деградационная гамма-модель; функция надёжности; ускоренные испытания; метод максимального правдоподобия; критерии согласия.

В современном мире большое внимание уделяется вопросам контроля качества и исследования надежности технических устройств, особенно если от их работоспособности зависит жизнь и здоровье человека. Если речь идет о высоконадежных изделиях, то данных об отказах таких изделий может быть недостаточно для оценки функции надежности, поскольку в период проведения эксперимента наступление отказов наблюдается крайне редко. Существует два возможных способа получить дополнительную информацию о надежности изделий: первый заключается в проведении ускоренных испытаний, когда изделия подвергаются повышенным нагрузкам, в результате чего отказы наступают раньше; второй способ состоит в измерении значений некоторого показателя, характеризующего процесс деградации (старения) изделия. При этом момент времени, когда значение деградационного показателя достигает критического уровня, считается временем наступления отказа. Оба подхода можно совместить, наблюдая процессы деградации и наступление отказов изделий, эксплуатирующихся при повышенных нагрузках. В качестве нагрузок могут выступать температура, давление, напряжение, механические нагрузки и др.

Проанализировав полученные данные о деградации изделия, для получения оценки надежности (прогноза) и проведения дальнейших исследований необходимо построить деградационную модель. При построении моделей деградации учитывается распределение и функция тренда приращений показателя старения, а также функция влияния объясняющих переменных – функция от ковариат. В большинстве работ по исследованиям деградационных процессов в качестве распределения приращений показателя старения рассматриваются либо гамма-распределение (гамма-процесс) [1], либо нормальное распределение (винеровский процесс) [2,3]. Это обусловлено тем, что данные распределения являются устойчивыми относительно суммирования (воспроизводимыми), и за счет этого можно легко определить распределение исследуемой случайной величины – показателя деградации в некоторый момент времени, а затем оценить вероятность безотказной работы.

В настоящей статье рассматриваются вопросы построения деградационной гамма-модели как наиболее часто используемой при описании реальных данных. Так, например, в [4] сравниваются деградационные гамма- и винеровская модели на примере исследования надежности арсенид-галлиевых лазеров; авторы [5] используют деградационную гамма-модель для описания износа автомобильных тормозных колодок; в [6] рассматривается пример моделирования зависимости износа автомобильной шины от нагрузок с использованием гамма-модели. Различные виды деградационных моделей на основе гамма-распределения представлены в работе [7].

Основной проблемой использования деградационной гамма-модели является отсутствие математического аппарата для проверки статистической гипотезы о виде модели, в то время как проверка гипотезы о согласии является обязательным этапом построения вероятностных моделей. В данной работе предлагается подход к проверке статистической гипотезы о виде деградационной гамма-модели надежности с учетом влияния объясняющих переменных, предусматривающий исследование методами компьютерного моделирования распределений статистик критериев согласия в интерактивном режиме проверки гипотезы. В качестве критериев согласия предлагается использовать непараметрические критерии типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга. Приводится пример построения деградационной гамма-модели надежности по данным об углеродистых резисторах.

1. Деградационная гамма-модель надежности

Случайный процесс $Z(t)$, характеризующий процесс деградации исследуемых изделий, называется деградационным гамма-процессом с параметром формы $v(t)$ и параметром масштаба σ , если

- 1) $Z(0) = 0$;
- 2) $Z(t)$ является случайным процессом с независимыми приращениями;
- 3) приращения $\Delta Z(t) = Z(t + \Delta t) - Z(t)$ подчиняются гамма-распределению с функцией плотности

$$f_{Gamma}(t; \sigma, \Delta v(t)) = \left(\frac{t}{\sigma}\right)^{\Delta v(t)-1} \frac{e^{-t/\sigma}}{\sigma \cdot \Gamma(\Delta v(t))},$$

где $\Delta v(t) = v(t + \Delta t) - v(t)$ – параметр формы и $\sigma > 0$ – параметр масштаба.

Пусть деградационный процесс наблюдается при некоторой постоянной во времени нагрузке (ковариате) x , диапазон значений которой определяется условиями эксперимента и представляет собой отрезок числовой прямой. Влияние ковариаты x на изменение показателя деградации будем учитывать так же, как это делается в модели ускоренных испытаний [8]:

$$Z_x(t) = Z\left(\frac{t}{r(x; \beta)}\right),$$

где $r(x; \beta)$ – положительная функция от ковариат. Существует множество моделей функций от ковариат. Наиболее популярные из них [9]:

- логлинейная модель вида $r(x; \beta) = e^{\beta_0 + \beta_1 x}$ применяется, например, для анализа данных усталости при тестировании различных электронных компонент;
- модель правила мощности в форме $r(x; \beta) = e^{\beta_0 + \beta_1 \ln x}$ используется в случаях, когда воздействием являются напряжение, механическая нагрузка;
- модель Аррениуса вида $r(x; \beta) = e^{\beta_0 + \beta_1 / x}$ применяется, когда в качестве нагрузки выступает температура.

Обозначим условное математическое ожидание случайного процесса $Z_x(t)$ через

$$M(Z_x(t)) = m_x(t),$$

где $m_x(t) = \sigma v\left(\frac{t}{r(x; \beta)}\right)$ – положительно определённая, возрастающая функция. Будем называть ее функцией тренда показателя деградации. В качестве функции тренда могут использоваться такие параметрические модели:

$$1) m_x(t; \beta) = \frac{t}{r(x; \beta)};$$

$$2) m_x(t; \beta, \gamma) = \left(\frac{t}{r(x; \beta)} \right)^{\gamma_0}, \quad \gamma_0 > 0;$$

$$3) m_x(t; \beta, \gamma) = \gamma_0 \left(1 - e^{-\left(\frac{t}{r(x; \beta)} \right)^{\gamma_1}} \right), \quad \gamma_0 > 0, \quad \gamma_1 > 0.$$

Несложно показать, что при выполнении сформулированных предположений случайный процесс $Z_x(t)$ в некоторый фиксированный момент времени $t = t_k$ представляет собой случайную величину, имеющую гамма-распределение с параметром масштаба σ и параметром формы, равным $\frac{m_x(t_k)}{\sigma}$. Время безотказной работы, которое зависит от ковариаты x , представляет собой величину

$$T_x = \sup \{t : Z_x(t) < \tilde{z}\},$$

где $\tilde{z}$ – критическое значение показателя деградации, при достижении которого фиксируется отказ объекта. Тогда функция надёжности для рассматриваемой деградационной гамма-модели принимает вид

$$S_x(t) = P\{T_x > t\} = P\{Z_x(t) < \tilde{z}\} = F_{Gamma}\left(\tilde{z}; \sigma, \frac{m_x(t; \beta, \gamma)}{\sigma}\right). \quad (1)$$

Пусть для каждого из n случайно отобранных из генеральной совокупности объектов известно изменение показателя деградации во времени в виде случайного процесса $Z^i(t)$, а также соответствующая величина нагрузки (ковариаты) x^i , при которой эксплуатировался i -й объект, $i = \overline{1, n}$. Обозначим измерения показателя деградации для i -го объекта через

$$Z^i = \{(0, Z_0^i), (t_1^i, Z_1^i), \dots, (t_{k_i}^i, Z_{k_i}^i)\},$$

где k_i – число измерений деградационного показателя во времени. Без потери общности будем считать, что начальное значение показателя старения $Z_0^i = 0$, $i = \overline{1, n}$.

Обозначим выборку приращений через

$$\mathbf{X}_n = \left\{ \left(\{X_j^1 = Z_j^1 - Z_{j-1}^1, j = \overline{1, k_1}\}, x^1 \right), \dots, \left(\{X_j^n = Z_j^n - Z_{j-1}^n, j = \overline{1, k_n}\}, x^n \right) \right\}.$$

Предполагая, что наблюдаемые случайные процессы $Z_{x^i}^i(t)$, $i = \overline{1, n}$, подчиняются деградационной гамма-модели с математическим ожиданием $m_x(t; \beta, \gamma)$, по выборке $\mathbf{X}_n$ можно оценить неизвестные параметры модели (параметр масштаба σ , параметры функции тренда γ (если таковые присутствуют в модели) и регрессионные параметры β) и построить прогноз времени безотказной работы с заданной вероятностью при заданных значениях ковариаты. Оценка максимального правдоподобия (ОМП) вектора параметров вычисляется в результате максимизации функции правдоподобия:

$$\ln L(\mathbf{X}_n) = \sum_{i=1}^n \sum_{j=1}^{k_i} \ln f_{Gamma}(X_j^i; \sigma, p_j^i) \rightarrow \max_{\sigma, \gamma, \beta}, \quad (2)$$

где $p_j^i = \frac{m_{x^i}(t_j^i; \gamma, \beta) - m_{x^i}(t_{j-1}^i; \gamma, \beta)}{\sigma}$ – параметр формы гамма-распределения.

2. Проверка статистической гипотезы о согласии

Обязательным этапом построения деградационной гамма-модели надёжности (впрочем, как и любой другой вероятностной модели) является проверка статистической гипотезы о согласии:

$$H_0: X_j^i \sim F_{Gamma}(t; \hat{\sigma}, \hat{p}_j^i), i = \overline{1, n}, j = \overline{1, k}, \quad (3)$$

где $\hat{p}_j^i = \frac{m_{x^i}(t_j^i; \hat{\gamma}, \hat{\beta}) - m_{x^i}(t_{j-1}^i; \hat{\gamma}, \hat{\beta})}{\hat{\sigma}}$, символ $\sim$ обозначает то, что случайная величина имеет указанное распределение вероятности.

Для проверки гипотез о согласии по выборкам независимых одинаково распределенных случайных величин существует целый ряд критериев, таких как критерии типа хи-квадрат, непараметрические критерии согласия Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга и многие другие. Однако к выборке приращений X_n классические критерии согласия неприменимы, поскольку элементы данной выборки в общем случае не являются одинаково распределенными.

Введём следующее преобразование приращений деградационного показателя:

$$R_j^i = F_{Gamma}(X_j^i; \hat{\sigma}, \hat{p}_j^i), i = \overline{1, n}, j = \overline{1, k_i}.$$

При справедливости гипотезы H_0

$$R_j^i \sim Uniform(0, 1), i = \overline{1, n}, j = \overline{1, k_i}.$$

Таким образом, задача проверки гипотезы H_0 сводится к проверке гипотезы о равномерном распределении случайных величин $R_j^i, i = \overline{1, n}, j = \overline{1, k_i}$. Для проверки данной гипотезы применим критерий согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга.

Обозначим через $R_{(1)}^* \leq R_{(2)}^* \leq \dots \leq R_{(N)}^*, N = \sum_{i=1}^n k_i$, элементы вариационного ряда, построенного по выборке

$$\mathbf{R}_N = \{R_j^i, i = \overline{1, n}, j = \overline{1, k_i}\}.$$

В критерии типа Колмогорова в качестве расстояния между эмпирическим и теоретическим законами распределения используется статистика с поправкой Большева [10] вида

$$S_k = \frac{6ND_N + 1}{6\sqrt{N}},$$

где $D_N = \max(D_N^+, D_N^-)$, $D_N^+ = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)}^* \right\}$, $D_N^- = \max_{1 \leq i \leq N} \left\{ R_{(i)}^* - \frac{i-1}{N} \right\}$.

В критерии Крамера–Мизеса–Смирнова используется статистика вида

$$S_\omega = N\omega_N^2 = \frac{1}{12N} + \sum_{i=1}^N \left\{ R_{(i)}^* - \frac{2i-1}{2N} \right\}^2,$$

а в критерии типа Андерсона–Дарлинга – статистика в форме

$$S_\Omega = -N - 2 \sum_{i=1}^N \left\{ \frac{2i-1}{2N} \ln R_{(i)}^* + \left(1 - \frac{2i-1}{2N} \right) \ln(1 - R_{(i)}^*) \right\}.$$

Необходимо учитывать, что проверяемая гипотеза является сложной, поскольку неизвестные параметры модели оцениваются по тем же данным, по которым проверяется согласие. При проверке сложных гипотез условные распределения данных статистик зависят от ряда факторов: метода оценивания параметров, типа оцениваемого параметра и числа оцениваемых параметров, а в случае гамма-распределения – от конкретного значения параметра формы [10]. Поэтому для вычисления достигнутого уровня значимости и принятия решения о гипотезе H_0 условное распределение статистики критерия

согласия $G(S|H_0)$ может быть оценено только в интерактивном режиме проверки гипотезы в соответствии со следующим алгоритмом:

1. На основе деградационной гамма-модели, с которой проверяется гипотеза о согласии, сгенерировать выборку приращений X_n в соответствии с заданным планом эксперимента (при заданных значениях ковариаты, тех же количествах объектов в группах с разными значениями ковариаты, а также с теми же моментами времени замера показателя деградации).

2. По полученной выборке X_n оценить параметры модели методом максимального правдоподобия (2).

3. Сформировать выборку R_N .

4. По выборке R_N вычислить значение статистики критерия согласия (статистики S_K , S_ω или S_Ω).

5. Повторяя пункты 1–4 M раз, получим выборку статистик объема M , на основе которой построить эмпирическую функцию распределения $G_M(S|H_0)$.

По полученному эмпирическому распределению $G_M(S|H_0)$ вычисляется оценка достигнутого уровня значимости $\alpha_N = 1 - G_M(S_N|H_0)$, где S_N – значение соответствующей статистики, полученное по исходной выборке, по которой проверяется гипотеза о согласии H_0 . Если α_N не превышает заданного уровня значимости α , то гипотеза H_0 отвергается.

3. Исследование распределений статистик критериев согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для деградационной гамма-модели надежности

Как уже отмечалось в предыдущем разделе, в случае проверки сложной гипотезы о согласии распределения статистик рассматриваемых критериев зависят от ряда факторов, в том числе и от значения параметра формы в случае гамма-распределения. Параметр формы деградационной гамма-модели (1) представляет собой значение функции тренда, которая, в свою очередь, зависит от функции от ковариат $r(x;\beta)$. Следовательно, распределения статистик критериев типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга при проверке сложной гипотезы о согласии (3) будут зависеть от вида параметрической функции тренда, вида параметрической функции от ковариат, плана эксперимента (моментов времени замера показателя деградации и значений ковариат, при которых наблюдались объекты).

С использованием методов имитационного моделирования исследуем зависимость распределений статистик критериев согласия от данных факторов. Результаты исследования распределений статистик рассматриваемых критериев согласия приведем на примере критерия типа Колмогорова. На рис. 1 представлены эмпирические функции распределения статистики критерия типа Колмогорова $G_M(S_K|H_0)$, полученные при различных вариантах выбора моментов времени замера деградационного показателя: $T_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $T_2 = \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$, $T_3 = \{0, 2, 4, 6, 8, 10\}$. Выборки приращений X_n объема $n = 30$ генерировались на основе гамма-модели со степенной функцией тренда и логлинейной функцией от ковариат. План эксперимента включает пять групп, соответствующих различным значениям скалярной ковариаты $x \in \{0; 0,2; 0,4; 0,6; 0,8\}$. Значения параметров модели: $\sigma = 1$, $\gamma_0 = 0,7$, $\beta_0 = 1$, $\beta_1 = 0,5$. Объем моделирования $M = 16\,600$.

Как видно из рис. 1, распределение статистики критерия существенно зависит как от числа замеров показателя деградации, так и от выбора самих моментов времени замера. Аналогичным образом выявлена зависимость распределений статистик от числа групп (количества различных значений ковариаты) в плане эксперимента.

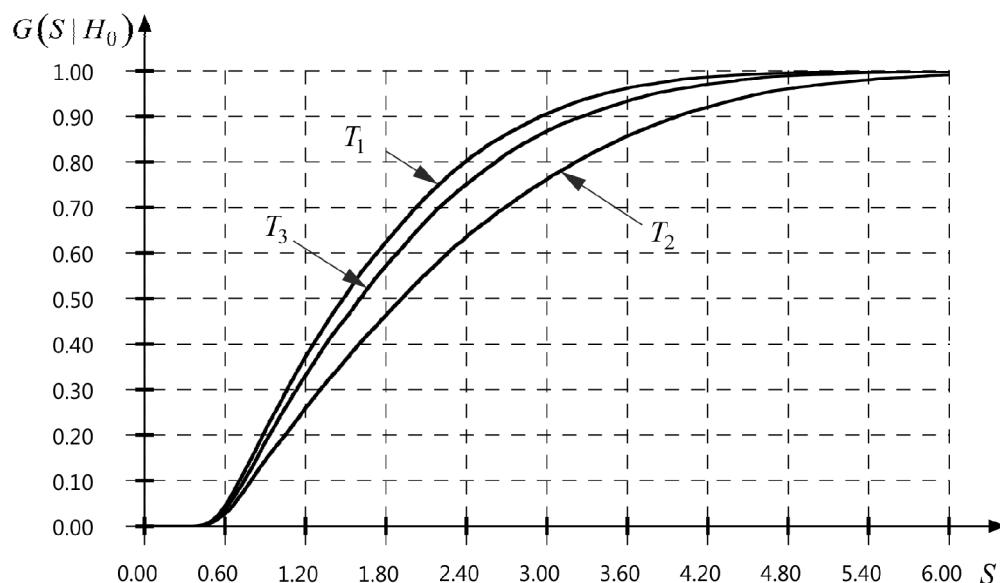


Рис. 1. Эмпирические функции распределения статистики критерия типа Колмогорова при проверке сложной гипотезы для гамма-модели в зависимости от выбора моментов замера показателей старения и их количества

Рассмотрим зависимость распределения статистики критерия типа Колмогорова от выбора функции от ковариат. На рис. 2 представлены эмпирические функции распределения статистики критерия типа Колмогорова $G_M(S_K | H_0)$, полученные для деградационных гамма-моделей с различными функциями от ковариат: логлинейной, Аррениуса и функции правила мощности. Выборки приращений X_n объема $n = 30$ генерировались на основе модели со степенной функцией тренда. План эксперимента включает пять групп, соответствующих различным значениям скалярной ковариаты $x \in \{0; 0,2; 0,4; 0,6; 0,8\}$, и замеры по времени $T = \{0,1,2,3,4,5,6,7,8,9,10\}$. Значения параметров: $\sigma = 1$, $\gamma_0 = 0,7$, регрессионные параметры в случае логлинейной модели и модели Аррениуса $\beta_0 = 1$, $\beta_1 = 0,5$ и для модели правила мощности $\beta_0 = 1$, $\beta_1 = -0,5$. Объем моделирования $M = 16\,600$.

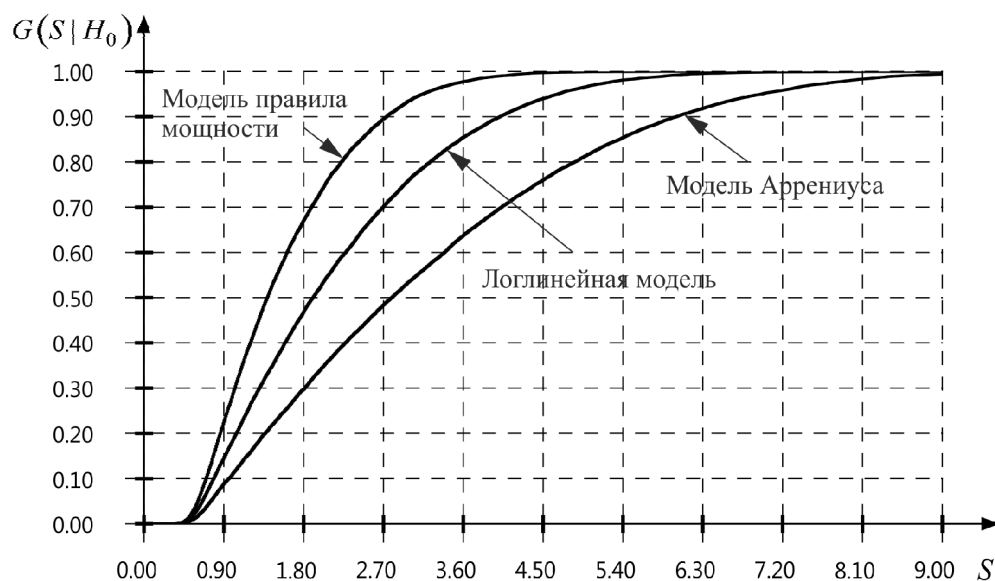


Рис. 2. Эмпирические функции распределения статистики критерия типа Колмогорова при проверке сложной гипотезы для гамма-модели в зависимости от выбора функции влияния объясняющих переменных

Как видно из рис. 2, выбор функции от ковариат для гамма-модели оказывает существенное влияние на распределение статистики критерия типа Колмогорова при проверке сложной гипотезы. Аналогичные исследования были проведены для различных функций тренда показателя деградации: линейной и степенной. Показано, что выбор функции тренда для деградационной гамма-модели также существенно влияет на распределение статистики критерия типа Колмогорова.

Аналогичные результаты исследований распределений статистик были получены для критериев согласия типа Крамера–Мизеса–Смирнова и Андерсона–Дарлинга.

4. Исследование мощности критериев согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для деградационной гамма-модели

Мощностью критерия называется вероятность отвергнуть неверную гипотезу при уровне значимости $\alpha: 1 - G(S_\alpha | H_1)$. Понятно, что чем больше мощность критерия, тем выше его способность различать близкие конкурирующие гипотезы.

С использованием методов имитационного моделирования проведем исследование мощности критериев типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга при проверке сложной гипотезы H_0 , соответствующей деградационной гамма-модели с линейной функцией тренда и логлинейной функцией от ковариат.

Рассмотрим три различные конкурирующие гипотезы:

1. H_1^1 : Общая деградационная модель [11] с линейной функцией тренда и логлинейной функцией от ковариат, с распределением приращений по закону Вейбулла.

2. H_1^2 : деградационная гамма-модель со степенной функцией тренда с параметром $\gamma_0 = 1,3$ и логлинейной функцией от ковариат.

3. H_1^3 : деградационная гамма-модель с линейной функцией тренда и функцией Аррениуса в качестве функции от ковариат.

Моделирование будем проводить для $M = 10\,000$, число объектов $n = 20$ и $n = 40$, число замеров деградационного показателя для каждого объекта равно $k_i = 10$ и $k_i = 20$ соответственно, $i = 1, \dots, n$; скалярная ковариата $x \in \{1, 2\}$.

В табл. 1, 2 и 3 представлены оценки мощности рассматриваемых критериев согласия при проверке сложной гипотезы H_0 против конкурирующих гипотез H_1^1 , H_1^2 и H_1^3 соответственно.

Как видно из табл. 1–3, с ростом объема выборки мощность критериев увеличивается, при этом более предпочтительными по мощности для всех рассмотренных пар конкурирующих гипотез оказались критерии типа Крамера–Мизеса–Смирнова и Андерсона–Дарлинга.

Т а б л и ц а 1

Мощность критериев в случае конкурирующей гипотезы H_1^1

Критерий согласия	$n = 20, k = 10$	$n = 20, k = 20$	$n = 40, k = 10$
Колмогорова	0,483	0,711	0,725
Крамера–Мизеса–Смирнова	0,578	0,836	0,843
Андерсона–Дарлинга	0,614	0,792	0,821

Т а б л и ц а 2

Мощность критериев в случае конкурирующей гипотезы H_1^2

Критерий согласия	$n = 20, k = 10$	$n = 20, k = 20$	$n = 40, k = 10$
Колмогорова	0,766	0,883	0,892
Крамера–Мизеса–Смирнова	0,827	0,921	0,929
Андерсона–Дарлинга	0,836	0,976	0,981

Мощность критериев в случае конкурирующей гипотезы H_1^3

Критерий согласия	$n = 20, k = 10$	$n = 20, k = 20$	$n = 40, k = 10$
Колмогорова	0,624	0,782	0,794
Крамера–Мизеса–Смирнова	0,683	0,855	0,860
Андерсона–Дарлинга	0,716	0,827	0,846

4. Пример построение деградационной гамма-модели по данным об углеродистых резисторах

Углеродистые резисторы (Carbon Film resistors) применяются в радиотехнической и электронной аппаратуре. Данные резисторы отличаются тем, что в качестве проводящего слоя они используют пленку пиролитического углерода. Отказ наступает в тот момент, когда процент повышения сопротивления относительно начального значения достигает критического уровня $\tilde{z}$ (2 или 5% в зависимости от типа резисторов).

В [12] представлены результаты ускоренных испытаний на надежность 29 резисторов в течение 8 084 ч. План эксперимента представляет собой три группы по 9, 10 и 10 резисторов соответственно. Исследования проводились при повышенной температуре: для первой группы резисторов температура составила 83°C, для второй – 133°C, для третьей – 173°C. На основе полученных данных построим деградационную гамма-модель с линейной функцией тренда и функцией Аррениуса в качестве функции от ковариат. В результате решения задачи (2) найдены ОМП параметров деградационной гамма-модели: $\hat{\sigma} = 2,7131$, $\hat{\beta}_0 = 13,1012$, $\hat{\beta}_1 = -7,9512$.

Проверим гипотезу о согласии с гамма-моделью для критериев типа Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга. Для критерия типа Колмогорова получены значение статистики $S_k = 0,882$ и достигнутый уровень значимости $\alpha_N = 0,418$; для критерия типа Крамера–Мизеса–Смирнова $S_\omega = 0,073$, $\alpha_N = 0,733$; для критерия Андерсона–Дарлинга $S_\Omega = 0,66$, $\alpha_N = 0,593$. Так как полученный достигнутый уровень значимости $\alpha_N > \alpha = 0,05$ для всех рассмотренных критериев, гипотеза о согласии с выбранной моделью не отвергается.

Рассчитаем прогноз времени безотказной работы для данных резисторов при различной фиксированной температуре и различных критических уровнях превышения сопротивления при допустимом проценте превышения. Результаты расчета времени прогноза для различных условий проведения эксперимента представлены в табл. 4.

Т а б л и ц а 4

Оценка вероятности безотказной работы при различных условиях проведения эксперимента

Температура, °C	Критический уровень $\tilde{z}$, %	Время безотказной работы t_k , ч	Оценка надежности $S_x(t_k)$
125	2	1 000	0,990
125	5	3 000	0,989
60	2	8 760 (1 год)	0,998

Как видно из табл. 4, с вероятностью около 0,99 при температуре 125°C время безотказной работы в случае критического повышения сопротивления $\tilde{z} = 2\%$ составит 1 000 ч, а для $\tilde{z} = 5\%$ – 3 000 ч. При температуре 60°C и $\tilde{z} = 2\%$ время безотказной работы составит 8 760 ч (примерно 1 год) с вероятностью 0,998.

Заключение

В данной работе рассмотрены вопросы построения деградационной гамма-модели надежности по результатам измерений деградационного показателя при различных постоянных во времени нагрузках. Предложен метод проверки сложной гипотезы о согласии для данной модели с использованием непара-

метрических критериев согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга. В результате исследований мощности рассматриваемых критериев можно утверждать, что предложенный метод проверки гипотезы о согласии позволяет проверять предположения как о виде распределения приращений, так и о виде функции тренда и функции от ковариат.

Рассмотрен пример построения деградационной гамма-модели надежности по данным об углеродистых резисторах. Проведена проверка гипотезы о согласии и на основе полученной модели рассчитаны вероятности безотказной работы для некоторых значений времени наработки при различных температурах.

ЛИТЕРАТУРА

1. Bordes L., Paroissin C., Salami A. Parametric inference in a perturbed gamma degradation process // Preprint/Statistics & Probability Letters. Pau, 2010. P. 13.
2. Liao C.M., Tseng S.-T. Optimal design for step-stress accelerated degradation test // IEEE Trans. Reliab. 2006. V. 55. P. 59–66.
3. Tang L.C., Yang L.C., Xie M. Planning of step-stress accelerated degradation test // Los Angeles: Reliability and Maintainability Annual Symposium, 2004.
4. Tsai C.-C., Tseng S.-T., Balakrishnan N. Mis-specification analyses of gamma and Wiener degradation processes // Journal of Statistical Planning and Inference. 2011. No. 12. P. 25–35.
5. Crowder M., Lawless, J. On a scheme for predictive maintenance // European J. Oper. Res. 2007. V. 16. P. 1713–1722.
6. Антонов А.В., Никулин М.С. Статистические модели в теории надежности. М. : Абрис, 2012. 390 с.
7. Park C., Padgett W.J. Accelerated degradation models for failure based on geometric Brownian motion and gamma process // Lifetime Data Analysis. 2005. V. 11. P. 511–527.
8. Nikulin M., Bagdonavicius N. Accelerated Life Models: Modeling and Statistical Analysis // Boca Raton: Chapman & Hall/CRC. 2001. P. 334.
9. Галанова Н.С., Лемешко Б.Ю., Чимитова Е.В. Применение непараметрических критериев согласия к проверке адекватности моделей ускоренных испытаний // Автометрия. 2012. № 6. С. 53–68.
10. Лемешко Б.Ю. и др. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход. Новосибирск : Изд-во НГТУ, 2011. 888 с.
11. Chimitova E., Chetvertakova E. Alternatives for Wiener and gamma degradation models: method of selection // Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control – AMSA'2013. Novosibirsk, 25–27 Sept. 2013 : proc. of the intern. workshop. Novosibirsk : NSTU Publisher., 2013. P. 77–82.
12. Meeker W.Q., Escobar L.A. Statistical Methods for Reliability Data. N.Y. : John Wiley and Sons, 1998.

Чимитова Екатерина Владимировна, канд. техн. наук, доцент. E-mail: ekaterina.chimitova@gmail.com

Четвертакова Евгения Сергеевна. E-mail: evgenia.chetvertakova@gmail.com

Новосибирский государственный технический университет

Поступила в редакцию 2 октября 2014 г.

Chimitova Ekaterina V., Chetvertakova Evgenia S. (Novosibirsk State Technical University, Russian Federation).

The construction of the gamma degradation model with covariates.

Keywords: gamma degradation model; reliability function; accelerated testing; maximum likelihood; goodness-of-fit tests.

In this paper, we consider the problem of constructing the gamma degradation model as the most frequently used one for description of a degradation process and prediction of non-failure operation time.

A stochastic process $Z(t)$ characterizing degradation process is referred to as gamma degradation process with a shape parameter $v(t)$ and scale parameter σ , if

1. $Z(0) = 0$.

2. $Z(t)$ is a stochastic process with independent increments.

3. the increments $\Delta Z(t) = Z(t + \Delta t) - Z(t)$ follows the gamma distribution with the probability density function

$$f_{\text{Gamma}}(t; \sigma, \Delta v(t)) = \left(\frac{t}{\sigma}\right)^{\Delta v(t)-1} \frac{e^{-t/\sigma}}{\sigma \cdot \Gamma(\Delta v(t))},$$

where $\Delta v(t) = v(t + \Delta t) - v(t)$ is the shape parameter and $\sigma > 0$ is the scale parameter.

The gamma distribution is a repeatable distribution (if random variables ξ_1 and ξ_2 follow the gamma distribution with a scale parameter σ and shape parameters v_1 and v_2 , then $\xi_1 + \xi_2$ follows the gamma distribution with the scale parameter σ and the shape parameter $v_1 + v_2$) that explains the fact of using the gamma distribution as a distribution of increments.

Non-failure operation time, which depends on covariate x , is equal to

$$T_x = \sup\{t : Z_x(t) < \tilde{z}\},$$

where $\tilde{z}$ is the critical value of degradation index, when the failure is fixed.

Then, the reliability function for gamma degradation model is equal to

$$S_x(t) = P\{T_x > t\} = P\{Z_x(t) < \tilde{z}\} = F_{Gamma}\left(\tilde{z}; \sigma, \frac{m_x(t)}{\sigma}\right),$$

where $m_x(t) = \sigma v\left(\frac{t}{r(x;\beta)}\right)$ is a positive monotone increasing trend function, $r(x;\beta)$ is a positive covariate function.

The main problem of using the gamma degradation model is the absence of mathematical methods for testing the statistical hypothesis of goodness-of-fit for the model. We propose an approach to testing goodness-of-fit of the gamma degradation model with covariates, which implies the investigation of test statistic distributions with computer simulation methods in interactive mode of testing hypothesis. The non-parametric goodness-of-fit tests of Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling are recommended for testing this hypothesis.

In this paper, we have carried out the research of statistics distributions and the power of the considered goodness-of-fit tests for the gamma degradation model by means of computer simulation methods. It has been shown that the distributions of the test statistics in the case of composite hypotheses depend on the type of the parametric trend function, the type of the parametric covariate function, and the design of experiment (the moments of time, in which degradation index was measured, and the values of covariates, for which objects were observed). We have carried out an empirical analysis of the power of considered tests for various pairs of competing hypotheses. It has been shown, that the proposed method of testing the goodness-of-fit hypothesis enables to test the assumption of a distribution of degradation increments, as well as the assumptions about the trend function and the covariate function. We have also considered an example of construction of the gamma degradation model for Carbon Film resistors data.

REFERENCES

1. Bordes L., Paroissin C., Salami A. *Parametric inference in a perturbed gamma degradation process*. In: *Statistics & Probability Letters*. Pau, 2010, pp. 13.
2. Liao C.M., Tseng S.-T. Optimal design for step-stress accelerated degradation test. *IEEE Trans. Reliab.*, 2006, vol. 55, pp. 59-66.
3. Tang L.C., Yang L.C., Xie M. *Planning of step-stress accelerated degradation test*. Los Angeles: Reliability and Maintainability Annual Symposium, 2004.
4. Tsai C.-C., Tseng S.-T., Balakrishnan N. Mis-specification analyses of gamma and Wiener degradation processes. *Journal of Statistical Planning and Inference*, 2011, no.12. pp. 25-35. DOI: 10.1016/j.jspi.2011.06.008
5. Crowder, M., Lawless, J. On a scheme for predictive maintenance. *European Journal Oper. Res.*, 2007, vol. 16, pp. 1713-1722. DOI: 10.1016/j.ejor.2005.10.051
6. Antonov A.V., Nikulin M.S (eds.) *Statisticheskie modeli v teorii nadezhnosti* [Statistical models in reliability theory]. Moscow: Abris Publ., 2012.
7. Park, C., Padgett, W.J. Accelerated degradation models for failure based on geometric Brownian motion and gamma process. *Lifetime Data Analysis*, 2005, vol. 11, pp. 511-527. DOI: 10.1007/s10985-005-5237-8
8. Nikulin M., Bagdonavicius N. *Accelerated Life Models: Modeling and Statistical Analysis*. Boca Raton: Chapman & Hall/CRC, 2001.
9. Galanova N.S., Lemesko B.Yu., Chimitova E.V. Using nonparametric goodness-of-fit tests to validate accelerated failure time models. *Optoelectronics, Instrumentation and Data Processing*, 2012, vol. 48. no. 6. pp. 580-592. DOI: 10.3103/S8756699012060064
10. Lemesko B.Yu., Lemesko S.B., Postovalov S.N., Chimitova E.V. *Statisticheskiy analiz dannykh, modelirovanie i issledovanie veroyatnostnykh zakonov*. Kom-p'yuternyy podkhod [Statistical data analysis, simulation and study of probability regularities computer approach]. Novosibirsk: NSTU Publ., 2011. 888 p.
11. Chimitova E., Chetvertakova E. [Alternatives for Wiener and gamma degradation models: method of selection]. *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control – AMSA'2013*. Proc. of the intern. workshop. Novosibirsk: NSTU Publ., 2013, pp. 77-82. (In Russian).
12. Meeker W.Q., Escobar L.A. *Statistical Methods for Reliability Data*. New York: John Wiley and Sons, 1998.

Б.М. Шумилов, У.С. Ыманов

СПЛАЙН-ВЕЙВЛЕТЫ, ОРТОГОНАЛЬНЫЕ МНОГОЧЛЕНАМ, И ОПТИМИЗАЦИЯ ВЫЧИСЛЕНИЙ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 14-31-50353-мол_нр.

Для случая эрмитовых сплайнов произвольной нечетной степени исследовано построение системы базисных сплайн-мультивейвлетов, реализующих условия ортогональности многочленам той же степени. Рассмотрены варианты построения и обращения блока фильтров. Приведены результаты численных экспериментов.

Ключевые слова: мультивейвлеты нечетной степени; ортогональность многочленам.

Кубические сплайны гладкости C^2 завоевали общее признание в среде инженеров-дорожников как адекватный способ математического представления пикетного метода трассирования реконструируемых автомобильных дорог [1–4]. Перспективы использования здесь эрмитовых сплайнов состоят в том, что они позволяют в явном виде, через значения сплайн-коэффициентов, учесть геометрические ограничения как на контрольные точки (пикеты автомобильной трассы), так и на тангенсы трассы (заданные направления касательных при въезде на мостовое сооружение или примыкание) и радиусы закруглений (заданные значения кривизны на разгонных и тормозных участках трассы). В работах [5, 6] были построены мультивейвлеты эрмитовых сплайнов пятой и третьей степени, которые ортогональны любым многочленам той же степени. В представленной статье получено матричное решение для коэффициентов мультивейвлетов произвольной нечетной степени, ортогональных многочленам той же степени. Предложена эффективная его модификация, которая позволяет воспользоваться для численного решения хорошо разработанным методом матричной прогонки [7. С. 103].

1. Построение системы базисных сплайн-мультивейвлетов на конечном отрезке

Пусть на отрезке $[a, b]$ задана вложенная последовательность равномерных сеток $\Delta^L: x_i = a + h \cdot i, i = 0, 1, \dots, 2^L, h = (b - a) / 2^L, L \geq 0$. Если базисные функции $N_{i,k}^L(x) = \varphi_k(v - i), k = 0, 1, \dots, r \forall i$, где $v = (x - a) / h$ с центрами в узлах сетки Δ^L порождены сжатиями и сдвигами $r+1$ функций вида [8. С. 82]:

$$\varphi_k(t) = \begin{cases} (-1)^k \omega_k(-t) & \text{при } -1 \leq t \leq 0, \\ \omega_k(t) & \text{при } 0 \leq t \leq 1, \end{cases}$$

где $\omega_k(t) = (1-t)^{r+1} \sum_{\beta=0}^{r-k} \frac{(r+\beta)!}{k!\beta!r!} t^{k+\beta}, k = 0, 1, \dots, r$, то, при условии отсечения выступающих за концы от-

резка половинок функций $\varphi_0(t), \dots, \varphi_r(t)$, полученное пространство V_L является пространством эрмитовых сплайнов степени $2r+1$ гладкости C^r .

Поскольку сетка $\Delta^{L-1}, L \geq 1$, получена из Δ^L посредством удаления каждого второго узла, то соответствующее пространство V_{L-1} с базисными функциями $N_{i,k}^{L-1}$ с носителями, в два раза большими по ширине, и центрами в четных узлах сетки Δ^L вложено в V_L . Остаток W_{L-1} от разности пространств V_L и V_{L-1} размерности $(r+1) \cdot (2^L + 1 - 2^{L-1} - 1) = (r+1)2^{L-1}$ называется пространством мультивейвлетов. Будем искать базисные мультивейвлеты $M_{i,k}^L(x), k = 0, 1, \dots, r \forall i$, как линейные комбинации базисных эрмитовых сплайнов на сетке Δ^{L+1} , удовлетворяющие условиям ортогональности многочленам $2r+2$ -го порядка, т.е.

$$\int_a^b M_{i,k}^L(x) x^m dx = 0, k = 0, 1, \dots, r \forall i (m = 0, 1, \dots, 2r+1), \quad (1)$$

и имеющие минимальные из возможных носители.

Теорема. Пусть задан отрезок $[0, 2^{L+1}]$, $L \geq 0$, с сеткой $\Delta^{L+1} : x_i = i, i = 0, 1, \dots, 2^{L+1}$, с единичным шагом $h = 1$, и разложения базисных мультивейвлетов имеют вид

$$M_{i,k}^L(x) = \sum_{l=0}^r \sum_{j=0}^p \alpha_{j,l}^l \varphi_l(2t-j), t = (x - x_{2i}), \quad -1 \leq t \leq p+1. \quad (2)$$

Тогда при условии $p = 2$ три матрицы R^0, R^1, R^2 размерности $(2r+2) \times (r+1)$ заданы элементами

$$R_{m,l}^j = \int_{j-1}^{j+1} \varphi_l(2t-j) t^m dt, j = 0, 1, 2, l = 0, 1, \dots, r, m = 0, 1, \dots, 2r+1,$$

каждый из $r+1$ столбцов блочной матрицы $\begin{bmatrix} A_0^{\text{inner}} / A_2^{\text{inner}} \end{bmatrix} = -[R^0 \mid R^2]^{-1} R^1$ дает значения коэффициентов $\alpha_{j,l}^l, j = 0, 2, l = 0, 1, \dots, r$, соответствующего k -го базисного мультивейвлета, полностью лежащего внутри отрезка $[0, 2^{L+1}]$, $L \geq 1$, а коэффициенты $\alpha_1^l = \{1, l = k; 0, l \neq k\}$. При $L = 0$ элементы матриц R^0, R^2 вычисляются в пределах интервала $[0, 2]$, а матрица коэффициентов $\alpha_{j,l}^l, j = 0, 2, l = 0, 1, \dots, r$, обозначены как $\begin{bmatrix} A_0^{\text{center}} / A_2^{\text{center}} \end{bmatrix} = -[R^0 \mid R^2]^{-1} R^1$.

При $L > 0$ для крайних слева базисных мультивейвлетов элементы матрицы R^0 вычисляются по укороченному слева интервалу

$$R_{m,l}^0 = \int_0^1 \varphi_l(2t) t^m dt, l = 0, 1, \dots, r, m = 0, 1, \dots, 2r+1,$$

а коэффициенты разложения (2) $\alpha_{j,l}^l, j = 1, 2, l = 0, 1, \dots, r$, даются значениями столбцов матрицы $\begin{bmatrix} A_1^{\text{left}} / A_2^{\text{left}} \end{bmatrix} = -[R^1 \mid R^2]^{-1} R^0$ при условии, что коэффициенты $\alpha_0^l = \{1, l = k; 0, l \neq k\}$. Для крайних справа базисных мультивейвлетов элементы матрицы R^2 вычисляются по укороченному справа интервалу

$$R_{m,l}^2 = \int_1^2 \varphi_l(2t-2) t^m dt, l = 0, 1, \dots, r, m = 0, 1, \dots, 2r+1,$$

а коэффициенты разложения (2) $\alpha_{j,l}^l, j = 0, 1, l = 0, 1, \dots, r$, даются значениями столбцов матрицы $\begin{bmatrix} A_0^{\text{right}} / A_1^{\text{right}} \end{bmatrix} = -[R^0 \mid R^1]^{-1} R^2$ при условии, что коэффициенты $\alpha_2^l = \{1, l = k; 0, l \neq k\}$.

Система функций $M_{i,k}^L(x), k = 0, 1, \dots, r, i = 1, 2, \dots, 2^L$, удовлетворяет условиям (1) с носителями не более чем из двух шагов сетки Δ^{L+1} и образует базис в пространстве $W_L, L \geq 0$.

Доказательство. а) Пусть $L \geq 1$ и носители мультивейвлетов $[x_{2i-1}, x_{2i+p+1}], p \geq 1$, полностью располагаются внутри отрезка $[0, 2^{L+1}]$. При $p \geq 1$ для всех $i = 1, 2, \dots, 2^L - 1 - [p/2]$ имеет место разложение (2). Подставим его в (1) и вычислим все необходимые интегралы, учитывая при этом, что подынтегральные выражения обращаются в нуль вне носителей вейвлетов. Условия ортогональности (1) определяют однородную систему $2r+2$ уравнений метода моментов [9. С. 42] относительно коэффициентов мультивейвлета $M_{i,k}^L(x)$. В силу линейной независимости на интервалах $[x_{2i-1}, x_{2i+p+1}]$ пробных функций $N_{2i+l,k}^{L+1}(x), l = 0, \dots, r, j = 0, 1, \dots, p$, и поверочных функций $x^m, m = 0, 1, \dots, 2r+1$, ранг полученной системы равняется $\min[2r+2, (r+1)(p+1)]$. Если предположить, что носитель мультивейвлета равен трем шагам

сетки Δ^{L+1} , т.е. $p = 1$, то однородная система становится квадратной и, будучи невырожденной, может иметь только тривиальное решение. Поэтому будем далее считать, что носитель мультивейвлета равен четырем шагам сетки Δ^{L+1} , т.е. мультивейвлет построен из $3r+3$ базисных сплайнов. В этом случае ранг матрицы равен $2r+2$. Поэтому однородная система $2r+2$ уравнений с $3r+3$ неизвестными имеет $r+1$ линейно независимых частных решений. При этом количество полученных для каждого номера i мультивейвлетов, полностью лежащих внутри отрезка $[0, 2^{L+1}]$, равно $(r+1)(2^L - 2) = (r+1)2^L - 2(r+1)$, что на $2(r+1)$ меньше разности между размерностями пространств V_{L+1} и V_L : $(r+1)(2^{L+1} + 1) - (r+1)(2^L + 1) = (r+1)2^L$.

Вблизи концов отрезка интервалы интегрирования не выходят за пределы отрезка. Поэтому с учетом того, что по краям отрезка от выступающих за концы отрезка функций $\varphi_0(t), \dots, \varphi_r(t)$ остается по половинке, граничные мультивейвлеты отличаются от мультивейвлетов, предложенных выше, при условии ортогональности многочленам $2r+1$ -й степени. Аналогичные рассуждения приводят к двум системам линейно независимых функций $M_{i,k}^L(x)$, $k = 0, 1, \dots, r$, отдельно на левом и на правом концах, дополняющих до базиса построенную выше мультивейвлет-систему.

б) Пусть $L = 0$. Вычисления снова дают $r+1$ линейно независимых частных решений, которые образуют базис в пространстве $V_1 - V_0$ с размерностью $(r+1) \cdot 3 - (r+1) \cdot 2 = (r+1)$.

в) Окончательно, в соответствии с правилами линейной алгебры выписываются матричные формулы для вычисления коэффициентов эрмитовых мультивейвлетов нечетной степени. Теорема доказана.

2. Построение и обращение блока фильтров

Если упорядочить базисные сплайн-функции в виде единой матрицы-строки $\varphi^L = [N_{0,0}^L, N_{0,1}^L, \dots, N_{0,r}^L, N_{1,0}^L, N_{1,1}^L, \dots, N_{2^L,r}^L]$, то можно записать функции φ^{L-1} в виде линейных комбинаций функций φ^L : $\varphi^{L-1} = \varphi^L P^L$. Здесь блоки матрицы P^L составлены из коэффициентов масштабных соотношений (также для единичного шага) [10]:

$$\begin{bmatrix} \varphi_0(t) \\ \varphi_1(t) \\ \vdots \\ \varphi_r(t) \end{bmatrix} = \sum_{k=0}^2 H_k \begin{bmatrix} \varphi_0(2t-k) \\ \varphi_1(2t-k) \\ \vdots \\ \varphi_r(2t-k) \end{bmatrix},$$

где $H_2 = U^{-1} \Lambda U$, $H_1 = \text{diag}(1, 2^{-1}, \dots, 2^{-r})$, $H_0 = S H_2 S^{-1}$ и матрица U размерности $(r+1) \times (r+1)$ задана элементами

$$U_{k,j} = (-1)^{r+1+k-j} \frac{(r+1+k)!}{(r+1+k-j)!}, \quad k, j = 0, 1, \dots, r,$$

а $\Lambda = \text{diag}(2^{-r-1}, \dots, 2^{-2r-1})$, $S = \text{diag}(1, -1, \dots, (-1)^{-r})$.

Аналогично запишем базисные мультивейвлет-функции $2r+1$ -й степени на уровне разрешения L в виде матрицы-строки $\psi^L = [M_{1,0}^L, M_{1,1}^L, \dots, M_{1,r}^L, \dots, M_{2^L,r}^L]$. Тогда для уровня разрешения $L-1$ можно выразить функции ψ^{L-1} в виде линейных комбинаций функций φ^L : $\psi^{L-1} = \varphi^L Q^L$, где блоки матрицы Q^L составлены из коэффициентов разложений (2). Соответствующие коэффициенты сплайна будем собирать в вектор $C^L = [C_{0,0}^{L,0}, C_{0,0}^{L,1}, \dots, C_{0,0}^{L,r}, C_{1,0}^{L,0}, \dots, C_{2^L,r}^{L,r}]^T$, и соответствующие мультивейвлет-коэффициенты – в вектор $D^L = [D_{1,0}^{L,0}, D_{1,0}^{L,1}, \dots, D_{1,0}^{L,r}, \dots, D_{2^L,r}^{L,r}]^T$. Тогда с использованием обозначений для блочных матриц процесс получения C^L из C^{L-1} и D^{L-1} может быть записан так [11. С. 101]:

$$C^L = [P^L | Q^L] \left[\frac{C^{L-1}}{D^{L-1}} \right]. \quad (3)$$

Разрешимость системы (3) относительно C^{L-1} , D^{L-1} гарантирована линейной независимостью базисных функций. Для упрощения вычислений матрицу $[P^L | Q^L]$ предлагается сделать блочной трехдиагональной, изменив порядок неизвестных так, чтобы блоки матриц P^L и Q^L перемежались:

$$\mathbf{K}^L u^{L-1} = C^L,$$

$$u^L = \left[C_0^{L,0}, C_0^{L,1}, \dots, C_0^{L,r}, D_1^{L,0}, D_1^{L,1}, \dots, D_1^{L,r}, C_1^{L,0}, \dots, D_{2^L}^{L,r}, \dots, C_{2^L}^{L,r} \right]^T,$$

$$\mathbf{K}^1 = \begin{bmatrix} H_1 & A_0^{\text{center}} & O \\ H_2^T & I & H_0^T \\ O & A_2^{\text{center}} & H_1 \end{bmatrix},$$

$$\mathbf{K}^L = \begin{pmatrix} H_1 & I & 0 & 0 & \dots & 0 & 0 \\ H_2^T & A_1^{\text{left}} & H_0^T & 0 & \dots & 0 & 0 \\ 0 & A_2^{\text{left}} & H_1 & A_0^{\text{inner}} & \ddots & \vdots & 0 \\ 0 & 0 & H_2^T & I & \ddots & 0 & \vdots \\ 0 & 0 & 0 & A_2^{\text{inner}} & \ddots & A_0^{\text{right}} & 0 \\ \vdots & \vdots & \vdots & \ddots & H_2^T & A_1^{\text{right}} & H_0^T \\ 0 & 0 & 0 & \dots & 0 & I & H_1 \end{pmatrix}, L > 1.$$

Здесь O обозначает матрицу $r+1$ -го порядка с нулевыми коэффициентами, тогда как I – единичная матрица $r+1$ -го порядка, многоточия обозначают повторяющиеся блоки матрицы $\mathbf{K}^L$. При этом с целью компенсации единичного шага сетки в уравнениях (3) в качестве исходных C^L нужно использовать значения функции и производных, домноженные на h в соответствующей степени: $\{f^{(k)}(i \cdot h) \cdot h^k, k = 0, 1, \dots, r, i = 0, 1, \dots, 2^L\}$, всего $(r+1) \cdot (2^L+1)$ чисел.

Процедуру разбиения C^L на часть C^{L-1} , соответствующую низшему разрешению, и уточняющие коэффициенты D^{L-1} можно применить рекурсивно и к самой части C^{L-1} . Следовательно, исходные значения C^L можно представить в виде иерархии все более грубых версий с разрешениями $C^0, C^1, \dots, C^{L-1}$ и уточняющих деталей $D^0, D^1, \dots, D^{L-1}$. При этом по величине вейвлет-коэффициентов $D^j, j = 0, 1, \dots, L-1$, можно судить об их значимости для результирующей сплайн-аппроксимации. Незначимые убираются с целью сжатия информации.

3. Пример сравнения мультивейвлетов степени 7 и 11 при одном и том же числе базисных функций 12

Для $x \in [0, 1]$ рассмотрим в качестве примера приближение функции Хартена [12]:

$$f(x) = \begin{cases} \frac{1}{2} \sin(3\pi x), & x \leq \frac{1}{3}, \\ |\sin(4\pi x)|, & \frac{1}{3} < x \leq \frac{2}{3}, \\ -\frac{1}{2} \sin(3\pi x), & x > \frac{2}{3}. \end{cases}$$

Если для случая $r=5$ аннулировать все вейвлет-коэффициенты, оставив только сплайн-коэффициенты $C^0 = [0, 2968; -32, 14; 2149; -7, 73 \cdot 10^4; 1, 854 \cdot 10^6; -3, 122 \cdot 10^7; 0, 3851; 55, 15; 3146; 1, 222 \cdot 10^5; 3, 31 \cdot 10^6; 6, 49 \cdot 10^7]^T$, то получится некоторый сглаживающий многочлен 11-й степени, весьма незначительно отличающийся от МНК-решения (рис. 1), представленного коэффициентами $[0, 1818; -22, 93; 1694; -6, 193 \cdot 10^4; 1, 485 \cdot 10^6; -2, 48 \cdot 10^7; 0, 2333; 42, 72; 2514; 9, 995 \cdot 10^4; 2, 74 \cdot 10^6; 5, 407 \cdot 10^7]^T$.

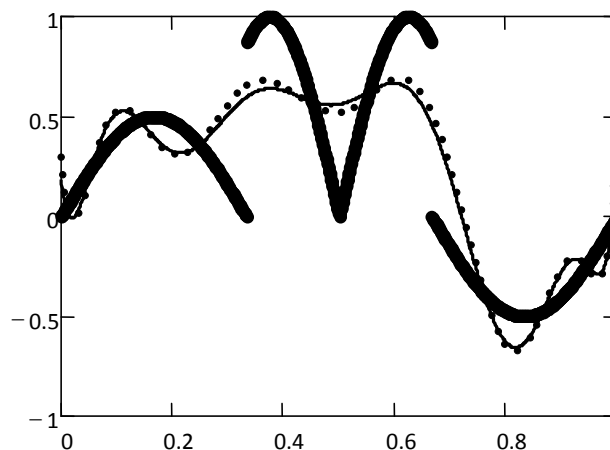


Рис. 1. Сравнение значений сглаживающего многочлена 11-й степени (точки) и МНК-многочлена 11-й степени (сплошная линия), построенных по значениям функции Хартена (жирная линия)

Для случая $r = 3$ остаются только сплайн-коэффициенты $C^I = [0,359; -16,22; 448,8; -6532; 0,293; -1,104; 53,74; 570,8; 0,57; 32,56; 755,3; 1,137 \cdot 10^4]^T$. В результате получается сглаживающий сплайн 7-й степени с одним узлом посередине отрезка. Отличие от МНК-решения с коэффициентами $[0,2312; -12,22; 377,1; -5711; 0,3952; -0,2757; 38,69; 229,2; 0,2653; 19,64; 463,6; 7208]^T$ выглядит более значительным при заметном улучшении подгонки в окрестности излома функции (рис. 2).

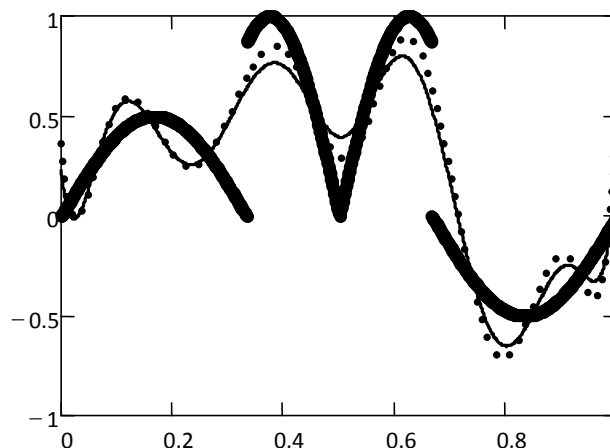


Рис. 2. Сравнение значений сглаживающего сплайна 7-й степени (точки) и МНК-сплайна 7-й степени (сплошная линия), построенных по значениям функции Хартена (жирная линия)

Заключение

В работе представлена общая схема построения эрмитовых сплайн-вейвлетов, ортогональных многочленам. Полученные результаты предоставляют широкие возможности для оптимизации методов обработки численной информации, которые допускают и параллельную реализацию [13. С. 139; 14].

ЛИТЕРАТУРА

1. Бойков В.Н., Шумилов Б.М. Сплайны в трассировании автомобильных дорог. Томск : Изд-во ГУ Томский ЦНТИ, 2001. 164 с.
2. Система проектирования IndorCAD. Проектирование автомобильных дорог: руководство пользователя / И.В. Кривых, Д.А. Петренко, В.Н. Бойков и др. 2-е изд., испр. Томск : Изд-во Том. ун-та, 2010. 250 с.

3. Бойков В.Н. САПР автодорог – перспективы развития // САПР и ГИС автомобильных дорог. 2013. № 1(1). С. 6–9.
4. Петренко Д.А. Новое поколение программных продуктов ИндорСофт // САПР и ГИС автомобильных дорог. 2013. № 1(1). С. 10–17.
5. Шумилов Б.М., Кудуев А.Ж. Новый тип мультивейвлетов пятой степени, ортогональных многочленам пятой степени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4(21). С. 108–116.
6. Шумилов Б.М. Мультивейвлеты эрмитовых сплайнов третьей степени, ортогональные кубическим многочленам // Математическое моделирование. 2013. Т. 25, № 4. С. 17–28.
7. Самарский А.А., Николаев Е.С. Методы решения сеточных уравнений : учеб. пособие. М. : Наука, 1978. 591 с.
8. Завьялов Ю.С., Квасов Б.И., Мирошников В.Л. Методы сплайн-функций. М. : Наука, 1980. 352 с.
9. Флетчер К. Численные методы на основе метода Галёркина : пер. с англ. М. : Мир, 1988. 352 с.
10. Strang G., Strela V. Short wavelets and matrix dilation equations // IEEE Trans. Signal Processing. 1995. V. 43, No. 1. P. 108–115.
11. Столниц Э., ДеРоуз Т., Салезин Д. Вейвлеты в компьютерной графике : пер. с англ. Ижевск : Регулярная и хаотическая динамика, 2002. 272 с.
12. Arándiga F., Baeza A., Donat R. Discrete multiresolution based on hermite interpolation: computing derivatives // Communications in Nonlinear Science and Numerical Simulation. 2004. No. 9. P. 263–273.
13. Ортега Дж. Введение в параллельные и векторные методы решения линейных систем. М. : Мир, 1991. 367 с.
14. Высокопроизводительные вычисления на кластерах : учеб. пособие / под ред. А.В. Старченко. Томск : Изд-во Том. ун-та, 2008. 198 с.

Шумилов Борис Михайлович, д-р физ.-мат. наук, профессор. E-mail: sbm@tsuab.ru

Томский государственный архитектурно-строительный университет

Ыманов Улукбек Сайдакромович. E-mail: YmanV8106@rambler.ru

Ошский государственный университет

Поступила в редакцию 8 сентября 2014 г.

Shumilov Boris M., Ymanov Ulukbek S. (Tomsk State University. Russian Federation, Osh State University. Kyrgyzstan).

Spline-wavelets, orthogonal to polynomials, and optimization of calculations of wavelet-transformation.

Keywords: Multiwavelets of odd degree, orthogonality to polynomials.

For the space of the Hermitian splines of any odd degree $2r+1$ of a kind $S^L(x) = \sum_{k=0}^r \sum_{i=0}^{2^L} C_{i,k}^{L,k} N_{i,k}^L(x)$, $a \leq x \leq b$, with a uniform grid of nodes $\Delta^L: u_i = a + (b-a) i / 2^L$, $i = 0, 1, \dots, 2^L$, $L \geq 0$, and the basic functions $N_{i,k}^{L(l)}(u_j) = \delta_i^j \cdot \delta_k^l$, $l = 0, 1, \dots, r$, with the centers in integers, it is proposed to use as wavelets the functions $M_{i,k}^L(x)$, satisfying the conditions of orthogonality to all polynomials of the $2r+2$ -nd order, i.e. $\int_a^b M_{i,k}^L(x) x^m dx = 0$, $k = 0, 1, \dots, r \forall i$ ($m = 0, 1, \dots, 2r+1$). For wavelets with the centers in even integers and the supports, which are equal to the supports of basic splines on a grid Δ^L , the formulas for calculating the coefficients $D_1^{L-1,0}, D_1^{L-1,1}, \dots, D_1^{L-1,r}, \dots, D_{2^{L-1}}^{L-1,r}$ in the thinned grid Δ^{L-1} from the spline coefficients $C_0^{L,0}, C_0^{L,1}, \dots, C_0^{L,r}, C_1^{L,0}, \dots, C_{2^L}^{L,r}$ in a dense grid Δ^L in the form of the solution of the linear algebraic equations system with a block three-diagonal matrix of a kind

$$\mathbf{K}^L \mathbf{u}^{L-1} = \mathbf{C}^L,$$

$$\mathbf{u}^L = \left[C_0^{L,0}, C_0^{L,1}, \dots, C_0^{L,r}, D_1^{L,0}, D_1^{L,1}, \dots, D_1^{L,r}, C_1^{L,0}, \dots, D_{2^L}^{L,r}, \dots, C_{2^L}^{L,r} \right]^T,$$

$$\mathbf{K}^1 = \begin{bmatrix} H_1 & A_0^{\text{center}} & O \\ H_2^T & I & H_0^T \\ O & A_2^{\text{center}} & H_1 \end{bmatrix},$$

$$\mathbf{K}^L = \begin{pmatrix} H_1 & I & 0 & 0 & \dots & 0 & 0 \\ H_2^T & A_1^{\text{left}} & H_0^T & 0 & \dots & 0 & 0 \\ 0 & A_2^{\text{left}} & H_1 & A_0^{\text{inner}} & \ddots & \vdots & 0 \\ 0 & 0 & H_2^T & I & \ddots & 0 & \vdots \\ 0 & 0 & 0 & A_2^{\text{inner}} & \ddots & A_0^{\text{right}} & 0 \\ \vdots & \vdots & \vdots & \ddots & H_2^T & A_1^{\text{right}} & H_0^T \\ 0 & 0 & 0 & \dots & 0 & I & H_1 \end{pmatrix}, L > 1.$$

are received. Here $H_2 = U^{-1} \Lambda U$, $H_1 = \text{diag}(1, 2^{-1}, \dots, 2^{-r})$, $H_0 = S H_2 S^{-1}$, and the matrix U of dimensions $(r+1) \times (r+1)$ is set by the elements $U_{k,j} = (-1)^{r+1+k-j} \frac{(r+1+k)!}{(r+1+k-j)!}$, $k, j = 0, 1, \dots, r$, and $\Lambda = \text{diag}(2^{-r-1}, \dots, 2^{-2r-1})$, $S = \text{diag}(1, -1, \dots, (-1)^{-r})$, O designates a matrix of the $r+1$ -st order with zero coefficients whereas I is a single matrix of the $r+1$ -st order, dots designate the repeating blocks of the matrix $\mathbf{K}^L$. The blocks of a matrix $\begin{bmatrix} A_0^{\text{inner}} & A_2^{\text{inner}} \end{bmatrix} = -[R^0 \mid R^2]^{-1} R^1$ are set by the elements, respectively

$$R_{m,l}^j = \int_{j-1}^{j+1} \varphi_l(2t-j) t^m dt, \quad j = 0, 1, 2, \quad l = 0, 1, \dots, r, \quad m = 0, 1, \dots, 2r+1,$$

where

$$\varphi_k(t) = \begin{cases} (-1)^k \omega_k(-t) & \text{при } -1 \leq t \leq 0, \\ \omega_k(t) & \text{при } 0 \leq t \leq 1, \end{cases}$$

and $\omega_k(t) = (1-t)^{r+1} \sum_{\beta=0}^{r-k} \frac{(r+\beta)!}{k! \beta! r!} t^{k+\beta}$, $k = 0, 1, \dots, r$.

Results of numerical experiments for different degree of a spline at the same number of basic functions in comparison with a classical method of the least squares (MLS) are presented. The improvement of adjustment on condition of coincidence of a spline knot to a break of the approximated function is revealed. It is shown that for lack of spline knots the MLS-solution and the wavelet-decision are close to each other. In the presence of spline knot the MLS-solution and the wavelet-decision differ from each other more considerably.

REFERENCES

1. Boykov V.N., Shumilov B.M. *Splayny v trassirovaniy avtomobil'nykh dorog* [Splines in tracing of highways]. Tomsk: State Administration of Tomsk Centre of Scientific Research Publ., 2001. 164 p.
2. Krivykh I.V., Petrenko D.A., Boykov V.N. et al. *Sistema proektirovaniya IndorCAD. Proektirovanie avtomobil'nykh dorog: rukovodstvo pol'zovatelya* [System of design IndorCAD. Design of highways: user's guide]. Tomsk: Tomsk State University Publ., 2010. 250 p.
3. Boykov V.N. CAD systems for roads – evolution prospects. *SAPR i GIS avtomobil'nykh dorog – CAD and GIS for roads*, 2013, no. 1(1), pp. 6-9. (In Russian).
4. Petrenko D.A. The new generation of IndorSoft software products. *SAPR i GIS avtomobil'nykh dorog – CAD and GIS for roads*, 2013, no. 1(1), pp. 10-17. (In Russian).
5. Shumilov B.M., Kuduev A.Zh. New type multiwavelets of the fifth degree orthogonal to quintic polynomials. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2012, no. 4(21), pp. 108-116. (In Russian).
6. Shumilov B.M. Multiwavelets of the third-degree Hermitian splines orthogonal to cubic polynomials. *Matematicheskoe modelirovanie – Mathematical Models and Computer Simulations*, 2013, vol. 25, no. 4, pp. 17-28. DOI: 10.1134/S2070048213060100
7. Samarskiy A.A., Nikolaev E.S. *Metody resheniya setochnykh uravneniy* [Methods of solving grid equations]. Moscow: Nauka Publ., 1978. 591 p.
8. Zav'yalov Yu.S., Kvasov B.I., Miroshnichenko V.L. *Metody splayn-funktsiy* [Methods of Spline Functions]. Moscow: Nauka Publ., 1980. 352 p.
9. Fletcher K. *Chislennyye metody na osnove metoda Galerkina* [Computational Galerkin Methods]. Translated from English. Moscow: Mir Publ., 1988. 352 p.
10. Strang G., Strela V. Short wavelets and matrix dilation equations. *IEEE Trans. Signal Processing*. 1995, vol. 43, no. 1, pp. 108–115. DOI: 10.1109/78.365291
11. Stollnitz E.J., DeRose T.D., Salesin D.H. *Wavelets for Computer Graphics: Theory and Applications*. San Francisco: Morgan Kaufmann, 1996. 245 p.
12. Arándiga F., Baeza A., Donat R. Discrete multiresolution based on hermite interpolation: computing derivatives. *Communications in Nonlinear Science and Numerical Simulation*, 2004, no. 9, pp. 263-273. DOI: 10.1016/S1007-5704(03)00116-3
13. Ortega J.M. *Introduction to Parallel and Vector Solutions of Linear Systems*. New York: Premium Press, 1988. 305 p.
14. Starchenko A.V. (ed.) *Vysokoproizvoditel'nye vychisleniya na klasterakh* [High-performance calculations on clusters]. Tomsk: Tomsk State University Publ., 2008. 198 p.

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.652.8

А.М. Бабанов

ПРАВИЛА ПОРОЖДЕНИЯ ОГРАНИЧЕНИЙ В СЕМАНТИЧЕСКИХ МОДЕЛЯХ ДАННЫХ ORM И ERMM

Сравниваются средства декларативного задания ограничений целостности семантических моделей данных «Объект – Роль» (OR-модели) [1] и «Сущность – Связь – Отображение» (ERM-модели) [2]. На конкретных примерах показано, что большинство известных типов ограничений целостности можно выразить с помощью обобщенных понятий логики и математики, используемых в ERM-модели.

Ключевые слова: семантическая модель данных; OR-модель; ERM-модель; ограничения целостности.

На семантические модели возложена очень важная задача – первичное формальное представление информации о предметной области (ПрО). Она еще более усложняется, если мы хотим построить хоть и высокоуровневое, но полное описание семантики моделируемого мира.

Из трех компонентов модели данных – правил структуризации данных, правил задания ограничений целостности (ОЦ) и операций над данными – наибольший вклад в решение задачи проектирования полной схемы ПрО вносит аппарат ОЦ. Действительно, многочисленные структурные понятия имеющихся моделей данных являются, по сути, вариациями известных математических форм данных – множество и отношение. А операции над данными для семантических моделей не актуальны (ERMM – исключение), так как физическая реализация БД в этих моделях не предполагается, поскольку нет и вряд ли появятся поддерживающие их СУБД.

О роли ОЦ в задаче фиксации семантики ПрО редко вспоминают, говорят в основном об их участии в решении проблемы поддержания БД в актуальном состоянии. Д. Цикритзис и Ф. Лоховски так определяют назначение ОЦ: «Ограничения вводятся в модели данных в целях повышения их семантической и расширения возможностей поддержания целостности данных. Первый аспект связан с адекватностью отражения реального мира в схеме, а второй – с возможностями СУБД обеспечивать соответствие порождаемых схемой состояний данных требованиям, выражаемым ограничениями» [3].

Именно поэтому представляет интерес современное состояние этого компонента семантических моделей данных. Далее в статье будут представлены возможности задания ОЦ двух наиболее выразительных семантических моделей – OR-модели и ERM-модели.

1. Классификации ограничений целостности ORM и ERMM

OR-модель признается многими наиболее оснащенной разнообразными средствами определения ОЦ. Халпин в своей модели предлагает следующие разновидности ограничений целостности [1, 4–6]:

- внутренние и внешние ограничения уникальности ролей;
- ограничения обязательности ролей;
- ограничения включающего «или»;
- ограничения исключения;
- ограничения исключающего «или»;
- ограничения сравнения множеств на подмножество, равенство и исключение (указанные выше ограничения могут применяться как к одной роли, так и к группе ролей или всей ассоциации в целом);
- ограничения на специализации;
- ограничения частоты появления объекта в роли;

- кольцевые ограничения;
- ограничения на значения для типов объектов и ролей;
- ограничения на мощности типов объектов и ролей.

Два последних вида ОЦ достаточно просты и традиционны и не будут рассматриваться в дальнейшем.

Особенностью аппарата ограничений целостности модели Халпина является, то, что для любой полезной особенности данных в нем изобретен свой специальный термин и графический примитив, что приводит к громоздкости определения модели и сложности ее применения.

Теперь посмотрим, как обстоят дела с правилами задания ограничений целостности в ERM-модели.

Первую группу ограничений целостности составляют правила, определяющие области значений отображений-характеристик. Область значений такого отображения можно задать определением синтаксиса составляющих его значений, перечислением допустимых значений (экстенсионалом), операцией сравнения, логическим выражением.

Вторая группа ограничений целостности определяет количественные характеристики (глобальное минимальное кардинальное число, реальное минимальное кардинальное число, максимальное кардинальное число) и тип (неограниченное, полностью определенное, частичное функциональное, полное функциональное) каждого отображения-характеристики и обратного ему отображения.

Третья группа ограничений целостности определяет количественные характеристики и тип реляционных отображений.

Четвертая группа ограничений целостности определяет количественные характеристики и тип отображений, сопутствующих атрибутным отображениям множеств связей.

Пятая группа ограничений целостности определяет получаемые отображения (образуемые с помощью операций над отображениями), а также их количественные характеристики и тип.

Шестая группа ограничений целостности определяет взаимоотношения между классами и между отображениями.

Последняя группа ограничений целостности связана со специализациями. Для каждой из них необходимо указать, является она полной или частичной, пересекающейся или непересекающейся, а также определить отображение, которое является основанием деления суперкласса.

Отличительной особенностью аппарата ограничений целостности ERM-модели является, то, что в ней предлагается ограниченный универсальный набор конструкций, с помощью которых можно представить многие (если не все) законы взаимоотношения данных.

Далее на примерах будет показано, как специализированные виды ОЦ OR-модели представляются с помощью одних и тех же обобщенных понятий ERM-модели. Из-за лимитированного объема статьи мы ограничимся наиболее важными классами ОЦ модели Халпина и рассмотрим лишь несколько примеров.

2. Ограничения уникальности и обязательности ролей

Внутренние ограничения уникальности указывают, какие роли или комбинации ролей одной ассоциации должны иметь уникальные вхождения. Внешние ограничения уникальности действуют аналогично, но применяются к ролям различных предикатов, концептуально соединенных друг с другом. Линии около прямоугольников ролей изображают внутренние ограничения уникальности (рис. 1).

В ORM-схеме, чтобы перейти от одного предиката к другому, мы должны пройти через тип объекта, выполняя концептуальное соединение на типе объектов. По умолчанию условие соединения состоит в том, что экземпляр объекта остается тем же самым, когда мы проходим через тип объектов. Это называют концептуальным внутренним соединением. Оно подобно реляционному естественному внутреннему соединению, за исключением того, что концептуальные соединения требуют, чтобы концептуальный объект был тем же самым, вместо того чтобы устанавливать соответствие значений атрибутов. Прямоугольники ролей, к которым применяется внешнее ограничение уникальности, присоединяются пунктирами к кружку уникальности, внутри которого указан знак «—» (см. рис. 1).

Ограничение обязательности роли означает, что каждый экземпляр класса, для которого она определена, должен играть эту роль. Роли, для которых отсутствуют такие ограничения, называются необязательными или опциональными. Чтобы указать, что роль является обязательной, мы добавляем точку обязательности роли к линии, которая соединяет роль с ее типом объектов (рис. 1). Эта точка может быть помещена с обоих концов линии роли [1].

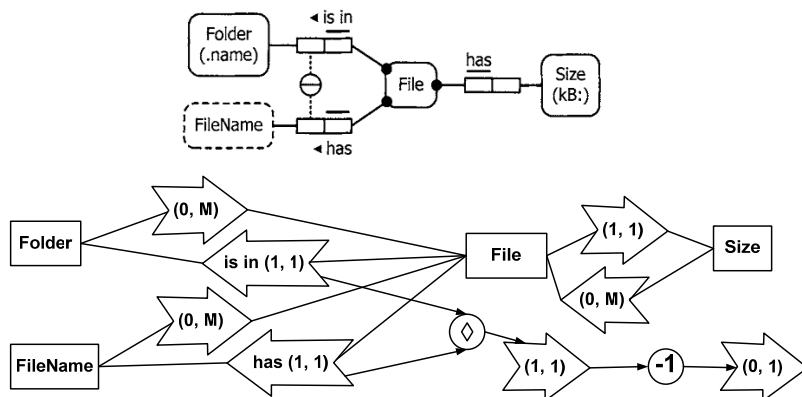


Рис. 1. Фрагменты OR-схемы и ERM-схемы ПрО «Файловая система»

Анализ OR-схемы и ERM-схемы на рис. 1 позволяет сделать следующие выводы:

- ограничение уникальности эквивалентно утверждению о функциональности отображения (в случае внутреннего ограничения – отображения, определяемого ролью или группой ролей, в случае внешнего ограничения – отображения, обратного отображению-агрегату);
- ограничение обязательности эквивалентно утверждению о полной определенности отображения, определяемого ролью.

3. Ограничения включающего «или», исключения и исключающего «или»

Ограничение включающего «или» указывает, что дизъюнкция двух или более ролей является обязательной. Ограничение исключения запрещает одновременное включение ролей ассоциаций. То есть экстенционалы этих ролей взаимно исключительны. Можно определить два ограничения одновременно – ограничение обязательности и взаимной исключительности ролей. Оба эти ограничения вместе обеспечивают так называемое ограничение исключающего «или».

Ограничение включающего «или» показывается точкой обязательности, взятой в кружок. Символ «X» в кружке указывает на ограничение исключения. Ограничение исключающего «или» изображается наложением символов исключения и обязательности в кружке и формированием символа, напоминающего спасательный круг (рис. 2). Во всех этих случаях кружок связывается пунктирными линиями с ролями, охватываемыми ограничением [1].

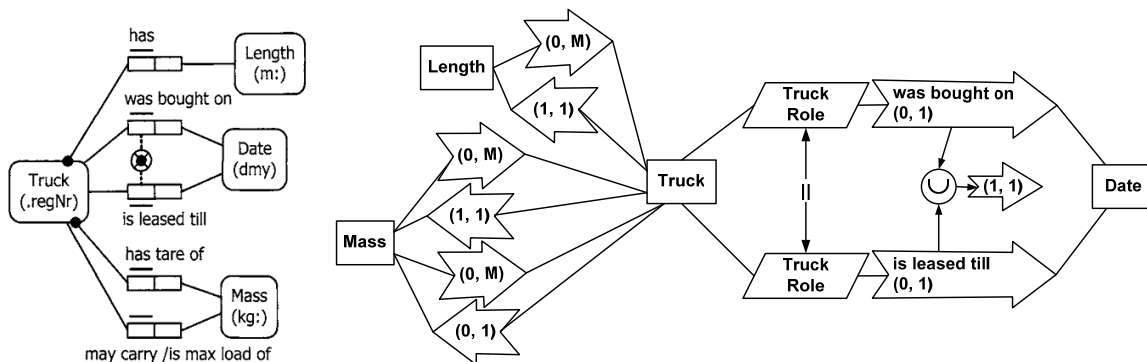


Рис. 2. Фрагменты OR-схемы и ERM-схемы ПрО «Гараж»

В приведенном на рис. 2 примере наряду с другими ОЦ показано ограничение исключаящего «или», декларирующее одновременно два бизнес-правила: «грузовик должен обязательно либо быть куплен, либо взят в аренду» и «нельзя один и тот же грузовик и купить, и взять в аренду одновременно».

В ERM-модели это ограничение выражается с помощью указаний на то, что реальные области отображений «was bought on» и «is leased till» не пересекаются (дуга между параллелограммами ролей со знаком в виде вертикальных отрезков) и что объединение этих отображений полностью определено (глобальное минимальное кардинальное число отображения равно 1).

4. Ограничения сравнения множеств

Ограничения сравнения множеств определяют бизнес-правила, касающиеся отношений между экстенционалами отдельных ролей, их групп или ассоциаций в целом. Рассматриваются следующие отношения, которыми два множества могут быть связаны: подмножество, равенство, исключение.

Ограничение подмножества задается символом подмножества в кружке на пунктирной стрелке, идущей от роли подмножества к роли супермножества. Ограничение равенства обозначается символом равенства «=», взятым в кружок и связанным пунктирами с релевантными ролями. Символ «X», взятый в кружок, указывает на ограничение исключения между ролями, с которыми кружок соединяется пунктирами [1].

Ограничение исключения между ролями уже показано на рис. 2. Следующий пример (рис. 3) демонстрирует применение ограничения подмножества.

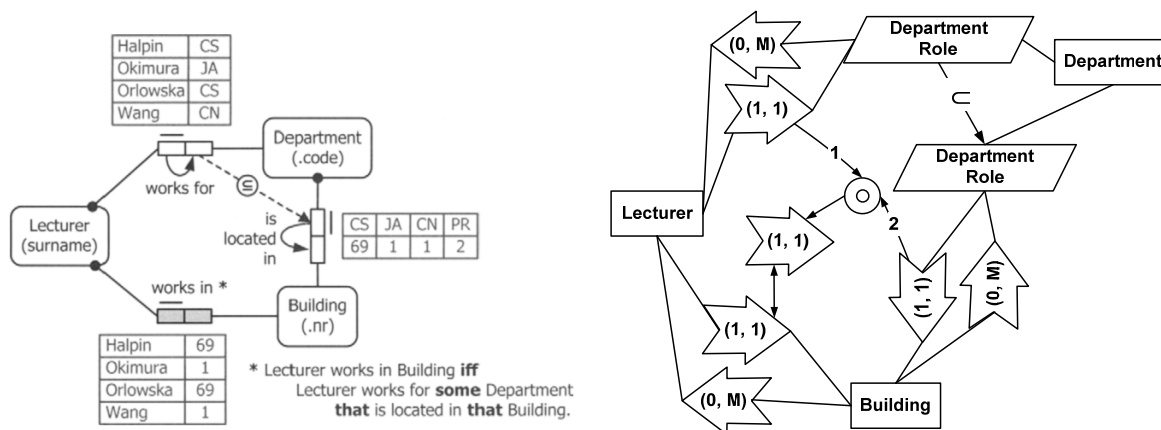


Рис. 3. Фрагменты OR-схемы и ERM-схемы ПрО «Университет»

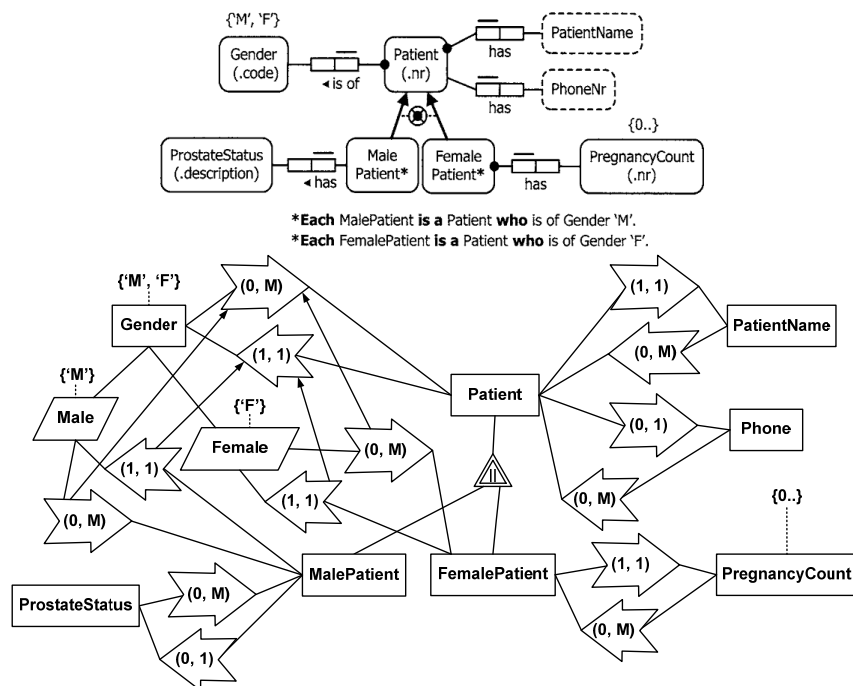
Как видно из рис. 3, в ERM-модели есть универсальное отношение включения между классами объектов (наряду с ним используются отношения равенства и непересекаемости), в данном случае оно применено к реальным областям определения отображений, которые олицетворяют параллелограммы ролей множества сущностей Department.

На этом же рисунке показана особенность данных, не имеющая прямого отношения к сравнению множеств и вообще к ОЦ. Заштрихованными прямоугольниками ролей в OR-модели показываются так называемые получаемые ассоциации. Дополнительно их прочтение снабжается знаком «*». В примечании справа внизу дается определение этого предиката через логическое выражение, включающее базовые (хранимые) предикаты. В отличие от OR-модели в ERM-модели декларация подобных элементов схемы также представляется в графической нотации с использованием производных отображений, получаемых из других отображений с помощью операций (в примере использована композиция). Построенное таким образом отображение соединяется дугами отношений следствия, эквивалентности (как в данном случае) и несовместности с другими отображениями.

5. Ограничения на специализации

Взаимно-исключительные подтипы показываются в ORM использованием символа исключения «X» в кружке и точечными линиями, идущими от него к релевантным связям подтипов. Отсутствие символа исключения указывает, что экстенционалы типов могут пересекаться.

Для того чтобы указать, что два или более подтипов являются исчерпывающими для их супертипа (т.е. их объединение равняется супертипу), используется символ ограничения включающего «или» (точка в кружке), который связывается точечными линиями с релевантными связями подтипов.



Этот же рисунок иллюстрирует и другой класс ОЦ – на значения для типов объектов и ролей. Так, практически схожими элементами диаграмм в обеих моделях показываются допустимые значения ('M' и 'F' для Gender), интервалы допустимых значений («0..» для PregnancyCount) типов объектов. Можно также указывать аналогичные ОЦ ролей ('M' для Male) и ('F' для Female).

«Чтобы указать, что каждое значение в столбце факта должно появиться там точно n раз, число n записывается рядом с ролью. Это пример простого ограничения частоты (также называемого ограничением частоты появления объекта в роли).

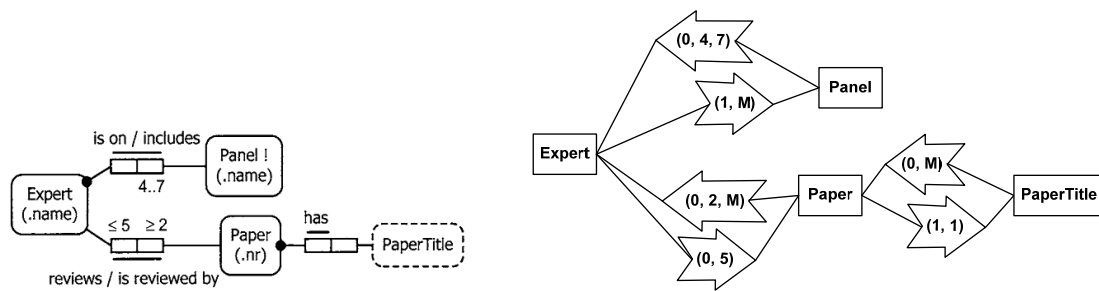


Рис. 5. Фрагменты OR-схемы и ERM-схемы ПрО «Конференция»

Максимальное кардинальное число отображения – это наибольшее из кардинальных чисел элементов ООД. На диаграмме эти числа последовательно записываются в стрелке отображения. В случае, когда реальное минимальное кардинальное число отображения равно 1, оно может быть опущено.

7. Кольцевые ограничения

Ограничения целостности этого типа довольно редко встречаются в моделях данных. Пожалуй, никто, кроме Халпина, их не рассматривает. Тем не менее определенный интерес они представляют.

Когда обе роли в бинарном предикате исполняют объекты одного и того же типа, путь от типа объектов через пару ролей и назад к типу объектов формирует кольцо. Если роли играют подтипами с общим супертипом, путь от одного из подтипов через предикат и супертип к самому себе также формирует кольцо. Обычно пару таких ролей формирует бинарный предикат, но они могут также быть частью более длинного предиката или даже могут браться из различных предикатов по пути концептуального соединения. Кольцевые ограничения могут применяться только к таким парам ролей [1].

Халпин рассматривает следующие виды кольцевых ограничений бинарных отношений:

- рефлексивность на экстенсionale;
- полная рефлексивность;
- симметричность;
- транзитивность;
- нерефлексивность;
- асимметричность;
- антисимметричность;
- нетранзитивность;
- ацикличность.

Приведем примеры некоторых из них.

Рассмотрим ПрО, где каждый, кто любит человека, также любим этим человеком. Если это всегда имеет место, мы говорим, что предикат «любит» симметричен (рис. 6) [1].

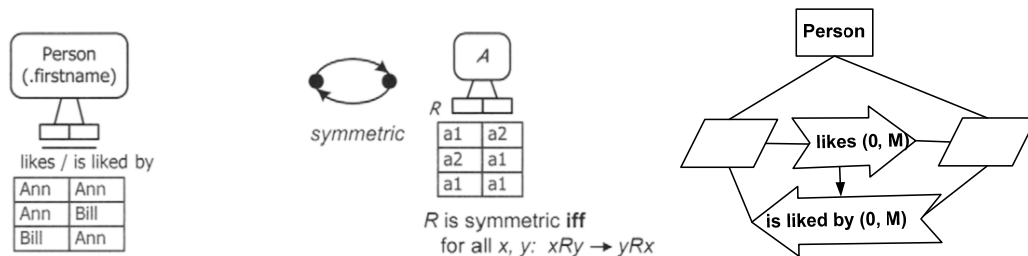


Рис. 6. Фрагменты OR-схемы и ERM-схемы с симметричным кольцевым ограничением

Выражение этого ограничения в ERM-модели представляет собой «практически дословный» перевод приведенного на рис. 6 логического условия симметричности и показано в правой части рисунка. С помощью отношения следствия между отображениями предиката «любит» мы как раз определяем включение экстенсionale прямого отображения в экстенсionale обратного отображения.

Теперь рассмотрим ПрО, где если один человек любит второго и второй человек любит третьего, то первый человек должен любить третьего. В этом случае мы говорим, что предикат «любит» транзитивен (рис. 7) [1].

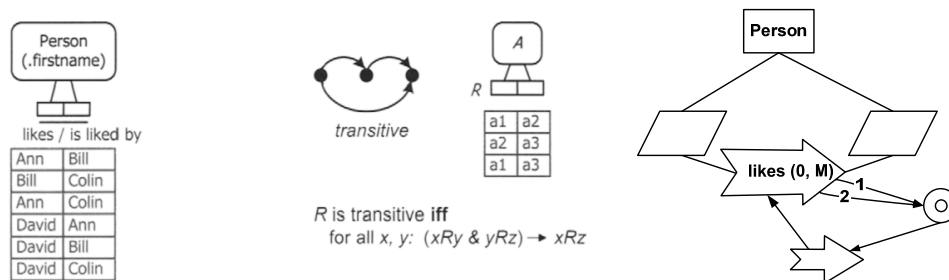


Рис. 7. Фрагменты OR-схемы и ERM-схемы с транзитивным кольцевым ограничением

ERM-диаграмма, приведенная на рис. 7 справа, выражает отношение транзитивности как тот факт, что отображение «любит» является следствием композиции этого отображения с самим собой.

Рассмотрим предикат «являться родителем» (рис. 8). Если мы принимаем первую строку таблицы как истинный факт, мы должны отклонить вторую. Если Энн – родитель Билла, то не может быть верно, что Билл является родителем Энн. Отношение «родитель» асимметрично [1].

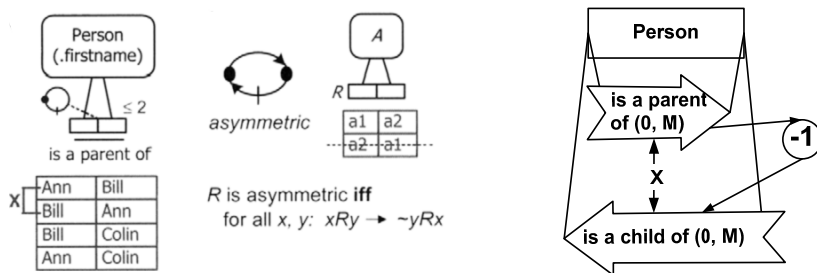


Рис. 8. Фрагменты OR-схемы и ERM-схемы с асимметричным кольцевым ограничением

Как видно на ERM-диаграмме, приведенной в правой части рисунка, ограничение асимметричности отношения определяется как несовместность отображения «являться родителем» и обратного ему отображения «являться ребенком».

Заключение

Понимая необходимость построения выразительной семантической модели данных, авторы ORM и ERM справедливо уделяют большое внимание правилам задания ОЦ, с помощью которых в схеме данных декларируются бизнес-правила ПрО. Этот аспект схем данных до сих пор был развит недостаточно, что приводило к неизбежной ручной реализации требуемых бизнес-правил в процедурном виде на языке программирования сервера БД. Если удастся полностью описать ПрО в семантической схеме, можно ставить вопрос об автоматизации преобразования этой схемы в столь же полное описание ПрО на языке СУБД.

Как показывает анализ средств задания ОЦ, проведенный в статье, авторы описанных моделей выбрали разные пути для решения проблемы повышения выразительности схем данных.

Особенностью аппарата ограничений целостности модели Халпина является то, что для любой полезной особенности данных в нем изобретены свои специальный термин и графический примитив, что приводит к громоздкости определения модели и сложности ее применения.

В отличие от OR-модели в ERM-модели предлагается ограниченный универсальный набор конструкций, с помощью которых можно представить многие (если не все) законы взаимоотношения данных. По крайней мере, все типы ограничений целостности OR-модели можно выразить предлагаемыми в ERM-модели средствами.

ЛИТЕРАТУРА

1. Halpin T., Morgan T. Information Modeling and Relational Databases. Second Edition. Morgan Kaufman, 2008. 943 p.
2. Бабанов А.М. Семантическая модель «Сущность – Связь – Отображение» // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2007. № 1. С. 77–91.
3. Цикритзис Д., Лоховски Ф. Модели данных : пер. с англ. М. : Финансы и статистика, 1985. 344 с.
4. Halpin T. ORM 2. On the Move to Meaningful Internet Systems, OTM 2005 Workshops // Springer Lecture Notes in Computer Science. 2005. V. 3762. P. 676–687.
5. Halpin T., Curland M. Recent Enhancements to ORM. On the Move to Meaningful Internet Systems, OTM 2013 Workshops // Springer Lecture Notes in Computer Science. 2013. V. 8186. P. 467–476.
6. Halpin T., Curland M. Enriched Support for Ring Constraints. On the Move to Meaningful Internet Systems, OTM 2011 Workshops // Springer Lecture Notes in Computer Science. 2011. V. 7046. P. 309–318.

Бабанов Алексей Михайлович, канд. техн. наук, доцент. E-mail: babanov2000@mail.ru
Томский государственный университет

Поступила в редакцию 2 июля 2014 г.

Babanov Alexey M. (Tomsk State University, Russian Federation).

Constraint specifications generating rules in semantic models ORM and ERMM.

Keywords: semantic data model; OR-model; ERM-model; integrity constraints.

Any application of database (DB) technology needs DB scheme designing. If an application domain (AD) is complex, its definition in language of the database management system (DBMS) is not easy task. In such cases, designers usually resort to a semantic design technique providing initial formalization of AD information within the framework of a semantic model and its subsequent transformation in DBMS data model.

Widespread semantic models, such as ER and EER, possess insufficiently expressive means of AD semantics description. So, such DB scheme contains only definitions of data structure and declarative integrity constraints, which can be presented in DBMS Data Definition Language. In this case, the designer has to subsequently append the missing elements of DB scheme (views, triggers, etc.) in the language of DB server and then again to analyze AD.

Thus, very important task is entrusted to semantic data models; this is an initial formal presentation of AD information. This task becomes complicated if we want to obtain high-level but complete description of AD semantics.

Among three data model components (structure data rules, integrity constraint (IC) rules, data operations) an IC instruments make an important contribution to solution of the complete AD scheme design task. Although constraint role in semantics fixation task is rare recollected; all people say about their participation in solution of actual DB state support problem. Tschritzis and Lochovsky define IC purpose in the following way: «Constraints are introduced in data models with a view to enhancement of their semantic expressiveness and to enhancement of data integrity support. The first aspect concerns adequacy of real world reflection in data scheme and the second – ability of DBMS to guarantee correspondence of database states to requirements, expressed by constraints».

Therefore, the modern situation with this component of semantic data models is of interest. In this article, the facilities of integrity constraint declarative definition are analyzed by the examples of two more expressive semantic data models – «Object – Role» (OR-model) and «Entity – Relationship – Mapping» (ERM-model).

Authors of these models understand the necessity of construction of an expressive semantic data model and so they attend to constraint definition rules, with help of which AD business rules are declared in data scheme. Heretofore, this aspect was insufficiently developed and manual realizations of many business rules in DBMS programming language were required. If complete high-level description of AD semantics is a success, a question about an automatization of transformation such schemata to complete DBMS schemata may bring up.

Following OR-constraint classes are the basis of the article's content:

- internal and external role uniqueness constraints,
- mandatory role constraints,
- inclusive-or constraints,
- exclusion constraints,
- exclusive-or constraints,
- set comparison constraints (subset, equality and exclusion),
- subtyping constraints,
- occurrence frequencies constraints,
- ring constraints.

Each IC class is illustrated by example from Halpin's book; the same example is presented in ERM notation. Thus, underlying model principles are clearly demonstrated. Analysis, carried out in article, shows that described model's authors chose different ways to solution of the semantic expressiveness enhancement problem.

Feature of Halpin's model IC instruments is that special term and graphic primitive are contrived for any useful property of data. Model unhandiness and complexity of its use are results of such approach.

Unlike OR-model ERM-model offers limited set of universal constructions, with help of which many (if not all) data relation rules can be presented. At least all kinds of OR-model IC can be expressed by ERM IC tools. The ERM basis is formed by universal concepts of logic and mathematics.

REFERENCES

1. Halpin T., Morgan T. *Information Modeling and Relational Databases*. Burlington: Morgan Kaufman, 2008. 943 p.
2. Babanov A.M. Semantic Model “Entity – Relationship – Mapping”. *Tomsk State University Journal of Control and Computer Science*, 2007, no. 1(1), pp. 77-91. (In Russian).
3. Tsichritzis D., Lochovsky F. *Modeli dannykh* [Data Models]. Translated from English. Moscow: Finansy i Statistika Publ, 1982.
4. Halpin T. On the Move to Meaningful Internet Systems. OTM 2005 Workshops. *Springer Lecture Notes in Computer Science*, 2005, vol. 3762, pp. 676-87.
5. Halpin T., Curland M. Recent Enhancements to ORM. On the Move to Meaningful Internet Systems. OTM 2013 Workshops. *Springer Lecture Notes in Computer Science*, 2013, vol. 8186, pp. 467-476.
6. Halpin T., Curland M. Enriched Support for Ring Constraints. On the Move to Meaningful Internet Systems. OTM 2011 Workshops. *Springer Lecture Notes in Computer Science*, 2011, vol. 7046, pp. 309-318.

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

УДК 519.7

А.С. Твардовский, Н.В. Евтушенко

К МИНИМИЗАЦИИ АВТОМАТОВ С ВРЕМЕННЫМИ ОГРАНИЧЕНИЯМИ

Рассматривается решение задачи минимизации для автомата, поведение которого зависит от времени поступления входного символа. Показано, что для полностью определённого автомата с временными ограничениями существует единственная (с точностью до изоморфизма) минимальная форма. Предлагается алгоритм минимизации полностью определённого автомата с временными ограничениями на основе построения разбиений k -эквивалентных состояний.

Ключевые слова: временной автомат; эквивалентность; минимальная форма; k -эквивалентные состояния.

Модели конечного автомата широко используются при анализе и синтезе дискретных систем [1]. Однако в ряде случаев при описании поведения реальных систем необходимо учитывать временные аспекты, и, соответственно, понятие конечного автомата расширяется добавлением таймаутов и различных временных ограничений [2, 3]. В данной статье мы используем понятие временного автомата из [3], в котором вводятся временные ограничения при подаче входных символов с добавлением возможных задержек по обработке входных символов. Поскольку сложность решения многих задач в теории автоматов существенно зависит от числа состояний рассматриваемого автомата, большое внимание уделяется задаче минимизации автомата, т.е. построению эквивалентного автомата с наименьшим числом состояний.

В работе [3] авторы предлагают описывать поведение временного автомата с использованием перехода к классическому конечному автомату. Соответственно, такой переход можно использовать для минимизации временного автомата, однако число входных и выходных символов классического автомата значительно возрастает по сравнению с временным автоматом. В настоящей работе мы предлагаем метод минимизации временного автомата без перехода к такой абстракции. Подобно [1], для минимизации временного автомата мы используем разбиение множества состояний автомата на классы k -эквивалентных состояний; по определению, состояния k -эквивалентны, если выходные реакции автомата в этих состояниях совпадают на любую входную последовательность длины не больше k . Мы также показываем, что минимальная форма временного детерминированного полностью и правильно определённого автомата единственна с точностью до изоморфизма.

1. Основные определения и обозначения

В данной работе под *временным автоматом* понимается четвёрка $S = (I, S, O, h_s)$, где I – входной алфавит, O – выходной алфавит, S – конечное непустое множество состояний, $h_s \subseteq (S \times I \times O \times S \times \Pi \times Z)$ – отношение переходов, Π – множество интервалов из промежутка $[0; \infty)$ и Z – множество целых неотрицательных чисел. Соответственно, *кортеж* (s, i, o, s', g, d) описывает переходы из состояния s в состояние s' под действием входного символа i , поступившего в момент времени t , $t \in g$, после перехода автомата в текущее состояние с выдачей выходного символа o через d тактов времени после поступления входного символа.

Временной автомат S называется *правильно определённым*, если для любых двух кортежей (s, i, o, s', g_1, d) , $(s, i, o, s', g_2, d) \in h_s$, справедливо, что g_1 и g_2 нельзя объединить в один интервал. Правильно определённый автомат достаточно просто может быть получен из произвольного временного автомата

выполнением всех возможных объединений кортежей, поэтому далее мы рассматриваем только правильно определённые автоматы. Если для любых двух кортежей $(s, i, o_1, s_1, g_1, d_1), (s, i, o_2, s_2, g_2, d_2) \in h_s$ справедливо соотношение $g_1 \cap g_2 = \emptyset$, то временной автомат называется *детерминированным*.

Временной автомат называется *полностью определённым по входным символам*, если для любой пары $(s, i) \in S \times I$, т.е. для любого входного символа i , поступающего на вход автомата в состоянии s , существует кортеж $(s, i, o, s', g, d) \in h_s$. Временной автомат называется *полностью определённым*, если автомат полностью определён по входным символам и для каждой пары $(s, i) \in S \times I$ объединение всех временных интервалов в состоянии s по входному символу i равно $[0; \infty)$. В настоящей статье мы рассматриваем задачу минимизации только для полностью определённых детерминированных временных автоматов, которые являются правильно определёнными.

Временным входным символом называется пара (i, t) , где i – символ входного алфавита, t – время поступления входного символа после перехода автомата в текущее состояние. *Временным выходным символом* называется пара (o, d) , где o – символ выходного алфавита, d – число тактов между подачей входного символа и выдачей выходного символа. В автомате существует переход $s \rightarrow (i, t)/(o, d) \rightarrow s'$, если и только если существует кортеж $(s, i, o, s', g, d) \in h_s$, такой что $t \in g$. Пусть s – состояние полностью определённого детерминированного временного автомата S , и (i, t) – временной входной символ. Тогда (i, t) -*преемником* состояния s будем называть состояние s' , такое что во множестве h_s существует элемент $(s, i, o, s', g, d), t \in g$. Для состояния s и входного символа i автомата S мы далее через $\Pi_{(s, i)}$ обозначаем множество интервалов g для всех кортежей $(s, i, o, s', g, d) \in h_s$. Будем говорить, что интервалу $g \in \Pi_{(s, i)}$ соответствует выходной символ (o, d) и (i, t) -*преемник* s' , $t \in g$, если во множестве h_s существует элемент (s, i, o, s', g, d) .

Последовательность временных входных символов $(i_1, t_1), (i_2, t_2) \dots (i_n, t_n)$ называется *временной входной последовательностью*; последовательность временных выходных символов $(o_1, d_1), (o_2, d_2) \dots (o_n, d_n)$ называется *временной выходной последовательностью*. Временная выходная последовательность, соответствующая поступившей на автомат, находящийся в состоянии s , временной входной последовательности α , называется *(выходной) реакцией* автомата в состоянии s на последовательность α .

Состояния s и p полностью определённых детерминированных временных автоматов S и P называются *k-эквивалентными*, если реакции автоматов в этих состояниях совпадают на любую временную входную последовательность длины k . Если состояния s и p не являются *k-эквивалентными*, то они называются *k-различимыми*. Состояния s и p полностью определённых временных автоматов S и P называются *эквивалентными*, если эти состояния *k-эквивалентны* для любого k , т.е. реакции автоматов в этих состояниях на любую временную входную последовательность совпадают. Аналогичным образом определяются отношения эквивалентности и *k-эквивалентности* для состояний одного автомата. Временной автомат называется *приведённым*, если любые состояния автомата попарно различимы.

Отношение *k-эквивалентности* на множестве состояний временного автомата индуцирует разбиение на множестве состояний, которое называется *k-разбиением* множества состояний временного автомата и обозначается E_k . Любые два состояния, принадлежащие одному классу разбиения E_k , являются *k-эквивалентными*; любые два состояния, принадлежащие различным классам разбиения E_k , являются *k-различимыми*. Через E далее обозначается разбиение множества состояний временного автомата S , соответствующее отношению эквивалентности. Так же как и для классических конечных автоматов, эквивалентные состояния во временном автомате удовлетворяют следующему свойству.

Утверждение 1. Если во временных автоматах S и P существуют цепочки переходов $s \rightarrow (i_1, t_1)/(o_1, d_1), \dots, (i_l, t_l)/(o_l, d_l) \rightarrow s', p \rightarrow (i_1, t_1)/(o_1, d_1), \dots, (i_l, t_l)/(o_l, d_l) \rightarrow p'$, в которых состояния s и p являются эквивалентными, то состояния s' и p' тоже являются эквивалентными.

Доказательство. Пусть состояния s и p временных автоматов S и P эквивалентны, т.е. реакции автоматов в этих состояниях на любую временную входную последовательность $(i_1, t_1), (i_2, t_2), \dots, (i_k, t_k)$, $k \geq 1, \dots$ совпадают. Тогда в состояниях s' и p' , таких что $s \rightarrow (i_1, t_1)/(o_1, d_1), \dots, (i_l, t_l)/(o_l, d_l) \rightarrow s', p \rightarrow (i_1, t_1)/(o_1, d_1), \dots, (i_l, t_l)/(o_l, d_l) \rightarrow p'$, реакции автоматов на любую временную входную последовательность также совпадают. Таким образом, состояния s' и p' эквивалентны.

Далее мы показываем, что в приведенном правильно определенном временном автомате множества временных интервалов, описывающих поведение автомата при поступлении входного символа, для эквивалентных состояний совпадают при любом входном символе.

Утверждение 2. Если состояния s и p временных детерминированных полностью и правильно определенных приведенных автоматов S и P эквивалентны, то для любого входного символа $i \in I$ справедливо $\Pi_{(s, i)} = \Pi_{(p, i)}$.

Доказательство. Предположим, что множества интервалов $\Pi_{(s, i)}$ и $\Pi_{(p, i)}$ в состояниях s и p для входного символа i не совпадают. Тогда существуют $[a_1, b_1] \subseteq \Pi_{(s, i)}$ и $[a_2, b_2] \subseteq \Pi_{(p, i)}$, такие что $[a_1, b_1] \neq [a_2, b_2]$ и $[a_1, b_1] \cap [a_2, b_2] \neq \emptyset$. Рассмотрим временной входной символ (i, t) , где $t \in [a_1, b_1] \cap [a_2, b_2]$, и соответствующие ему переходы $s \rightarrow (i, t)/(o_1, d_1) \rightarrow s'$, $p \rightarrow (i, t)/(o_2, d_2) \rightarrow p'$. Если $o_1 \neq o_2$, $d_1 \neq d_2$ или состояния s' и p' не являются эквивалентными, то состояния s и p также не являются эквивалентными в силу определения отношения эквивалентности и утверждения 1. Пусть теперь $o_1 = o_2$, $d_1 = d_2$ и состояния s' и p' эквивалентны. Тогда с учётом $[a_1, b_1] \neq [a_2, b_2]$ и в силу полной определённости автоматов существует интервал $[a_3, b_3]$, такой что либо $[a_3, b_3] \cap [a_2, b_2] \neq \emptyset$, $[a_3, b_3] \cap [a_1, b_1] = \emptyset$ и $[a_3, b_3] \cup [a_1, b_1]$ образуют интервал, либо $[a_3, b_3] \cap [a_1, b_1] \neq \emptyset$, $[a_3, b_3] \cap [a_2, b_2] = \emptyset$ и $[a_3, b_3] \cup [a_2, b_2]$ образует интервал. Выберем t таким образом, что $t \in [a_3, b_3] \cap [a_2, b_2]$ ($t' \in [a_3, b_3] \cap [a_1, b_1]$). Тогда для переходов $s \rightarrow (i, t)/(o_3, d_3) \rightarrow s''$, $p \rightarrow (i, t)/(o_2, d_2) \rightarrow p'$ ($s \rightarrow (i, t')/(o_1, d_1) \rightarrow s'$ и $p \rightarrow (i, t')/(o_3, d_3) \rightarrow p''$) в силу правильной определённости автомата S (автомата P), $o_2 \neq o_3$, $d_2 \neq d_3$, или $s'' \neq s'$ ($o_1 \neq o_3$, $d_1 \neq d_3$ или $p'' \neq p'$). Таким образом, ввиду приведённости автоматов S и P состояния s' и s'' (p' и p'') не являются эквивалентными, т.е. состояния s'' и p' (s' и p'') также не являются эквивалентными, а значит, состояния s и p не являются эквивалентными.

Следствие. Если для состояний s и p полностью и правильно определённых приведенных временных автоматов S и P существует входной символ i , такой что $\Pi_{(s, i)} \neq \Pi_{(p, i)}$, то состояния s и p различимы.

Два временных автомата S и P эквивалентны, если для каждого состояния автомата S существует эквивалентное состояние в автомате P и для каждого состояния автомата P существует эквивалентное состояние в автомате S .

Минимальной (или приведенной) формой полностью определённого детерминированного временного автомата S называется приведенный правильно определенный временной автомат, эквивалентный S .

Временные автоматы S и P с одинаковыми входными и выходными алфавитами *изоморфны*, если между множествами состояний этих автоматов можно установить взаимно однозначное соответствие $H: S \rightarrow P$, такое что элемент $(s_i, i, o, s_j, g, d) \in h_s$, если и только если элемент $(H(s_i), i, o, H(s_j), g, d) \in h_p$.

2. Алгоритм минимизации временного автомата

В настоящем разделе описывается предлагаемый алгоритм минимизации временных автоматов на основе построения классов k -эквивалентных состояний.

Алгоритм построения приведенной формы временного детерминированного полностью и правильно определённого автомата.

Вход: временной детерминированный полностью и правильно определённый автомат S .

Выход: приведенный автомат B , эквивалентный S .

Шаг 1: Состояния s_1 и s_2 принадлежат одному блоку разбиения E_1 , если и только если для любого входного символа $i \in I$ и любых кортежей $(s_1, i, o_1, s'_1, g_1, d_1)$, $(s_2, i, o_2, s'_2, g_2, d_2) \in h_s$, таких что $g_1 \cap g_2 \neq \emptyset$, справедливо $o_1 = o_2$ и $d_1 = d_2$;

$k := 1$.

/* Разбиение E_1 есть разбиение множества состояний автомата на 1-эквивалентные состояния.

Шаг 2: Состояния s_1 и s_2 из одного блока разбиения E_k принадлежат одному блоку разбиения E_{k+1} , если и только если для любого входного символа $i \in I$ и любых кортежей $(s_1, i, o, s'_1, g_1, d)$, $(s_2, i, o, s'_2, g_2, d) \in h_s$, таких что $g_1 \cap g_2 \neq \emptyset$, состояния s'_1 и s'_2 принадлежат одному блоку разбиения E_k .

/* Разбиение E_k есть разбиение множества состояний автомата на k -эквивалентные состояния.

Если $E_{k+1} = E_k$, то $E = E_k$ и **Шаг 3**.

/* Разбиение E есть разбиение множества состояний автомата на эквивалентные состояния.

Иначе $k++$ и **Шаг 2**.

Шаг 3. Входной и выходной алфавиты автомата V совпадают с таковыми для автомата S , состояния $b_1, b_2, \dots, b_l$ соответствуют блокам $B_1, B_2, \dots, B_l$ разбиения E . Множество переходов h_B автомата V формируем следующим образом: для состояний b_i и b_j существует кортеж $(b_i, i, o, b_j, g, d) \in h_B$ тогда и только тогда, когда существуют $s_i \in B_i$ и $s_j \in B_j$, такие что кортеж $(s_i, i, o, s_j, g_1, d) \in h_s$.

КОНЕЦ алгоритма.

Утверждение 3. Для любого временного детерминированного полностью и правильно определенного автомата S с n состояниями вышеописанный алгоритм (шаги 1 и 2) доставляет разбиение E множества состояний автомата S на эквивалентные состояния не более чем за n шагов.

Доказательство. Поскольку автомат S имеет n состояний и при построении разбиения E_k (шаг 2) по крайней мере один блок разбиения $E_{(k-1)}$ разбивается на два блока разбиения E_k , то при некотором $k < n$ получим $E_{k+1} = E_k$.

Докажем справедливость построения разбиения E_1 на шаге 1 алгоритма. Рассмотрим два состояния s_1 и s_2 из одного блока разбиения E_1 , временной входной символ (i, t) и кортежи $(s_1, i, o_1, s'_1, g_1, d_1)$, $(s_2, i, o_2, s'_2, g_2, d_2) \in h_s$, $t \in g_1 \cap g_2$. По построению E_1 , $o_1 = o_2$ и $d_1 = d_2$, так как состояния являются 1-эквивалентными. Если состояния s_1 и s_2 не принадлежат одному блоку разбиения E_1 , то существует временной входной символ (i, t) и кортежи $(s_1, i, o_1, s'_1, g_1, d_1)$, $(s_2, i, o_2, s'_2, g_2, d_2) \in h_s$, $t \in g_1 \cap g_2$, такие что $o_1 \neq o_2$ и $d_1 \neq d_2$, т.е. состояния s_1 и s_2 не являются 1-эквивалентными.

Пусть $k > 0$. Рассмотрим два состояния s_1 и s_2 из одного блока разбиения E_k , входной символ i , кортежи $(s_1, i, o_1, s'_1, g_1, d_1)$, $(s_2, i, o_2, s'_2, g_2, d_2) \in h_s$ и $t \in g_1 \cap g_2$. По построению E_k , $o_1 = o_2$ и $d_1 = d_2$, так как состояния s_1 и s_2 являются 1-эквивалентными. Поскольку s'_1 и s'_2 принадлежат одному блоку разбиения E_k , то в этих состояниях автомат имеет одинаковые выходные реакции на любую входную временную последовательность длины k и в состояниях s_1 и s_2 автомат имеет одинаковые выходные реакции на любую входную временную последовательность длины $(k+1)$, т.е. состояния s_1 и s_2 являются $(k+1)$ -эквивалентными.

Если состояния s'_1 и s'_2 не принадлежат одному блоку разбиения E_k , то существует входная временная последовательность α длины k , на которую отличаются реакции автомата в этих состояниях. Таким образом, автомат имеет различные выходные реакции на входную последовательность $(i, t). \alpha$ и, соответственно, состояния s_1 и s_2 не являются $(k+1)$ -эквивалентными.

Утверждение 4. Пусть V – автомат, построенный для автомата S по вышеописанному алгоритму. Автомат V является детерминированным полностью и правильно определенным автоматом, и состояние b_i автомата V эквивалентно состоянию s автомата S , если и только если $s \in B_i$.

Доказательство. Автомат V является полностью и правильно определенным автоматом, поскольку автомат S обладает этими свойствами. Детерминированность автомата V следует из того факта, что любые два состояния s, s' , принадлежащие одному блоку B_i , являются эквивалентными, т.е. для любого временного входного символа (i, t) справедливо: 1) $\Pi_{(s, i)} = \Pi_{(s', i)}$; 2) каждому интервалу $g \in \Pi_{(s, i)}$ соответствует один и тот же выходной символ (o, d) и 3) (i, t) -преемники s и s' принадлежат одному блоку разбиения E . Свойство 1 следует из утверждения 1; свойства 2 и 3 следуют из того факта, что эквивалентные состояния являются k -эквивалентными при любом $k > 0$.

Покажем, что состояния $s_i \in B_i$ и b_i эквивалентны. Для любого кортежа $(s_i, i, o, s'_i, g_i, d) \in h_s$ существует кортеж $(b_i, i, o, b'_i, g'_i, d) \in h_B$, где $g_i \subseteq g'_i$ в силу условий формирования h_B , т.е. реакции автоматов V и S в состояниях b_i и s_i на любой временной входной символ совпадают. А поскольку $s'_i \in B'_i$, то для состояний s'_i и b'_i и для всех последующих их (i, t) -преемников можем повторить аналогичные рассуждения. Таким образом, реакции автоматов V и S в состояниях b_i и s_i на любую входную последовательность совпадают. Покажем теперь, что состояния $s_j \in B_j$ и b_i , где $i \neq j$, не эквивалентны. Состояние $s_i \in B_i$ эквивалентно состоянию b_i по первой части доказательства. Состояния s_i и s_j не эквивалентны, так как лежат в разных блоках E -разбиения. Таким образом, состояния s_j и b_i не эквивалентны.

Следствие 1. Автомат В, построенный по вышеописанному алгоритму, есть приведённый автомат.

Следствие 2. Автомат В, построенный по вышеописанному алгоритму, эквивалентен автомату S. Таким образом, на основании следствий 1 и 2 справедлива следующая теорема.

Теорема 1. Пусть S – полностью определённый детерминированный временной автомат. Полностью определённый детерминированный временной автомат В, построенный по вышеописанному алгоритму, является приведённой формой автомата S.

Пример. Для демонстрации полученных алгоритмов рассмотрим временной автомат S (рис. 1), в котором $I = \{i_1, i_2\}$, $O = \{o_1, o_2\}$, $S = \{s_1, s_2, s_3, s_4, s_5\}$. Построим разбиение E множества состояний автомата S на эквивалентные состояния, используя вышеописанный алгоритм.

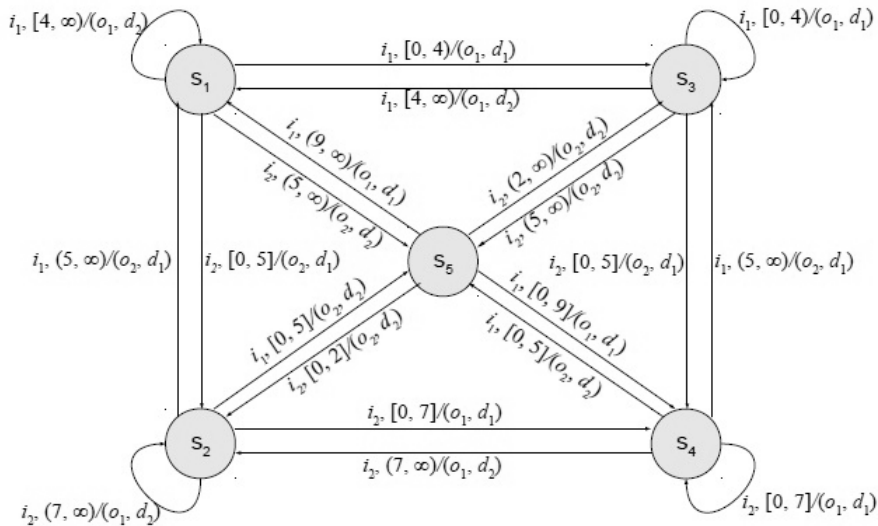


Рис. 1. Автомат S

Шаг 1. Построим разбиение E_1 на 1-эквивалентные состояния. Непосредственной проверкой можно убедиться, что состояния s_1 и s_3 и s_2 и s_4 являются 1-эквивалентными, т.е. $E_1 = \{\{s_1, s_3\}, \{s_2, s_4\}, \{s_5\}\}$.

Шаг 2. Проанализировав переходы из состояний s_1 и s_3 (s_2 и s_4), убеждаемся, что для любого временного входного символа (i, t) соответствующие (i, t) -преемники находятся в одном блоке разбиения E_1 .

Шаг 3. Поскольку $E_1 = E_2$, то разбиение E_1 есть разбиение E состояний автомата S на эквивалентные состояния.

Шаг 4. Построим по разбиению E приведенную форму автомата S, поставив в соответствие блоку $\{s_1, s_3\}$ состояние b_1 приведенной формы, блоку $\{s_2, s_4\}$ – состояние b_2 , блоку $\{s_5\}$ – состояние b_3 . Например, кортежу $(s_3, i_2, o_2, s_4, [0, 5], d_1)$ исходного автомата соответствует кортеж $(b_1, i_2, o_2, b_2, [0, 5], d_1)$ приведенной формы. В результате получим приведенную форму В автомата S (рис. 2).

Далее покажем, что две приведенные формы временного детерминированного полностью и правильно определенного приведенного автомата изоморфны, т.е. приведенная форма такого автомата единственна с точностью до изоморфизма.

Теорема 2. Детерминированные полностью и правильно определенные приведенные временные автоматы S и P эквивалентны, если и только если эти автоматы изоморфны.

Необходимость. Пусть S и P – детерминированные полностью и правильно определённые приведенные временные автоматы, которые являются эквивалентными. Покажем, что автоматы S и P изоморфны. Рассмотрим отображение $H: S \rightarrow P$, при котором $H(s)$ есть состояние автомата P, эквивалентное состоянию s . Отображение H является взаимно однозначным в силу приведённости автоматов S и P. Кроме того, в силу эквивалентности S и P и утверждений 1 и 2 выполняется следующее свойство. Для любого кортежа (s, i, o, s', g, d) , $s \in S$, $i \in I$, $g \in \Pi_{(s, i)}$ существует кортеж $(H(s), i, o, H(s'), g, d)$. Таким образом, автоматы S и P изоморфны.

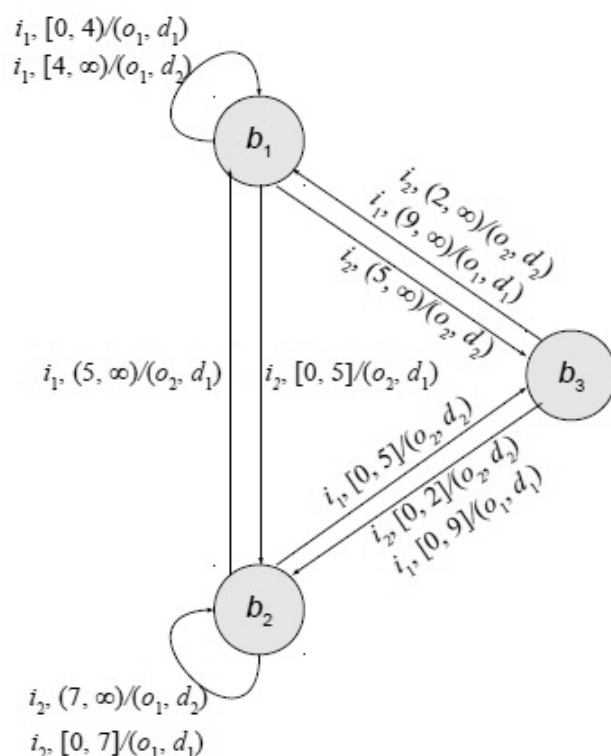


Рис. 2. Приведённая форма В автомата S

Достаточность. Поскольку изоморфные автоматы совпадают с точностью до переименования состояний, то изоморфные автоматы эквивалентны.

Пусть Q и P – приведенные формы детерминированного полностью и правильно определённого временного автомата S. Поскольку отношение эквивалентности транзитивно, то автоматы Q и P эквивалентны, т.е. изоморфны. Таким образом, справедливо следующее утверждение.

Следствие из теоремы 2. Любые две приведенные формы детерминированного полностью и правильно определённого временного автомата S изоморфны, т.е. приведенная форма детерминированного полностью и правильно определённого временного автомата S единственна с точностью до изоморфизма.

Заключение

В настоящей статье предложен алгоритм минимизации детерминированного полностью и правильно определённого временного автомата. Показано, что приведенная (минимальная) форма такого автомата единственна с точностью до изоморфизма. Тот факт, что разработанные алгоритмы принимают на вход только правильно определённые временные автоматы, не ограничивает применимость разработанного алгоритма, поскольку произвольный автомат преобразуется в эквивалентный правильно определённый автомат объединением соответствующих переходов. Следует отметить, что в случае, когда в автомате, кроме временных ограничений, допускаются тайм-ауты [4], приведенные формы автомата не обязательно являются изоморфными [5].

ЛИТЕРАТУРА

1. Гилл А. Введение в теорию конечных автоматов. М. : Наука, 1966. 272 с.
2. Кондратьева О.В. Разработка методов синтеза и анализа композиций временных автоматов : магистерская дис. на соискание степени магистра радиофизики. Томск, 2012. 72 с.
3. El-Fakih K., Gromov M., Shabaldin N., Yevtushenko N. Distinguishing Experiments for Timed Nondeterministic Finite State Machines // Acta Cybernetica. 2013. № 2. С. 205–222.

4. Bresolin D., El-Fakih K., Villa T., Yevtushenko N. Timed Finite State Machines: Equivalence checking and expressive power // Proc. of the 5th Intern. Symposium on Games, Automata, Logics and Formal Verification, CANDALF'2014. P. 203–216.
5. Кондратьева О.В. Минимизация временных автоматов с таймаутами // Материалы конференции «Новые информационные технологии в исследовании сложных структур». Томск : Издательский дом ТГУ, 2014. С. 132.

Твардовский Александр Сергеевич. E-mail: tvardal@mail.ru

Евтушенко Нина Владимировна, д-р техн. наук, профессор. E-mail: yevtushenko@sibmail.com

Поступила в редакцию 5 августа 2014 г.

Tvardovskiy Alexander S., Yevtushenko Nina V. (Tomsk State University, Russian Federation).

Minimizing Timed Finite State Machines.

Keywords: Timed Finite State Machine (TFSM); equivalence relation; minimal (reduced) form; k -equivalent states.

Finite State Machines (FSM) are widely used for analysis and synthesis of discrete event systems. However, for systems for which timed aspects are important, the notion of a classical FSM has to be extended. One of such extensions is to take into account time guards for inputs and timeouts for outputs. Correspondingly, the behavior of such FSM depends on a time instance when an input is applied; and moreover, the output timeout specifies the time duration that is needed for processing a given input. In this paper, we solve the problem of minimizing such timed FSM (TFSM). We discuss appropriate properties of TFSMs; the notion of k -equivalent and equivalent states and a reduced (minimal) form are introduced. The TFSM minimization is based on the deriving partitions for k -equivalent states which can be refined at the next step. In other words, the clauses of the partition for $(k+1)$ -equivalent states are subsets of those for k -equivalent states. If the refined partition equals the previous one then the derived partition corresponds to the equivalence relation and its clauses can be selected as states of the corresponding reduced TFSM. We show that a reduced form for a TFSM is unique up to isomorphism.

REFERENCES

1. Gill A. *Vvedenie v teoriyu konechnykh avtomatov* [Introduction to the Theory of Finite-State Machines]. Translated from English. Moscow: Nauka Publ., 1966. 272 p.
2. Kondratyeva O. *Razrabotka metodov sinteza i analiza kompozitsiy vremennykh avtomatov* : masterskaya dis. na soiskanie stepeni magistra radiofiziki [Developing methods for the synthesis and analysis of timed finite state machines. Master thesis]. Tomsk: Tomsk State University Publ., 2012. 72 p.
3. El-Fakih K., Gromov M., Shabaldin N., Yevtushenko N. Distinguishing Experiments for Timed Nondeterministic Finite State Machines. *Acta Cybernetica*, 2013, no. 2, pp. 205-222.
4. Bresolin D., El-Fakih K., Villa T., Yevtushenko N. Timed Finite State Machines: Equivalence checking and expressive power. *Proc. of the 5th Intern. Symposium on Games, Automata, Logics and Formal Verification, CANDALF'2014*, pp. 203-216.
5. Kondratyeva O. [Minimization of finite state machines with timeouts]. *Materialy konferentsii "Novye informatsionnye tekhnologii v issledovanii slozhnykh struktur"* [Proc. of the 10th Russian conference with the international participation "Novel information technologies for studying complex structures"]. Tomsk: Tomsk State University Publ., 2014, pp. 132. (In Russian).

СВЕДЕНИЯ ОБ АВТОРАХ

АШУРОВ Михаил Фаритович – аспирант кафедры прикладной информатики, факультет информатики Томского государственного университета. E-mail: therevenge.amf@gmail.com

БАБАНОВ Алексей Михайлович – доцент, кандидат технических наук, доцент кафедры программной инженерии факультета информатики Томского государственного университета. E-mail: babanov2000@mail2000.ru

БЕРЕСТ Олег Борисович – ассистент факультета электроники и информационных технологий Сумского государственного университета. E-mail: Berest_Oleg@mail.ru

ГУМЕНЮК Александр Степанович – доцент, кандидат технических наук, доцент кафедры «Информатика и вычислительная техника» факультета информационных технологий и компьютерных систем Омского государственного технического университета. E-mail: gumas45@mail.ru

ДОВБЫШ Анатолий Степанович – профессор, доктор технических наук, заведующий кафедрой компьютерных наук факультета электроники и информационных технологий Сумского государственного университета. E-mail: kras@id.sumdu.edu.ua

ЕВТУШЕНКО Нина Владимировна – профессор, доктор технических наук, зав. кафедрой информационных технологий в исследовании дискретных структур радиофизического факультета Томского государственного университета. E-mail: yevtushenko@sibmail.com

МУХИНА Оксана Олеговна – аспирантка кафедры прикладной математики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: oksm7@sibmail.com

ПОЗДНИЧЕНКО Николай Николаевич – аспирант кафедры «Информатика и вычислительная техника» факультета информационных технологий и компьютерных систем Омского государственного технического университета. E-mail: nick670@yandex.ru

СМАГИН Валерий Иванович – профессор, доктор технических наук, профессор кафедры прикладной математики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: vsm@mail.tsu.ru

ТВАРДОВСКИЙ Александр Сергеевич – студент радиофизического факультета Томского государственного университета. E-mail: tvardal@mail.ru

ФЁДОРОВА Екатерина Александровна – аспирант факультета прикладной математики и кибернетики Томского государственного университета. E-mail: moiskate@mail.ru

ЧЕТВЕРТАКОВА Евгения Сергеевна – аспирант кафедры программных систем и баз данных Новосибирского государственного технического университета. E-mail: evgenia.chetvertakova@gmail.com

ЧИМИТОВА Екатерина Владимировна – кандидат технических наук, доцент, докторант кафедры программных систем и баз данных Новосибирского государственного технического университета. E-mail: ekaterina.chimitova@gmail.com

ШПЫНОВ Станислав Николаевич – доктор медицинских наук, ведущий научный сотрудник, руководитель лаборатории НИИ эпидемиологии и микробиологии им. Н.Ф. Гамалеи. E-mail: stan63@inbox.ru

ШУМИЛОВ Борис Михайлович – профессор, доктор физико-математических наук, профессор кафедры прикладной математики общеобразовательного факультета Томского государственного архитектурно-строительного университета. E-mail: sbm@tsuab.ru

ЫМАНОВ Улукбек Сайдакромович – старший преподаватель кафедры информационных технологий и автоматизированных систем факультета математики и информационных технологий Омского государственного университета, в настоящее время стажер ТГАСУ по гранту РФФИ № 14-31-50353-мол_нр. E-mail: YmanV8106@rambler.ru

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2014. № 4 (29)

Редактор Н.А. Афанасьева
Оригинал-макет А.И. Лелоюр
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцова

Подписано к печати 17.12.2014 г. Формат 60x84¹/₈.

Гарнитура Times.

Печ. л. 10,7; усл. печ. л. 9,9.

Тираж 250 экз. Заказы № 715.

Журнал отпечатан на полиграфическом оборудовании
Издательского Дома Томского государственного университета
634050, г. Томск, пр. Ленина, 36
Тел. 8+(382-2)–53-15-28
Сайт: <http://publish.tsu.ru>
E-mail: rio.tsu@mail.ru