

# Информатика и её применения

Том 5 Выпуск 4 Год 2011

## СОДЕРЖАНИЕ

Стохастические информационные технологии для исследования нелинейных круговых стохастических систем <b>И. Н. Синицын</b>	2
Система массового обслуживания с ненадежным прибором в дискретном времени <b>А. В. Печинкин, И. А. Соколов</b>	6
Об одном классе марковских систем обслуживания <b>Я. А. Сатин, А. И. Зейфман, А. В. Коротышева, С. Я. Шоргин</b>	18
Максимумы активности в безмасштабных случайных сетях с тяжелыми хвостами <b>А. В. Лебедев</b>	25
Аналитическая модель расчета эффективности плана распределения вычислительных ресурсов многопроцессорной системы при решении специального класса задач <b>М. Я. Агаларов</b>	29
Декомпозиция при частично известном распределении ошибки <b>В. Г. Ушаков, Н. Г. Ушаков</b>	36
Задачи анализа и оптимизации для модели пользовательской активности. Часть 1. Анализ и прогнозирование <b>А. В. Босов</b>	40
Решение проблем взаимодействия с СУБД в кроссплатформенной библиотеке EFFIDB <b>А. В. Янушко, А. В. Бабанин, О. А. Кузнецова, С. В. Петрушенко</b>	53
Вероятностно-статистическая оценка адекватности информационных объектов <b>Л. А. Кузнецов</b>	64
Управление инфокоммуникационными проектами: «своевременность–производительность–информация» <b>А. А. Зацаринный, А. П. Шабанов</b>	76
Моделирование процессов формирования экспертных знаний для мониторинга программно-целевой деятельности <b>И. М. Зацман, А. А. Дурново</b>	84
Трансформационные модели языковых структур для французско-русского машинного перевода <b>Ю. И. Морозова</b>	99
The strategies of syntactic analysis based on head-driven grammars and the methods of their implementation in information systems <b>Е. В. Kozerenko and P. V. Ermakov</b>	107
Abstracts	114
Об авторах	118
About Authors	119
Авторский указатель за 2011 г.	120
2011 Author Index	123

# СТОХАСТИЧЕСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ИССЛЕДОВАНИЯ НЕЛИНЕЙНЫХ КРУГОВЫХ СТОХАСТИЧЕСКИХ СИСТЕМ\*

И. Н. Синицын<sup>1</sup>

**Аннотация:** Статья посвящена стохастическим (корреляционным и спектрально-корреляционным) информационным технологиям аналитического и статистического анализа и моделирования процессов в нелинейных круговых стохастических системах на базе методов круговой статистической линейаризации «намотанным» нормальным распределением. В основу технологий положены методы, алгоритмы и инструментальное программное обеспечение (ИПО) CStS-ANALYSIS в среде MATLAB.

**Ключевые слова:** аналитическое моделирование; круговой стохастический процесс; круговая стохастическая система; круговая статистическая линейаризация; компьютерная поддержка статистических научных исследований; MATLAB; корреляционные уравнения; спектрально-корреляционные уравнения; стохастические информационные технологии; статистическое моделирование

## 1 Введение

Компьютерная поддержка научных исследований (КПНИ) как неотъемлемая часть автоматизации научных исследований становится все более характерным признаком современных научных исследований (НИ) и оказывает сильное влияние на их интенсивность и эффективность, превращается в важнейший фактор дальнейшего прогресса науки [1, 2]. Современный этап развития КПНИ характеризуется интенсивным проникновением ее в новые сферы исследований и разработок, расширением контингента пользователей, охватом всех этапов исследований от сбора и первичной обработки данных, управления экспериментами до анализа и перспективного планирования основных направлений НИ и их информационных технологий.

Под информационной технологией обычно понимают совокупность систематических и массовых способов создания, накопления, обработки, хранения, передачи и распределения информации (данных, знаний) с помощью средств вычислительной техники и связи.

На практике обычно создается ИТ, рассчитанная на выполнение с ее помощью некоторой основной функции, что связано с необходимостью решения нескольких типовых задач исследований. Перечень основных функций довольно ограничен, а с другой стороны, выполнение этих функций может потребоваться во многих применениях. Это делает целесообразным выделение функционально-

ориентированных, предметно-ориентированных и проблемно-ориентированных ИТ [1].

На примере статистических НИ в [1] рассмотрены современные принципы подходы и задачи КПНИ, сформулированы требования к стохастическим ИТ (СтИТ) анализа, моделирования и синтеза оптимальных, субоптимальных и условно-оптимальных фильтров для обработки информации, описано ИПО, а также некоторые приложения. В качестве основных математических моделей в СтИТ принимались стохастические дифференциальные, интегральные и смешанные уравнения в евклидовом пространстве, а также их разностные аналоги. Для круговых, сферических, кватернионных и других гипергеометрических стохастических уравнений, относящихся к системам на многообразиях [3], известные методы анализа, моделирования и синтеза требуют развития. Однако при этом основные принципы, подходы и задачи статистических НИ сохраняются.

Обзор зарубежного универсального методического и программного обеспечения для математической статистики круговых случайных величин и функций дан в [4]. Отдельные прикладные задачи решены, например, в [1–8]. В ИПИ РАН начиная с 2010 г. в рамках тем, поддерживаемых РФФИ, ведутся работы по созданию методического обеспечения для анализа, моделирования и синтеза фильтров для обработки информации в круговых стохастических системах (КСтС) [9, 10].

\*Работа выполнена при финансовой поддержке РФФИ (проект № 10-07-00021) и программы ОНИТ РАН «Информационные технологии и анализ сложных систем» (проект 1.5).

<sup>1</sup>Институт проблем информатики Российской академии наук, sinitsin@dol.ru

Рассмотрим полезные для практики простые квазилинейные, основанные на эквивалентной круговой статистической линейзации (КСЛ), корреляционные и спектрально-корреляционные методы, алгоритмы и ИПО для офлайн-анализа и моделирования круговых стохастических процессов (КСтП) в нелинейных КСтС.

## 2 Статистическая линейзация нелинейных преобразований круговых случайных величин

Пусть сначала  $X$  и  $Y$  — скалярные круговые случайные величины (КСВ), связанные между собой детерминированной нелинейной зависимостью

$$Y = \varphi(X). \quad (1)$$

Согласно принципу эквивалентной статистической линейзации заменим нелинейную зависимость (1) приближенной линейной зависимостью:

$$\varphi(X) \approx U = \varphi_0 + k_1(X - \mu), \quad (2)$$

где  $\mu = \mu_x$  — круговое среднее направление КСВ  $X$ . Параметры  $\varphi_0$  и  $k_1$  находят из критерия минимума безусловного риска для выбранной функции потерь  $\ell(X, U)$ :

$$R = M[\ell(X, U)] = \min, \quad (3)$$

где  $M$  — символ математического ожидания.

Если выбрать эквивалентное одномерное распределение (ЭР) КСВ  $X$  и функцию потерь в виде

$$\ell(X, U) = (e^{iX} - e^{iU})^2, \quad (4)$$

то после подстановки (2) в (3) и (4) и приравнивания нулю частных производных  $\partial R/\partial \varphi_0$  и  $\partial R/\partial k_1$  получим одно комплексное уравнение для неизвестных параметров  $\varphi_0$  и  $k_1$ :

$$M_{\text{ЭР}} \exp \{i\varphi(X)\} = M_{\text{ЭР}} \exp \{i[\varphi_0 + k_1(X - \mu)]\}, \quad (5)$$

где  $M_{\text{ЭР}}$  — символ математического ожидания по ЭР; коэффициенты КСЛ  $\varphi_0$  и  $k_1$  зависят от вероятностных характеристик КСВ  $X$ .

Принимая в качестве ЭР для КСВ  $X$  намотанное нормальное распределение с параметрами  $\mu$  и  $\sigma$ , т. е.  $WN(\mu, \sigma)$  [4, 7], перепишем комплексное уравнение (5) в виде двух действительных уравнений:

$$\varphi_0(\mu, \sigma) = \psi(\mu, \sigma);$$

$$k_1(\mu, \sigma) = \frac{\sqrt{-2 \ln r(\mu, \sigma)}}{\sigma},$$

где введено следующее обозначение:

$$r e^{i\psi} = M_{WN} \exp \{-i\varphi(X)\}.$$

Для скалярного нелинейного преобразования векторного аргумента

$$Y = \varphi(X_1, \dots, X_n) \quad (6)$$

при условии, что ЭР вектора КСВ  $X = [X_1, \dots, X_n]^T$  является известным намотанным нормальным распределением [4, 7], уравнения принципа статистической линейзации по критерию (4) имеют следующий вид:

$$\varphi(X) \approx U = \varphi_0 + \sum_{h=1}^n k_{1h} X_h^0.$$

Здесь  $\varphi_0$  — первый векторный коэффициент КСЛ, равный

$$\varphi_0 = M_{WN} \varphi(X),$$

$k_{1h}$  ( $h = 1, \dots, n$ ) — второй векторный коэффициент КСЛ, который определяется путем решения алгебраической системы уравнений

$$\sum_{j=1}^n k_{1h} K_{jh} = M_{WN} X_j^0 X_h^0 \varphi(X),$$

где  $K_{1h} = M_{WN} X_j^0 X_h^0$  ( $j, h = 1, \dots, n$ ).

Аналогично выписываются формулы для коэффициентов КСЛ для векторных и матричных нелинейных преобразований, а также посредством канонических представлений [1].

Для типовых нелинейных преобразований (1) и (6) составлены таблицы и разработано ИПО CStS-ANALYSIS [11].

## 3 Основные результаты

**Теорема 3.1.** Пусть нестационарная дифференциальная система

$$\dot{Y} = \varphi(Y, t) + V, \quad Y(t_0) = Y_0$$

удовлетворяет следующим допущениям:

- (1)  $n$ -мерный круговой (на  $[0, 2\pi]$ ) СтП  $Y = Y(t)$  обладает конечными вероятностными моментами второго порядка;
- (2)  $n$ -мерный круговой белый шум, понимаемый в строгом смысле, ( $V = \dot{W}$ ,  $W$  — КСтП с независимыми приращениями на  $[0, 2\pi]$  и матрицей интенсивности  $G(t)$ );

(3) детерминированное нелинейное преобразование  $\varphi(Y, t)$  не обладает памятью и допускает КСЛ согласно алгоритмам разд. 2, причем статистически линеаризованная система для  $Y^0 = Y - m_y$ :

$$\dot{Y}^0 = k_1 Y^0 + V \quad (m_y = MY) \quad (7)$$

асимптотически устойчива. Тогда корреляционное уравнение квазилинейного анализа и аналитического моделирования имеют следующий вид:

$$\begin{aligned} \dot{m}_y &= \varphi_0(m_y, K_y, t); \quad m_y(t_0) = m_{y0}; \\ \dot{K}_y &= k_1(m_y, K_y, t)K_y + \\ &\quad + K_y k_1^T(m_y, K_y, t) + G(t); \\ K_y(t_0) &= K_{y0}; \\ \frac{\partial K_y(t_1, t_2)}{\partial t_2} &= K_y(t_1, t_2)k_1^T(m_y, K_y, t_2); \\ K_y(t_1, t_2) &= \begin{cases} K_y(t_1, t_2) & \text{при } t_2 > t_1; \\ K_y(t_2, t_1)^T & \text{при } t_2 < t_1, \end{cases} \end{aligned}$$

где  $m_y$ ,  $K_y(t)$  и  $K_y(t_1, t_2)$  — соответственно вектор математического ожидания, ковариационная матрица и матрица ковариационных функций КСмП  $\tilde{Y}(t)$ .

**Теорема 3.2.** В условиях теоремы 3.1 при стационарных функциях  $\varphi(Y, t) = \varphi(Y)$ ,  $G(t) = G$  корреляционные уравнения анализа и аналитического моделирования для КСмП  $\tilde{Y}(t)$  имеют вид:

$$\varphi_0(\tilde{m}_y, \tilde{K}_y) = 0;$$

$$k_1(\tilde{m}_y, \tilde{K}_y)\tilde{K}_y + \tilde{K}_y k_1(\tilde{m}_y, \tilde{K}_y) + G = 0 \quad (8)$$

$$\left. \begin{aligned} \frac{d\tilde{k}_y(\tau)}{d\tau} &= k_1(\tilde{m}_y, \tilde{K}_y)\tilde{k}_y(\tau); \\ k_y(\tau) &= \tilde{K}_y(t_1, t_1 + \tau), \end{aligned} \right\} \quad (9)$$

где  $\tilde{m}_y$ ,  $\tilde{K}_y$  и  $\tilde{k}_y(\tau)$  — соответственно математическое ожидание, ковариационная матрица и матрица ковариационных функций ( $\tau = t_1 - t_2$ ) стационарного КСмП  $\tilde{Y}(t)$ .

**Теорема 3.3.** В условиях теоремы 3.2 уравнения (8) и (9), если вместо ковариационной матрицы  $k_y(\tau)$  использовать спектральную плотность  $s_y(\omega)$ , допускают следующее спектральное представление:

$$\begin{aligned} \tilde{K}_y &= \int_{-\infty}^{\infty} s_y(\omega; \tilde{m}_y, \tilde{K}_y) d\omega; \\ k_y(\tau) &= \int_{-\infty}^{\infty} e^{i\omega\tau} s_y(\omega; \tilde{m}_y, \tilde{K}_y) d\omega, \end{aligned}$$

где  $s_y(\omega; \tilde{m}_y, \tilde{K}_y)$  — матрица спектральных плотностей:

$$s_y(\omega; \tilde{m}_y, \tilde{K}_y) = \Phi(i\omega; \tilde{m}_y, \tilde{K}_y) \frac{G}{2\pi} I_n \Phi(i\omega; \tilde{m}_y, \tilde{K}_y)^*;$$

$\Phi(i\omega; \tilde{m}_y, \tilde{K}_y)$  — передаточная функция статистически линеаризованной системы (7):

$$\Phi(i\omega; \tilde{m}_y, \tilde{K}_y) = [k_1(\tilde{m}_y, \tilde{K}_y) - Ii\omega]^{-1}; \quad I = I_n;$$

\* — символ эрмитова сопряжения;  $I_n$  — единичная ( $n \times n$ )-матрица.

**Замечание.** Рассмотренные в [1] другие схемы статистической линеаризации очевидным образом обобщаются на круговой случай. При этом могут быть использованы различные модели КСтС [9].

Алгоритмы теорем 3.1–3.3 и их дискретных версий лежат в основе СтИТ анализа аналитического моделирования. Они реализованы в ИПО CStS-ANALYSIS в среде MATLAB [9–11]. Инструментальное программное обеспечение имеет возможность реализовать также и статистическое моделирование КСтС для следующих «намотанных» круговых распределений КСВ: решетчатого, нормального, Мизеса, равномерного, пуассонова, кардиоидного, треугольного, Коши и других устойчивых распределений [4, 7]. Точность алгоритмов анализа и аналитического моделирования проверялась на радиотехнических примерах [6], а также методом статистического моделирования.

## 4 Заключение

Принципы, подходы и задачи статистических научных исследований, развитые в [1] для стохастических систем в евклидовом пространстве, сохраняются и для круговых систем. Методическое и алгоритмическое обеспечение, основанное на статистической линеаризации для эквивалентного «намотанного» нормального распределения, даются теоремами 3.1–3.3. Разработано и испытано на ряде тестовых примеров универсальное ИПО CStS-ANALYSIS в среде MATLAB для анализа, аналитического и статистического моделирования.

## Литература

1. Сеницын И. Н. Канонические представления случайных функций и их применения в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009.

2. Босов А. В., Будзко В. И., Захаров В. Н., Козмиди-ди В. А., Корепанов Э. Р., Синицын И. Н., Шоргин С. Я., Ушмаев О. С. Информатика: состояние, проблемы, перспективы / Под ред. И. А. Соколова. — М.: ИПИ РАН, 2009.
3. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы / Пер. с англ. под ред. А. Н. Ширяева. — М.: Наука, 1986.
4. Rao Jammalamadaka S., Sen Gupta A. Topics in circular statistics. — Singapore: World Scientific, 2001.
5. Леви П. Стохастические процессы и броуновское движение / Пер. с фр. под ред. Н. Н. Ченцова. — М.: Наука, 1972.
6. Тихонов В. И., Миронов М. А. Марковские процессы. — М.: Сов. радио, 1977.
7. Мардиа К. Статистический анализ угловых наблюдений / Пер. с англ. под ред. Л. Н. Большева. — М.: Наука, 1978.
8. Морозов А. Н., Назолин А. Л. Динамические системы с флуктуирующим временем. — М.: МГТУ им. Н. Э. Баумана, 2001.
9. Синицын И. Н. Канонические разложения случайных функций и их применение в стохастических информационных технологиях научных исследований: Курс лекций // Распознавание образов и анализ изображений: новые информационные технологии — РОАИ-10-2010: Мат-лы Междунар. конф. — СПб., 2010.
10. Синицын И. Н., Корепанов Э. Р., Белоусов В. В. и др. Развитие компьютерной поддержки статистических научных исследований систем высокой точности и доступности // Системы и средства информатики, 2011. Вып. 21. № 1. С. 3–33.
11. Sinitsyn I. N., Belousov V. V., Konashenkova T. D. Software tools for circular stochastic systems analysis / 29th Seminar (International) on Stability Problems for Stochastic Models and 5th Workshop “Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems” (APTP + MS’2011) Book of Abstracts. — М.: IPI RAS, 2011. P. 86–87.

# СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С НЕНАДЕЖНЫМ ПРИБОРОМ В ДИСКРЕТНОМ ВРЕМЕНИ\*

А. В. Печинкин<sup>1</sup>, И. А. Соколов<sup>2</sup>

**Аннотация:** Для функционирующей в дискретном времени однолинейной системы массового обслуживания (СМО)  $Geo/G/1/\infty$  с ненадежным прибором, который может отказывать как в рабочем, так и в свободном состоянии, причем с разными вероятностями отказов и распределениями времен ремонта, найдены основные стационарные характеристики обслуживания. В качестве примера в последнем разделе рассмотрен один из вариантов СМО  $\bar{Geo}_2/\bar{G}_2/1/\infty$  с двумя типами заявок и абсолютным приоритетом.

**Ключевые слова:** система массового обслуживания; дискретное время; ненадежный прибор

## 1 Введение

В системах обслуживания, в том числе в современных инфотелекоммуникационных системах, часто возникает ситуация, когда прибор подвержен различного рода поломкам и неисправностям, что приводит к ухудшению обслуживания и даже полным остановкам устройств. Поэтому системам с ненадежными приборами уделялось и уделяется большое внимание в различных областях науки и техники. В частности, этой тематике посвящено значительное число работ в теории массового обслуживания.

Следует отметить, что в подавляющем числе работ по изучению СМО с ненадежными приборами рассматриваются системы с непрерывным временем. Некоторое представление о таких работах можно найти в [1]. Однако в последнее время в связи с применением новых технологий, особенно цифровых методов передачи данных, существенно возрос интерес к СМО, функционирующим в дискретном времени. Здесь можно упомянуть работы [2–11].

В настоящей статье для определенной в следующем разделе функционирующей в дискретном времени однолинейной СМО с ненадежным прибором найдены основные стационарные характеристики обслуживания. В отличие от других работ, предполагается, что отказы прибора могут возникать не только при обслуживании заявок, но и в свободной системе. Это предположение в дискретном времени приводит к дополнительным трудностям из-за возможности одновременного появления сразу нескольких событий.

В разд. 3 получены соотношения для вычисления в терминах преобразования Лапласа–Стилтье-

са (ПЛС) распределения общего времени пребывания заявки на приборе (с учетом возможных прерываний на ремонт прибора).

В разд. 4 вводится вложенная цепь Маркова и определяются стационарные вероятности этой цепи.

В разд. 5 находятся стационарные распределения числа заявок по времени и по моментам поступления в систему.

В разд. 6 вычисляется стационарное распределение времени пребывания заявки в системе.

В разд. 7 рассмотрен один из вариантов СМО  $\bar{Geo}_2/\bar{G}_2/1/\infty$  с двумя типами заявок и абсолютным приоритетом. Показано, как с помощью полученных в предыдущих разделах результатов можно вычислить некоторые стационарные характеристики обслуживания неприоритетных заявок в системе, в которую поступает два независимых геометрических потока. Отметим, что наличие нескольких вариантов системы с приоритетом обусловлено следующим обстоятельством. В отличие от непрерывного времени, в дискретном времени входящий геометрический поток заявок двух типов можно определять по-разному. Действительно, суперпозиция двух независимых геометрических потоков не является геометрическим потоком (точнее говоря, является геометрическим неординарным потоком, в котором одновременно могут прийти либо одна, либо две заявки), и наоборот: после разреживания геометрического потока на два, хотя каждый из полученных потоков и будет геометрическим, но в совокупности потоки будут зависимыми (поскольку появление заявки одного потока исключает появление заявки второго потока).

Наконец, в разд. 8 приведен пример расчета стационарных средних характеристик рассматрива-

\* Работа выполнена при поддержке РФФИ (грант 11-07-00112).

<sup>1</sup> Институт проблем информатики Российской академии наук, apetchinkin@ipiran.ru

<sup>2</sup> Институт проблем информатики Российской академии наук, isokolov@ipiran.ru

емой СМО, проведенного с помощью полученных формул.

Отметим, что примененная в данной работе методика была использована в [12, 13] для анализа некоторых СМО с отрицательными заявками и в [14] для исследования СМО  $MAR_K/G_K/1$  с обобщенной дисциплиной преимущественного разделения прибора.

Далее в этой статье для сокращения записи вероятность дополнительного события будем снабжать чертой сверху. Например,  $\bar{a} = 1 - a$ .

## 2 Описание системы и обозначения

Рассмотрим функционирующую в дискретном времени СМО  $Geo/G/1/\infty$ .

Входящий в систему поток является геометрическим: в конце каждого такта с вероятностью  $a$ , не зависящей от всей предыстории функционирования системы до данного такта, поступает новая заявка, которая тут же переходит на прибор, если прибор свободен, и становится в очередь в противном случае. Среднее число  $\hat{a}$  тактов между поступлениями заявок определяется формулой:

$$\hat{a} = \sum_{k=1}^{\infty} k \bar{a}^{k-1} a = \frac{1}{a}.$$

В системе имеется один прибор, который может находиться как в исправном, так и в неисправном состоянии.

Число тактов обслуживания заявки исправным прибором (которое далее будем называть длиной заявки) распределено по закону  $\{b_k, k \geq 0\}$  (предполагается, что длина заявки не может быть нулевой, т. е.  $b_0 = 0$ ). Производящую функцию (ПФ) и среднее значение длины заявки будем обозначать через

$$\beta(z) = \sum_{k=0}^{\infty} b_k z^k; \quad \hat{b} = \beta'(1) = \sum_{k=0}^{\infty} k b_k.$$

Опишем подробно функционирование прибора, который может отказывать (в конце такта) как в свободном состоянии, так и при обслуживании заявки. Пусть  $c^*$  — вероятность отказа на такте, если прибор свободен, и  $c$  — если на приборе находится заявка. Введем ПФ

$$\sigma(z) = \sum_{k=1}^{\infty} \bar{c}^{k-1} c z^k = \frac{cz}{1 - \bar{c}z};$$

$$\sigma^*(z) = \sum_{k=1}^{\infty} (\bar{c}^*)^{k-1} c^* z^k = \frac{c^* z}{1 - \bar{c}^* z}$$

соответственно для геометрических распределений времени до момента отказа занятого и свободного прибора.

Если прибор отказывает в свободном состоянии, то он ремонтируется случайное время, распределенное по закону  $\{c_j^*, j \geq 0\}$  (предполагается, что  $c_0^* = 0$ ). При этом первая заявка, поступившая в свободную систему в момент ремонта прибора, становится на прибор, но ее обслуживание начинается только после окончания ремонта. Остальные заявки скапливаются в очереди.

Если же в момент отказа прибора на нем находится заявка, то ее обслуживание прекращается, а прибор ремонтируется случайное время, распределенное по закону  $\{c_j, j \geq 0\}$  (предполагается, что  $c_0 = 0$ ). При этом заявка продолжает находиться на приборе, но ее обслуживание возобновляется только после окончания ремонта, причем суммарная обслуженная длина засчитывается при продолжении обслуживания (дисциплина с дообслуживанием). Поступающие заявки, как и прежде, становятся в очередь.

Обозначим ПФ и среднее значение распределения  $\{c_j, j \geq 0\}$  через

$$\chi(z) = \sum_{j=0}^{\infty} z^j c_j;$$

$$\hat{c} = \chi'(1) = \sum_{k=0}^{\infty} k c_k,$$

а ПФ и среднее значение распределения  $\{c_j^*, j \geq 0\}$  — через

$$\chi^*(z) = \sum_{j=0}^{\infty} z^j c_j^*;$$

$$\hat{c}^* = \chi^{*'}(1) = \sum_{k=0}^{\infty} k c_k^*.$$

Примем дополнительно следующие правила:

- очередной отказ прибора не может наступить сразу же в момент окончания ремонта (прибор обязательно должен находиться в исправном состоянии хотя бы один такт);
- если перед некоторым моментом в системе нет заявок, прибор исправен и в этот момент поступает новая заявка, то она сразу же переходит на прибор, причем прибор с вероятностью  $c$  отказывает и ремонтируется случайное время, распределенное по закону  $\{c_j, j \geq 0\}$ ;
- если перед некоторым моментом в системе нет заявок, (свободный) прибор неисправен, а в

этот момент заканчивается ремонт прибора и поступает новая заявка, то она сразу же переходит на прибор и начинает обслуживаться;

- если оканчивается последний такт обслуживания заявки на приборе, в очереди есть еще заявки или в систему поступает новая заявка, то заявка с прибора покидает систему, на прибор переходит следующая заявка, а прибор с вероятностью  $c$  отказывает и ремонтируется случайное время, распределенное по закону  $\{c_j, j \geq 0\}$ ;
- если оканчивается обслуживание заявки на приборе, в очереди больше нет заявок и в систему не поступает новая заявка, то прибор отказывает с вероятностью  $c^*$ , а ремонтируется случайное время, распределенное по закону  $\{c_j^*, j \geq 0\}$ .

Положим:

$$C_i = \sum_{j=i}^{\infty} c_j, \quad i \geq 0;$$

$$C_i^* = \sum_{j=i}^{\infty} c_j^*, \quad i \geq 0.$$

Выбор заявок из очереди на обслуживание осуществляется в порядке поступления в систему.

Наряду с предположением о конечности средних времен обслуживания и ремонта, которое, как будет видно из дальнейшего, является необходимым для существования стационарного режима функционирования данной системы, при вычислении стационарных средних значений числа заявок в системе и времени пребывания заявки в системе будем считать, что конечными являются также вторые моменты времен обслуживания и ремонта.

### 3 Время пребывания заявки на приборе

Исследование СМО Geo/G/1/∞ с ненадежным прибором начнем с вычисления распределения общего времени пребывания заявки на приборе (с учетом возможных прерываний на ремонт прибора, в том числе перед началом обслуживания). Здесь возможны два случая: либо заявка поступает в свободную систему, либо в систему, в которой уже имеются (по крайней мере, непосредственно перед моментом поступления) другие заявки.

Начнем со второго случая. Здесь необходимым условием поступления заявки на прибор является окончание обслуживания предыдущей заявки, и поэтому в момент поступления заявки на прибор он

обязательно должен находиться в исправном состоянии (но сразу же после поступления заявки прибор с вероятностью  $c$  может отказать и тогда ремонтируется случайное время, распределенное по закону  $\{c_j, j \geq 0\}$ ). Обозначим через  $\{d_j, j \geq 0\}$  распределение времени пребывания на приборе такой заявки.

Пусть длина заявки равна  $i, i \geq 0$ , тактов. За это время с вероятностью  $\binom{i}{k} c^k \bar{c}^{i-k}, k = \overline{0, i}$ , произойдет ровно  $k$  прерываний обслуживания (напомним, что число  $k$  учитывает возможное прерывание обслуживания в момент постановки данной заявки на прибор). Поскольку времена прерываний — независимые одинаково распределенные случайные величины с распределением  $\{c_j, j \geq 0\}$ , то общее время прерываний при условии, что произошло ровно  $k$  прерываний, имеет ПФ  $\chi^k(z)$ . С учетом собственной длины  $i$ -й заявки получаем для ПФ  $\delta(z) = \sum_{j=0}^{\infty} z^j d_j$  времени пребывания на приборе заявки, поступившей из очереди, выражение:

$$\begin{aligned} \delta(z) &= \sum_{i=0}^{\infty} z^i b_i \sum_{k=0}^i \chi^k(z) \binom{i}{k} c^k \bar{c}^{i-k} = \\ &= \sum_{i=0}^{\infty} [cz\chi(z) + \bar{c}z]^i b_i = \beta(cz\chi(z) + \bar{c}z). \end{aligned}$$

Среднее время пребывания на приборе такой заявки задается формулой:

$$\hat{\delta} = \delta'(1) = \beta'(1)[1 + c\chi'(1)] = \hat{b}(1 + c\hat{c}).$$

Перейдем теперь к первому случаю.

Предположим сначала, что в систему вообще не поступают заявки, и рассмотрим два процесса восстановления.

Интервал между соседними восстановлениями у обоих процессов восстановления представляет собой число тактов между соседними моментами окончания ремонта прибора, или, что с точки зрения распределения то же самое, между соседними моментами отказа прибора, и состоит из двух независимых частей. Первая часть — время до момента первого отказа после окончания ремонта, имеющее геометрическое распределение с параметром  $c^*$ . Вторая часть — время ремонта неисправного прибора, имеющее распределение  $\{c_j^*, j \geq 0\}$ .

Поэтому ПФ  $\varphi(z) = \sum_{j=0}^{\infty} z^j f_j$  распределения  $\{f_j, j \geq 0\}$  интервала между соседними восстановлениями представима в виде:

$$\varphi(z) = \sigma^*(z)\chi^*(z) = \frac{c^*z}{1 - \bar{c}^*z}\chi^*(z).$$

Первый момент восстановления первого процесса восстановления с вероятностью  $\bar{c}^*$  равен нулю и с вероятностью  $c^*c_j^*$  равен  $j$ ,  $j \geq 1$ . Первый момент восстановления второго процесса восстановления имеет (начиная с нуля!) геометрическое распределение с параметром  $c^*$ .

Введенные таким образом процессы восстановления имеют очень простой смысл. Предположим, что в начальный момент 0 система освобождается от заявок (уходит с прибора последняя заявка) и полностью прекращается поступление новых заявок. Тогда моменты восстановления первого процесса восстановления представляют собой моменты окончания ремонта прибора (сюда же относится и момент 0, если в этот момент прибор не отказывает), а моменты восстановления второго процесса восстановления совпадают с моментами отказа прибора.

Обозначим через  $\{h_{1,k}, k \geq 0\}$  и  $\{h_{2,k}, k \geq 0\}$  ряды восстановления этих процессов. Эти ряды будут иметь ПФ

$$\begin{aligned} \mathfrak{a}_1(z) &= \sum_{k=0}^{\infty} z^k h_{1,k} = \frac{\bar{c}^* + c^* \chi^*(z)}{1 - \varphi(z)}; \\ \mathfrak{a}_2(z) &= \sum_{k=0}^{\infty} z^k h_{2,k} = \frac{c^*}{(1 - \bar{c}^* z)[1 - \varphi(z)]}. \end{aligned}$$

Очевидно также, что

$$h_{1,0} = \mathfrak{a}_1(0) = \bar{c}^*.$$

Вернемся к исходной СМО и предположим, что в свободную систему поступает заявка. В этом случае она сразу же попадает на прибор, но прибор может находиться на ремонте, и поэтому еще до фактического начала обслуживания заявка может провести на приборе некоторое время. Распределение времени пребывания на приборе заявки, поступившей в свободную систему, обозначим через  $\{d_k^*, k \geq 0\}$ .

Для вычисления  $\{d_k^*, k \geq 0\}$  предположим, что с момента освобождения прибора прошло  $i \geq 1$  тактов и в этот момент в систему поступила заявка.

Теперь возможны следующие варианты:

- на  $m$ -м,  $m = \overline{0, i-1}$ , такте закончился ремонт прибора (с вероятностью  $h_{1,m}$ ) и больше за  $i-1-m$  тактов прибор не отказывал (с вероятностью  $(\bar{c}^*)^{i-1-m}$ ). Тогда заявка сразу же переходит на прибор, но прибор с вероятностью  $c$  начинает ремонтироваться, и время пребывания заявки на приборе имеет распределение  $\{d_k, k \geq 0\}$ ;
- окончание ремонта прибора происходит в момент  $i$  (с вероятностью  $h_{1,i}$ )  $i \geq 1$ , совпадающий с моментом поступления заявки. В этом

варианте она также переходит на прибор, но, в отличие от предыдущего варианта, сразу же начинает обслуживаться, и распределение  $\{d_{0,k}^*, k \geq 0\}$  времени ее пребывания на приборе имеет ПФ

$$\begin{aligned} \delta_0^*(z) &= \sum_{k=0}^{\infty} z^k d_{0,k}^* = \\ &= \sum_{k=1}^{\infty} z^k b_k \sum_{j=0}^{k-1} \chi^j(z) \binom{k-1}{j} c^j \bar{c}^{k-1-j} = \\ &= z \sum_{k=1}^{\infty} [cz\chi(z) + \bar{c}z]^{k-1} b_k = \frac{z\beta(cz\chi(z) + \bar{c}z)}{cz\chi(z) + \bar{c}z} = \\ &= \frac{\delta(z)}{c\chi(z) + \bar{c}} \end{aligned}$$

со средним значением

$$\begin{aligned} \hat{\delta}_0^* &= \delta_0^{*'}(1) = \\ &= \frac{\delta'(1)}{c\chi(1) + \bar{c}} - \delta(1) \frac{c\chi'(1)}{[c\chi(1) + \bar{c}]^2} = \hat{\delta} - c\hat{c}; \end{aligned}$$

- ремонт прибора начался за  $m$ ,  $m = \overline{1, i}$ , тактов до поступления заявки в систему (с вероятностью  $h_{2,i-m}$ ) и не закончился в момент ее поступления (с вероятностью  $C_{m+1}^*$ ). Тогда заявка пребывает на приборе время  $k$ ,  $k \geq 1$ , если сначала она ожидает (на приборе)  $j$ ,  $j = \overline{1, k}$ , тактов окончания ремонта прибора (с вероятностью  $c_{m+j}^*/C_{m+1}^*$ ) и после окончания ремонта находится на приборе еще  $k-j$  тактов (с вероятностью  $d_{0,k-j}^*$ ). Значит, в этом варианте вероятность  $d_{1,k}^*(m)$  того, что заявка будет находиться на приборе  $k$  тактов, определяется формулой:

$$d_{1,k}^*(m) = \sum_{j=1}^k \frac{c_{m+j}^*}{C_{m+1}^*} d_{0,k-j}^*, \quad k \geq 1, \quad m = \overline{1, i}.$$

Таким образом, распределение  $\{d_k^*(i), k \geq 0\}$ ,  $i \geq 1$ , времени нахождения заявки на приборе при условии, что заявка поступила на такте  $i$ ,  $i \geq 1$ , после освобождения прибора, определяется формулой:

$$\begin{aligned} d_k^*(i) &= \sum_{m=0}^{i-1} h_{1,m} (\bar{c}^*)^{i-1-m} d_k + \\ &+ h_{1,i} d_{0,k}^* + \sum_{m=1}^i h_{2,i-m} \sum_{j=1}^k c_{m+j}^* d_{0,k-j}^*, \\ & \quad k \geq 0, \quad i \geq 1, \end{aligned}$$

или в терминах ПФ  $\delta^*(z; i) = \sum_{k=0}^{\infty} z^k d_k^*(i)$ :

$$\delta^*(z; i) = \sum_{m=0}^{i-1} h_{1,m}(\bar{c}^*)^{i-1-m} \delta(z) + h_{1,i} \delta_0^*(z) + \sum_{k=0}^{\infty} z^k \sum_{m=1}^i h_{2,i-m} \sum_{j=1}^k c_{m+j}^* d_{0,k-j}^*, \quad i \geq 1.$$

Поскольку число тактов от момента освобождения системы до момента поступления следующей заявки в свободную систему имеет геометрическое распределение с параметром  $a$ , то, применяя формулу полной вероятности, получаем для ПФ  $\delta^*(z) = \sum_{k=0}^{\infty} z^k d_k^*$  безусловного распределения  $\{d_k^*, k \geq 0\}$  после элементарных преобразований следующее выражение:

$$\delta^*(z) = \sum_{i=1}^{\infty} a \bar{a}^{i-1} \delta^*(z; i) = a \left( \delta(z) \mathfrak{a}_1(\bar{a}) \frac{1}{1 - a \bar{c}^*} + \delta_0^*(z) \frac{\mathfrak{a}_1(\bar{a}) - h_{1,0}}{\bar{a}} + \delta_0^*(z) \mathfrak{a}_2(\bar{a}) \frac{\bar{a} \chi^*(\bar{a}) - z \chi^*(\bar{a})}{\bar{a}(z - \bar{a})} \right).$$

Среднее время  $\hat{\delta}^*$  пребывания на приборе заявки, поступившей в свободную систему, вычисляется по формуле:

$$\hat{\delta}^* = \delta^{*'}(1) = a \left( \hat{\delta} \mathfrak{a}_1(\bar{a}) \frac{1}{1 - a \bar{c}^*} + \hat{\delta}_0^* \frac{\mathfrak{a}_1(\bar{a}) - h_{1,0}}{\bar{a}} + \hat{\delta}_0^* \mathfrak{a}_2(\bar{a}) \frac{\bar{a} - \chi^*(\bar{a})}{a \bar{a}} + \mathfrak{a}_2(\bar{a}) \left[ \frac{\bar{a} \bar{c}^* - \chi^*(\bar{a})}{a \bar{a}} - \frac{\bar{a} - \chi^*(\bar{a})}{a^2 \bar{a}} \right] \right),$$

или после несложных, но утомительных преобразований:

$$\hat{\delta}^* = \hat{b}(1 + c\hat{c}) - \frac{c^*[(1 + ac\hat{c})(1 - \chi^*(\bar{a})) - ac\hat{c}^*]}{a[1 - \bar{c}^* \bar{a} - c^* \bar{a} \chi^*(\bar{a})]}.$$

#### 4 Стационарные вероятности состояний по вложенной цепи Маркова

Пусть  $\tau_n, n \geq 1$ , — последовательные моменты уходов заявок из системы. Обозначим через  $\nu_n$  число заявок в системе сразу же после момента  $\tau_n$  (отметим, что  $\nu_n$  определяется после ухода обслуженной заявки, а если в момент  $\tau_n$  в систему поступает новая заявка — то после ее прихода). Очевидно, что  $\{\nu_n, n \geq 1\}$  — (однородная) цепь Маркова.

Положим

$$\tilde{p}_j = \sum_{k=j}^{\infty} d_k \binom{k}{j} a^j \bar{a}^{k-j} = \frac{a^j}{j!} \delta^{(j)}(\bar{a}), \quad j \geq 0;$$

$$\tilde{p}_j^* = \sum_{k=j}^{\infty} d_k^* \binom{k}{j} a^j \bar{a}^{k-j} = \frac{a^j}{j!} \delta^{*(j)}(\bar{a}), \quad j \geq 0.$$

Как нетрудно видеть,  $\tilde{p}_j$  и  $\tilde{p}_j^*$  представляют собой вероятности того, что за время пребывания на приборе заявки, поступившей в систему при наличии других заявок, или заявки, поступившей в свободную систему, в систему придет еще  $j$  заявок.

Выпишем переходные вероятности цепи  $\nu_n$ .

Если  $i \geq 1$ , то переходные вероятности  $p_{i,j}$  определяются точно так же, как и для системы Geo/G/1/∞ с надежным прибором, за исключением того, что вместо распределения  $\{b_k, k \geq 0\}$  нужно взять распределение  $\{d_k, k \geq 0\}$ . Поэтому

$$p_{i,j} = \sum_{k=j-i+1}^{\infty} d_k \binom{k}{j-i+1} a^{j-i+1} \bar{a}^{k-j+i-1} = \tilde{p}_{j-i+1}, \quad i \geq 1, j \geq i-1.$$

Однако, если  $i = 0$ , то, в отличие от СМО Geo/G/1/∞, поступающая в свободную систему заявка пребывает на приборе время, распределенное по закону  $\{d_k^*, k \geq 0\}$ . Значит,

$$p_{0,j} = \sum_{k=j}^{\infty} d_k^* \binom{k}{j} a^j \bar{a}^{k-j} = \tilde{p}_j^*, \quad j \geq 0.$$

Окончательно получаем, что матрица переходных вероятностей вложенной цепи Маркова  $\{\nu_n, n \geq 0\}$  имеет вид:

$$P = (p_{ij}) = \begin{pmatrix} \tilde{p}_0^* & \tilde{p}_1^* & \tilde{p}_2^* & \cdots \\ \tilde{p}_0 & \tilde{p}_1 & \tilde{p}_2 & \cdots \\ 0 & \tilde{p}_0 & \tilde{p}_1 & \cdots \\ 0 & 0 & \tilde{p}_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Будем предполагать, что средние времена  $\hat{c}$  и  $\hat{c}^*$  ремонта прибора и средняя длина  $\hat{b}$  заявки конечны и, кроме того, выполнено условие  $\rho = a\hat{d} < 1$ . Тогда существуют предельные (стационарные) вероятности состояний  $p_i^+ = \lim_{n \rightarrow \infty} \mathbf{P}\{\nu_n = i\}, i \geq 0$ , по вложенной цепи Маркова  $\nu_n$ . Необходимость вышеперечисленных условий для эргодичности вложенной цепи будет показана ниже, достаточность их нетрудно получить, воспользовавшись критерием Мустафы.

Для определения  $p_i^+, i \geq 0$ , выпишем систему уравнений равновесия (СУР):

$$p_i^+ = p_0^+ \tilde{p}_i^* + \sum_{k=1}^{i+1} p_k^+ \tilde{p}_{i-k+1}, \quad i \geq 0. \quad (1)$$

Система (1) легко решается рекуррентно:

$$p_i^+ = p_0^+ r_i, \quad i \geq 1,$$

где  $r_i$  определяются рекуррентной формулой:

$$r_1 = \frac{1}{\tilde{p}_0} (1 - \tilde{p}_0^*);$$

$$r_i = \frac{1}{\tilde{p}_0} \left[ r_{i-1} - \tilde{p}_{i-1}^* - \sum_{k=1}^{i-1} r_k \tilde{p}_{i-k} \right], \quad i \geq 2.$$

Более рациональный алгоритм вычисления  $p_i^+$  можно получить, если при составлении СУР воспользоваться уравнениями частичного баланса. Действительно, рассматривая вероятности перехода за один шаг цепи Маркова из множества состояний с номерами, не меньшими  $i, i \geq 1$ , в множество состояний с номерами, меньшими  $i$ , и обратно, получаем систему уравнений:

$$p_i^+ \tilde{p}_0 = p_0^+ \tilde{P}_i^* + \sum_{j=1}^{i-1} p_j^+ \tilde{P}_{i-j}, \quad i \geq 1,$$

где

$$\tilde{P}_i^* = \sum_{j=i}^{\infty} \tilde{p}_j^*, \quad \tilde{P}_i = \sum_{j=i}^{\infty} \tilde{p}_j, \quad i \geq 1,$$

которая также решается рекуррентно, но, в отличие от описанного ранее алгоритма, при решении не содержит вычитаний, что позволяет во многих случаях существенно повысить точность вычислений.

Воспользуемся теперь ПФ  $P^+(z) = \sum_{i=0}^{\infty} p_i^+ z^i$ .

Умножая  $i$ -е уравнение системы (1) на  $z^i$  и суммируя, получаем:

$$P^+(z) = p_0^+ \delta^*(\bar{a} + az) + \frac{1}{z} [P^+(z) - p_0^+] \delta(\bar{a} + az),$$

откуда находим:

$$P^+(z) = \frac{\delta(\bar{a} + az) - z \delta^*(\bar{a} + az)}{\delta(\bar{a} + az) - z} p_0^+. \quad (2)$$

Для определения  $p_0^+$ , как обычно, воспользуемся условием нормировки:

$$\sum_{i=0}^{\infty} p_i^+ = P^+(1) = 1.$$

Применяя правило Лопиталья, имеем:

$$\frac{a\hat{\delta} - 1 - a\hat{\delta}^*}{a\hat{\delta} - 1} p_0^+ = 1,$$

что приводит к следующему выражению для  $p_0^+$ :

$$p_0^+ = \frac{1 - a\hat{\delta}}{1 - a\hat{\delta} + a\hat{\delta}^*} = \frac{1 - \rho}{1 - \rho + a\hat{\delta}^*}.$$

Из полученной формулы, в частности, следует, что условия  $\rho < 1$  и  $\hat{c}^* < \infty$  являются необходимыми для эргодичности вложенной цепи Маркова.

Дифференцируя формулу (2) в точке  $z = 1$  соответствующее число раз, можно получить моменты любого порядка стационарного распределения числа заявок в системе по вложенной цепи Маркова. Так, среднее число заявок задается формулой:

$$N^+ = P^{+'}(1) = \frac{a^2 \tilde{\delta}}{2(1 - \rho)} - \frac{a^2 \tilde{\delta} - 2a\hat{\delta}^* - a^2 \tilde{\delta}^*}{2(1 - \rho + a\hat{\delta}^*)},$$

где

$$\tilde{\delta} = \tilde{b}(1 + c\hat{c})^2 + \hat{b}c(2\hat{c} + \tilde{c});$$

$$\tilde{\delta}^* = \delta^{*''}(1) = a \left( \tilde{\delta} \mathfrak{a}_1(\bar{a}) \frac{1}{1 - a\tilde{c}^*} + \right.$$

$$\left. + \tilde{\delta}_0^* \frac{\mathfrak{a}_1(\bar{a}) - \tilde{c}^*}{\bar{a}} + \tilde{\delta}_0^* \mathfrak{a}_2(\bar{a}) \frac{\bar{a} - \chi^*(\bar{a})}{a\bar{a}} + \right.$$

$$\left. + 2\tilde{\delta}_0^* \mathfrak{a}_2(\bar{a}) \left[ \frac{a\hat{c}^* - 1 + \chi^*(\bar{a})}{a^2} \right] + \right.$$

$$\left. + \mathfrak{a}_2(\bar{a}) \left[ \frac{\tilde{c}^*}{a} - 2 \frac{a\hat{c}^* - 1 + \chi^*(\bar{a})}{a^3} \right] \right);$$

$$\tilde{b} = \beta''(1) = \sum_{k=1}^{\infty} k(k-1)b_k;$$

$$\tilde{c} = \chi''(1) = \sum_{k=1}^{\infty} k(k-1)c_k;$$

$$\tilde{\delta}_0^* = \delta_0^{*''}(1) = \tilde{\delta} - 2c\hat{c} - c\tilde{c} + 2c\hat{c}^2;$$

$$\tilde{c}^* = \chi^{*''}(1) = \sum_{k=1}^{\infty} k(k-1)c_k^*.$$

Далее при вычислении стационарных вероятностей числа заявок по моментам поступления в систему удобно пользоваться другой цепью Маркова  $\{\nu_n^-, n \geq 1\}$ , отличающейся от введенной выше только тем, что если в момент  $\tau_n$  в систему поступает новая заявка, то число  $\nu_n^-$  определяется до ее прихода.

Обозначая через  $p_i^-, i \geq 0$ , стационарные вероятности цепи  $\nu_n^-$ , получаем следующие соотношения, связывающие  $p_i^-$  и  $p_i^+$ :

$$p_0^+ = \bar{a}p_0^-; \quad (3)$$

$$p_i^+ = \bar{a}p_i^- + ap_{i-1}^-, \quad i \geq 1. \quad (4)$$

Вводя ПФ  $P^-(z) = \sum_{i=0}^{\infty} z^i p_i^-$ , получаем из равенств (3) и (4)

$$P^-(z) = \frac{1}{az + \bar{a}} P^+(z).$$

Приведем также формулу для среднего числа заявок  $N^-$  по цепи Маркова  $\nu_n^-$ :

$$N^- = P^{-'}(1) = N^+ - a.$$

## 5 Стационарные вероятности состояний по времени и по моментам поступления заявок

Пусть в некоторый момент в свободную систему (на свободный прибор, который, возможно, в этот момент уже ремонтируется или только начинает ремонтироваться) поступает заявка, обслуживание которой (с учетом возможных прерываний обслуживания на ремонт прибора) закончится через  $k$  тактов.

Обозначим через  $t_i(k)$ ,  $k \geq 1$ ,  $i = \overline{0, k-1}$ , среднее число таких тактов до момента  $k$  окончания обслуживания выделенной заявки, когда в системе в очереди будет находиться  $i$  заявок. Через  $t_i$ ,  $i \geq 0$ , обозначим аналогичное среднее, но для заявки произвольной длины, поступающей в тот момент, когда заканчивается обслуживание заявки на приборе (и в очереди отсутствуют другие заявки), а через  $t_i^*$ ,  $i \geq 0$ , — для заявки произвольной длины, поступающей в свободную систему (т. е. систему, в которой, по крайней мере, уже на предыдущем такте отсутствовали заявки). Наконец, через  $s_0^*$  обозначим среднюю длину свободного периода, т. е. среднее время от момента освобождения системы до момента поступления следующей заявки. Тогда

$$t_i(k) = \sum_{j=i}^{k-1} \binom{j}{i} a^i \bar{a}^{j-i}, \quad k \geq 1, \quad i = \overline{0, k-1};$$

$$t_i = \sum_{k=i+1}^{\infty} t_i(k) d_k = \sum_{k=i+1}^{\infty} \sum_{j=i}^{k-1} \binom{j}{i} a^i \bar{a}^{j-i} d_k, \quad i \geq 0;$$

$$t_i^* = \sum_{k=i+1}^{\infty} t_i(k) d_k^* = \sum_{k=i+1}^{\infty} \sum_{j=i}^{k-1} \binom{j}{i} a^i \bar{a}^{j-i} d_k^*, \quad i \geq 0;$$

$$s_0^* = \sum_{k=1}^{\infty} k a \bar{a}^{k-1} = \frac{1}{a}.$$

Заметим, что ПФ  $T(z) = \sum_{i=0}^{\infty} t_i z^i$  и  $T^*(z) = \sum_{i=0}^{\infty} t_i^* z^i$  последовательностей  $\{t_i, i \geq 0\}$  и  $\{t_i^*, i \geq 0\}$  имеют вид:

$$T(z) = \sum_{i=0}^{\infty} z^i \sum_{k=i+1}^{\infty} \sum_{j=i}^{k-1} \binom{j}{i} a^i \bar{a}^{j-i} d_k = \frac{1 - \delta(az + \bar{a})}{a(1-z)};$$

$$T^*(z) = \sum_{i=0}^{\infty} z^i \sum_{k=i+1}^{\infty} \sum_{j=i}^{k-1} \binom{j}{i} a^i \bar{a}^{j-i} d_k^* = \frac{1 - \delta^*(az + \bar{a})}{a(1-z)}.$$

Введем обозначения:

$t$  — среднее число тактов между соседними моментами изменения состояний вложенной цепи Маркова в стационарном режиме функционирования системы;

$s_i$ ,  $i \geq 0$ , — среднее число таких тактов между соседними моментами изменения состояний вложенной цепи Маркова в стационарном режиме функционирования системы, когда в системе находится  $i$  заявок.

Тогда

$$t = p_0^+(s_0^* + \hat{\delta}^*) + (1 - p_0^+) \hat{\delta} = \frac{1}{a};$$

$$s_0 = p_0^+ s_0^*;$$

$$s_i = p_0^+ t^*_{i-1} + \sum_{j=1}^i p_j^+ t_{i-j}, \quad i \geq 1.$$

Стационарное распределение  $\{p_i, i \geq 0\}$  числа заявок в системе по времени (т. е. по интервалам времени между соседними моментами окончания тактов) находится по формуле:

$$p_i = \frac{s_i}{t} = a s_i, \quad i \geq 0.$$

Переходя к ПФ  $P(z) = \sum_{i=0}^{\infty} p_i z^i$  и производя элементарные преобразования, получаем:

$$P(z) = \frac{1}{t} \left( p_0^+ \left[ s_0^* + \sum_{i=1}^{\infty} z^i t_{i-1}^* \right] + \sum_{i=1}^{\infty} z^i \sum_{j=1}^i p_j^+ t_{i-j} \right) = P^+(z). \quad (5)$$

Таким образом, стационарное распределение числа заявок в системе по времени совпадает со стационарным распределением числа заявок по цепи Маркова  $\nu_n$ , порожденной моментами уходов заявок из системы. Иными словами, для рассматриваемой системы выполнен закон стационарной очереди Хинчина. В частности, совпадают моменты любых порядков стационарных распределений числа заявок в системе по времени и по цепи Маркова  $\nu_n$ . Например, среднее число заявок в системе  $N$  по времени имеет вид:

$$N = P'(1) = N^+.$$

Стационарные вероятности  $p_k^*$ ,  $k \geq 0$ , того, что в момент поступления заявки в систему в ней будет  $k$  заявок (напомним, что в соответствии с принятыми правилами в число  $k$  не входит заявка, обслуживание которой заканчивается в этот момент) можно найти, например, следующим образом.

Введем обозначения:

$$D_i = \sum_{j=i+1}^{\infty} d_j, \quad i \geq 0;$$

$$D_i^* = \sum_{j=i+1}^{\infty} d_j^*, \quad i \geq 0.$$

В момент 0 (прихода выделенной заявки) в системе будут отсутствовать заявки, если в момент  $-i$ ,  $i \geq 0$ , закончилось обслуживание единственной заявки (с вероятностью  $p_0^-/t$ ) и за время  $i$  новые заявки не поступят (с вероятностью  $\bar{a}^i$ ). Поэтому

$$p_0^* = \frac{1}{t} p_0^- \sum_{i=0}^{\infty} \bar{a}^i = \frac{1}{at} p_0^- = p_0^-.$$

В системе в момент 0 (прихода выделенной заявки) будет  $k$ ,  $k \geq 1$ , других заявок в одном из трех случаев:

- (1) в момент  $-i$ ,  $i \geq k$ , закончилось обслуживание единственной заявки (с вероятностью  $p_0^-/t$ ), и в этот же момент в систему (на прибор) поступила новая заявка (с вероятностью  $a$ ), которая будет обслуживаться более  $i$  тактов (с вероятностью  $D_i$ ) и за  $i-1$  тактов, оставшихся до момента 0, поступит еще  $k-1$  заявок (с вероятностью  $\binom{i-1}{k-1} a^{k-1} \bar{a}^{i-k}$ );
- (2) в некоторый предыдущий момент произошло последнее до момента 0 окончание обслуживания единственной заявки (с вероятностью  $p_0^-/t$ ), но, в отличие от предыдущего случая, в этот момент новая заявка в систему не поступила (с вероятностью  $\bar{a}$ ). Затем в момент  $-i$ ,  $i \geq k$ , в свободную систему (на прибор) поступила новая заявка, которая будет обслуживаться более  $i$  тактов (с вероятностью  $D_i^*$ ) и за  $i-1$  тактов, оставшихся до момента 0, поступит еще  $k-1$  заявок (с вероятностью  $\binom{i-1}{k-1} a^{k-1} \bar{a}^{i-k}$ );
- (3) в момент  $-i$ ,  $i \geq k-j$ , закончилось обслуживание заявки на приборе и в системе осталось  $j$ ,  $j = \bar{1}, k$ , других заявок (с вероятностью  $p_j^-/t$ ). Новая заявка, перешедшая из очереди на прибор, будет обслуживаться более  $i$  тактов (с вероятностью  $D_i$ ), и за  $i$  тактов, оставшихся до момента 0, поступит еще  $k-j$  заявок

(с вероятностью  $\binom{i}{k-j} a^{k-j} \bar{a}^{i-k+j}$ ). Таким образом,

$$p_k^* = \frac{1}{t} \left( p_0^- a \sum_{i=k}^{\infty} D_i \binom{i-1}{k-1} a^{k-1} \bar{a}^{i-k} + \right. \\ \left. + p_0^- \bar{a} \sum_{i=k}^{\infty} D_i^* \binom{i-1}{k-1} a^{k-1} \bar{a}^{i-k} + \right. \\ \left. + \sum_{j=1}^k p_j^- \sum_{i=k-j}^{\infty} D_i \binom{i}{k-j} a^{k-j} \bar{a}^{i-k+j} \right), \quad k \geq 1.$$

В терминах ПФ  $P^*(z) = \sum_{k=0}^{\infty} z^k p_k^*$  стационарное распределение числа заявок в системе по моментам поступления заявок определяется выражением:

$$P^*(z) = \frac{1}{1-z} \left[ \frac{\delta(az + \bar{a}) - z\delta^*(az + \bar{a})}{az + \bar{a}} p_0^+ + \right. \\ \left. + [1 - \delta(az + \bar{a})] P^-(z) \right],$$

или после элементарных преобразований

$$P^*(z) = P^-(z),$$

т.е. совпадает со стационарным распределением числа заявок по цепи Маркова  $\nu_n^-$ .

Соответственно, среднее число заявок  $N^*$  по моментам поступления совпадает со средним числом заявок  $N^-$  по цепи Маркова  $\nu_n^-$ .

## 6 Стационарное распределение времени пребывания заявки в системе

Рассмотрим сначала моменты после окончания очередного такта функционирующей в стационарном режиме системы и введем следующие стационарные вероятности:

- $w_0$  — вероятность того, что прибор свободен и исправен;
- $w_{0,i}$ ,  $i \geq 1$ , — вероятность того, что прибор свободен, но неисправен, причем до окончания ремонта осталось  $i$  тактов;
- $w_i$ ,  $i \geq 1$ , — вероятность того, что в системе имеются заявки и суммарная работа по обслуживанию всех этих заявок (с учетом возможных прерываний) равна  $i$ .

Тогда

$$w_0 = \bar{a}(\bar{c}^* w_0 + w_{0,1} + \bar{c}^* w_1); \quad (6)$$

$$w_{0,i} = \bar{a}w_{0,i+1} + \bar{a}c^*c_i^*(w_0 + w_1), \quad i \geq 1; \quad (7)$$

$$w_i = \bar{a}w_{i+1} + ad_iw_0 + \sum_{j=1}^i aw_{0,j}d_{0,i-j+1}^* + \sum_{j=1}^i aw_jd_{i-j+1}, \quad i \geq 1. \quad (8)$$

Равенства (7) и (8) в терминах ПФ  $w_0(z) = \sum_{i=1}^{\infty} w_{0,i}z^i$  и  $w(z) = \sum_{i=1}^{\infty} w_i z^i$  могут быть записаны в виде:

$$w_0(z) = \bar{a}\frac{1}{z}(w_0(z) - zw_{0,1}) + \bar{a}c^*\chi^*(z)(w_0 + w_1); \quad (9)$$

$$w(z) = \bar{a}\frac{1}{z}(w(z) - zw_1) + a\delta(z)w_0 + a\frac{1}{z}[w_0(z)\delta_0^*(z) + w(z)\delta(z)]. \quad (10)$$

Из формулы (9) имеем:

$$w_0(z) = \frac{\bar{a}z[c^*\chi^*(z)(w_0 + w_1) - w_{0,1}]}{z - \bar{a}}. \quad (11)$$

Поскольку в точке  $z = \bar{a}$  знаменатель правой части равенства (11) обращается в нуль, то в силу непрерывности функции  $w_0(z)$  на отрезке  $[0, 1]$  в этой точке в нуль должен обращаться и числитель, что дает

$$w_{0,1} = c^*\chi^*(\bar{a})(w_0 + w_1). \quad (12)$$

Формулы (6), (11) и (12) позволяют выразить  $w_1$  и  $w_0(z)$  через  $w_0$ :

$$w_1 = \left[ \frac{1}{\bar{a}[\bar{c}^* + c^*\chi^*(\bar{a})]} - 1 \right] w_0;$$

$$w_0(z) = \frac{\bar{a}c^*z[\chi^*(z) - \chi^*(\bar{a})](w_0 + w_1)}{z - \bar{a}} = \frac{c^*z[\chi^*(z) - \chi^*(\bar{a})]}{[\bar{c}^* + c^*\chi^*(\bar{a})](z - \bar{a})} w_0. \quad (13)$$

Подставляя найденные значения  $w_1$  и  $w_0(z)$  в (10), имеем:

$$w(z) = \frac{z}{z - \bar{a} - a\delta(z)} \left( a\delta(z) - \frac{1}{\bar{c}^* + c^*\chi^*(\bar{a})} + \bar{a} + ac^*\delta_0^*(z) \frac{\chi^*(z) - \chi^*(\bar{a})}{[\bar{c}^* + c^*\chi^*(\bar{a})](z - \bar{a})} \right) w_0. \quad (14)$$

Для определения  $w_0$  воспользуемся условием нормировки, которое в данном случае записывается так:

$$w_0 + w_0(1) + w(1) = 1.$$

Подставляя в это равенство вместо  $w_0(1)$  и  $w(1)$  их значения, вычисленные по формулам (13) и (14), применяя правило Лопиталья и производя элементарные преобразования, получаем:

$$w_0 = \frac{1 - \rho}{1 - \rho + a\hat{\delta}^*} \frac{a[\bar{c}^* + c^*\chi^*(\bar{a})]}{1 - \bar{a}c^* - \bar{a}c^*\chi^*(\bar{a})}.$$

Естественно,

$$w_0 + w_0(1) = \frac{1 - \rho}{1 - \rho + a\hat{\delta}^*} = p_0 = p_0^+.$$

Введем теперь следующие стационарные вероятности, связанные с временем ожидания начала обслуживания заявки:

$w_0^-$  — вероятность того, что поступающая заявка сразу же попадает на прибор (который, возможно, ремонтируется);

$w_i^-, i \geq 1$ , — вероятность того, что поступающая заявка будет ожидать попадания на прибор (который, возможно, ремонтируется)  $i$  тактов;

$w_0^+$  — вероятность того, что поступающая заявка сразу же начнет обслуживаться (прибор свободен и исправен);

$w_i^+, i \geq 1$ , — вероятность того, что поступающая заявка начнет обслуживаться (на исправном приборе) через  $i$  тактов.

Тогда

$$w_0^- = w_0 + \sum_{j=1}^{\infty} w_{0,j} + w_1 = w_0 + w_0(1) + w_1 = \frac{1}{\bar{a}}p_0;$$

$$w_i^- = w_{i+1}, \quad i \geq 1;$$

$$w_0^+ = w_0\bar{c} + w_{0,1} + w_1\bar{c} = \frac{1 - \rho}{1 - \rho + a\hat{\delta}^*} \cdot \frac{a[\bar{c} + c^*\chi^*(\bar{a})]}{\bar{a}[1 - \bar{a}c^* - \bar{a}c^*\chi^*(\bar{a})]};$$

$$w_i^+ = w_0cc_i + w_{0,i+1} + w_{i+1}\bar{c} + \sum_{j=0}^{i-1} w_{j+1}cc_{i-j}, \quad i \geq 1.$$

В терминах ПФ  $w^-(z) = \sum_{i=0}^{\infty} z^i w_i^-$  стационарного распределения времени ожидания момента поступления заявки на прибор и ПФ  $w^+(z) = \sum_{i=0}^{\infty} z^i w_i^+$  стационарного распределения времени ожидания

начала обслуживания заявки на приборе последние формулы приобретают вид:

$$w^-(z) = w_0 + w_0(1) + \frac{1}{z}w(z);$$

$$w^+(z) = \bar{c}w_0 + c\chi(z)w_0 + \frac{1}{z}w_0(z) + \quad (15)$$

$$+ \frac{\bar{c}}{z}w(z) + \frac{c}{z}\chi(z)w(z). \quad (16)$$

Наконец, стационарная вероятность  $v_i, i \geq 0$ , того, что заявка будет находиться в системе  $i$  тактов, определяется соотношением:

$$v_i = \sum_{j=0}^i w_j^+ d_{0,i-j}^*, \quad i \geq 0,$$

или в терминах ПФ  $v(z) = \sum_{i=0}^{\infty} z^i v_i$  стационарного распределения общего времени пребывания заявки в системе:

$$v(z) = w^+(z)\delta_0^*(z). \quad (17)$$

Как обычно, дифференцируя полученные в этом разделе формулы, можно найти моменты соответствующих характеристик. В частности, стационарные средние значения времени ожидания момента поступления заявки на прибор, времени ожидания начала обслуживания заявки на приборе и общего времени пребывания заявки в системе определяются выражениями:

$$\begin{aligned} \hat{w}^- &= w'^-(1) = w'(1) - w(1); \\ \hat{w}^+ &= w'^+(1) = \hat{c}w_0 + w'_0(1) - w_0(1) + \\ &\quad + w'(1) - (1 - \hat{c})w(1); \\ \hat{v} &= v'(1) = \hat{w}^+ + \hat{\delta}_0^*, \end{aligned}$$

где

$$\begin{aligned} w_0(1) &= \frac{c^*[1 - \chi^*(\bar{a})]}{a[\bar{c}^* + c^*\chi^*(\bar{a})]}w_0; \\ w'_0(1) &= w_0(1) + \\ &\quad + \frac{1}{\bar{c}^* + c^*\chi^*(\bar{a})} \left( \frac{c^*\hat{c}^*}{a} - \frac{c^*[1 - \chi^*(\bar{a})]}{a^2} \right) w_0; \\ w(1) &= \frac{a\hat{\delta}^*}{1 - \rho + a\hat{\delta}^*}; \\ w'(1) &= w(1) + \frac{1}{2(1 - \rho)} a\tilde{\delta}w(1) + \\ &\quad + \frac{1}{2(1 - \rho)} \left( a\tilde{\delta} + \frac{c^*}{\bar{c}^* + c^*\chi^*(\bar{a})} \left[ \tilde{\delta}_0^*[1 - \chi^*(\bar{a})] + \right. \right. \\ &\quad \left. \left. + 2\hat{\delta}_0^*\hat{c}^* + \bar{c}^* - 2\frac{1}{a} \left( \hat{\delta}_0^*[1 - \chi^*(\bar{a})] + \right. \right. \right. \\ &\quad \left. \left. \left. + \bar{c}^* - \frac{1 - \chi^*(\bar{a})}{a} \right) \right] \right) w_0. \end{aligned}$$

Производя арифметические преобразования, приходим к равенству:

$$N^+ = a\hat{v},$$

которое представляет собой формулу Литтла для данной системы.

## 7 Система Geo<sub>2</sub>/G<sub>2</sub>/1/∞ с двумя типами заявок и абсолютным приоритетом

Рассмотрим теперь СМО Geo<sub>2</sub>/G<sub>2</sub>/1/∞, в которую поступают два независимых геометрических потока заявок.

Заявки второго потока, или неприоритетные заявки, поступают на такте с вероятностью  $a$  и обслуживаются в соответствии с распределением  $\{b_k, k \geq 0\}$ . Заявки первого потока, или приоритетные заявки, поступают на такте с вероятностью  $c$  и обслуживаются в соответствии с распределением  $\{g_k, k \geq 0\}$ . Заявки первого потока имеют абсолютный приоритет перед заявками второго потока, т. е. если поступающая заявка первого потока застает на приборе заявку второго потока, то она прерывает ее обслуживание и сама становится на прибор. Заявка второго потока с прерванным обслуживанием становится в очередь неприоритетных заявок и, как только система освобождается от всех приоритетных заявок, возвращается на прибор и продолжает обслуживаться. Заявки одного потока обслуживаются в порядке поступления.

Покажем, что те формулы, которые были получены выше для СМО Geo/G/1/∞ с ненадежным прибором, могут быть применены и к системе Geo<sub>2</sub>/G<sub>2</sub>/1/∞ с абсолютным приоритетом для нахождения стационарных распределений числа заявок в системе и времени пребывания заявки каждого потока в системе.

Заметим, прежде всего, что неприоритетные заявки не оказывают никакого влияния на обслуживание приоритетных заявок. Поэтому распределения числа приоритетных заявок в системе и времени обслуживания приоритетной заявки определяются как и для обычной системы Geo/G/1/∞, в которую поступают только приоритетные заявки.

При определении стационарных характеристик для неприоритетных заявок предположим, что неприоритетные заявки обслуживаются «вслепую», т. е. они не знают, из-за чего происходит прерывание обслуживания. Тогда с их точки зрения система представляет собой просто систему Geo/G/1/∞ с ненадежным прибором, в которой вероятность поступления заявки на такте равна вероятности  $a$

поступления неприоритетной заявки, вероятности отказов прибора в свободном и занятом состояниях совпадают и равны  $c$ , а время обслуживания заявки имеет распределение  $\{b_k, k \geq 0\}$ . Кроме того, совпадают распределения  $\{c_k, k \geq 0\}$  и  $\{c_k^*, k \geq 0\}$  времен ремонта прибора. Однако прерывания происходят не на время обслуживания приоритетной заявки, а на то время, пока система полностью не освободится от всех приоритетных заявок. Поэтому распределения  $\{c_k, k \geq 0\}$  и  $\{c_k^*, k \geq 0\}$  равны не  $\{g_k, k \geq 0\}$ , а распределению периода занятости системы приоритетными заявками. Напомним, что ПФ  $\gamma(z)$  периода занятости системы  $\vec{G}_{02}/\vec{G}_2/1/\infty$  приоритетными заявками, совпадающая с ПФ периода занятости системы  $Geo/G/1/\infty$ , в которую поступают только приоритетные заявки, определяется из уравнения:

$$\gamma(z) = g(cz + \bar{c}z\gamma(z)),$$

где  $g(z)$  — ПФ распределения  $\{g_k, k \geq 0\}$ :

$$g(z) = \sum_{k=0}^{\infty} z^k g_k.$$

Подставляя эти значения в формулы (5), (15) и (17), получаем ПФ стационарных распределений числа неприоритетных заявок в системе, времени ожидания неприоритетной заявкой начала обслуживания и полного времени пребывания неприоритетной заявки в системе.

Используя результаты предыдущих разделов, нетрудно показать, что необходимым и достаточным условием существования стационарного режима функционирования является

$$\rho = a\hat{b} + c\hat{g} < 1,$$

где  $\hat{b} = \sum_{k=0}^{\infty} kb_k$  и  $\hat{g} = \sum_{k=0}^{\infty} kg_k$  — средние длины приоритетных и неприоритетных заявок соответственно, а  $\rho$  — суммарная загрузка системы, а также найти стационарные средние значения  $N$ ,  $w$  и  $v$  для неприоритетных заявок.

## 8 Заключение

Таким образом, в настоящей статье получены аналитические соотношения, позволяющие вычислять основные стационарные характеристики обслуживания функционирующей в дискретном времени однолинейной СМО с ненадежным прибором, который может отказывать как в рабочем, так и в свободном состояниях. Отметим, что отказ свободного прибора и последующий ремонт можно трактовать как проведение случайной профилактики, что и сделаем в следующем примере.

**Пример.** Пусть рассматриваемая СМО имеет следующие параметры:

средняя длина заявки  $\hat{b} = 4,6$ ;

факториальный момент второго порядка длины заявки  $\hat{b} = 17,0$ ;

вероятность отказа занятого прибора  $c = 0,3$ ;

среднее время ремонта прибора, отказавшего в занятом состоянии,  $\hat{c} = 3,9$ ;

факториальный момент второго порядка времени ремонта прибора, отказавшего в занятом состоянии,  $\hat{c} = 16,1$ ;

вероятность отказа свободного прибора (начала профилактики)  $c^* = 0,11$ ;

параметр распределения времени ремонта (профилактики) прибора, отказавшего в свободном состоянии,  $\lambda = 0,4$ .

Предполагается, что время профилактики имеет сдвинутое на единицу распределение Паскаля с показателем 2 и параметром  $\lambda$  (это распределение представляет собой распределение суммы двух независимых геометрических случайных величин с параметром  $\lambda$ ).

В табл. 1 и 2 для некоторых значений вероятности  $a$  поступления заявки на такте приведены результаты расчетов загрузки  $\rho$ , среднего числа заявок  $N$  и среднего времени  $\hat{v}$  пребывания заявки в системе для 4-х вариантов:

(AA) — прибор в рабочем состоянии не отказывает, профилактика не проводится;

(AB) — прибор в рабочем состоянии не отказывает, профилактика проводится;

(BA) — прибор в рабочем состоянии отказывает, профилактика не проводится;

(BB) — прибор в рабочем состоянии отказывает, профилактика проводится.

Как видно из таблиц, для данного примера наличие или отсутствие профилактики при прочих

**Таблица 1** Результаты расчетов вариантов AA и AB

a	$\rho$	N		$\hat{v}$	
		AA	AB	AA	AB
0,03	0,138	0,1469	0,1694	4,8958	5,6470
0,05	0,230	0,2576	0,2954	5,1520	5,9070
0,07	0,322	0,3834	0,4365	5,4776	6,2362
0,09	0,414	0,5315	0,6001	5,9055	6,6673
0,11	0,506	0,7142	0,7983	6,4927	7,2576
0,13	0,598	0,9553	1,0551	7,3486	8,1165
0,15	0,690	1,3069	1,4225	8,7129	9,4832
0,17	0,782	1,9088	2,0402	11,228	12,001
0,19	0,874	3,3093	3,4566	17,418	18,192
0,21	0,966	11,991	12,154	57,100	57,877

Таблица 2 Результаты расчетов вариантов ВА и ВВ

$a$	$\rho$	$N$		$\hat{v}$	
		ВА	ВВ	ВА	ВВ
0,01	0,0998	0,1061	0,1101	10,610	11,006
0,02	0,1996	0,2279	0,2360	11,394	11,799
0,03	0,2995	0,3721	0,3845	12,402	12,816
0,04	0,3993	0,5498	0,5667	13,745	14,167
0,05	0,4991	0,7812	0,8026	15,624	16,053
0,06	0,5989	1,1062	1,1324	18,437	18,873
0,07	0,6987	1,6180	1,6491	23,114	23,558
0,08	0,7986	2,5942	2,6302	32,427	32,878
0,09	0,8984	5,4033	5,4445	60,036	60,494
0,10	0,9982	314,98	315,03	3149,8	3150,3

равных параметрах практически не влияет на характеристики функционирования. В то же время замена ненадежного прибора безотказным значительно улучшает качество функционирования СМО (в частности, пропускаемая нагрузка — максимальное значение вероятности  $a$  поступления заявки на такте — увеличивается более чем в два раза).

## Литература

1. Печинкин А. В., Соколов И. А., Чаплыгин В. В. Многолинейная система массового обслуживания с групповым отказом приборов // Информатика и её применения, 2009. Т. 3. Вып. 3. С. 4–15.
2. Gaver D. P. A waiting line with interrupted service, including priorities // J. Roy. Stat. Soc. B, 1962. Vol. 24. P. 73–90.
3. Dimitrov B., Dokev Ch. The single server queue system with non-reliable server in discrete time. Non-stationary characteristics // Ann. of Univ. of Sofia. Ser. Math., 1981. Vol. 70. P. 175–190.
4. Altiook. T. Queueing modeling of a single processor with failures // Performance Evaluation, 1989. Vol. 9. No. 2. P. 93–102.
5. Takagi H. Queueing analysis. Vol. 3: Discrete-time systems. — Amsterdam: North-Holland, 1993.
6. Duan-Shin Lee. Analysis of a single server queue with semi-Markovian service interruption // Queueing Syst.: Theory Appl., 1997. Vol. 27. No. 1–2. P. 153–178.
7. Fiems D., Bruneel H. Analysis of a discrete-time queueing system with timed vacations // Queueing Syst., 2002. Vol. 42. P. 243–254.
8. Fiems D., Steyaert B., Bruneel H. Analysis of a discrete-time GI–G–1 queueing model subjected to bursty interruptions // Computers Operations Res., 2003. Vol. 30. P. 139–153.
9. Moreno P. A discrete-time retrial queue with unreliable server and general server lifetime // J. Math. Sci., 2006. Vol. 132. P. 643–655.
10. Atencia I., Moreno P. A discrete-time Geo/G/1 retrial queue with the server subject to starting failures // Annals Operations Res., 2006. Vol. 141. No. 1. P. 85–107.
11. Demoor T., Fiems D., Walraevens J., Bruneel H. The preemptive repeat hybrid server interruption model. — Berlin: Springer-Verlag, 2010.
12. Bocharov P. P., D'Apice C., Manzo R., Pechinkin A. V. Analysis of the multi-server Markov queueing system with unlimited buffer and negative customers // Automation Remote Control, 2007. Vol. 68. No. 1. P. 85–94.
13. Manzo R., Cascone N., Razumchik R. V. Exponential queueing system with negative customers and bunker for ousted customers // Automation Remote Control, 2008. Vol. 69. No. 9. P. 1542–1551.
14. D'Apice C., Manzo R., Pechinkin A. V. A finite MAP<sub>K</sub>/G<sub>K</sub>/1 queueing system with generalized foreground-background processor-sharing discipline // Automation Remote Control, 2004. Vol. 65. No. 11. P. 114–121.

## ОБ ОДНОМ КЛАССЕ МАРКОВСКИХ СИСТЕМ ОБСЛУЖИВАНИЯ\*

Я. А. Сатин<sup>1</sup>, А. И. Зейфман<sup>2</sup>, А. В. Коротышева<sup>3</sup>, С. Я. Шоргин<sup>4</sup>

**Аннотация:** Рассматриваются модели обслуживания, описываемые конечными марковскими цепями с непрерывным временем. При этом предполагается, что интенсивности поступления и обслуживания требований не зависят от числа требований в системе. Получены оценки скорости сходимости и устойчивости различных характеристик таких систем.

**Ключевые слова:** нестационарные марковские системы обслуживания; скорость сходимости; устойчивость; оценки

## 1 Введение

Классы систем массового обслуживания, описываемых процессами рождения и гибели (стационарными и нестационарными, с катастрофами) изучались начиная с 1970-х гг. многими авторами (см., например, [1–6]). С помощью методов, разработанных одним из авторов настоящей статьи (подробное изложение этих методов приведено в [7–9]), для таких систем удалось получить точные оценки скорости сходимости и устойчивости.

Оказывается, этот же подход можно применить и к существенно более общему классу систем обслуживания.

Рассмотрим систему массового обслуживания, число требований в которой описывается нестационарной марковской цепью с непрерывным временем и конечным пространством состояний, причем требования могут поступать и обслуживаться группами.

Пусть  $X = X(t)$ ,  $t \geq 0$ , — число требований в системе обслуживания ( $0 \leq X(t) \leq r$ ).

Обозначим через

$$p_{ij}(s, t) = \Pr \{X(t) = j | X(s) = i\}, \\ i, j \geq 0, 0 \leq s \leq t,$$

переходные вероятности процесса  $X = X(t)$ , а через  $p_i(t) = \Pr \{X(t) = i\}$  — его вероятности состояний.

Будем предполагать, что интенсивности поступления и обслуживания  $k$  требований в момент  $t$  в системе обслуживания ( $\lambda_k(t)$  и  $\mu_k(t)$  соответственно) не зависят от числа требований, находящихся в системе в момент  $t$ , являются локально интегрируемыми на  $[0, \infty)$  функциями времени  $t$  и, кроме

того,  $\lambda_{k+1}(t) \leq \lambda_k(t)$  и  $\mu_{k+1}(t) \leq \mu_k(t)$  при всех  $k$  и почти при всех  $t \geq 0$ .

Тогда для описания вероятностной динамики процесса получаем прямую систему Колмогорова в виде

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}(t), \quad (1)$$

где

$$A(t) = \begin{pmatrix} a_{00}(t) & \mu_1(t) & \mu_2(t) & \mu_3(t) & \mu_4(t) & \cdots & \mu_r(t) \\ \lambda_1(t) & a_{11}(t) & \mu_1(t) & \mu_2(t) & \mu_3(t) & \cdots & \mu_{r-1}(t) \\ \lambda_2(t) & \lambda_1(t) & a_{22}(t) & \mu_1(t) & \mu_2(t) & \cdots & \mu_{r-2}(t) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_r(t) & \lambda_{r-1}(t) & \lambda_{r-2}(t) & \cdots & \lambda_2(t) & \lambda_1(t) & a_{rr}(t) \end{pmatrix},$$

причем

$$a_{ii}(t) = - \sum_{k=1}^i \mu_k(t) - \sum_{k=1}^{r-i} \lambda_{r-k}(t).$$

Далее будем обозначать через  $\|\bullet\|$   $l_1$ -норму, т. е.  $\|x\| = \sum |x_i|$ , а  $\|B\| = \max_j \sum_i |b_{ij}|$ , если  $B = (b_{ij})_{i,j=0}^r$ . Тогда, в частности, имеем

$$\|A(t)\| \leq 2 \sum_{k=1}^r (\lambda_k(t) + \mu_k(t))$$

при всех  $t \geq 0$ .

Через

$$E(t, k) = E \{X(t) | X(0) = k\}$$

будем далее обозначать математическое ожидание процесса (среднее число требований) в момент  $t$

\* Исследование поддержано РФФИ, гранты 11-07-00112-а и 11-01-12026-офи-м.

<sup>1</sup> Вологодский государственный педагогический университет, yasovi@mail.ru

<sup>2</sup> Вологодский государственный педагогический университет; Институт проблем информатики Российской академии наук; Институт социально-экономического развития территорий Российской академии наук, a\_zeifman@mail.ru

<sup>3</sup> Вологодский государственный педагогический университет, a\_korotysheva@mail.ru

<sup>4</sup> Институт проблем информатики Российской академии наук, SShorgin@ipiran.ru

при условии, что в нулевой момент времени он находится в состоянии  $k$ , а через  $E_{\mathbf{p}}(t)$  обозначим математическое ожидание процесса в момент  $t$  при начальном распределении вероятностей состояний  $\mathbf{p}(0) = \mathbf{p}$ .

Тогда  $X(t)$  слабо эргодичен, при любых начальных условиях  $\mathbf{p}^*(s)$ ,  $\mathbf{p}^{**}(s)$  и любых  $s, t, 0 \leq s \leq t$ , справедлива оценка

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq \frac{8G}{d} e^{-\int_s^t \alpha(u) du}. \quad (3)$$

## 2 Оценки скорости сходимости

Рассмотрим вспомогательную последовательность положительных чисел  $\{d_i\}, i = 1, \dots, r$ .

Положим

$$d = \min_{1 \leq i \leq r} d_i; \quad G = \sum_{i=1}^r d_i; \quad W = \min_k \frac{d_k}{k}.$$

Рассмотрим величины

$$\begin{aligned} \alpha_i(t) = & -a_{ii}(t) + \lambda_{r-i+1}(t) - \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \\ & - \mu_i(t)) \frac{d_k}{d_i} - \sum_{k=1}^{r-i} (\lambda_k(t) - \lambda_{i+r-1}(t)) \frac{d_{k+i}}{d_i}, \\ \alpha(t) = & \min_{1 \leq i \leq r} \alpha_i(t). \end{aligned}$$

**Теорема 1.** Пусть существует последовательность положительных чисел  $\{d_j\}$  такая, что

$$\int_0^{\infty} \alpha(t) dt = +\infty. \quad (2)$$

Кроме того,  $X(t)$  имеет предельное среднее  $\phi(t)$  и при любых  $k$  и  $t \geq 0$  справедливо неравенство:

$$|E(t, k) - \phi(t)| \leq \frac{4G}{W} e^{-\int_0^t \alpha(u) du}. \quad (4)$$

Доказательство.

Пользуясь предложенным в предыдущих работах способом, выразим

$$p_0 = 1 - \sum_{1 \leq i \leq r} p_i.$$

Тогда получим неоднородное уравнение:

$$\frac{dz}{dt} = B(t)z(t) + \mathbf{f}(t), \quad (5)$$

где  $\mathbf{f}(t) = (\lambda_1, \lambda_2, \dots, \lambda_r)^T$ ;

$$B = \begin{pmatrix} a_{11} - \lambda_1 & \mu_1 - \lambda_1 & \mu_2 - \lambda_1 & \mu_3 - \lambda_1 & \dots & \dots & \mu_{r-1} - \lambda_1 \\ \lambda_1 - \lambda_2 & a_{22} - \lambda_2 & \mu_1 - \lambda_2 & \mu_2 - \lambda_2 & \dots & \dots & \mu_{r-2} - \lambda_2 \\ \lambda_2 - \lambda_3 & \lambda_1 - \lambda_3 & a_{33} - \lambda_3 & \mu_1 - \lambda_2 & \dots & \dots & \mu_{r-3} - \lambda_3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \lambda_{r-1} - \lambda_r & \lambda_{r-2} - \lambda_r & \dots & \dots & \lambda_2 - \lambda_r & \lambda_1 - \lambda_r & a_{rr} - \lambda_r \end{pmatrix}.$$

Рассмотрим треугольную матрицу

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \dots & d_1 \\ 0 & d_2 & d_2 & \dots & d_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & d_r \end{pmatrix}$$

и соответствующую норму  $\|\mathbf{z}\|_D = \|D\mathbf{z}\|_1$ .

Тогда имеем:

$$DBD^{-1} = \begin{pmatrix} a_{11} - \lambda_r & (\mu_1 - \mu_2) \frac{d_1}{d_2} & (\mu_2 - \mu_3) \frac{d_1}{d_3} & \dots & (\mu_{r-1} - \mu_r) \frac{d_1}{d_r} \\ (\lambda_1 - \lambda_r) \frac{d_2}{d_1} & a_{22} - \lambda_{r-1} & (\mu_1 - \mu_3) \frac{d_2}{d_3} & \dots & (\mu_{r-2} - \mu_r) \frac{d_2}{d_r} \\ (\lambda_2 - \lambda_r) \frac{d_3}{d_1} & (\lambda_1 - \lambda_{r-1}) \frac{d_3}{d_2} & a_{33} - \lambda_{r-2} & \dots & (\mu_{r-3} - \mu_r) \frac{d_3}{d_r} \\ \dots & \dots & \dots & \dots & \dots \\ (\lambda_{r-1} - \lambda_r) \frac{d_r}{d_1} & (\lambda_{r-2} - \lambda_{r-1}) \frac{d_r}{d_2} & (\lambda_{r-3} - \lambda_{r-2}) \frac{d_r}{d_3} & \dots & a_{rr} - \lambda_1 \end{pmatrix}.$$

Далее, оценивая логарифмическую норму оператора  $B(t)$  (см., например, подробное рассмотрение в [8–10]), получаем

$$\begin{aligned} \gamma(B(t))_{1D} &= \gamma(DB(t)D^{-1})_1 = \\ &= \max \left( a_{ii}(t) - \lambda_{r-i+1}(t) + \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \mu_i(t)) \frac{d_k}{d_i} + \sum_{k=1}^{r-i} (\lambda_k(t) - \lambda_{i+r-1}(t)) \frac{d_{k+i}}{d_i} \right) = \\ &= -\min \alpha_i(t) = -\alpha(t). \end{aligned}$$

Тогда

$$\|z^*(t) - z^{**}(t)\|_{1D} \leq e^{-\int_s^t \alpha(u) du} \|z^*(s) - z^{**}(s)\|_{1D}$$

для всех  $0 \leq s \leq t$  и любых начальных условий  $z^*(s), z^{**}(s)$ .

Теперь, учитывая оценки для сравнения норм (см., например, [8]), получаем:

$$\begin{aligned} \|p^*(t) - p^{**}(t)\| &\leq 2\|z^*(t) - z^{**}(t)\| \leq \\ &\leq \frac{4}{d} \|z^*(t) - z^{**}(t)\|_{1D} \leq \\ &\leq \frac{4}{d} e^{-\int_s^t \alpha(u) du} \|z^*(s) - z^{**}(s)\|_{1D} \leq \\ &\leq \frac{4G}{d} e^{-\int_s^t \alpha(u) du} \|z^*(s) - z^{**}(s)\| \leq \\ &\leq \frac{4G}{d} e^{-\int_s^t \alpha(u) du} \|p^*(s) - p^{**}(s)\| \leq \frac{8G}{d} e^{-\int_s^t \alpha(u) du} \end{aligned}$$

для любых начальных условий  $p^*(s), p^{**}(s)$  и любых  $s, t, 0 \leq s \leq t$ .

Из слабой эргодичности процесса с конечным пространством состояний вытекает существование предельного среднего, начальные условия для которого можно в общем случае выбрать произвольно. Для оценки средних воспользуемся неравенством, приведенным в параграфе 2.3 из [8]:

$$\begin{aligned} \|z\|_{1D} &= d_0 \left| \sum_{i=1}^{\infty} p_i \right| + d_1 \left| \sum_{i=2}^{\infty} p_i \right| + \dots \geq \\ &\geq W \sum_{k \geq 1} k \left| \sum_{i \geq k} p_i \right| \geq \frac{W}{2} \sum_{k \geq 1} k |p_k|. \end{aligned}$$

Получаем теперь

$$\begin{aligned} |E(t, k) - \phi(t)| &\leq \frac{2}{W} \|p^*(t) - p^{**}(t)\|_{1D} \leq \\ &\leq \frac{2}{W} e^{-\int_0^t \alpha(u) du} \|e_k - p^{**}(0)\|_{1D} \leq \frac{4G}{W} e^{-\int_0^t \alpha(u) du}, \end{aligned}$$

что и требовалось доказать.

**Замечание 1.** Положим в условиях теоремы 1

$$\beta(t) = \max_{1 \leq i \leq r} \alpha_i(t).$$

Тогда, пользуясь внедиагональной неотрицательностью матрицы  $DB(t)D^{-1}$  с помощью методики, описанной в [8, 10], получаем справедливость неравенства

$$\|p^*(t) - p^{**}(t)\| \geq \frac{d}{8G} e^{-\int_s^t \beta(u) du}$$

при любых  $s, t, 0 \leq s \leq t$  и уже не при любых начальных условиях  $p^*(s), p^{**}(s)$ , а таких, что  $D(p^*(s) - p^{**}(s)) \geq 0$ . Следовательно, оценки теоремы 1 будут заведомо иметь точный по времени порядок, если удастся выбрать вспомогательную последовательность  $\{d_i\}$  так, что  $\alpha(t) = \beta(t)$ , т.е. все  $\alpha_i(t)$  одинаковы (не зависят от индекса  $i$ ).

Введем теперь в рассмотрение величины

$$\begin{aligned} \zeta_i(t) &= -a_{ii}(t) + \lambda_{r-i+1}(t) + \\ &+ \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \mu_i(t)) \frac{d_k}{d_i} + \\ &+ \sum_{k=1}^{r-i} (\lambda_k(t) - \lambda_{i+r-1}(t)) \frac{d_{k+i}}{d_i}; \\ \chi(t) &= \max_{1 \leq i \leq r} \zeta_i(t). \end{aligned}$$

**Замечание 2.** В условиях теоремы 1 при любых начальных условиях  $p^*(s), p^{**}(s)$  и любых  $s, t, 0 \leq s \leq t$ , справедлива следующая двухсторонняя оценка скорости сходимости:

$$\begin{aligned} \frac{d}{4G} e^{-\int_s^t \chi(u) du} \|p^*(s) - p^{**}(s)\| &\leq \|p^*(t) - p^{**}(t)\| \leq \\ &\leq \frac{4G}{d} e^{-\int_s^t \alpha(u) du} \|p^*(s) - p^{**}(s)\|. \end{aligned}$$

Таким образом, можно оценить и сверху и снизу время вхождения системы обслуживания в предельный режим. Более подробно о получении нижних оценок см., например, в [10, 11].

Рассмотрим два частных случая теоремы.

**Следствие 1.** Пусть при выполнении остальных условий теоремы 1 вместо (2) выполняется условие  $\alpha(t) \geq \alpha > 0$  почти при всех  $t \geq 0$ . Тогда вместо (3) и (4) справедливы оценки:

$$\begin{aligned} \|p^*(t) - p^{**}(t)\| &\leq \frac{8G}{d} e^{-\alpha(t-s)}; \\ |E(t, k) - \phi(t)| &\leq \frac{4G}{W} e^{-\alpha t}. \end{aligned}$$

Положим

$$M_0 = \max_{|t-s| \leq 1} \int_s^t \alpha(u) du;$$

$$\alpha^* = \int_0^1 \alpha(t) dt; \quad M = e^{M_0 + \alpha^*}.$$

С учетом неравенства

$$e^{-\int_s^t \alpha(u) du} \leq M e^{-\alpha^*(t-s)}$$

получаем следующее утверждение.

**Следствие 2.** Пусть все  $\lambda_k(t)$  и  $\mu_k(t)$  1-периодичны, а при выполнении остальных условий теоремы 1 вместо (2) выполняется условие  $\alpha^* > 0$ . Тогда предельный режим (скажем,  $\mathbf{p}^*(t)$ ) и соответствующее ему предельное среднее  $\phi^*(t)$  можно выбрать 1-периодическими, а вместо (3) и (4) справедливы оценки:

$$\|\mathbf{p}(t) - \mathbf{p}^*(t)\| \leq \frac{8GM}{d} e^{-\alpha^* t}$$

и, кроме того,

$$|E(t, k) - \phi^*(t)| \leq \frac{4GM}{W} e^{-\alpha^* t}$$

при любом  $k$  и  $t \geq 0$ .

### 3 Устойчивость

Рассмотрим также «возмущенный» процесс обслуживания  $\bar{X} = \bar{X}(t)$ ,  $t \geq 0$ , в котором интенсивности поступления и обслуживания требований также не зависят от числа требований в системе, обозначая его соответствующие характеристики теми же буквами с чертой сверху. Для простоты записи оценок будем предполагать, что возмущения «равномерно малы», т. е. выполняется неравенство  $\|A(t) - \bar{A}(t)\| \leq \varepsilon$ . Первые результаты для нестационарных цепей с непрерывным временем получены в [12], а детальное рассмотрение для более общего случая неравномерных оценок можно без труда провести так же, как это сделано в [13, 14]. Для получения требуемых равномерных оценок устойчивости необходима экспоненциальная эргодичность соответствующего процесса, т. е. существование положительных констант  $N$ ,  $a$  таких, что для правой части (3) справедливо неравенство:

$$e^{-\int_s^t \alpha(u) du} \leq N e^{-a(t-s)}. \quad (6)$$

Оценка (6) заведомо имеет место, в частности, если выполнены условия одного из следствий предыдущего параграфа.

**Теорема 2.** Пусть выполнены условия теоремы 1 и (6). Тогда при любых начальных условиях  $\mathbf{p}(s)$  и  $\bar{\mathbf{p}}(s)$  для процессов  $X(t)$  и  $\bar{X}(t)$  соответственно справедливы следующие оценки устойчивости:

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon(1 + \ln(4GN/d))}{a};$$

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}}(t)| \leq \frac{r\varepsilon(1 + \ln(4GN/d))}{a}.$$

Доказательство основано на подходе, введенном для стационарных процессов в [15] и описанном для нестационарной ситуации в [16]. Если при любых начальных условиях для исходного процесса справедлива оценка

$$\|\mathbf{p}(t) - \mathbf{p}^*(t)\| \leq ce^{-b(t-s)},$$

то, полагая

$$\beta(t, s) = \sup_{\|\mathbf{v}\|=1, \sum v_i=0} \|V(t, s)\mathbf{v}(t, s)\| =$$

$$= \frac{1}{2} \max_{i,j} \sum_k |p_{ik}(t, s) - p_{jk}(t, s)|,$$

где  $V(t, s)$  — матрица Коши уравнения (1), получаем в итоге следующее неравенство:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \begin{cases} \|\mathbf{p}(s) - \bar{\mathbf{p}}(s)\| + (t-s)\varepsilon, & 0 < t < b^{-1} \ln\left(\frac{c}{2}\right); \\ b^{-1} \left( \ln \frac{c}{2} + 1 - \frac{c}{2} e^{-b(t-s)} \right) \varepsilon + & \\ + \frac{c}{2} e^{-b(t-s)} \|\mathbf{p}(s) - \bar{\mathbf{p}}(s)\|, & \\ & t \geq b^{-1} \ln\left(\frac{c}{2}\right) \end{cases}$$

для любых начальных условий  $\mathbf{p}(s)$  и  $\bar{\mathbf{p}}(s)$ . Из неравенств (3) и (6) вытекает, что  $b = a$ ,  $c = 8GN/d$ . Устремив  $t \rightarrow \infty$  и взяв  $s = 0$ , получаем требуемые оценки.

**Замечание 3.** В полученную оценку устойчивости для математического ожидания процесса в качестве множителя входит размерность  $r$ , поэтому иногда лучший результат удается получить при помощи другого подхода, описанного в работе [16].

Положим

$$S = \max_{1 \leq i, j \leq r} \frac{d_i}{d_j},$$

и пусть числа  $K, L$  таковы, что

$$d_1 \lambda_1(t) + (d_1 + d_2) \lambda_2(t) + \dots + \left( \sum_{1 \leq i \leq r} d_i \right) \lambda_r(t) \leq K,$$

а

$$d_1(\lambda_1(t) - \bar{\lambda}_1(t)) + (d_1 + d_2)(\lambda_2(t) - \bar{\lambda}_2(t)) + \dots \\ \dots + \left( \sum_{1 \leq i \leq r} d_i \right) (\lambda_r(t) - \bar{\lambda}_r(t)) \leq L\varepsilon$$

почти при всех  $t \geq 0$ .

**Теорема 3.** Пусть выполнены условия теоремы 2 и, кроме того, при всех  $k$  и почти всех  $t \geq 0$   $\lambda_k(t) < \infty$ . Тогда при любых начальных условиях  $\mathbf{p}(s)$  и  $\bar{\mathbf{p}}(s)$  для процессов  $X(t)$  и  $\bar{X}(t)$  соответственно справедливо неравенство

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}}(t)| \leq \frac{N\varepsilon(La + 2KNS)}{Wa(a - 2\varepsilon S)}.$$

**Доказательство.** Перепишем исходную систему (5) для невозмущенного процесса в следующем виде:

$$\frac{d\mathbf{p}}{dt} = \bar{B}(t)\mathbf{p}(t) + \mathbf{f}(t) + (B(t) - \bar{B}(t))\mathbf{p}(t).$$

Тогда

$$\mathbf{p}(t) = \bar{U}(t, 0)\mathbf{p}(0) + \int_0^t \bar{U}(t, \tau)\mathbf{f}(\tau) d\tau + \\ + \int_0^t \bar{U}(t, \tau)(B(\tau) - \bar{B}(\tau))\mathbf{p}(\tau) d\tau; \\ \bar{\mathbf{p}}(t) = \bar{U}(t, 0)\bar{\mathbf{p}}(0) + \int_0^t \bar{U}(t, \tau)\mathbf{f}(\tau) d\tau,$$

где  $U(t, s)$  — матрица Коши для уравнения (5). В любой норме при одинаковых начальных условиях получаем следующую оценку:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \int_0^t \|\bar{U}(t, \tau)\| (\|B(\tau) - \bar{B}(\tau)\| \|\mathbf{p}(\tau)\| + \\ + \|\mathbf{f}(\tau) - \bar{\mathbf{f}}(\tau)\|) d\tau. \quad (7)$$

Имеем почти при всех  $t \geq 0$ :

$$\|B(t) - \bar{B}(t)\|_{1D} = \|D(B(t) - \bar{B}(t))D^{-1}\| \leq 2S\varepsilon;$$

$$\|\mathbf{f}(t)\|_{1D} \leq d_1 \lambda_1(t) + (d_1 + d_2) \lambda_2(t) + \dots + \\ + \left( \sum_{1 \leq i \leq r} d_i \right) \lambda_r(t) \leq K, \quad \|\mathbf{f}(\tau) - \bar{\mathbf{f}}(\tau)\|_{1D} \leq L\varepsilon.$$

А тогда

$$\gamma(\bar{B}(t))_{1D} \leq \gamma(DB(t)D^{-1}) + \|B(t) - \bar{B}(t)\|_{1D} \leq \\ \leq -\alpha(t) + 2S\varepsilon.$$

Оценим теперь

$$\|\mathbf{p}(t)\|_{1D} \leq \|U(t)\mathbf{p}(0)\|_{1D} + \int_0^t \|U(t, \tau)\mathbf{f}(\tau)\|_{1D} d\tau \leq \\ \leq Ne^{-at} \|\mathbf{p}(0)\|_{1D} + \frac{KN}{a}.$$

Тогда с учетом (7) получаем:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\|_{1D} \leq N \int_0^t e^{-(a-2\varepsilon S)(t-\tau)} \times \\ \times \left( 2S\varepsilon(Ne^{-a\tau} \|\mathbf{p}(0)\|_{1D} + \frac{KN}{a}) + L\varepsilon \right) d\tau \leq \\ \leq o(1) + \frac{N\varepsilon(L + 2KNS/a)}{a - 2\varepsilon S}.$$

## 4 Примеры

### Пример 1.

Рассмотрим исходный процесс обслуживания с интенсивностями  $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = \lambda(t) = 3 + \sin 2\pi t$ ,  $\mu_1(t) = \mu_2(t) = \mu(t) = 2 + \cos 2\pi t$ ,  $\lambda_4(t) = \dots = \lambda_r(t) = \mu_3(t) = \dots = \mu_r(t) = 0$ . Выберем последовательность  $d_k = h^k$ , где  $0,82 < h < 1$ . Тогда имеем

$$d = h^r; \quad G \leq \frac{h}{1-h}; \quad W = \frac{h^r}{r}.$$

Будем предполагать, что возмущенный процесс имеет такую же структуру матрицы интенсивностей, причем  $|\lambda(t) - \bar{\lambda}(t)| \leq \varepsilon$  и  $|\mu(t) - \bar{\mu}(t)| \leq \varepsilon$  почти при всех  $t \geq 0$ . Отметим кстати, что при этом  $\|A(t) - \bar{A}(t)\| \leq 10\varepsilon$  почти при всех  $t \geq 0$ . Рассмотрим дальнейшие оценки:

$$S = \frac{1}{h^2}; \quad K = 4(3h + 2h^2 + h^3); \quad L = 3h + 2h^2 + h^3;$$

$$\alpha(t) \geq \lambda(t)(3 - h - h^2 - h^3) - \mu(t) \left( \frac{1}{h^2} + \frac{1}{h} - 2 \right);$$

$$\alpha^* = 3(3 - h - h^2 - h^3) - 2 \left( \frac{1}{h^2} + \frac{1}{h} - 2 \right);$$

$$M_0 \leq \int_0^1 |\alpha(t)| dt \leq 4(3 - h - h^2 - h^3) + 3 \left( \frac{1}{h^2} + \frac{1}{h} - 2 \right);$$

$$M = e^{\alpha^* + M_0}.$$

Если, например, взять  $h = 0,9$ , то  $\alpha^* = 0,992$ ,  $M_0 = 3,281$ ,  $M = 71,737$ .

Тогда получаем следующие оценки.

По следствию 2

$$\|\mathbf{p}(t) - \mathbf{p}^*(t)\| \leq \frac{8Me^{-\alpha^*t}}{h^{r-1}(1-h)};$$

$$|E_{\mathbf{p}}(t) - \phi^*(t)| \leq \frac{4Mre^{-\alpha^*t}}{h^{r-1}(1-h)}.$$

По теореме 2 ( $N = M$ ,  $a = \alpha^*$ ) с использованием оценок следствия 2

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon(1 + \ln(4M/(h^{r-1}(1-h))))}{\alpha^*};$$

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{r\varepsilon(1 + \ln(4M/(h^{r-1}(1-h))))}{\alpha^*}.$$

По теореме 3 с использованием оценок следствия 2

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{rM\varepsilon(3h + 2h^2 + h^3)(\alpha^*h^2 + 8M)}{h^r\alpha^*(\alpha^*h^2 - 2\varepsilon)}.$$

### Пример 2.

Рассмотрим процесс с интенсивностями  $\lambda_1(t) = \lambda_2(t) = \dots = \lambda_r(t) = \lambda(t) = 3 + \sin 2\pi t$ ;  $\mu_1(t) = \mu_2(t) = \mu(t) = 2 + \cos 2\pi t$ ;  $\mu_3(t) = \dots = \mu_r(t) = 0$ .

Будем предполагать, что возмущенный процесс имеет такую же структуру матрицы интенсивностей, причем  $|\lambda(t) - \bar{\lambda}(t)| \leq \varepsilon$  и  $|\mu(t) - \bar{\mu}(t)| \leq \varepsilon$  почти при всех  $t \geq 0$ . При этом будем иметь  $\|A(t) - \bar{A}(t)\| \leq 2r\varepsilon$  почти при всех  $t \geq 0$ .

Выберем последовательность  $d_k = 1$ . Тогда

$$d = 1; \quad G = r; \quad W = \frac{1}{r}; \quad S = 1;$$

$$K = \frac{4r(1+r)}{2}; \quad L = \frac{r(1+r)}{2};$$

$$\alpha(t) = \lambda(t); \quad \alpha = 2; \quad \alpha^* = 3; \quad M_0 \leq 4; \quad M \leq e^7.$$

И получаем следующие оценки.

По следствию 1

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 8re^{-2t};$$

$$|E_{\mathbf{p}}(t) - \phi(t)| \leq 4r^2e^{-2t}.$$

По следствию 2

$$\|\mathbf{p}(t) - \mathbf{p}^*(t)\| \leq 8re^{7-3t};$$

$$|E_{\mathbf{p}}(t) - \phi^*(t)| \leq 4r^2e^{7-3t}.$$

По теореме 2 ( $N = 1$ ,  $a = \alpha$ ) с учетом оценок следствия 1

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon(1 + \ln 4r)}{2};$$

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{r\varepsilon(1 + \ln 4r)}{2}.$$

По теореме 2 ( $N = M$ ,  $a = \alpha^*$ ) с учетом оценок следствия 2

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{\varepsilon(8 + \ln 4r)}{3};$$

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{r\varepsilon(8 + \ln 4r)}{3}.$$

По теореме 3 с учетом оценок следствия 1

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{5\varepsilon r^2(1+r)}{4(1-\varepsilon)}.$$

По теореме 3 с учетом оценок следствия 2

$$\limsup_{t \rightarrow \infty} |E_{\mathbf{p}}(t) - \bar{E}_{\bar{\mathbf{p}}(t)}| \leq \frac{\varepsilon e^7 r^2(1+r)(3+8e^7)}{6(3-2\varepsilon)}.$$

## Литература

1. Баруча-Руд А. Т. Элементы теории марковских процессов и их приложения. — М.: Наука, 1969.
2. Гнеденко Б. В., Макаров И. П. Свойства решений задачи с потерями в случае периодических интенсивностей // Дифф. уравнения, 1971. Вып. 7. С. 1696–1698.
3. Gnedenko D. B. On a generalization of Erlang formulae // Zastosow. Mat., 1971. Vol. 12. P. 239–242.
4. Саати Т. Л. Элементы теории массового обслуживания и ее приложения. — М.: Сов. радио, 1971.
5. Gnedenko B., Soloviev A. On the conditions of the existence of final probabilities for a Markov process // Math. Operations. Stat., 1973. P. 379–390.
6. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. — М.: Наука, 1987.

7. Granovsky B. L., Zeifman A. I. The N-limit of spectral gap of a class of birth-death Markov chains // Appl. Stoch. Models Business Ind., 2000. Vol. 16. P. 235–248.
8. Зейфман А. И., Бенинг В. Е., Соколов И. А. Марковские цепи и модели с непрерывным временем. — М.: Элекс-КМ, 2008.
9. Van Doorn E. A., Zeifman A. I., Panfilova T. L. Bounds and asymptotics for the rate of convergence of birth-death processes // Th. Prob. Appl., 2010. Vol. 54. P. 97–113.
10. Zeifman A. I. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // Stoch. Proc. Appl., 1995. Vol. 59. P. 157–173.
11. Granovsky B. L., Zeifman A. I. On the lower bound of the spectrum of some mean-field models // Theory Prob. Appl., 2005. Vol. 49. P. 148–155.
12. Zeifman A. I. Stability for continuous-time nonhomogeneous Markov chains // Lect. Notes Math., 1985. Vol. 1155. P. 401–414.
13. Zeifman A. Stability of birth and death processes // J. Math. Sci., 1998. Vol. 91. P. 3023–3031.
14. Андреев Д., Елесин М., Кузнецов А., Крылов Е., Зейфман А. Эргодичность и устойчивость нестационарных систем обслуживания // Теория вероятностей и математическая статистика, 2003. Т. 68. С. 1–11.
15. Mitrophanov A. Yu. Stability and exponential convergence of continuous-time Markov chains // J. Appl. Prob., 2003. Vol. 40. P. 970–979.
16. Зейфман А. И., Коротышева А. В., Панфилова Т. Л., Шоргин С. Я. Оценки устойчивости для некоторых систем обслуживания с катастрофами // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 27–33.

# МАКСИМУМЫ АКТИВНОСТИ В БЕЗМАСШТАБНЫХ СЛУЧАЙНЫХ СЕТЯХ С ТЯЖЕЛЫМИ ХВОСТАМИ\*

А. В. Лебедев<sup>1</sup>

**Аннотация:** Рассматриваются ориентированные степенные случайные графы как модели информационных сетей, где каждый узел обладает случайной информационной активностью, распределение которой имеет тяжелый (правильно меняющийся) хвост. Используется модель случайного графа, в которой входящие степени вершин независимы и имеют распределение со степенным хвостом. Выведены достаточные условия, при которых максимум суммарных активностей (по узлу и его входящим соседям) растет асимптотически так же, как и максимум индивидуальных активностей, и в силу этого для них имеет место предельный закон Фреше.

**Ключевые слова:** максимумы; случайные суммы; безмасштабные сети; степенной закон; случайный граф; тяжелый хвост; правильное изменение; распределение Фреше

## 1 Введение

Степенными (*power-law*) или безмасштабными (*free-scale*) называют случайные графы, у которых степени вершин подчиняются асимптотически степенному закону (с вероятностями  $p_k \sim ck^{-\beta}$ ,  $k \rightarrow \infty$ ,  $\beta > 1$ ). Активные исследования данного класса графов и их приложений в последнее десятилетие были инициированы работой [1], где приведен ряд интересных примеров (Интернет, электрическая сеть, социальная сеть киноактеров). С тех пор были предложены и изучены различные модели степенных графов. В одних степенной закон возникает благодаря некоторому случайному процессу [1, 2], в других он постулируется изначально [3, 4]. Следует отметить, что некоторые асимптотические свойства графов при одинаковом распределении степеней вершин могут оказаться общими, а другие зависят от выбора модели.

Степенной граф может служить моделью некоторой информационной сети. Например, исследования кириллического сегмента «Живого журнала» ([livejournal.com](http://livejournal.com)) [2] показывают, что он хорошо описывается степенным графом с  $\beta \approx 3$ . Пусть каждый узел этой сети обладает случайной информационной активностью (интенсивностью производства информации). Имеется в виду среднее количество информации, производимой узлом в единицу времени. Активность будем полагать индивидуальной характеристикой, присущей узлу. Например, речь может идти о пользователях, которые пишут сообщения в Интернет. Предположим, что активности узлов независимы и одинаково распределены, причем их распределение  $F$  имеет тяжелый (правильно

меняющийся) хвост, т. е.  $\bar{F}(x) \sim x^{-a}L(x)$ ,  $x \rightarrow \infty$ ,  $a > 0$ , где  $L(x)$  — медленно меняющаяся функция [5, § 8.8]. Такое предположение находится в русле современных представлений о распространенности степенных законов в природе, технике и человеческой деятельности. Активности и степени вершин (узлов) для простоты будем полагать независимыми.

Рассмотрим суммарную активность в узле (т. е. сумму его собственной и ближайших соседей). Например, в «Живом журнале» каждый пользователь может оставлять свои записи и читать записи своих друзей, объединяемые для удобства в общую «ленту друзей» (френдленту). Далее будем интересоваться вопросом: когда максимум суммарных активностей растет асимптотически так же, как и максимум индивидуальных активностей узлов? В этом случае для максимумов легко выводится предельный закон Фреше  $\Phi_a(x) = \exp\{-x^{-a}\}$ ,  $x > 0$  [5, § 8.8; 6, § 3.3.1].

Для модели степенного графа, введенной в [3], этот вопрос был изучен автором в [7]. Там число вершин степени  $k$  полагалось детерминированным и равным  $\lfloor e^\alpha/k^\beta \rfloor$ ,  $\alpha, \beta > 0$ ,  $1 \leq k \leq e^{\alpha/\beta}$ , а распределение на множестве графов, удовлетворяющих этому условию, равномерным. Были получены достаточные условия того, что максимум сумм с ростом числа узлов (при  $\alpha \rightarrow \infty$ ) растет асимптотически так же, как и максимум индивидуальных активностей:  $a < \beta - 3/2$ , если  $3/2 < \beta < 3$ , и  $a < \beta/2$ , если  $\beta \geq 3$ . При этом применялись ранее полученные автором результаты для общей схемы максимумов сумм независимых случайных величин [8].

\* Работа выполнена при поддержке РФФИ, грант 11-01-00050.

<sup>1</sup> Механико-математический факультет Московского государственного университета им. М. В. Ломоносова, [avlebed@yandex.ru](mailto:avlebed@yandex.ru)

Рассмотрим теперь модель ориентированного случайного графа, где направления ребер соответствуют направлениям передачи информации. Пусть имеется  $n$  вершин и заданы независимые неотрицательные целочисленные случайные величины  $K_1, \dots, K_n$ , имеющие одинаковое распределение, заданное вероятностями  $p_k \sim ck^{-\beta}$ ,  $k \rightarrow \infty$ ,  $\beta > 1$ . Положим  $D_i = \min\{K_i, n - 1\}$ . Для  $i$ -й вершины выберем случайным образом (равновероятно и независимо от выбора для других вершин)  $D_i$  различных вершин из числа остальных (кроме  $i$ -й) и выпустим из них ребра, направленные в  $i$ -ю вершину. Полученный в результате граф можно отнести к степенным в том смысле, что входящие степени вершин распределены асимптотически по степенному закону. Суммарной активностью в узле в данном случае будем считать сумму собственной активности узла и всех узлов, из которых в него поступает информация (его входящих соседей).

К сожалению, метод, использованный в [7], здесь не работает при  $\beta < 3$ , так как второй момент входящей степени вершины тогда растет слишком быстро при  $n \rightarrow \infty$ . Эта проблема решается с помощью урезания. При этом получают более сильные ограничения на параметры, что связано с более быстрым ростом максимальной (входящей) степени вершины в графе. Однако поскольку используются лишь достаточные условия, не исключено, что эти ограничения в будущем могут быть ослаблены.

Отметим, что асимптотическая эквивалентность хвостов распределений суммы и максимума конечного числа независимых одинаково распределенных случайных величин в случае тяжелых хвостов представляет собой давно известный факт [5, § 8.8], обусловленный тем, что основной вклад в сумму дает самое большое слагаемое (максимум), а сумма остальных слагаемых по сравнению с ним оказывается мала. Теперь обобщим это утверждение на модель, где имеется некоторый набор случайных сумм со случайными числами слагаемых и от сумм берется максимум. По-прежнему оказывается, что основной вклад (в одну или несколько сумм, а значит, и в их максимум) дает только одно, максимальное слагаемое. Однако для этого хвост распределения слагаемых должен быть достаточно тяжелым.

Проверка наличия подобного эффекта в реальных сетях, разумеется, требует экспериментального исследования, выходящего за рамки данной работы, которая имеет теоретический характер.

## 2 Основной результат

Будем рассматривать сети из  $n$  узлов, затем устремляя  $n$  к бесконечности. Обозначим через

$M(n)$  максимум суммарных активностей (самого узла и его входящих соседей), а через  $M_0(n)$  — максимум индивидуальных активностей узлов. Требуется определить условия, при которых

$$\frac{M(n)}{M_0(n)} \xrightarrow{P} 1, \quad n \rightarrow \infty. \quad (1)$$

Введем неотрицательную функцию  $u(s)$  такую, что  $s\bar{F}(u(s)) \rightarrow 1$ ,  $s \rightarrow \infty$ . Заметим, что  $u(s)$  заведомо существует и правильно меняется с показателем  $1/a$ , т. е.  $u(s) \sim s^{1/a}L_2(s)$ ,  $s \rightarrow \infty$ , где  $L_2(s)$  — медленно меняющаяся функция [9, § 1.5].

Тогда имеет место предельный закон для максимумов независимых случайных величин в случае правильно меняющихся хвостов [5, § 8.8; 6, § 3.3.1]:

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \frac{M_0(n)}{u(n)} \leq x \right) = \Phi_a(x), \quad x > 0,$$

что в сочетании с (1) дает

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \frac{M(n)}{u(n)} \leq x \right) = \Phi_a(x), \quad x > 0.$$

**Теорема 1.** Соотношение (1) выполняется при  $a < \beta - 2$ , если  $2 < \beta < 3$ , и при  $a < (\beta - 1)/2$ , если  $\beta \geq 3$ .

Доказательство теоремы будет приведено в разд. 4.

## 3 Общая схема максимумов сумм

Напомним введенную в [8] схему (немного изменив обозначения). Пусть заданы случайный процесс  $\Upsilon(t)$ ,  $t \in T$ , значениями которого являются конечные классы конечных подмножеств  $\mathbf{N}$ , и семейство  $\Xi = \{\xi_{i,t}, i \in \mathbf{N}, t \in T\}$  неотрицательных случайных величин, независимых и одинаково распределенных при любом фиксированном значении параметра  $t \in T$ . Полагаем, что  $\Upsilon$  и  $\Xi$  независимы.

Для любых  $A \subset \mathbf{N}$ ,  $t \in T$  обозначим максимум набора случайных величин  $\{\xi_{i,t}, i \in A\}$  через  $M_t(A)$ ,  $r$ -й максимум (т. е. число, стоящее  $r$ -м с конца в вариационном ряду) — через  $M_t^{(r)}$ , сумму — через  $S_t(A)$ . Пусть

$$U(t) = \bigcup_{A \in \Upsilon(t)} A.$$

Введем случайные процессы, порожденные  $\Upsilon$  и  $\Xi$ :

$$Z(t) = \sup_{A \in \Upsilon(t)} S_t(A); \quad \kappa(t) = \sup_{A \in \Upsilon(t)} |A|; \quad \nu(t) = |U(t)|;$$

$$\mu_1(t) = M_t(U(t)); \quad \mu_r(t) = M_t^{(r)}(U(t)),$$

где через  $|A|$  обозначен размер (число элементов) множества  $A$ . Через  $|\Upsilon(t)|$  обозначим число различных множеств  $A \in \Upsilon(t)$ .

Предполагается, что  $\nu(t) < \infty$  почти наверное (п. н.) при всех  $t \in T$ , откуда следует конечность п. н. всех процессов, введенных выше.

Рассмотрим предельное поведение  $Z(t)$  при  $t \rightarrow \infty$ .

Пусть существует случайный процесс  $\rho(t)$  со значениями в  $\mathbf{Z}_+$ , измеримый относительно  $\Upsilon$  и такой, что  $\rho(t) \geq 1$  при  $\nu(t) \geq 1$ ,  $\rho(t) \leq \nu(t)$  п. н. при всех  $t \in T$ .

Обозначим через  $\pi(t)$  вероятность того, что для множества  $B$ , равновероятно выбранного среди всех подмножеств  $U(t)$ , состоящих из  $\rho(t)$  элементов, имеет место  $\sup_{A \in \Upsilon(t)} |A \cap B| > 1$ .

**Теорема I.** Пусть выполнены условия

$$(\kappa(t) - 1) \frac{\mu_{\rho(t)}(t)}{\mu_1(t)} \xrightarrow{P} 0, \quad t \rightarrow \infty; \quad (2)$$

$$\pi(t) \rightarrow 0, \quad t \rightarrow \infty, \quad (3)$$

тогда

$$\frac{Z(t)}{\mu_1(t)} \xrightarrow{P} 1, \quad t \rightarrow \infty. \quad (4)$$

Доказано также следующее свойство порядковых статистик в случае правильно меняющихся хвостов. Пусть  $X_n$ ,  $n \geq 1$ , независимы и имеют распределение  $F$  с правильно меняющимся хвостом  $\bar{F}(x) \sim x^{-a}L(x)$ ,  $x \rightarrow \infty$ ,  $a > 0$ . Обозначим максимум  $X_1, \dots, X_n$  через  $\tilde{X}_n$  и  $r$ -й максимум через  $\tilde{X}_n^{(r)}$ .

**Следствие II.** Пусть  $r_n \sim n^\gamma$ ,  $n \rightarrow \infty$ ,  $0 < \gamma < 1$  и  $0 < \delta < \gamma/a$ , тогда

$$\frac{n^\delta \tilde{X}_n^{(r_n)}}{\tilde{X}_n} \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

## 4 Приложение к случайным графам

Адаптируем общую схему к изучению случайных графов. Рассмотрим процесс  $\Upsilon$  с дискретным временем, соответствующим числу узлов  $n$ . Случайные величины  $\xi_{i,n}$ ,  $1 \leq i \leq n$ , описывают информационные активности узлов. Обозначим через  $A_i$  множество из индекса  $i$  и индексов входящих соседей  $i$ -го узла, тогда набор множеств  $A$  получается из набора  $A_1, \dots, A_n$  удалением повторов (если они есть). Имеем  $|A_i| = D_i + 1$  и  $\kappa(n) = \max_{1 \leq i \leq n} D_i + 1$ . Очевидно,  $|\Upsilon(n)| \leq n$  и  $\nu(n) = n$ . Последовательность  $\rho(n)$

далее будем полагать детерминированной. Кроме того, в используемых обозначениях  $M(n) = Z(n)$ ,  $M_0(n) = \mu_1(n)$  и (4) эквивалентно (1).

Обозначим

$$Q(n, m) = n \mathbf{M}((D + 1) \mathbf{D} \mathbf{I}\{D \leq m - 1\}),$$

где  $D \stackrel{d}{=} D_1$ .

**Лемма 1.** При  $n > 2$  и  $\rho(n) < n$  верно неравенство

$$\pi(n) \leq \frac{\rho(n)(\rho(n) - 1)}{2(n - 1)(n - 2)} Q(n, m) + \mathbf{P}(\kappa(n) > m).$$

**Доказательство.** Событие  $\{\sup_{A \in \Upsilon(t)} |A \cap B| > 1\}$  представляет собой объединение событий  $\{|A_i \cap B| > 1\}$ ,  $1 \leq i \leq n$ . Пусть для простоты  $B = \{1, 2, \dots, \rho(n)\}$  (в противном случае можно перенумеровать  $A_i$ ). Зафиксируем входящие степени вершин  $d_1, \dots, d_n$ . Тогда для  $1 \leq i \leq \rho(n)$  один элемент множества  $B$  заведомо принадлежит  $A_i$  (а именно, индекс  $i$ ), а любой другой принадлежит с вероятностью  $d_i/(n - 1)$ . Для  $\rho(n) + 1 \leq i \leq n$  любая пара индексов из  $B$  принадлежит  $A_i$  с вероятностью  $d_i(d_i - 1)/((n - 1)(n - 2))$ , а всего таких пар  $\rho(n)(\rho(n) - 1)/2$ . Суммируя вероятности, получаем оценку сверху:

$$\frac{\rho(n) - 1}{n - 1} \sum_{i=1}^{\rho(n)} d_i + \frac{\rho(n)(\rho(n) - 1)}{2(n - 1)(n - 2)} \sum_{i=\rho(n)+1}^n d_i(d_i - 1).$$

Обозначим

$$q_1 = \mathbf{M}(\mathbf{D} \mathbf{I}\{D \leq m - 1\});$$

$$q_2 = \mathbf{M}(D(D - 1) \mathbf{I}\{D \leq m - 1\}).$$

Усредняя по наборам входящих степеней вершин в области  $\kappa(n) \leq m$ , получаем оценку сверху:

$$\begin{aligned} \frac{\rho(n) - 1}{n - 1} \rho(n) q_1 + \frac{\rho(n)(\rho(n) - 1)}{2(n - 1)(n - 2)} (n - \rho(n)) q_2 &\leq \\ &\leq \frac{\rho(n)(\rho(n) - 1)}{2(n - 1)(n - 2)} n (2q_1 + q_2) = \\ &= \frac{\rho(n)(\rho(n) - 1)}{2(n - 1)(n - 2)} Q(n, m). \end{aligned}$$

Учитывая также вероятность события  $\{\kappa(n) > m\}$ , получаем утверждение леммы.

**Лемма 2.** Пусть выполнены следующие условия:

(1) все  $\xi_{i,n}$  имеют одинаковое распределение  $F$  на  $\mathbf{R}_+$  с хвостом  $\bar{F}(x) \sim x^{-a}L(x)$ ,  $x \rightarrow \infty$ ,  $a > 0$ ;

(2)  $m \sim n^\delta$ ,  $n \rightarrow \infty$ ,  $\delta > 0$ ;

- (3)  $Q(n, m) = O(n^b), n \rightarrow \infty, 0 < b < 2;$
- (4)  $\kappa(n) = o_p(m), n \rightarrow \infty;$
- (5)  $a < (2 - b)/(2\delta).$

Тогда верно (1).

**Доказательство.** Можно выбрать  $\gamma \in (0, 1)$  так, чтобы выполнялось неравенство  $a\delta < \gamma < (2 - b)/2$ . Положим  $\rho(n) = [n^\gamma]$ , тогда по следствию II получаем (2), а по лемме 1 — (3), так что условия теоремы I выполняются и верно соотношение (4), эквивалентное (1).

**Доказательство теоремы 1.** Поскольку  $p_k \sim ck^{-\beta}, k \rightarrow \infty$ , то хвост распределения имеет асимптотику  $\bar{F}_K(k) \sim c_1 k^{-(\beta-1)}$ . Отсюда получаем  $\kappa(n) = O_p(n^{1/(\beta-1)}), n \rightarrow \infty$ , и, следовательно,  $\kappa(n) \sim o_p(m)$  при любом  $\delta = (1 + \varepsilon)/(\beta - 1), \varepsilon > 0$ . Имеем

$$Q(n, m) = n \sum_{k=1}^{m-1} (k+1)kp_k = \begin{cases} O(n^{1+\delta(3-\beta)}), & 1 < \beta < 3; \\ O(n \ln n), & \beta = 3; \\ O(n), & \beta > 3. \end{cases}$$

При  $2 < \beta < 3$  применяем лемму 2 с  $b = 2\delta > 1 + \delta(3 - \beta)$  и, устремляя  $\varepsilon$  к нулю, получаем достаточное условие  $a < \beta - 2$  для выполнения (1).

При  $\beta \geq 3$  применяем лемму 2 с  $b = 1 + \varepsilon$  и аналогично получаем достаточное условие  $a < (\beta - 1)/2$ .

## Литература

1. *Barabási A., Albert R.* Emergence of scaling in random networks // Science, 1999. Vol. 286. P. 509–512.
2. *Захаров П.* Народ-блогосец // Компьютерра, 2007. № 27–28. С. 36–39. <http://offline.computerra.ru/2007/695/327726>.
3. *Aiello W., Chung F., Lu L.* A random graph model for power law graphs // Experimental Math., 2001. Vol. 10. No. 1. P. 53–66.
4. *Reittu H., Norros I.* On the power-law random graph model of massive data network // Performance Evaluation, 2004. Vol. 55. P. 3–23.
5. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.
6. *Embrechts P., Klüppelberg C., Mikosh T.* Modelling extremal events for insurance and finance. — Springer-Verlag, 2003.
7. *Лебедев А. В.* Максимумы активности в случайных сетях в случае тяжелых хвостов // Проблемы передачи информации, 2008. Т. 44. № 2. С. 96–100.
8. *Лебедев А. В.* Общая схема максимумов сумм независимых случайных величин и ее приложения // Математические заметки, 2005. Т. 77. № 4. С. 544–550.
9. *Сенета Е.* Правильно меняющиеся функции. — М.: Наука, 1985.

# АНАЛИТИЧЕСКАЯ МОДЕЛЬ РАСЧЕТА ЭФФЕКТИВНОСТИ ПЛАНА РАСПРЕДЕЛЕНИЯ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ МНОГОПРОЦЕССОРНОЙ СИСТЕМЫ ПРИ РЕШЕНИИ СПЕЦИАЛЬНОГО КЛАССА ЗАДАЧ

М. Я. Агаларов<sup>1</sup>

**Аннотация:** Рассмотрена модель многопроцессорной системы, предназначенная для решения задач, распараллеливаемых на слабо зависимые вычисления. В качестве модели использована мультисервисная система массового обслуживания (СМО) с явными потерями, пуассоновскими входящими потоками и общими функциями распределения времени обслуживания заявок. Получены рекуррентные формулы для вычисления стационарного распределения вероятностей состояний и явные выражения для вероятностей отказов системы различным типам заявок. В рамках данной модели предложен метод оценки пропускной способности многопроцессорной системы при заданном статическом плане распределения вычислительных ресурсов.

**Ключевые слова:** многопроцессорная система; система массового обслуживания; мультисервисная система; распределение вычислительных ресурсов

## 1 Введение

Одной из основных проблем, возникающих при проектировании и эксплуатации многопроцессорных систем, является выработка рекомендаций по рациональному использованию ресурсов вычислительной среды. При оценке эффективности проектных решений по распределению ресурсов и потоков задач в многопроцессорных системах используются детерминированные и вероятностные модели. Использование той или иной модели зависит от архитектуры вычислительной системы, которая, в свою очередь, зависит от типа решаемых задач. Детерминированные модели основаны на теории расписаний, в рамках которой разработано немало методов определения очередности выполнения задач и расчета планов распределения вычислительных ресурсов [1–3]. Недостатком этих моделей является отсутствие учета случайных факторов, что вызвало широкое распространение вероятностных моделей, базирующихся на теории массового обслуживания (см., например, [4–10]).

Ниже рассматривается многопроцессорная система, используемая для решения задач, распараллеливаемых на независимые (слабо зависимые) вычисления (аналогично таким задачам, как восстановление криптографированного текста с помощью всех возможных ключей шифрования, поиск в больших объемах данных по ключевым признакам, поиск вариантов в пошаговых игровых

программах, множественное вычисление целевой функции в процедурах многомерной нелинейной оптимизации и т. д.) [3, 11, 12].

Для исследования рассматриваемой многопроцессорной системы предлагается модель, представленная в виде мультисервисной СМО с явными потерями и общими функциями распределения времени обслуживания заявок. Для предлагаемой модели получены формулы и алгоритмы расчета показателей производительности многопроцессорной системы.

## 2 Описание задачи

Рассматривается специализированный вычислительный комплекс (ВК), функциональная структура которого условно представлена четырьмя основными компонентами (рис. 1): поток задач, центр управления (ЦУ), линии передачи данных и вычислительные ресурсы. Вычислительные ресурсы — множество единиц ресурсов с заданными параметрами производительности и надежности. На ВК поступает поток задач, требующих выполнения на вычислительных ресурсах за указанный заранее для каждой задачи интервал времени (контрольное время). Задача представляет собой совокупность элементарных заданий, каждое из которых может независимо выполняться на любом ресурсе.

<sup>1</sup>Институт проблем информатики Российской академии наук, murad-agalarov@yandex.ru

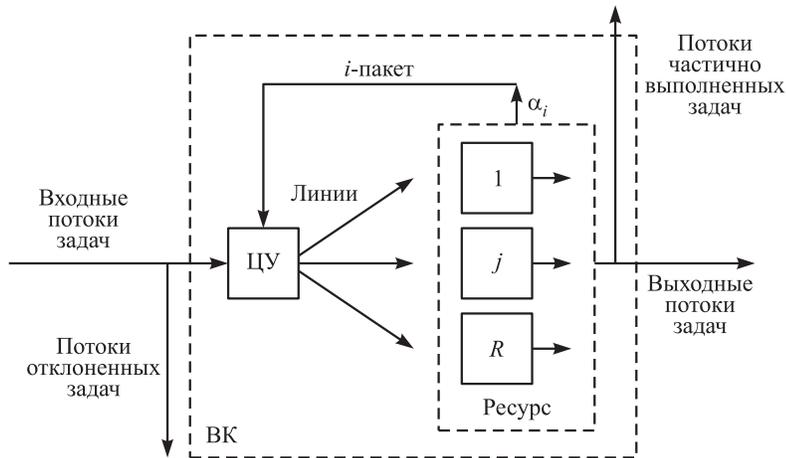


Рис. 1 Структурная схема ВК

Каждая задача может содержать одно элементарное задание, имеющее некоторый отличительный признак (называемый в дальнейшем ключом). Целью выполнения задачи является поиск ключа. Центр управления принимает решение о допуске и выполняет подготовку поступившей задачи к вычислениям, управляет коммуникационными ресурсами и процессом выполнения задачи на вычислительных ресурсах.

Обработка и выполнение задачи в ВК происходит следующим образом. Центр управления для каждой задачи с учетом контрольного для нее времени вычисляет по заданному алгоритму необходимое число ресурсов, которые до завершения выполнения задачи доступны только ей. Если в ВК отсутствует требуемое число свободных ресурсов, задача не принимается и теряется (или повторяется позже). В ЦУ принятая задача разбивается на блоки из элементарных заданий, называемые в дальнейшем пакетами. Пакеты направляются на ресурсы группами, называемыми посылками, в каждой из которых (за исключением, возможно, последней) количество пакетов совпадает с количеством выделенных для задачи ресурсов.

Посылка занимает и освобождает все выделенные ресурсы одновременно, при этом один пакет занимает один ресурс. Выполнение пакета считается успешным, если во время его выполнения не произошел сбой ресурса, иначе выполнение пакета считается неуспешным, и в составе какой-нибудь новой посылки задача заново направляется из ЦУ на выполнение. Выполнение посылки считается завершенным, если выполнены (успешно или неуспешно) все составляющие ее пакеты. Очередную посылку ЦУ направляет на ресурсы только после завершения выполнения предыдущей посылки. Считается, что ключ с равной вероятностью находится

в одном из пакетов задачи. Выполнение задачи завершается либо в момент достижения контрольного времени, либо в момент завершения выполнения очередной посылки, если обнаружен пакет с ключом, либо после успешного выполнения всех пакетов задачи. Задача покидает систему в момент завершения выполнения, освобождая одновременно все занятые ею ресурсы.

Требуется вычислить пропускную способность (интенсивность выходного потока задач) ВК и загрузженность ресурсов при заданном плане распределения ресурсов.

### 3 Модель вычислительного комплекса

В качестве модели ВК рассматривается мультисервисная СМО с явными потерями и произвольным временем обслуживания. В дальнейшем, чтобы описать работу модели в терминах теории массового обслуживания, вместо термина «задача» будем использовать термин «заявка». На вход системы, состоящей из  $R$  одинаковых общедоступных независимых каналов, поступает пуассоновский поток заявок интенсивности  $\lambda$ , имеющих случайную длину (случайное число элементарных заданий), распределенную равномерно на множестве  $L$ . Без потери общности будем считать, что  $L$  — конечное множество положительных чисел, т. е.  $|L| < \infty$ . Пусть  $L_j$  — полуинтервалы такие, что  $L_j \cap L_i = \emptyset$  при  $j \neq i, j, i = 1, \dots, M, \bigcap_{j=1}^M L_j = L, M \leq |L|$ . Поступающие заявки образуют  $M$  потоков по следующему правилу: заявка принадлежит  $j$ -му потоку (является  $j$ -заявкой, заявкой  $j$ -го типа), если ее

длина попадает в полуинтервал  $L_j$ ,  $j = 1, \dots, M$ .

Процесс обслуживания заявки состоит из двух этапов и происходит следующим образом. На первом этапе при наличии в системе необходимого числа свободных каналов поступившая  $j$ -заявка разбивается в системе на пакеты длиной  $lp_j$ . Время разбиения  $j$ -заявки на пакеты считается независимой случайной величиной с произвольной функцией распределения  $GZ_j(t)$  со средним значением  $\bar{t}z_j < t_{\text{contrz},j}$ . Считается, что обслуживание заявки на первом этапе происходит как в СМО типа  $M/G/\infty$ .

Необходимое заявкам число каналов определяется по следующему правилу. Для каждого  $j$ -потока вычисляется величина

$$m_j = \max \left\{ \text{целая часть} \left[ \frac{lz_j}{v(t_{\text{contrz},j} - \bar{t}z_j)} \right] + d_j, 1 \right\},$$

где  $m_j \leq R$ ;  $lz_j$  — максимальная длина  $j$ -заявки;  $v$  — пропускная способность канала (максимальная скорость передачи каналом элементарных заданий);  $t_{\text{contrz},j} > 0$  — контрольное время обслуживания  $j$ -заявки;  $d_j \geq 0$  — заданное целое число,  $j = 1, \dots, M$ . Поступившая  $j$ -заявка, если в системе есть не меньше  $m_j$  свободных каналов, занимает  $m_j$  каналов одновременно, иначе она теряется. Набор  $\bar{m} = \{m_j, j = 1, \dots, M\}$  будем называть планом распределения ресурсов.

На втором этапе обслуживания формируются наборы из  $m_j$  пакетов (посылка), которые последовательно передаются на выделенные  $m_j$  каналов. Каждый из этих каналов принимает на обслуживание по одному поступающему пакету. Один из пакетов заявки имеет некоторый признак, называемый ключом, при обнаружении которого обслуживание заявки заканчивается. Считается, что ключ может обнаружиться в любом из пакетов заявки с равной вероятностью. Каналы системы могут подвергаться самовосстанавливающимся коротким отказам (сбоям), которые вызывают в обслуживаемых пакетах ошибки. Вероятность появления ошибки при обслуживании пакета  $j$ -заявки — заданная величина, равная  $\alpha_j$ . Если пакет обслужен без ошибок (успешное обслуживание), то он удаляется из системы, иначе обслуживание пакета производится заново в составе какой-нибудь последующей посылки этой же заявки. Пустой пакет всегда обслуживается успешно.

Очередная посылка  $j$ -заявки направляется на каналы только после окончания обслуживания предыдущей посылки и при условии, что не завершилось обслуживание заявки. Посылка считается обслуженной, если обслужены все составляющие ее пакеты. Отметим, что в последней посылке заявки часть пакетов могут быть частично или полностью

пустыми. Обслуживание  $j$ -заявки завершается либо в момент превышения контрольного времени ее выполнения  $t_{\text{contrz},j}$ , либо в момент окончания обслуживания очередной посылки, если обнаружен ключ в одном из успешно обслуженных пакетов (пакеты с ошибками не проверяются на наличие ключа), либо успешно обслужены все пакеты задачи. Будем считать, что время обслуживания одной посылки  $j$ -заявки является независимой случайной величиной  $tp_j$  с произвольной функцией распределения  $G_j(t)$  с конечным средним значением  $\bar{t}p_j$ .

Обслуживание  $j$ -заявки называется успешным, если длительность обслуживания не превысила  $t_{\text{contrz},j}$ . Заявка после обслуживания покидает систему, освобождая одновременно все занятые ею каналы.

Как видно из описания модели, она относится к классу мультисервисных СМО с явными потерями, где для обслуживания различных потоков заявок выделяется различное число каналов.

Ставится задача расчета стационарных характеристик описанной модели: распределение вероятностей числа заявок в системе, загруженность каналов, интенсивность выходного потока успешно обслуженных заявок (обслуженных за контрольное время).

## 4 Решение задачи

Найдем вид функции распределения времени обслуживания (пребывания в системе) заявки  $j$ -го типа ( $j = 1, \dots, M$ ) в мультисервисной СМО, описанной в предыдущем разделе в качестве модели ВК.

Рассмотрим сначала случай  $t_{\text{contrz},j} = \infty$ ,  $j = 1, \dots, M$ . Фиксируем некоторый тип заявок  $j$ . Пусть заявка  $j$ -го типа обслуживается на  $m_j = m \geq 1$  каналах, состоит из  $n$  пакетов и  $\alpha_j = \alpha$ . Назовем шагом процесса обслуживания заявки отрезок времени обслуживания одной посылки заявки, состоянием заявки — число пакетов, оставшихся после очередного шага. Обозначим через  $X_k$  состояние заявки после  $k$ -го шага,  $Y$  — число шагов, выполненных за время обслуживания заявки. Так как процесс обслуживания отдельной посылки не зависит от процессов обслуживания других посылок заявки, то значение  $X_k$  зависит только от  $X_{k-1}$ , и поэтому процесс обслуживания заявки  $X = \{X_k, k = 1, 2, \dots\}$  является однородной (см. ниже матрицу вероятностей перехода) марковской цепью с поглощающим состоянием «0». Заметим, что для рассматриваемого случая ( $t_{\text{contrz},j} = \infty$ ) цепь находится в состоянии «0», если успешно обслужены пакеты в последней посылке заявки или

при обслуживании очередной посылки обнаружен ключ.

Переход процесса из состояния  $i$  в состояние  $j > 0$  за один шаг равносильно тому, что из  $m$  пакетов успешно завершили обслуживание  $(i - j)$  пакетов,  $0 \leq (i - j) \leq m$ , и в них не обнаружен ключ. Переход процесса  $X$  из состояния  $i$  в состояние «0» за один шаг равносильно тому, что из  $m$  пакетов успешно завершили обслуживание  $(i - j)$  пакетов,  $0 \leq (i - j) \leq m$ , и в одном из них обнаружен ключ. Событие, состоящее в том, что из  $m$  пакетов успешно завершили обслуживание  $(i - j)$  пакетов,  $0 \leq (i - j) \leq m$ , в данном случае то же самое, что

в схеме Бернулли с вероятностью успеха  $\alpha$  в  $m$  испытаниях произошло  $(i - j)$  успехов. Заметим, что переход процесса  $i$  в состояние  $j$  при  $i < j$  — невозможное событие. Тогда вероятность того, что на одном шаге из  $m$  пакетов успешно завершили обслуживание  $(i - j)$  пакетов,  $0 \leq i - j \leq m$ , равна

$$b_{ij} = \begin{cases} \binom{m}{i-j} (1-\alpha)^{i-j} \alpha^{m-(i-j)}, & \text{если } m \leq i \leq n, 0 \leq i-j \leq m; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Матрица вероятностей  $B = (b_{ij})$  имеет вид:

$$B = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ b_{10} & b_{11} & 0 & \dots & \dots & \dots & 0 \\ b_{20} & b_{21} & b_{22} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ b_{m-10} & b_{m-11} & \dots & \dots & b_{m-1m-1} & 0 & \dots & 0 \\ a_0 & a_1 & \dots & \dots & a_m & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & \dots & a_m & 0 & \dots & 0 \\ 0 & 0 & a_0 & a_1 & \dots & \dots & a_m & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & a_0 & a_1 & \dots & \dots & \dots & a_m & 0 \\ 0 & \dots & \dots & a_0 & a_1 & \dots & \dots & \dots & a_m & 0 \end{pmatrix},$$

где

$$a_j = b_{mj} = \binom{m}{m-j} (1-\alpha)^{m-j} \alpha^j, \quad 0 \leq j \leq m;$$

$$b_{ij} = a_{m-(i-j)}, \quad m \leq i \leq n, \quad 0 \leq i-j \leq m.$$

Так как ключ может находиться с равной вероятностью в любом успешно обслуженном пакете, то матрица переходов  $P = (p_{ji})$  марковской цепи  $X$  имеет вид:

$$p_{ji} = \begin{cases} \sum_{i=\max\{0, j-m\}}^j b_{ji} \frac{j-i}{j} & \text{при } i = 0; \\ b_{ji} \frac{i}{j} & \text{при } i = \max\{1, j-m\}, \dots, j \end{cases}$$

при  $j = 1, \dots, n$  и  $p_{00} = 1$ , где  $p_{ji}$  — вероятность перехода цепи из состояния  $j$  в состояние  $i$ . Следовательно, закон распределения случайной величины  $Y$  имеет вид:

$$P\{Y = k\} = \begin{cases} f(n; m; k) & \text{при } k = 1, 2, \dots; \\ 0 & \text{при } k = 0; \\ 0 & \text{при } k = \infty, \end{cases}$$

где  $f(n; m; k) = \sum_{j=1}^n P(j; n; m; k-1) p_{j0}$  — вероятность того, что впервые цепь попадет в погло-

щающее состояние (либо все пакеты обслужены, либо обнаружен ключ) на  $k$ -м шаге,  $P(i; n; m; k) = \sum_{j=i}^{i+m} P(j; n; m; k-1) p_{ji}$  — вероятность того, что через  $k$  шагов цепь окажется в состоянии  $i \geq 1$ ,  $P(i; n; m; k) = 0$ , если  $i < n - km$  и  $n - km > 0$ ;

$$P(i; n; m; 0) = \begin{cases} 1, & \text{если } i = n; \\ 0, & \text{если } i < n. \end{cases}$$

Обозначим через  $N_j$  число различных значений  $n_i$  величины  $n$  для  $j$ -заявки при различных  $l z_j \in L_j$ ,  $q_{ji}$  — вероятность того, что  $n = n_i$ ,  $i = 1, \dots, N_j$ . Заметим, что для равномерно распределенной на  $L_j$  длины  $j$ -заявки при фиксированных  $L_j$  и  $l p_j$  величины  $q_{ji}$ ,  $i = 1, \dots, N_j$ , легко вычисляются. Тогда функция распределения времени обслуживания  $j$ -заявки при  $t_{\text{contr}, j} = \infty$  для случая непрерывных случайных величин  $t z_j$  и  $t p_j$  имеет вид:

$$B_j(t) = GZ_j * \sum_{i=1}^{N_j} q_{ji} \sum_{k=1}^{\infty} f(n_i; m_j; k) \overbrace{G_j * \dots * G_j}^k(t), \quad 0 \leq t, \quad (1)$$

где \* — знак свертки функций;  $\overbrace{G_j * \dots * G_j}^k$  —  $k$ -кратная свертка функции  $G_j$ ;

$$\alpha = \alpha_j; n_i = \begin{cases} \frac{lz_i}{lp_j}, & \text{если } \frac{lz_i}{lp_j} \text{ — целое число;} \\ \text{целая часть } \left[ \frac{lz_i}{lp_j} \right] + 1 & \text{иначе.} \end{cases}$$

Среднее время обслуживания  $j$ -заявки в этом случае согласно формуле (1) будет равно

$$\bar{t}o_j = \bar{t}z_j + \bar{t}p_j \sum_{i=1}^{N_j} q_{ji} \sum_{k=1}^{\infty} k f(n_i; m_j; k). \quad (2)$$

Вернемся к случаю  $t_{\text{contrz},j} < \infty$ . Тогда функция распределения времени обслуживания  $j$ -заявки имеет вид:

$$\tilde{B}_j(t) = \begin{cases} B_j(t), & \text{если } t < t_{\text{contrz},j}; \\ 1, & \text{если } t \geq t_{\text{contrz},j}. \end{cases}$$

Для среднего значения времени обслуживания имеем:

$$\bar{t}o_j = \int_0^{t_{\text{contrz},j}} t dB_j(t) + [1 - B_j(t_{\text{contrz},j})] t_{\text{contrz},j}.$$

Для случая детерминированных величин  $tz_j$  и  $tp_j$  имеем:

$$\tilde{B}_j(t) = \begin{cases} \sum_{i=1}^{N_j} q_{ji} \sum_{k=1}^{K_j} f(n_i; m_j; k) \chi(t - tz_j - ktp_j), & \text{если } t < t_{\text{contrz},j}; \\ 1, & \text{если } t \geq t_{\text{contrz},j}; \end{cases}$$

$$\bar{t}o_j = tz_j + tp_j \sum_{i=1}^{N_j} q_{ji} \sum_{k=1}^{K_j} k f(n_i; m_j; k) + t_{\text{contrz},j} \left[ 1 - \sum_{i=1}^{N_j} q_{ji} \sum_{k=1}^{K_j} f(n_i; m_j; k) \right],$$

где  $K_j$  — целое число такое, что выполняются неравенства:

$$K_j tp_j < t_{\text{contrz},j} - tz_j \leq (K_j + 1) tp_j,$$

где  $\chi(t)$  — функция Хевисайда:

$$\chi(t) = \begin{cases} 0, & t < 0; \\ 1, & t \geq 0. \end{cases}$$

Для мультисервисной СМО в случае экспоненциальных времен обслуживания заявок известна

мультипликативная формула расчета стационарных вероятностей числа заявок в системе (вывод формул см., например, в [13–16]), которая согласно утверждениям, приведенным в некоторых работах (см., например, [13, 15–17]), инвариантна к виду функций распределения времени обслуживания. Обозначим через  $\bar{k} = \{k_1, \dots, k_M\}$  вектор числа заявок в системе, где  $k_j$  — число  $j$ -заявок в системе;  $P(\bar{k})$  — стационарная вероятность того, что вектор состояния системы равен  $\bar{k}$ ;  $\bar{\Omega}_j$  — множество всевозможных значений вектора  $\bar{k}$ , при которых  $j$ -заявка получает отказ (т. е.  $\bar{\Omega}_j = \{\bar{k} : R - m_j < \sum_{i=1}^M k_i m_i \leq R\}$ ).

В обозначениях, приведенных выше, для стационарных вероятностей состояний системы справедлива формула:

$$p(\bar{k}) = p(\bar{0}) \prod_{j=1}^M \frac{\rho_j^{k_j}}{k_j!},$$

где  $\rho_j = \lambda_j (\bar{t}o_j - tz_j)$ , вероятность  $p(\bar{0})$  вычисляется из условия нормировки и равна

$$p(\bar{0}) = \left[ \sum_{\bar{k} \in \Omega} \prod_{j=1}^M \frac{\rho_j^{k_j}}{k_j!} \right]^{-1}.$$

Согласно последней формуле вероятность того, что  $j$ -заявка получит отказ, равна:

$$\pi_j = \sum_{\bar{k} \in \bar{\Omega}_j} p(\bar{k}).$$

Получаем, что интенсивность выходного потока одновременно обслуженных заявок равна:

$$\Lambda_{\text{вых}} = \sum_{j=1}^M \lambda_j (1 - \pi_j) B_j(t_{\text{contrz},j}). \quad (3)$$

Коэффициент использования системы (загруженность системы) равен:

$$U = \sum_{d=1}^R \frac{dq(d)}{R}, \quad (4)$$

где  $q(d) = \sum_{\bar{k} \in D_d} p(\bar{k})$ ,  $D_d = \left\{ \bar{k} \in \Omega : \sum_{j=1}^M k_j m_j = d \right\}$ .

Вероятности  $p(\bar{k})$  и  $q(d)$  вычисляются по рекуррентным формулам (см., например, [13, 14, 16]).

## 5 Вычислительный эксперимент

В качестве примера использования модели были проведены расчеты значений выходных потоков для нескольких планов распределения ресурсов между потоками задач различной длины. Суммарный поток задач является пуассоновским с интенсивностью 0,5, распределение длины задачи считается равномерным на множестве значений {30, 60, 90, 120, 150}. Задачи разбиты на 5 типов в соответствии с длинами 30, 60, 90, 120 и 150, требуемое время выполнения (контрольное время) задачи любого типа равно 30. Согласно «справедливому» плану минимальное необходимое число ресурсов  $r_j = j$  для  $j$ -задачи,  $j = 1, \dots, 5$ . Другие исходные данные:  $v = 1$ ;  $R = 100$ ;  $\alpha_j = 1 - (1 - 0,001)^{lp_j}$ ;  $tz_j = 0,1$ ;  $tp_j = 0,1 * j * (lp_j + 1)$ ,  $j = 1, \dots, 5$ . В данном примере задача не снимается с обслуживания из-за превышения контрольного времени и среднее время обслуживания задачи считается по формуле (2).

В табл. 1 и на рис. 2 приведены зависимости интенсивности суммарного потока своевременно выполненных задач от плана распределения ресурсов

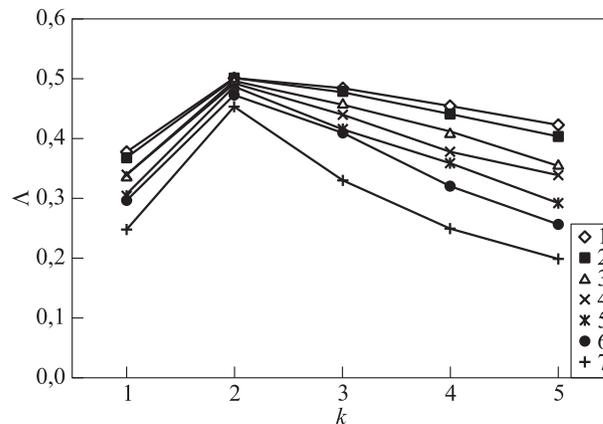
$m_j = kr_j$ ,  $j = 1, \dots, 5$  при различных длинах пакетов. Задача считается своевременно выполненной, если время обслуживания не превысило контрольное время. Ряды 1–7 соответствуют длинам пакетов  $lp_j = 1, 2, 4, 6, 8, 10, 15$ ,  $j = 1, \dots, 5$ .

## 6 Заключение

Как следует из описания данной модели, распределение каналов происходит в ЦУ после поступления задачи с учетом ее длины, заданного контрольного времени и производительности каналов, и при этом производительность ресурсов считается одинаковой. В действительности ВК может состоять из ресурсов, которые имеют различные производительность и вероятность сбоя. Очевидно, для более эффективного использования ВК распределение ресурсов следует производить с учетом их производительности и надежности. Отметим, что предложенная модель ВК может быть использована для оценки и оптимизации эффективности следующего плана распределения ресурсов, учитывающего производительности ресурсов. Согласно указанному плану множество ресурсов

**Таблица 1** Интенсивность выходного потока своевременно выполненных задач

$k$	$lp_j$						
	1	2	4	6	8	10	15
1	0,376290	0,365933	0,338158	0,337962	0,303518	0,297001	0,246217
2	0,498492	0,497666	0,494870	0,490414	0,484638	0,472359	0,451572
3	0,482897	0,476613	0,457543	0,437682	0,412980	0,408341	0,326782
4	0,454405	0,440319	0,410806	0,377289	0,358364	0,318597	0,247202
5	0,421171	0,402171	0,355661	0,336658	0,289217	0,254539	0,196318



**Рис. 2** Зависимости интенсивности суммарного потока своевременно выполненных задач от плана распределения ресурсов: цифрами обозначены ряды;  $\lambda$  — интенсивность выходного потока;  $m_1, m_2, m_3, m_4, m_5$  — число ресурсов, выделенных соответственно потокам 1-, 2-, 3-, 4-, 5-го типа,  $m_1 = k, m_2 = 2k, m_3 = 3k, m_4 = 4k, m_5 = 5k$ ;  $k$  — коэффициент увеличения числа ресурсов

разделено на подмножества так, что ресурсы в одном подмножестве имеют одинаковую производительность. Задано также правило распределения поступающих задач по указанным подмножествам и известны интенсивности потоков задач, поступающих на подмножества ресурсов. Для каждого подмножества распределение ресурсов происходит по такому же плану, что и в разработанной выше модели. Тогда рассчитываем интенсивность выходного потока задач для каждого подмножества ресурсов по формуле (3) и взяв их сумму, получим интенсивность выходного потока для всего множества ресурсов. По формуле (4) вычисляется загруженность ресурсов в каждом подмножестве.

Модель легко обобщается на случай, когда задача содержит несколько ключевых элементарных заданий, которые равновероятно обнаруживаются в любом пакете и задача снимается с выполнения при обнаружении любого из этих ключей. Модель также легко обобщается на случай, когда задача содержит ключи с заданной вероятностью.

Разработанная модель ВК может быть использована для оценки эффективности и оптимизации плана распределения вычислительных ресурсов на определенном выше множестве статических планов типа  $\overline{m}$  в смысле максимизации интенсивности выходного потока  $\Lambda_{\text{вых}}$  по параметрам  $L_j, lp_j, t_{\text{contrz},j}, m_j, j = 1, \dots, M$ .

## Литература

1. Головкин Б. А. Расчет характеристик и планирование параллельных вычислительных процессов. — М.: Радио и связь, 1983.
2. Гуз Д. С., Красовский Д. В., Фуругян М. Г. Эффективные алгоритмы планирования вычислений в многопроцессорных системах реального времени. — М.: ВЦ РАН, 2004.
3. Голосов П. Е., Козлов М. В., Малащенко Ю. Е., Назарова И. А., Ронжин А. Ф. Модель системы управления специализированным вычислительным комплексом // Сообщения по прикладной математике. — М.: ВЦ РАН, 2010.
4. Weiss G., Pinedo M. Scheduling tasks with exponential service times on non-identical processors to minimize various cost functions // J. Appl. Prob., 1980. No. 17. P. 187–202.
5. Лобанов Л. П., Кударенко А. А., Пивоваров Н. В., Терсков В. А. Метод анализа одного класса систем массового обслуживания для оценки производительности многопроцессорных вычислительных систем // Программирование, 1988. № 5. С. 6–12.
6. Borst S., Boxma O., Groote J. F., Mauw S. Task allocation in a multi-server system // J. Scheduling, 2003. No. 6. P. 423–436.
7. Фельдман Л. П., Михайлова Т. В. Использование аналитических методов для оценки эффективности многопроцессорных вычислительных систем // Электронное моделирование, 2007. Т. 29. № 2. С. 17–27.
8. Куланов С. А. Принципы оценки финансовых показателей ГРИД-систем // Jet Info, 2007. № 12. С. 16–27.
9. Koole G., Righter R. Resource allocation in grid computing // J. Scheduling, 2008. No. 11. P. 163–173.
10. Подкопаев И. В. Исследование и разработка методов повышения эффективности управления вычислительными мощностями в кластерах рабочих станций. Автореф. дисс. . . . канд. техн. наук. — М.: МИЭТ(ТУ), 2010.
11. Сайт distributed.net. — <http://www.distributed.net>.
12. Цирюлик О. QNX: кластерные вычисления // Программное обеспечение. Системы реального времени. СТА, 2004. № 3. С. 54–62.
13. Ross K. W. Multiservice loss models for broadband telecommunication networks. — Springer Verlag, 1995.
14. Башарин Г. П. Лекции по математической теории телетрафика. — М.: РУДН, 2007.
15. Степанов С. Н. Основы телетрафика мультисервисных сетей. — М.: Эко-Трендз, 2010.
16. Iversen V. B. Teletraffic engineering and network planning. DTU Course 34340. — Denmark: Technical University of Denmark, Revised January 24, 2011. <http://oldwww.com.dtu.dk/education/34340/material/telenook2011pdf.pdf>.
17. Васькин Ю. А., Пшеничников А. П., Степанов М. С. Распределение канального ресурса при обслуживании мультисервисного трафика // Т-КОММ — Телекоммуникации и транспорт, 2009. № 4. С. 46–48.

# ДЕКОМПОЗИЦИЯ ПРИ ЧАСТИЧНО ИЗВЕСТНОМ РАСПРЕДЕЛЕНИИ ОШИБКИ\*

В. Г. Ушаков<sup>1</sup>, Н. Г. Ушаков<sup>2</sup>

**Аннотация:** Рассмотрена проблема непараметрического оценивания распределения генеральной совокупности в случае, когда выборочные значения наблюдаются с некоторой случайной ошибкой. Предполагается, что распределение ошибки известно лишь частично. Исследуются проблемы идентифицируемости и состоятельного оценивания.

**Ключевые слова:** непараметрическое оценивание; декомпозиция

## 1 Введение

Модель декомпозиции возникает в ситуациях, когда случайная величина  $X$ , распределение которой оценивается, не наблюдается непосредственно, а с некоторой аддитивной ошибкой измерения, т. е. наблюдаются независимые реализации случайной величины

$$Y = X + \varepsilon, \quad (1)$$

где  $\varepsilon$  — ошибка измерения. В классической постановке задачи декомпозиции предполагается, что распределение ошибки известно. Однако в большинстве случаев такое предположение нереалистично. Один из путей преодоления данной трудности состоит в создании дополнительной выборки, элементы которой являются независимыми реализациями ошибки  $\varepsilon$ , и оценивании распределения ошибки по этой выборке. Но и этот путь не всегда возможен. В работах [1–3] исследовались ситуации, когда распределение  $X$  может быть оценено по выборке из распределения  $Y$  при наличии лишь некоторой информации о распределениях  $X$  и  $\varepsilon$ .

Первая проблема, которая возникает при данной постановке, — условия идентифицируемости, т. е. условия, при которых распределение  $X$  может быть однозначно определено по распределению  $Y$ . Вторая проблема — построение состоятельной оценки распределения  $X$ . В работах [1, 3] исследовался случай, когда распределение ошибки нормально. Были даны условия идентифицируемости и доказана состоятельность некоторых типов оценок. Более общие распределения ошибки  $\varepsilon$

рассматривались в [2], где, в частности, были получены некоторые условия идентифицируемости и состоятельности в случае устойчивого распределения ошибки.

В разд. 2, касающемся идентифицируемости, также предполагается, что распределение ошибки устойчиво. Будет получен простой критерий идентифицируемости, который в сочетании с широким набором известных свойств устойчивых распределений и свойств операции свертки позволяет указать ряд условий идентифицируемости. В разд. 3, касающемся построения оценок и их состоятельности, рассматривается общий случай произвольных распределений, для которых идентифицируемость имеет место, т. е. для которых распределение  $X$  однозначно определяется из распределения  $Y$ .

## 2 Идентифицируемость

В этом разделе будет получено необходимое и достаточное условие идентифицируемости распределения  $X$  в модели (1) в случае, когда распределение ошибки является симметричным устойчивым распределением. Симметричное устойчивое распределение с показателем  $\alpha$  ( $0 < \alpha \leq 2$ ) и параметром масштаба  $\gamma$  ( $\gamma > 0$ ) будем обозначать  $U_{\gamma, \alpha}$ . Таким образом,  $U_{\gamma, \alpha}$  — распределение с характеристической функцией  $\exp(-\gamma|t|^\alpha)$ . Множество всех симметричных устойчивых распределений с показателем  $\alpha$  обозначим  $\mathcal{U}_\alpha$ .

Обозначим через  $\mathcal{P}$  множество всех вероятностных распределений на действительной прямой.

\*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а и 11-07-00112а), а также Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

<sup>1</sup>Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vgushakov@mail.ru

<sup>2</sup>Институт проблем технологии микроэлектроники и особочистых материалов Российской академии наук, ushakov@math.ntnu.no

Пусть  $P, P_1 \in \mathcal{P}$ . Распределение  $P_1$  называется компонентой распределения  $P$ , если существует распределение  $P_2$  такое, что  $P = P_1 * P_2$  (\* обозначает свертку). Обозначим через  $\mathcal{P}_\alpha$  множество всех распределений, содержащих в качестве компоненты какое-либо симметричное устойчивое распределение с показателем  $\alpha$ :

$$\mathcal{P}_\alpha = \{P \in \mathcal{P} | \exists \gamma > 0, \exists P_1 \in \mathcal{P} : P = P_1 * U_{\gamma, \alpha}\}.$$

Дополнение множества  $\mathcal{P}_\alpha$  будем обозначать, как обычно,  $\mathcal{P}_\alpha^c$ .

**Теорема 1.** Для того чтобы распределение  $P_X$  случайной величины  $X$  в модели (1), где  $\varepsilon$  имеет симметричное устойчивое распределение  $U_{\gamma, \alpha}$ , было идентифицируемым, необходимо и достаточно, чтобы  $P_X \in \mathcal{P}_\alpha^c$ . Другими словами,

(а) если

$$P_1, P_2 \in \mathcal{P}_\alpha^c, \quad (2)$$

то из равенства

$$P_1 * U_{\gamma_1, \alpha} = P_2 * U_{\gamma_2, \alpha} \quad (3)$$

следует, что  $P_1 = P_2$  и  $\gamma_1 = \gamma_2$ ;

(б) если  $P_1 \notin \mathcal{P}_\alpha^c$  (т.е.  $P_1 \in \mathcal{P}_\alpha$ ), то существуют  $P_2 \in \mathcal{P}$  и  $\gamma_1, \gamma_2 > 0$  такие, что  $P_1 \neq P_2$ ,  $\gamma_1 \neq \gamma_2$  и  $P_1 * U_{\gamma_1, \alpha} = P_2 * U_{\gamma_2, \alpha}$ .

**Доказательство.** Достаточность. Предположим, что (3) имеет место, но  $\gamma_1 \neq \gamma_2$ . Пусть для определенности  $\gamma_1 < \gamma_2$ . Тогда

$$P_1 * U_{\gamma_1, \alpha} = P_2 * U_{\gamma_2 - \gamma_1, \alpha} * U_{\gamma_1}$$

и, следовательно,

$$P_1 = P_2 * U_{\gamma_2 - \gamma_1, \alpha}$$

(поскольку характеристические функции устойчивых законов распределения не обращаются в нуль), что противоречит (2). Противоречие доказывает, что

$$\gamma_1 = \gamma_2. \quad (4)$$

Равенство  $P_1 = P_2$  немедленно следует из (3) и (4).

**Необходимость.** Поскольку  $P_1 \in \mathcal{P}_\alpha$ , существуют  $P_0 \in \mathcal{P}$  и  $\gamma_0 > 0$  такие, что  $P_1 = P_0 * U_{\sigma_0, \alpha}$ . Пусть  $\gamma'_0$  и  $\gamma$  удовлетворяют условиям  $\gamma'_0 > 0$ ,  $\gamma > 0$ ,  $\gamma'_0 < \gamma_0$ . Положим  $\gamma_1 = \gamma + \gamma_0$ ,  $\gamma_2 = \gamma + \gamma'_0$  и  $P_2 = P_0 * U_{\gamma_0 - \gamma'_0}$ . Тогда  $\gamma_1 \neq \gamma_2$ ,  $P_1 \neq P_2$  и

$$\begin{aligned} P_1 * U_{\gamma_1, \alpha} &= P_0 * U_{\gamma + \gamma_0, \alpha} = P_0 * U_{\gamma_0 - \gamma'_0, \alpha} * U_{\gamma'_0 + \gamma, \alpha} \\ &= P_2 * U_{\gamma_2, \alpha}. \end{aligned}$$

Опираясь на полученный критерий, укажем несколько достаточных условий идентифицируемости.

**Условие А.** Носитель распределения  $X$  не совпадает со всей действительной прямой (в случае нормального распределения ошибки это условие получено в [3]).

**Условие Б.** Плотность распределения случайной величины  $X$  не является бесконечно дифференцируемой.

**Условие В.** Характеристическая функция  $\varphi(t)$  случайной величины  $X$  удовлетворяет условию

$$\lim_{|t| \rightarrow \infty} e^{\gamma|t|^\alpha} |\varphi(t)| = \infty$$

для любого  $\gamma > 0$  (частные случаи этого условия рассмотрены в [2]).

### 3 Оценка. Состоятельность

Пусть  $\mathcal{P}_0$  — некоторое множество распределений и распределение  $P^\varepsilon$  ошибки  $\varepsilon$  принадлежит этому множеству:  $P^\varepsilon \in \mathcal{P}_0$ . Обозначим через  $\mathcal{D}_0$  множество распределений (необязательно всех) случайной величины  $X$ , для которых модель (1) идентифицируема. Это значит, что если  $P_1, P_2 \in \mathcal{D}_0$ ,  $Q_1, Q_2 \in \mathcal{P}_0$ , и  $P_1 * Q_1 = P_2 * Q_2$ , то  $P_1 = P_2$  и  $Q_1 = Q_2$ . Если, например, распределение  $\varepsilon$  симметричное устойчивое с показателем  $\alpha$ , т.е.  $\mathcal{P}_0 = \mathcal{U}_\alpha$ , то можно взять  $\mathcal{D}_0 = \mathcal{P}_\alpha^c$ .

Пусть  $\rho(\cdot, \cdot)$  — метрика в пространстве  $\mathcal{P}$ , удовлетворяющая следующим условиям.

**У1.** В этой метрике любое выборочное распределение является сильно (с вероятностью 1) состоятельной оценкой соответствующего теоретического распределения. Другими словами,

$$P \left( \lim_{n \rightarrow \infty} \rho(\hat{P}_n, P) = 0 \right) = 1,$$

где  $P \in \mathcal{P}$  и  $\hat{P}_n$  — выборочное распределение, построенное по выборке объема  $n$  из генеральной совокупности с распределением  $P$ .

**У2.** Если  $\rho(P_n, P) \rightarrow 0$  и  $\rho(Q_n, Q) \rightarrow 0$ , то  $\rho(P_n * Q_n, P * Q) \rightarrow 0$ .

Примеры метрик, удовлетворяющих условиям У1 и У2, даны в следующем разделе. В частности, метрика, используемая в [3], удовлетворяет им.

Пусть  $Y_1, \dots, Y_n$  — случайная выборка в модели (1), т.е. это независимые одинаково распределенные случайные величины, имеющие одинаковое с  $Y$  распределение. Обозначим через  $\hat{P}_n^Y$  выборочное распределение, построенное по этой выборке. Оценку определим следующим образом.

Пару  $(P^X, P^\varepsilon)$  оценим таким элементом множества  $\mathcal{D}_0 \times \mathcal{P}_0$ , который минимизирует расстояние  $\rho(P' * P'', \hat{P}_n^Y)$ , и обозначим его  $(\hat{P}_n^X, \hat{P}_n^\varepsilon)$ . Другими словами,

$$\rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, \hat{P}_n^Y) \leq \rho(P' * P'', \hat{P}_n^Y)$$

для любых  $P' \in \mathcal{D}_0, P'' \in \mathcal{P}_0$ . Заметим, что если такая оценка существует для каждого  $n$ , то

$$P\left(\lim_{n \rightarrow \infty} \rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, P^Y) = 0\right) = 1. \quad (5)$$

Действительно,

$$\begin{aligned} \rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, P^Y) &\leq \rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, \hat{P}_n^Y) + \\ + \rho(\hat{P}_n^Y, P^Y) &\leq \rho(P^X * P^\varepsilon, \hat{P}_n^Y) + \rho(\hat{P}_n^Y, P^Y) = \\ &= 2\rho(\hat{P}_n^Y, P^Y), \end{aligned}$$

откуда (5) следует в силу условия У1. Кстати, попутно получено полезное неравенство

$$\rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, P^Y) \leq 2\rho(\hat{P}_n^Y, P^Y).$$

**Теорема 2.** Если множества  $\mathcal{D}_0$  и  $\mathcal{P}_0$  не пусты и компактны (в метрике  $\rho$ ), то  $\hat{P}_n^X$  и  $\hat{P}_n^\varepsilon$  существуют и являются сильно (почти наверное) состоятельными оценками  $P^X$  и  $P^\varepsilon$ , т. е.

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} \rho(\hat{P}_n^X, P^X) = 0\right) &= 1; \\ P\left(\lim_{n \rightarrow \infty} \rho(\hat{P}_n^\varepsilon, P^\varepsilon) = 0\right) &= 1. \end{aligned}$$

**Доказательство.** Существование очевидно в силу компактности множеств  $\mathcal{D}_0$  и  $\mathcal{P}_0$ . Пусть  $\omega$  — произвольное элементарное событие в исходном вероятностном пространстве, такое что

$$\lim_{n \rightarrow \infty} \rho(\hat{P}_n^X * \hat{P}_n^\varepsilon, P^Y) = 0. \quad (6)$$

Вероятность события, состоящего из всех таких элементарных событий, равна 1. Зафиксируем это  $\omega$ , и далее в доказательстве все случайные величины и функции будем рассматривать на нем.

Рассмотрим последовательности  $\hat{P}_n^X = \hat{P}_n^X(\omega)$  и  $\hat{P}_n^\varepsilon = \hat{P}_n^\varepsilon(\omega)$ . Предположим, что либо  $\rho(\hat{P}_n^X, P^X)$ , либо  $\rho(\hat{P}_n^\varepsilon, P^\varepsilon)$  (либо оба) не сходятся к нулю. Тогда существует подпоследовательность  $\{n_k^{(1)}\} \subseteq \{n\}$  такая, что

$$\rho(\hat{P}_{n_k^{(1)}}^X, P^X) \geq \delta,$$

или подпоследовательность  $\{n_k^{(2)}\} \subseteq \{n\}$  такая, что

$$\rho(\hat{P}_{n_k^{(2)}}^\varepsilon, P^\varepsilon) \geq \delta$$

для некоторого  $\delta > 0$ . Не ограничивая общности (и ради упрощения обозначений), можно считать, что подпоследовательности  $\{n_k^{(1)}\}$  и  $\{n_k^{(2)}\}$  совпадают с самой последовательностью  $\{n\}$ , т. е.

$$\rho(\hat{P}_n^X, P^X) \geq \delta \quad (7)$$

или

$$\rho(\hat{P}_n^\varepsilon, P^\varepsilon) \geq \delta. \quad (8)$$

Поскольку последовательность  $\{\hat{P}_n^X\}$  принадлежит компактному множеству  $\mathcal{D}_0$ , существуют подпоследовательность  $\{n_k\} \subseteq \{n\}$  и распределение  $P' \in \mathcal{D}_0$  такие, что

$$\lim_{k \rightarrow \infty} \rho(\hat{P}_{n_k}^X, P') = 0. \quad (9)$$

В свою очередь, поскольку подпоследовательность  $\{P_{n_k}^\varepsilon\}$  содержится в компактном множестве  $\mathcal{P}_0$ , существуют ее подпоследовательность  $\{n'_k\} \subseteq \{n_k\}$  и распределение  $P'' \in \mathcal{P}_0$  такие, что

$$\lim_{k \rightarrow \infty} \rho(\hat{P}_{n'_k}^\varepsilon, P'') = 0. \quad (10)$$

Из (9) и (10) в силу свойства У2 метрики  $\rho$  следует, что

$$\lim_{k \rightarrow \infty} \rho(\hat{P}_{n'_k}^X * \hat{P}_{n'_k}^\varepsilon, P' * P'') = 0.$$

С другой стороны, из (6) имеем:

$$\lim_{k \rightarrow \infty} \rho(\hat{P}_{n'_k}^X * \hat{P}_{n'_k}^\varepsilon, P^X * P^\varepsilon) = 0.$$

Таким образом,

$$P^X * P^\varepsilon = P' * P'', \quad (11)$$

причем

$$P^X, P' \in \mathcal{D}_0, P^\varepsilon, P'' \in \mathcal{P}_0 \quad (12)$$

и как минимум одно из неравенств (7), (8) имеет место. Если выполнено (7), то в силу (9)

$$P^X \neq P', \quad (13)$$

а (11) и (13) противоречат (12). Если имеет место (8), то в силу (10)

$$P^\varepsilon \neq P'' \quad (14)$$

и теперь (11) и (14) противоречат (12). Полученные противоречия доказывают теорему.

**Следствие.** Пусть ошибка  $\varepsilon$  имеет симметричное устойчивое распределение с показателем  $\alpha$  и параметром масштаба  $\gamma$ , а случайная величина  $X$  ограничена. Если известны постоянные  $a$  и  $b$  такие, что  $|X| \leq a$  и  $|\gamma| \leq b$ , то оценка  $(\hat{P}_n^X, \hat{P}_n^\varepsilon)$  является состоятельной для  $(P^X, P^\varepsilon)$ .

Другие подобные следствия могут быть легко получены с использованием условий А, Б и В. Предоставляем это читателю.

## 4 Выбор метрики

В этом разделе рассмотрим несколько метрик, которые могут использоваться в качестве метрики  $\rho(\cdot, \cdot)$ , т. е. удовлетворяют условиям У1 и У2. Пусть  $P$  и  $Q$  (с индексом или без индекса) — вероятностные распределения. Обозначим соответствующие функции распределения и характеристические функции как  $F(x)$ ,  $G(x)$  и  $\varphi(t)$ ,  $\psi(t)$ .

### 1. Метрика Колмогорова (равномерная метрика)

Метрика определяется как равномерное расстояние между функциями распределения:

$$\rho(P, Q) = \sup_{-\infty < x < \infty} |F(x) - G(x)|.$$

Условие У1 выполнено в силу теоремы Гливенко—Кантелли. Условие У2 следует из очевидного неравенства

$$\begin{aligned} & \sup_{-\infty < x < \infty} |F_n * G_n(x) - F * G(x)| \leq \\ & \leq \sup_{-\infty < x < \infty} |F_n(x) - F(x)| + \sup_{-\infty < x < \infty} |G_n(x) - G(x)|. \end{aligned}$$

### 2. Взвешенная равномерная метрика, основанная на характеристических функциях

Хорошо известно, что выборочная характеристическая функция почти наверное сходится к соответствующей теоретической характеристической функции равномерно на каждом ограниченном интервале. Более того, если  $\varphi(t)$  и  $\varphi_n(t)$  — соответственно теоретическая и выборочная характеристические функции ( $n$  — объем выборки), то почти наверное

$$\lim_{n \rightarrow \infty} \sup_{|t| \leq T_n} |\varphi_n(t) - \varphi(t)| = 0 \quad (15)$$

для любой последовательности положительных чисел  $\{T_n\}$ , удовлетворяющей условию:

$$\lim_{n \rightarrow \infty} \frac{\log T_n}{n} = 0.$$

Таким образом, можно использовать следующую метрику:

$$\rho(P, Q) = \sup_{-\infty < t < \infty} |\varphi(t) - \psi(t)| w(t),$$

где  $w(t)$  — положительная четная функция, удовлетворяющая условию  $\lim_{t \rightarrow \infty} w(t) = 0$ . Условие У1

следует из (15) (и даже просто из равномерной сходимости на каждом ограниченном интервале). Условие У2 следует из неравенства

$$\begin{aligned} & |\varphi_n(t)\psi_n(t) - \varphi(t)\psi(t)| \leq \\ & \leq |\varphi_n(t) - \varphi(t)| + |\psi_n(t) - \psi(t)|. \end{aligned} \quad (16)$$

### 3. Интегральная метрика, основанная на характеристических функциях

$$\rho(P, Q) = \int_{-\infty}^{\infty} |\varphi(t) - \psi(t)|^p w(t) dt,$$

где  $p > 0$ , а  $w(t)$  — плотность распределения вероятностей. Как и в предыдущем случае, (15) гарантирует выполнение условия У1. Условие У2 следует из (16). Если  $p = 1$ , эта метрика совпадает с метрикой, использованной в [3].

### 4. Метрика Леви

$$\begin{aligned} \rho(P, Q) = L(F, G) = \inf \{ h \geq 0 : F(x - h) - h \leq \\ \leq G(x) \leq F(x + h) + h, -\infty < x < \infty \}. \end{aligned}$$

Сходимость в метрике Леви слабее, чем сходимость в метрике Колмогорова, поэтому условие У1 выполнено в силу теоремы Гливенко—Кантелли. Условие У2 следует из неравенства

$$L(F_n * G_n, F * G) \leq L(F_n, F) + L(G_n, G),$$

которое можно найти в любой книге по вероятностным метрикам.

## Литература

1. *Matias C.* Semiparametric deconvolution with unknown noise variance // *ESIAM Probab. Statist.*, 2002. Vol. 6. P. 271–292.
2. *Butucea C., Matias C.* Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model // *Bernoulli*, 2005. Vol. 11. No. 2. P. 309–340.
3. *Schwarz M., Van Bellegem S.* Consistent density deconvolution under partially known error distribution // *Stat. Probab. Lett.*, 2010. Vol. 80. No. 3–4. P. 236–241.

# ЗАДАЧИ АНАЛИЗА И ОПТИМИЗАЦИИ ДЛЯ МОДЕЛИ ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ.

## ЧАСТЬ 1. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ

А. В. Босов<sup>1</sup>

**Аннотация:** Предложена математическая модель описания активности пользователей, формирующих запросы в некоторой информационной системе. Исследованы свойства модели, предложены процедуры прогнозирования и идентификации параметров уравнений. Представлены результаты численных экспериментов, в том числе с использованием реальных данных.

**Ключевые слова:** информационная система; моделирование; эргодический случайный процесс; параметрическая идентификация; прогнозирование

### 1 Введение

Уровень развития телекоммуникационных средств, совершенство информационных технологий, доступность и востребованность программного обеспечения в последние десятилетия не только привели к повсеместному распространению различных информационных систем, но и оформили их в качестве перспективного объекта применения самых разных математических методов. Современные информационные системы объединяют разнотипные аппаратные ресурсы, коммуникационные средства, программные системы, а также пользователей разного уровня. Масштабные характеристики их постоянно растут, что является стимулом к совершенствованию используемого для их анализа математического аппарата, созданию новых моделей, постановкам новых задач оптимизации.

Наиболее популярное сложившееся направление исследований в этой области составляют задачи оптимизации процессов функционирования крупных информационно-телекоммуникационных систем. Так, весьма развиты и широко используются различные методы моделирования и анализа потоков в телекоммуникационных сетях, основанные на теории массового обслуживания (см., например, [1, 2]). Другой существенный класс задач, рассматриваемых в рамках данного направления, связан с управлением потоками информации, в том числе в Интернете, осуществляемым с помощью различных версий протокола сетевого взаимодействия TCP (Transfer Control Protocol) [3, 4]. Постановки последних чаще всего основаны на моделях управляемых конечных цепей Маркова (см., например, [5–7]).

Эти и другие методы позволяют получать существенные результаты, но имеют и определенные ограничения применимости. Так, в [8] отмечается: (1) неадекватность моделей массового обслуживания, возникающая в связи с возрастающей сложностью структур и дисциплин обслуживания; (2) необходимость наличия детального описания структуры сети; (3) недостаточность моделей, использующих в качестве входной информацию о наблюдаемых (статистических) показателях функционирования сети. Сюда же следует добавить (4) отсутствие моделей и постановок, учитывающих специфику работы программного обеспечения, в частности в связи с влиянием на систему действий пользователей.

Последнее обстоятельство представляется не вполне объяснимым, ведь хорошо известны многие практические примеры оптимизации функционирования программного обеспечения. К ним можно отнести и управление страничным файлом операционной системы, и оптимизацию запросов реляционной системы управления базами данных [9], и алгоритмы диспетчеризации задач в многопроцессорных средах [10], и многие другие. Однако в целом существенных примеров успешного применения математического аппарата к таким задачам немного.

В данной работе объектом исследования является именно процесс функционирования программной системы, причем в качестве основного фактора влияния рассматривается пользовательская активность. Несмотря на достаточно общий характер постановок задач, рассматриваются они в связи с функционированием вполне конкретного программного обеспечения — Информационного

<sup>1</sup>Институт проблем информатики Российской академии наук, AVBosov@ipiran.ru

веб-портала [11], представляющего собой достаточно сложное программное средство, реализующее детально описанные алгоритмы обработки информации. Это позволяет не просто предложить новые модели и сформулировать для них задачи оптимизации, но и проанализировать работоспособность предлагаемых процедур на реальных данных.

Основная задача, решаемая далее, связана с моделированием процесса активности пользователей, использующих портал: предложена модель процесса в форме динамической стохастической системы специального вида, выяснены ее свойства, разработаны процедуры прогнозирования и идентификации параметров модели. В последующих работах будут представлены постановки и решения задач оптимизации расходования вычислительных ресурсов, используемых порталом, на основе модели пользовательской активности.

## 2 Используемые обозначения

Далее в работе будут использованы следующие обозначения:

- $\triangleq$  — равенство по определению;
- $\mathbf{M}[x]$  и  $\mathbf{M}[x|\mathfrak{S}]$  — соответственно безусловное математическое ожидание случайной величины  $x$  и условное математическое ожидание  $x$  относительно  $\sigma$ -алгебры  $\mathfrak{S}$ ;
- $\mathbf{P}(\cdot)$  и  $\mathbf{P}(\cdot|\mathfrak{S})$  — соответственно безусловная и условная относительно  $\sigma$ -алгебры  $\mathfrak{S}$  вероятностные меры, заданные на подходящем измеримом пространстве;
- $\mathbf{B}(\mathbf{R}^1)$  —  $\sigma$ -алгебра борелевских множеств на числовой прямой  $\mathbf{R}^1$ ;
- $x^T$  — операция транспонирования вектора (матрицы)  $x$ ;
- $\text{col}(x_1, \dots, x_n) \triangleq (x_1, \dots, x_n)^T$  — вектор-столбец с элементами  $x_1, \dots, x_n$ ;
- $\text{row}(x_1, \dots, x_n) \triangleq (x_1, \dots, x_n)$  — вектор-строка с элементами  $x_1, \dots, x_n$ ;
- $e_k = (0, \dots, 0, 1, 0, \dots, 0)^T$  — единичный вектор в пространстве  $\mathbf{R}^n$ , все координаты которого равны 0, а  $k$ -я координата равна 1,  $1 \leq k \leq n$ ;
- $\mathfrak{S}_t^y \triangleq \sigma\{y_\tau, \tau \leq t\}$  —  $\sigma$ -алгебра, порожденная наблюдениями  $y_\tau, \tau \leq t$ ;
- $\psi_x(x, t)$  — безусловная плотность вероятности случайного процесса  $x_t$ ;
- $\bar{\psi}_x(x, t)$  — условная плотность вероятности  $x_t$  относительно  $\mathfrak{S}_{t-1}^y$ ;

- $\bar{\psi}_x(x, t, j)$ ,  $j = 0, 1, \dots$  — условная плотность вероятности  $x_{t+j}$  относительно  $\sigma$ -алгебры  $\mathfrak{S}_{t-1}^y$  (таким образом,  $\bar{\psi}_x(x, t, 0) \triangleq \bar{\psi}_x(x, t)$ );
- $\hat{\psi}_x(x, t)$  — условная плотность вероятности  $x_t$  относительно  $\mathfrak{S}_t^y$ ;
- $\mathbf{R}[a; b]$  — равномерное распределение на отрезке  $[a; b]$ .

## 3 Модель активности пользователей портала

Сделаем вначале два замечания. Во-первых, об алгоритмах, реализуемых Информационным веб-порталом. Во-вторых, о базовых предположениях, используемых при формировании модели.

Основная функциональность портала обеспечивает выполнение пользовательских запросов, формируемых средствами веб-сайта, множеством информационных источников — «внешних» по отношению к portalу информационных систем, подключенных к нему с целью предоставления пользователю унифицированного доступа к разнородным данным, формируемым федеративной средой некоторой распределенной информационной системы [12]. Портал в этой системе выполняет функции центрального веб-узла, единой точки доступа к данным и сервисам. Для целей данной работы достаточно учитывать следующую упрощенную цепочку действий, выполняемых программным обеспечением портала при обслуживании пользовательского запроса (подробнее реализуемые порталом алгоритмы описаны в [13, 14]).

Для выполнения запроса на поиск данных в информационных источниках, подключенных к portalу, пользователь заполняет поисковую форму на сайте портала, формируя тем самым пользовательский HTTP-запрос. Полученный запрос порталными подсистемами вначале ассоциируется с набором команд — вызовов функций отбора данных одного из поддерживаемых порталом типов данных. Затем из каждой команды формируется набор запросов к нескольким источникам, поддерживающим выбранный пользователем тип данных. Запросы к источникам выполняются непосредственно их адаптерами. Число пользовательских запросов зависит, очевидно, от текущего числа пользователей портала (пользовательской активности). Каждому пользовательскому запросу может быть поставлено в соответствие любое число команд (это обеспечивается шаблонами сайта портала, позволяющими объединять в одном URL запросы к произвольному набору типов данных портала). Число источников, участвующих в выполнении команды,

также заранее неизвестно, так как зависит, с одной стороны, от выбранного типа данных, с другой — от заданного пользователем состава опрашиваемых источников.

При формировании модели пользовательской активности необходимо учесть, во-первых, что, как и подавляющее большинство веб-систем, портал использует протокол HTTP, взаимодействие по которому не является непрерывным (disconnected) и не имеет состояния сеанса (stateless). Это означает, что текущее число пользователей неизвестно и судить о нем можно лишь по косвенным данным, например по поступающему числу пользовательских запросов, формируемых команд и/или выполняемых запросов к источникам.

Кроме того, в модели нужно учесть и целый набор неконтролируемых факторов, формируемых окружением портала, например производительность и доступность информационных систем источников, объемные характеристики поддерживаемых хранилищ, семантику поступающих запросов и проч. Наибольший вклад в неопределенность портального окружения, естественно, вносят его пользователи. Их численность представляется наиболее важным фактором, влияющим на функционирование портала, так как обслуживание пользовательских запросов является его основной задачей. В конечном итоге именно сведения о числе обслуживаемых пользователей должны стать основным фактором, используемым при оптимизации работы программного обеспечения портала.

Описывая далее пользовательскую активность, прежде всего постараемся учесть хорошо известный факт наличия спонтанных всплесков пользовательской активности (так называемые «slashdot-эффект» или «хабраэффект», получившие название от соответствующих сайтов [15]). Такие всплески могут быть вызваны, например, появлением контента, вызывающего повышенный интерес, повышением деловой активности в связи с особыми событиями, наконец, распорядком рабочих/свободных часов. Терминологически такие обстоятельства предлагается характеризовать предположением о наличии нескольких режимов функционирования портала. В каждом режиме число пользователей будем описывать простой линейной моделью. Скачкообразный характер рассматриваемого показателя будем учитывать предложением о смене моделей при изменениях режимов. Причем основанием для смены режима будем полагать достижение текущим показателем некоторого заданного порогового значения.

Хорошо известны модели числовых рядов, обладающие таким характером, — линейные модели с порогом TAR, TARMA и другие их разновидности

[16]. Кроме того, этот подход уже применялся в задаче оптимизации функционирования портала при взаимодействии с подключенными информационными источниками [17]. Этим же принципом воспользуемся и для формирования модели пользовательской активности.

Будем считать, что интервал работы портала  $[t_0; +\infty)$  разбит на отрезки (можно считать равные)  $t_0 < t_1 < t_2 < \dots < t_{n-1}$ , на текущем отрезке  $(t_{n-1}; t_n]$  формируется значение показателя пользовательской активности  $x_{t_n}$ . Практически значение данного показателя может задавать либо среднее число пользователей, обратившихся с запросами к portalу в течение заданного промежутка времени, либо их суммарное число. Дискретизация показателя, а также выбор достаточно большой длины интервала  $(t_{n-1}; t_n]$  (2, 5, 10 или даже 30 мин) объясняются необходимостью ограничения затрат вычислительных ресурсов, выделенных портальным приложениям, на оптимизацию собственной работы. Представляется очевидным, что использование непрерывной модели приведет к чрезмерным накладным расходам и будет оказывать слишком большое влияние на выполнение прямой функциональности портала.

Для упрощения обозначений далее будем считать, что  $t_0 = 0, t_1 = 1, \dots, t_n = t, \dots$ , показатель пользовательской активности описывается процессом  $x_t, t = 0, 1, \dots$

В качестве наблюдений за состоянием  $x_t$  можно воспользоваться представляющим достаточно обоснованным предположением о линейной зависимости между числом активных пользователей  $x_t$  и числом команд  $y_t$ , выполненных порталом за интервал наблюдения:

$$y_t = cx_t + \sigma w_t, \quad (1)$$

где параметр  $c$  определяет среднее число команд, формируемых одним пользователем за интервал наблюдения  $(t - 1; t]$ ;  $w_t$  — возмущение (шум), моделирующее отклонения числа команд от заданного среднего уровня (далее предполагается, что  $\{w_t\}$  — стандартный дискретный белый шум в узком смысле, сечения которого имеют плотность вероятности  $\varphi_w(\cdot)$ );  $\sigma$  — характеристика величины разброса этого возмущения.

Для моделирования спонтанных всплесков будем предполагать, что заданы  $n > 1$  характерных режимов пользовательской активности, каждому из которых соответствует интервал значений процесса  $x_t$ . Иными словами, будем предполагать, что область значений  $x_t$  разбита на непересекающиеся интервалы  $\Delta_k$ :

$$-\infty = a_1 < a_2 < \dots < a_n < a_{n+1} = +\infty;$$

$$\Delta_k = (a_k, a_{k+1}], \quad k = 1, \dots, n-1, \quad \Delta_n = (a_n, +\infty).$$

Для описания текущего режима пользовательской активности определим индикаторную функцию  $\Theta(x)$ :

$$\Theta(x) = \text{col}(I_{\Delta_1}(x), \dots, I_{\Delta_n}(x));$$

$$I_{\Delta_k}(x) = \begin{cases} 1, & \text{если } x \in \Delta_k, \\ 0, & \text{если } x \notin \Delta_k. \end{cases} \quad (2)$$

Например, можно предположить, что есть три режима пользовательской активности:

- (1) режим слабой активности;
- (2) режим повседневной активности (описывается средним характерным числом пользователей, регулярно использующих портал в рабочее время);
- (3) режим повышенной активности.

Для каждого режима зададим набор параметров  $a = \text{row}(a_1, \dots, a_n)$ ,  $q = \text{row}(q_1, \dots, q_n)$ ,  $b = \text{row}(b_1, \dots, b_n)$ , предполагая, что эволюция  $x_t$  в  $k$ -м режиме описывается простейшей линейной системой — авторегрессией первого порядка:

$$x_t = a_k x_{t-1} + q_k + b_k v_t.$$

Если каждая из авторегрессий устойчива, то параметры  $a_k$ ,  $q_k$  для  $k$ -го режима определяют некоторое характерное среднее число пользователей, параметр  $b_k$  характеризует величину разброса, моделируемого возмущением  $v_t$ .

Таким образом, с учетом введенных обозначений параметров и соотношений (2) получена следующая модель для описания показателя пользовательской активности:

$$x_t = a\Theta(x_{t-1})x_{t-1} + q\Theta(x_{t-1}) + b\Theta(x_{t-1})v_t, \quad t = 1, 2, \dots \quad (3)$$

Отметим, что согласно (3) изменения режимов активности продуцируются непосредственно текущим значением показателя  $x_t$ : при достижении им границы интервала, определенной для текущего режима, характер динамики (номер авторегрессии) изменяется, и если при этом стационарные средние значения соответствующих устойчивых авторегрессий существенно отличаются, то за несколько последующих шагов (переходный процесс) произойдет и смена характера динамики  $x_t$ , и резкая смена текущего числа пользователей.

## 4 Свойства показателя пользовательской активности

В данном разделе исследованы свойства случайного процесса  $x_t$ , описываемого уравнениями (2) и (3). Уточняя модель (3), будем предполагать, что  $\{v_t\}$  — стандартный дискретный белый шум в узком смысле, сечения которого имеют плотность вероятности  $\varphi_v(\cdot)$ ,  $x_0$  — случайная величина, не зависящая от  $\{v_t\}$  и имеющая плотность вероятности  $\psi_0(\cdot)$ .

Обозначим переходную вероятность для процесса  $x_t$ :  $\mathbf{P}(x, t, B) \triangleq \mathbf{P}(x_t \in B | x_0 = x)$ ,  $x \in \mathbf{R}^1$ ,  $B \in \mathbf{B}(\mathbf{R}^1)$ .

**Теорема 1.** Пусть выполнены следующие условия:

- (1)  $\varphi_v(x) > 0$ ,  $\psi_0(x) > 0 \quad \forall x \in \mathbf{R}^1$ ;
- (2)  $\varphi_v(\cdot)$ ,  $\psi_0(\cdot)$  непрерывны;
- (3)  $|a_1| < 1$ ,  $|a_n| < 1$ ;
- (4)  $b_k > 0$ ,  $k = 1, \dots, n$ .

Тогда процесс  $x_t$  является эргодическим, т. е. существует единственное вероятностное распределение  $\pi^*(\cdot)$  на  $(\mathbf{R}^1, \mathbf{B}(\mathbf{R}^1))$ :

$$\sup_{B \in \mathbf{B}(\mathbf{R}^1)} |\mathbf{P}(x, t, B) - \pi^*(B)| \rightarrow 0 \quad \text{при } t \rightarrow \infty, \quad (4)$$

где сходимость имеет место  $\forall x \in \mathbf{R}^1$  и  $\forall \psi_0(\cdot)$ . При этом предельная мера  $\pi^*(\cdot)$  абсолютно непрерывна относительно меры Лебега.

**Доказательство.** Заметим вначале, что в условиях теоремы маргинальная плотность вероятности  $\psi_x(x, t)$  существует и описывается следующими рекуррентными соотношениями:

$$\left. \begin{aligned} \psi_x(x, t) &= \int_{\mathbf{R}^1} \psi_x(y, t-1) \rho(x, y, t) dy = \\ &= \sum_{k=1}^n \int_{\Delta_k} \psi_x(y, t-1) \frac{1}{b_k} \varphi_v\left(\frac{x - a_k y - q_k}{b_k}\right) dy; \\ \psi_x(x, 0) &= \psi_0(x), \end{aligned} \right\} \quad (5)$$

где переходная плотность  $x_t$  за один шаг

$$\rho(x, y, t) = \sum_{k=1}^n I_{\Delta_k}(y) \frac{1}{b_k} \varphi_v\left(\frac{x - a_k y - q_k}{b_k}\right).$$

Обозначим  $p_y(B) \triangleq \mathbf{P}(y, 1, B)$ . Тогда

$$p_y(B) = \int_B \rho(x, y, 1) dx =$$

$$= \int_B \frac{1}{b_k} \varphi_v\left(\frac{x - a_k y - q_k}{b_k}\right) dx$$

для  $k : y \in \Delta_k$ . Отсюда и из неотрицательности плотностей  $\varphi_v(\cdot)$ ,  $\psi_0(\cdot)$  следует, что  $p_y(B) > 0 \forall y \in \mathbf{R}^1$  и для любого борелевского  $B$ , имеющего ненулевую меру Лебега. Отсюда следует, что у цепи  $x_t$  нет циклов и она неразложима.

Сходимость (4) обеспечивается условиями (3) и (4) теоремы 1 (см. замечание 3.1.2 к теореме 3.1 в [18]).

Заметим теперь, что из (5) с учетом непрерывности  $\varphi_v(\cdot)$ ,  $\psi_0(\cdot)$  вытекает непрерывность  $\psi_x(x, t) \forall t$ , т.е. меры  $\mathbf{P}(x, t, B)$  абсолютно непрерывны и имеют непрерывные плотности  $\forall x \in \mathbf{R}^1$ . Таким образом, сходимость (4) гарантирует абсолютную непрерывность предельной меры  $\pi^*(\cdot)$  и непрерывность предельной плотности  $\psi^*(x)$ . Теорема доказана.

**Следствие.** Если предельная плотность вероятности определяет распределение начального условия  $x_0$ , т.е.  $\psi_0(x) = \psi^*(x)$ , то процесс  $x_t$  является стационарным в узком смысле. Если при этом  $M[v_t^2] < \infty$ , то процесс  $x_t$  является стационарным в широком смысле.

**Пример 1.** Вычислительные эксперименты для предложенной модели пользовательской активности проводились как на модельных данных, позволяющих в полном объеме протестировать предлагаемые далее оптимизационные процедуры, так и на реальных данных, позволяющих сделать некоторые выводы о возможностях практического применения. В статье приведены два варианта расчетов. Вначале рассмотрим модельный вариант определения параметров. Воспользуемся предположением разд. 3 о трех уровнях пользовательской активности и зададим следующие три интервала:  $\Delta_1 = (-\infty; 3]$ ;  $\Delta_2 = (3; 7]$ ;  $\Delta_3 = (7; +\infty)$ . Параметры уравнения (3) и некоторые числовые характеристики процесса  $x_t$  приведены в табл. 1.

Будем предполагать, что распределения, определяемые плотностями вероятности возмущений

**Таблица 1** Параметры модели

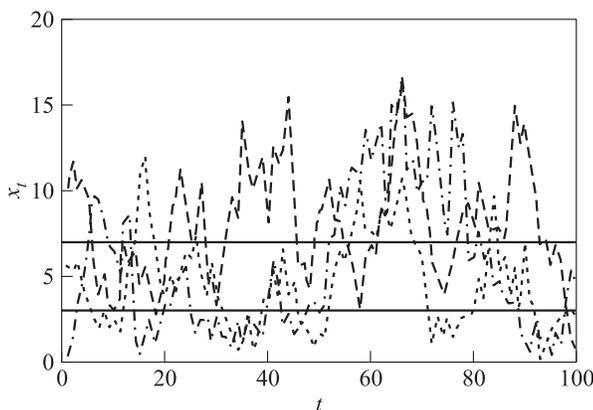
$a_1$	$a_2$	$a_3$	$q_1$	$q_2$	$q_3$	$b_1$	$b_2$	$b_3$
0,3	0,4	0,7	1,4	3,0	3,0	0,9	1,5	2,5
$M[x_t]$					$D[x_t]$			
5,61					19,77			
$M[\Theta^T(x_t)e_1]$			$M[\Theta^T(x_t)e_2]$			$M[\Theta^T(x_t)e_3]$		
0,3328			0,3746			0,2926		

$\varphi_v(\cdot)$  и  $\varphi_w(\cdot)$ , являются стандартными гауссовскими, начальное условие  $x_0$  также предполагалось гауссовским со средним  $M[x_t]$  и дисперсией  $D[x_t]$ , указанными в табл. 1, т.е. равными соответствующим моментам предельного распределения.

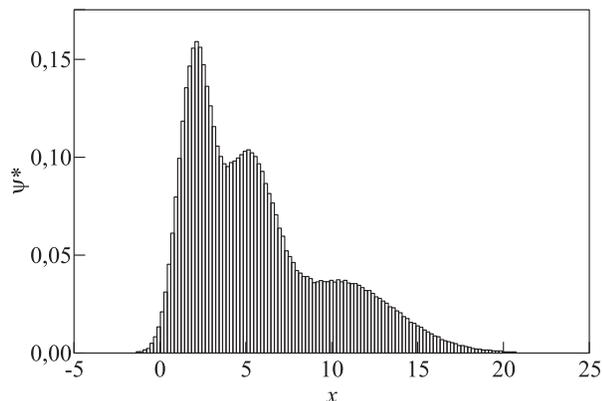
Примеры траекторий  $x_t$  с указанными параметрами приведены на рис. 1, оценка (гистограмма) предельной плотности  $\psi^*(x)$  — на рис. 2. В целом можно констатировать, что траектории иллюстрируют ожидаемое поведение, а вид предельной плотности вполне согласуется с переключающимся характером процесса.

Второй вариант выполненных расчетов опирается на реальные данные — наблюдения за активностью пользователей портала РАН [www.ras.ru](http://www.ras.ru). Накоплена значительная статистика, включающая данные о числе команд, выполненных порталом на 10-минутных интервалах. На рис. 3 представлены произвольно выбранные 5000 наблюдений. Отметим, что, вообще говоря, эти данные характеризуют пользовательскую активность лишь косвенно. Высказанные в разд. 3 предположения о модели наблюдений означают, что на рис. 3 представлены реализации процесса  $y_t$  из (3). Однако линейная зависимость в (3) позволяет предполагать (по крайней мере, для целей первичного анализа), что динамический характер обоих процессов сходен.

Нетрудно видеть, что в имеющихся данных присутствует регулярный периодический тренд (линия тренда нанесена на рисунке темным цветом). Пе-



**Рис. 1** Выборочные функции



**Рис. 2** Предельная плотность

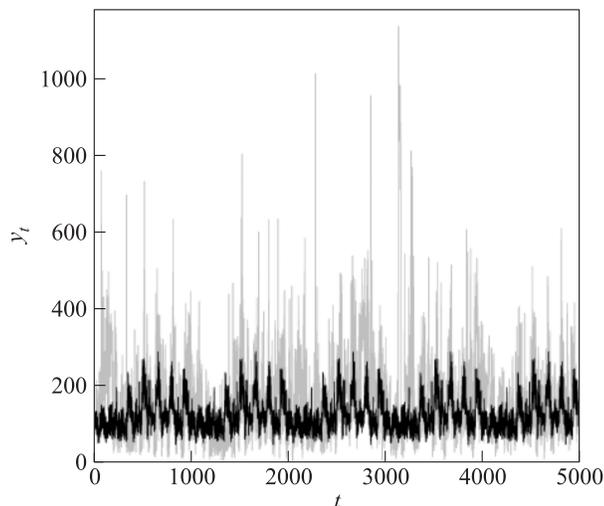


Рис. 3 Данные сайта www.ras.ru

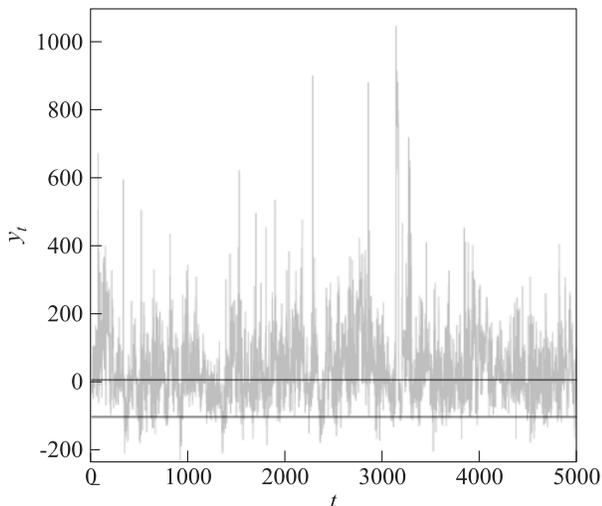


Рис. 4 Траектория с исключенным трендом

риодичность, как выяснилось, связана с естественными недельными колебаниями активности. Этот детерминированный компонент нетрудно было выделить. С этой целью имеющаяся выборка была разделена на недельные интервалы, тренд сформирован из медианных оценок, вычисленных в каждый момент времени (использование медианы в качестве оценки объясняется наличием выбросов в анализируемых данных). После этого из имеющихся наблюдений тренд был исключен (что объясняет, в частности, наличие отрицательных значений у процесса  $x_t$  далее). В результате сформировался

новый набор данных из приращений, фрагмент которых приведен на рис. 4. Эти данные позволили сформировать экспертные предположения о значениях параметров модели и некоторых ее числовых характеристиках (подробно использованные экспертом предложения описаны в [19]). Как и в предыдущем примере, определены три интервала пользовательской активности:  $\Delta_1 = (-\infty; -100]$ ;  $\Delta_2 = (-100; 10]$ ;  $\Delta_3 = (10; +\infty)$ ; параметры и характеристики  $x_t$  приведены в табл. 2.

Примеры траекторий  $x_t$  приведены на рис. 5, гистограмма предельной плотности  $\psi^*(x)$  — на

Таблица 2 Параметры модели

$a_1$	$a_2$	$a_3$	$q_1$	$q_2$	$q_3$	$b_1$	$b_2$	$b_3$
0,361	0,353	0,725	-67,78	-26,26	39,52	36,43	31,41	87,04
$M[x_t]$				$D[x_t]$				
25,49				10672,87				
$M[\Theta^T(x_t)e_1]$			$M[\Theta^T(x_t)e_2]$			$M[\Theta^T(x_t)e_3]$		
0,0534			0,5668			0,3798		

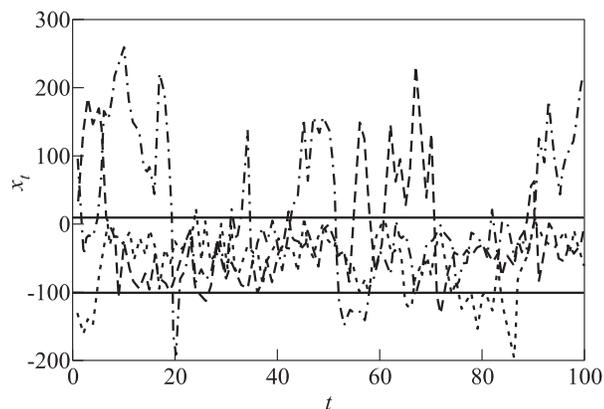


Рис. 5 Выборочные функции

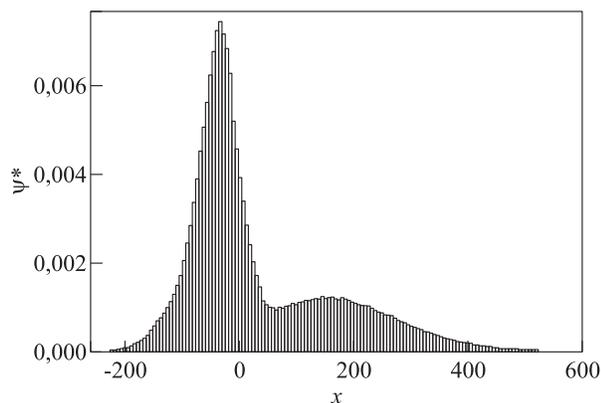


Рис. 6 Предельная плотность

рис. 6. Заметим, что по сравнению с предыдущим примером, в котором вклад всех трех режимов в предельной плотности ярко выражен, в данном случае первый режим (слабая активность) на форме распределения  $\psi^*(x)$  сказывается незначительно, что связано, очевидно, с незначительным (около 5%) временем пребывания процесса в этом режиме.

## 5 Прогнозирование

Стохастическую систему наблюдения (1)–(3) в дальнейшем предлагается использовать в постановках различных задач оптимизации распределением ресурсов, используемых программной системой (порталом) для обслуживания запросов пользователей. Для целей данной работы достаточно заметить, что для широкого класса возможных оптимизационных постановок выработка управляющего воздействия в конечном итоге будет основываться на наилучшем в среднем квадратическом прогнозе числа активных пользователей. Поэтому в качестве первого шага в решении задач оптимизации распределения ресурсов следует рассматривать задачу прогнозирования. Поскольку в дальнейшем и критерии оптимизации будут квадратичными, то для существования необходимых моментных характеристик всюду далее будем предполагать, что выполнены условия  $\mathbf{M}[v_t^2] < \infty$ ,  $\mathbf{M}[w_t^2] < \infty$ ,  $\mathbf{M}[x_0^2] < \infty$ .

В случае портала цель прогнозирования можно уточнить: требуется иметь возможность прогнозировать число поступающих для выполнения

команд. Таким образом, требуется по наблюдениям  $y_\tau$ ,  $\tau \leq t - 1$ , сформировать оценки  $\bar{y}_{t+j,t-1}$  будущих наблюдений  $y_{t+j}$ ,  $j = 0, 1, \dots$

**Теорема 2.** Пусть для системы наблюдения (1)–(3)  $\min_{1 \leq k \leq n} b_k > 0$ ,  $\sigma > 0$ . Тогда оптимальный в среднем квадратическом прогноз  $\bar{y}_{t+j,t-1}$  выхода  $y_{t+j}$  имеет вид:

$$\left. \begin{aligned} \bar{y}_{t+j,t-1} &= \sum_{k=1}^n \int_{\Delta_k} c(a_k \xi + q_k) \bar{\psi}_x(\xi, t-1, j) d\xi, \\ & \qquad \qquad \qquad j = 1, 2, \dots; \\ \bar{y}_{t,t-1} &= \sum_{k=1}^n \int_{\Delta_k} c(a_k \xi + q_k) \hat{\psi}_x(\xi, t-1) d\xi, \end{aligned} \right\} (6)$$

где прогнозирующие плотности вероятности определяются соотношениями:

$$\left. \begin{aligned} \bar{\psi}_x(x, t, j) &= \sum_{k=1}^n \frac{1}{b_k} \int_{\Delta_k} \bar{\psi}_x(\xi, t, j-1) \times \\ & \qquad \times \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) d\xi; \quad j = 1, 2, \dots; \\ \bar{\psi}_x(x, t, 0) &\triangleq \bar{\psi}_x(x, t) = \\ &= \sum_{k=1}^n \frac{1}{b_k} \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) d\xi. \end{aligned} \right\} (7)$$

**Замечание.** Соотношения (7) требуется дополнить выражением для  $\hat{\psi}_x(x, t)$  — условной плотности вероятности  $x_t$  относительно  $\mathfrak{S}_t^y$  (решение соответствующей задачи фильтрации подробно рассмотрено в [13, 14]):

$$\hat{\psi}_x(x, t) = \frac{\varphi_w((y_t - cx)/\sigma) \sum_{k=1}^n b_k^{-1} \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v((x - a_k \xi - q_k)/b_k) d\xi}{\sum_{k=1}^n b_k^{-1} \int_{\mathbf{R}^1} \varphi_w((y_t - cx)/\sigma) \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v((x - a_k \xi - q_k)/b_k) d\xi dx}.$$

**Доказательство.** Для условной функции распределения вектора  $z_t = \text{col}(x_t, y_t)$  относительно  $\mathfrak{S}_{t-1}^y$  имеем:

$$\begin{aligned} \mathbf{P}(x_t \leq x, y_t \leq y | \mathfrak{S}_{t-1}^y) &= \mathbf{P} \left( \begin{array}{l} a\Theta(x_{t-1})x_{t-1} + q\Theta(x_{t-1}) + b\Theta(x_{t-1})v_t \leq x, \\ cx_t + \sigma w_t \leq y \end{array} \middle| \mathfrak{S}_{t-1}^y \right) = \\ &= \mathbf{P} \left( v_t \leq \frac{x - a\Theta(x_{t-1}) - q\Theta(x_{t-1})}{b\Theta(x_{t-1})}, w_t \leq \frac{y - cx_t}{\sigma} \middle| \mathfrak{S}_{t-1}^y \right) = \\ &= \mathbf{P} \left( \begin{array}{l} v_t \leq \sum_{k=1}^n \frac{x - a_k x_{t-1} - q_k}{b_k} I_{\Delta_k}(x_{t-1}), \\ w_t \leq \frac{y - c(a\Theta(x_{t-1})x_{t-1} + q\Theta(x_{t-1}) + b\Theta(x_{t-1})v_t)}{\sigma} \end{array} \middle| \mathfrak{S}_{t-1}^y \right) = \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k=1}^n \mathbf{P} \left( v_t \leq \sum_{k=1}^n \frac{x - a_k x_{t-1} - q_k}{b_k}, x_{t-1} \in \Delta_k, \left| \mathfrak{S}_{t-1}^y \right. \right) = \\
 &= \sum_{k=1}^n \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \int_{-\infty}^{(x - a_k \xi - q_k)/b_k} \varphi_v(v) \int_{-\infty}^{(y - c(a_k \xi + q_k + b_k v))/\sigma} \varphi_w(w) dw dv d\xi,
 \end{aligned}$$

где учтены предположение о существовании условной плотности вероятности  $\hat{\psi}_x(x, t-1)$  и независимость величин  $v_t$  и  $w_t$  друг от друга и от  $\sigma$ -алгебры  $\mathfrak{S}_{t-1}^y$ . Последнее соотношение следует непосредственно из существования и свойств регулярных условных распределений для случайных последовательностей [20]. Дифференцируя полученное выражение с переменными верхними пределами по  $x$  и по  $y$ , получаем условную плотность вероятности  $\bar{\psi}_x(z, t)$ ,  $z = \text{col}(x, y)$ :

$$\begin{aligned}
 \bar{\psi}_z(z, t) &= \\
 &= \sum_{k=1}^n \frac{1}{\sigma b_k} \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) \times \\
 &\times \varphi_w \left( \frac{1}{\sigma} \left( y - c \left( a_k \xi + q_k + b_k \frac{x - a_k \xi - q_k}{b_k} \right) \right) \right) d\xi.
 \end{aligned}$$

Далее очевидные упрощения дают:

$$\begin{aligned}
 \bar{\psi}_z(z, t) &= \frac{1}{\sigma} \varphi_w \left( \frac{y - cx}{\sigma} \right) \times \\
 &\times \sum_{k=1}^n \frac{1}{b_k} \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) d\xi,
 \end{aligned}$$

откуда путем интегрирования по  $x$  получается маргинальная условная плотность  $\bar{\psi}_y(y, t)$ :

$$\begin{aligned}
 \bar{\psi}_y(y, t) &= \frac{1}{\sigma} \sum_{k=1}^n \frac{1}{b_k} \int_{\mathbf{R}^1} \varphi_w \left( \frac{y - cx}{\sigma} \right) \times \\
 &\times \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) d\xi dx.
 \end{aligned}$$

Отсюда получаем выражение для прогноза  $\bar{y}_{t,t-1}$  выхода  $y_t$  по наблюдениям  $y_\tau$ ,  $\tau \leq t-1$

(условное математическое ожидание  $y_t$  относительно  $\mathfrak{S}_{t-1}^y$  [21]):

$$\begin{aligned}
 \bar{y}_{t,t-1} &= \int_{\mathbf{R}^1} y \bar{\psi}_y(y, t) dy = \\
 &= \sum_{k=1}^n \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \left( \frac{1}{b_k} \int_{\mathbf{R}^1} \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) \times \right. \\
 &\quad \left. \times \left( \frac{1}{\sigma} \int_{\mathbf{R}^1} y \varphi_w \left( \frac{y - cx}{\sigma} \right) dy \right) dx \right) d\xi,
 \end{aligned}$$

откуда после непосредственного вычисления внутренних интегралов получается требуемое выражение (6) для  $\bar{y}_{t,t-1}$ . Рекуррентные соотношения для  $\hat{\psi}_x(x, t)$  получаем по формуле Байеса для плотностей вероятности [21] как отношение  $\bar{\psi}_z(z, t)$  и  $\bar{\psi}_y(y, t)$  при  $y = y_t$ .

Далее заметим, что уравнение (7) для плотности вероятности  $\bar{\psi}_x(x, t) = \bar{\psi}_x(x, t, 0)$  одношагового прогноза  $x_t$  относительно  $\mathfrak{S}_{t-1}^y$  получается из выражения для  $\bar{\psi}_z(z, t)$  путем интегрирования по  $y$ :

$$\begin{aligned}
 \bar{\psi}_x(x, t) &= \sum_{k=1}^n \frac{1}{b_k} \int_{\Delta_k} \hat{\psi}_x(\xi, t-1) \times \\
 &\times \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) \left( \frac{1}{\sigma} \int_{\mathbf{R}^1} \varphi_w \left( \frac{y - cx}{\sigma} \right) dy \right) d\xi,
 \end{aligned}$$

где внутренний интеграл по  $y$ , очевидно, равен 1.

Далее для вычисления прогноза  $\bar{y}_{t+j,t-1}$  запишем:

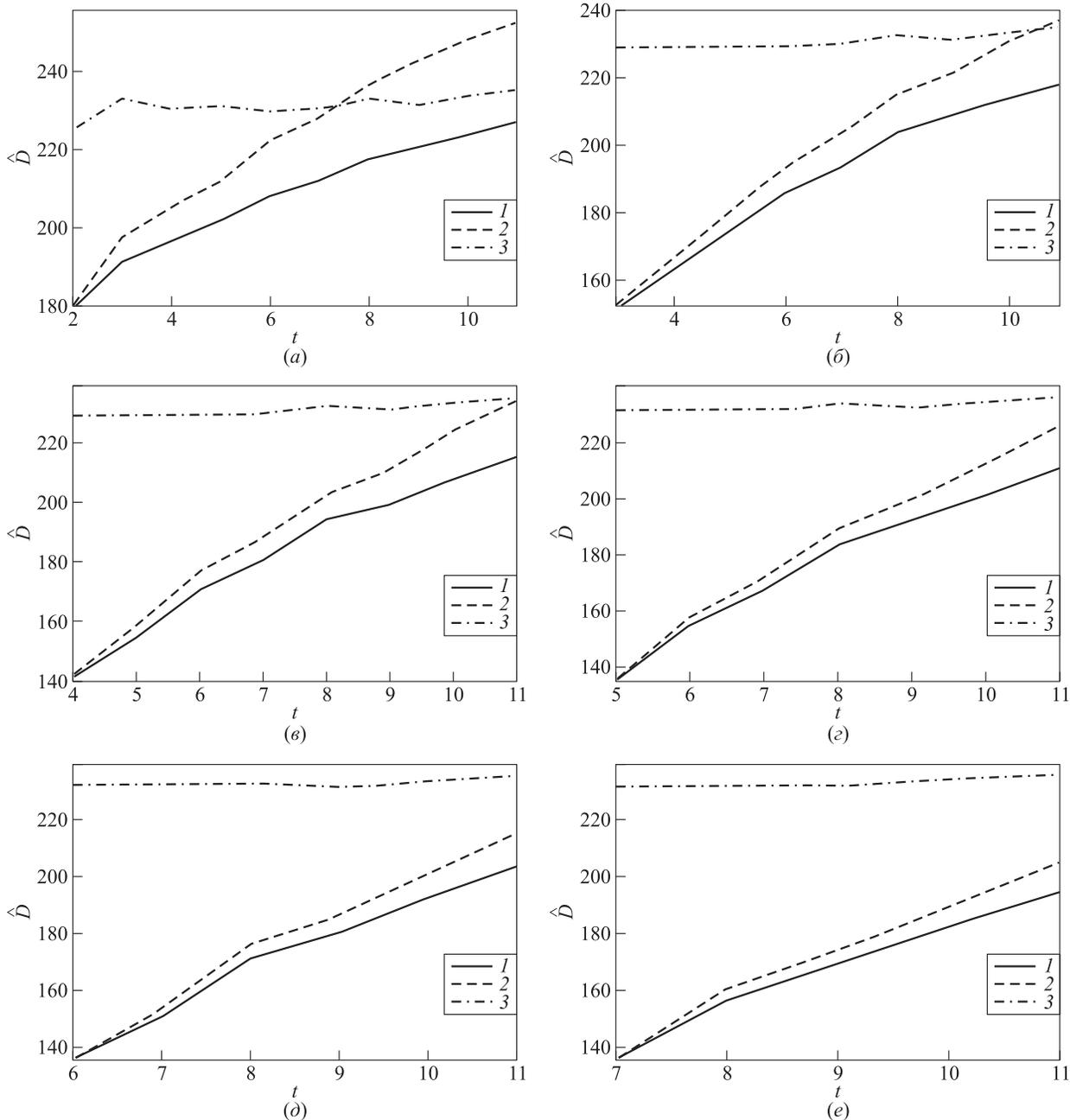
$$\mathbf{P} \left( x_{t+j} \leq x, y_{t+j} \leq y \mid \mathfrak{S}_{t-1}^y \right) = \mathbf{P} \left( \begin{array}{l} a\Theta(x_{t+j-1})x_{t+j-1} + q\Theta(x_{t+j-1}) + b\Theta(x_{t+j-1})v_{t+j} \leq x, \\ cx_{t+j} + \sigma w_{t+j} \leq y \end{array} \mid \mathfrak{S}_{t-1}^y \right)$$

и выполним с последним выражением преобразования, полностью аналогичные проделанным выше при выводе уравнения для плотности вероятности  $\bar{\psi}_z(z, t)$ .

В результате после дифференцирования получим следующее выражение для прогнозирующей плотности вероятности  $\bar{\psi}_z(z, t, j)$  вектора  $z_{t+j}$ ,  $j = 1, 2, \dots$ , относительно  $\mathcal{S}_{t-1}^y$ :

$$\bar{\psi}_z(z, t, j) = \frac{1}{\sigma} \varphi_w \left( \frac{y - cx}{\sigma} \right) \times \sum_{k=1}^n \frac{1}{b_k} \int_{\Delta_k} \bar{\psi}_k(\xi, t, j-1) \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) d\xi$$

в предположении, что уравнение (7) имеет место для  $\bar{\psi}_x(x, t, j-1)$ . Уравнение для плотности вероят-



**Рис. 7** Точность прогноза: 1 –  $\hat{D}[\bar{y}_{t+j,t-1} - y_{t+j}]$ ; 2 –  $\hat{D}[\tilde{y}_{t+j,t-1} - y_{t+j}]$ ; 3 –  $\hat{D}[y(t)]$ , где  $j = 1$  (а); 2 (б); 3 (в); 4 (г); 5 (д);  $j = 6$  (е)

ности  $\bar{\psi}_x(x, t, j)$  состояния  $x_{t+j}$  относительно  $\mathfrak{Z}_{t-1}^y$  получается отсюда путем интегрирования по  $y$ , выражение для прогнозирующей плотности вероятности  $\bar{\psi}_y(y, t, j)$  выхода  $y_{t+j}$  относительно  $\mathfrak{Z}_{t-1}^y$  — путем интегрирования по  $x$ , а сам прогноз  $\bar{y}_{t+j, t-1}$  представляется в виде:

$$\begin{aligned} \bar{y}_{t+j, t-1} &= \int_{\mathbf{R}^1} y \bar{\psi}_y(y, t, j) dy = \\ &= \sum_{k=1}^n \int_{\Delta_k} \bar{\psi}_x(\xi, t, j) \left( \frac{1}{b_k} \int_{\mathbf{R}^1} \varphi_v \left( \frac{x - a_k \xi - q_k}{b_k} \right) \times \right. \\ &\quad \left. \times \left( \frac{1}{\sigma} \int_{\mathbf{R}^1} y \varphi_w \left( \frac{y - cx}{\sigma} \right) dy \right) dx \right) d\xi. \end{aligned}$$

Непосредственное вычисление внутренних интегралов в полученном равенстве дает выражение (6) для  $\bar{y}_{t+j, t-1}$ . Теорема доказана.

**Пример 2.** Анализ точности предложенной процедуры прогнозирования проведен на модельных данных из примера 1, параметры наблюдений (1):  $c = 3,0$ ,  $\sigma = 9,1$ .

Расчеты по формулам (6) и (7) проводились для:

- 10 шагов траектории:  $t = 1, 2, \dots, 10$ ;
- точность прогнозирования анализировалась по пучку из 10 000 траекторий;
- для интегрирования в (6) и (7) использовался метод Монте-Карло, для чего моделировались 1000 случайных величин, равномерно распределенных на интервале  $[-5; 25]$ .

Рассчитывались прогнозы  $\bar{y}_{\tau+j, \tau-1}$  для  $y_{\tau+j}$ ,  $\tau = 2, 3, \dots, 11$ ,  $j = 0, 1, \dots, 11 - \tau$ . Анализировались величины ошибок оценок  $\bar{y}_{\tau+j, \tau-1}$ , для чего по пучку траекторий оценивались дисперсии ошибок  $\bar{y}_{\tau+j, \tau-1} - y_{\tau+j}$ . Для анализа качества фильтрации оценивалась также дисперсионная функция процесса  $y_t$  и дисперсия ошибок  $\tilde{y}_{\tau+j, \tau-1} - y_{\tau+j}$  для тривиальной оценки  $\tilde{y}_{\tau+j, \tau-1}$ , вычисляемой в соответствии с уравнениями (1)–(3):

$$\begin{aligned} \tilde{y}_{\tau+j, \tau-1} &= c\tilde{x}_{\tau+j, \tau-1}; \\ \tilde{x}_{\tau+j, \tau-1} &= \\ &= a\Theta(\tilde{x}_{\tau+j-1, \tau-1})\tilde{x}_{\tau+j-1, \tau-1} + q\Theta(\tilde{x}_{\tau+j-1, \tau-1}). \end{aligned}$$

Некоторые результаты расчетов приведены на рис. 7.

Из приведенных результатов видно, что точность оценивания, обеспечиваемая оптимальным прогнозом  $\bar{y}_{\tau+j, \tau-1}$ , в сравнении с точностью прогноза  $\tilde{y}_{\tau+j, \tau-1}$  на первых трех–четырёх шагах ( $j = 0, 1, 2, 3$ ) хотя и выше, но отличия сравнительно

небольшие (не более 10%). Разница в качестве оценивания возрастает с увеличением шага прогнозирования, но вместе с тем дисперсия ошибки оценки оптимального прогноза приближается к дисперсии процесса, как и следовало ожидать. Кроме того, надо отметить недостаток тривиального прогноза  $\tilde{y}_{\tau+j, \tau-1}$  — с увеличением шага прогнозирования (см. рис. 7, а–в) обеспечиваемая этим прогнозом точность становится неудовлетворительной, так как дисперсия ошибки превосходит дисперсию самого процесса.

## 6 Идентификация параметров

Заметим, что задача оценивания решена в теореме 2 в предположении, что параметры модели известны точно. На практике же можно предложить лишь их экспертную оценку [19], судить о точности которой без достоверной процедуры идентификации невозможно. В данном разделе предлагается процедура оценивания параметров  $a$ ,  $q$ ,  $b$ ,  $c$  и  $\sigma$  модели (1)–(3) на основе байесовского подхода.

Будем предполагать, что для некоторого неизвестного случайного параметра  $\gamma = \text{col}(\gamma^1, \dots, \gamma^m) \in \mathbf{R}^m$  параметры модели могут быть записаны в виде  $a = a(\gamma)$ ,  $q = q(\gamma)$ ,  $b = b(\gamma)$ ,  $c = c(\gamma)$  и  $\sigma = \sigma(\gamma)$ . Предположив, что существует и известна плотность вероятности параметра  $\gamma$ , можно распространить результат теоремы 2 на новую модель наблюдения, решив параллельно и задачу идентификации, формулируемую теперь как задача оценивания векторного случайного параметра  $\gamma$  по наблюдениям  $y_\tau$ ,  $\tau \leq t$ . Отметим, что такая модель идентификации позволяет предложить оценку для любого набора параметров модели за счет изменения размерности вектора  $\gamma$  и комбинирования в выборе функций  $a(\gamma)$ ,  $q(\gamma)$ ,  $b(\gamma)$ ,  $c(\gamma)$  и  $\sigma(\gamma)$ .

Дополним уравнения состояния (3) соотношением:

$$\gamma_t = \gamma_{t-1}, \quad t = 1, 2, \dots; \quad \gamma_0 = \gamma, \quad (8)$$

где  $\gamma = \text{col}(\gamma^1, \dots, \gamma^m)$  — случайный вектор, не зависящий от  $\{w_t\}$ ,  $\{v_t\}$  и  $x_0$  и имеющий плотность вероятности  $\psi_\gamma(\cdot)$

**Теорема 3.** Пусть для системы наблюдения (1)–(3), (8) выполнено:

- (1)  $\min_{1 \leq k \leq n} b_k(g) > 0 \forall g \in \mathbf{R}^m : \psi_\gamma(g) > 0$ ;
- (2)  $\sigma(g) > 0 \forall g \in \mathbf{R}^m : \psi_\gamma(g) > 0$ .

Тогда условная плотность вероятности  $\hat{\psi}_{x, \gamma}(x, g, t)$  существует и описывается уравнениями:

$$\begin{aligned} \hat{\psi}_{x,\gamma}(x, g, t) &= \\ &= \frac{\bar{\psi}_{z,\gamma}(z, g, t)}{\int_{\mathbf{R}^{m+1}} \bar{\psi}_{z,\gamma}(z, g, t) dx dg}, \quad z = \text{col}(x, y_t); \quad (9) \end{aligned}$$

$$\begin{aligned} \bar{\psi}_{z,\gamma}(z, g, t) &= \sum_{k=1}^n \frac{1}{\sigma(g)b_k(g)} \varphi_w \left( \frac{y - c(g)x}{\sigma(g)} \right) \times \\ &\times \int_{\Delta_k} \hat{\psi}_{x,\gamma}(\xi, g, t - 1) \times \\ &\times \varphi_v \left( \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) d\xi. \quad (10) \end{aligned}$$

**Доказательство.** Для условной функции распределения вектора  $\text{col}(x_t, y_t, \gamma_t)$  относительно  $\mathfrak{Z}_{t-1}^y$ , проделав преобразования, полностью аналогичные выполненным при доказательстве теоремы 2, получаем:

$$\begin{aligned} P(x_t \leq x, y_t \leq y, \gamma_t \leq g | \mathfrak{Z}_{t-1}^y) &= \\ &= \sum_{k=1}^n \int_{\Delta_k} \int_{-\infty}^{g_1} \cdots \int_{-\infty}^{g_m} \hat{\psi}_x(\xi, \mu, t - 1) \times \\ &\times \int_{-\infty}^{(x - a_k(\mu)\xi - q_k(\mu))/b_k(\mu)} \varphi_v(v) \times \\ &\times \int_{-\infty}^{(y - c(\mu)(a_k(\mu)\xi + q_k(\mu) + b_k(\mu)v))/\sigma(\mu)} \varphi_w(w) dw d\mu_1 \cdots d\mu_m d\xi \end{aligned}$$

в предположении существования условной плотности вероятности  $\hat{\psi}_{x,\gamma}(x, g, t - 1)$ . Отсюда, дифференцируя интегралы с переменными верхними пределами по  $x$ , по  $y$  и по  $g$ , получаем, что существует условная плотность вероятности  $\bar{\psi}_{z,\gamma}(z, g, t)$  следующего вида:

$$\begin{aligned} \bar{\psi}_{z,\gamma}(z, g, t) &= \sum_{k=1}^n \frac{1}{\sigma(g)b_k(g)} \times \\ &\times \int_{\Delta_k} \hat{\psi}_{x,\gamma}(\xi, g, t - 1) \varphi_v \left( \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) \times \\ &\times \varphi_w \left( \frac{1}{\sigma(g)} \left( y - c(g) \left( a_k(g)\xi + q_k(g) + \right. \right. \right. \\ &\left. \left. \left. + b_k(g) \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) \right) \right) d\xi. \end{aligned}$$

Отсюда уравнение (10) получается после очевидных преобразований, а также путем интегрирования по  $x$  и по  $g$  получается маргинальная условная плотность  $\bar{\psi}_y(y, t)$ . Окончательно соотношение (9) получаем по формуле Байеса для плотностей веро-

ятности [21] как отношение  $\bar{\psi}_{z,\gamma}(z, g, t)$  и  $\bar{\psi}_y(y, t)$ , где  $y = y_t$ . Теорема доказана.

**Замечание 1.** В соотношениях (9) и (10) используется начальное условие  $\hat{\psi}_{x,\gamma}(x, g, 0) = \psi_0(x)\psi_\gamma(g)$ . Эти соотношения полностью описывают рекуррентную процедуру пересчета искомой условной плотности вероятности вектора  $\text{col}(x_t, \gamma)$  относительно  $\mathfrak{Z}_{t-1}^y$  — решения задачи оптимальной фильтрации.

**Замечание 2.** Уравнения (9) и (10) могут быть преобразованы к виду формулы Байеса [21], если использовать в них ненормированные условные плотности вероятности [22]. Пусть  $t = 1$ . Тогда с учетом начального условия из (10) получаем:

$$\begin{aligned} \bar{\psi}_{z,\gamma}(z, g, 1) &= \sum_{k=1}^n \frac{1}{\sigma(g)b_k(g)} \varphi_w \left( \frac{y_1 - c(g)x}{\sigma(g)} \right) \times \\ &\times \psi_\gamma(g) \int_{\Delta_k} \psi_0(\xi) \varphi_v \left( \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) d\xi. \end{aligned}$$

Обозначим через  $\tilde{\psi}_x(x, 1|g)$  ненормированную условную плотность вероятности  $x_t, t = 1$ , относительно  $\mathfrak{Z}_1^y$  и условия  $\gamma_1 = g$ :

$$\begin{aligned} \tilde{\psi}_x(x, 1|g) &= \sum_{k=1}^n \frac{1}{\sigma(g)b_k(g)} \varphi_w \left( \frac{y_1 - c(g)x}{\sigma(g)} \right) \times \\ &\times \int_{\Delta_k} \psi_0(\xi) \varphi_v \left( \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) d\xi. \end{aligned}$$

Используя  $\tilde{\psi}_x(x, 1|g)$  в качестве начального условия, запишем следующие рекуррентные уравнения:

$$\begin{aligned} \tilde{\psi}_x(x, t|g) &= \sum_{k=1}^n \frac{1}{\sigma(g)b_k(g)} \varphi_w \left( \frac{y - c(g)x}{\sigma(g)} \right) \times \\ &\times \int_{\Delta_k} \tilde{\psi}_x(\xi, t - 1|g) \varphi_v \left( \frac{x - a_k(g)\xi - q_k(g)}{b_k(g)} \right) d\xi. \quad (11) \end{aligned}$$

Тогда (9) приобретает вид:

$$\hat{\psi}_{x,\gamma}(x, g, t) = \frac{\tilde{\psi}_x(x, t|g)\psi_\gamma(g)}{\int_{\mathbf{R}^{m+1}} \tilde{\psi}_x(x, t|g)\psi_\gamma(g) dx dg}. \quad (12)$$

При практическом применении алгоритма идентификации использование уравнений (11) и (12) вместо (9) и (10) облегчает численную реализацию за счет сокращения вычислительных затрат.

**Пример 3.** Предложенную в теореме 3 процедуру идентификации применим к реальным данным, а именно: воспользуемся экспертной моделью из примера 1, предложенной для наблюдений за активностью пользователей портала РАН [www.ras.ru](http://www.ras.ru).

Экспертные оценки параметров уравнения (3) из табл. 1 дополним предположением о параметрах уравнения (1):  $c = 1, \sigma = 1$ .

Рассматривались четыре варианта моделей идентификации параметров:

- (1) в первом случае неизвестными считались три параметра  $a_1, q_1$  и  $b_1$ , определяющих динамику для первого уровня пользовательской активности;
- (2) во втором — выполнялось оценивание параметров  $a_2, q_2$  и  $b_2$ , определяющих динамику для второго уровня пользовательской активности;
- (3) в третьем — выполнялось оценивание параметров  $a_3, q_3$  и  $b_3$ , определяющих динамику для третьего уровня пользовательской активности;
- (4) в четвертом — выполнялось оценивание параметров  $c$  и  $\sigma$ , определяющих уравнения наблюдений (1).

Полученные результаты в первых трех случаях содержательно идентичны, поэтому далее рассматриваются только первый и четвертый варианты. В каждом из них априорные распределения подлежащих идентификации параметров предполагались равномерными, параметры приведены в табл. 3 и 4. Расчеты выполнялись с использованием ненормированных распределений по формулам (11) и (12) для имеющихся 5000 наблюдений. Для интегрирования методом Монте-Карло моделировались 1000 случайных величин для интегрирования по  $d\xi(dx)$  (пределы интегрирования заданы интер-

**Таблица 3** Параметры идентификации — вариант 1

Идентифицируемые параметры	Оценка эксперта	Априорное распределение
$a_1$	0,361	$R[0,1; 0,9]$
$q_1$	-67,784	$R[-100,0; -50,0]$
$b_1$	36,432	$R[0,0; 50,0]$

**Таблица 4** Параметры идентификации — вариант 4

Идентифицируемые параметры	Оценка эксперта	Априорное распределение
$c$	1,0	$R[0,1; 5,1]$
$\sigma$	1,0	$R[0,1; 5,1]$

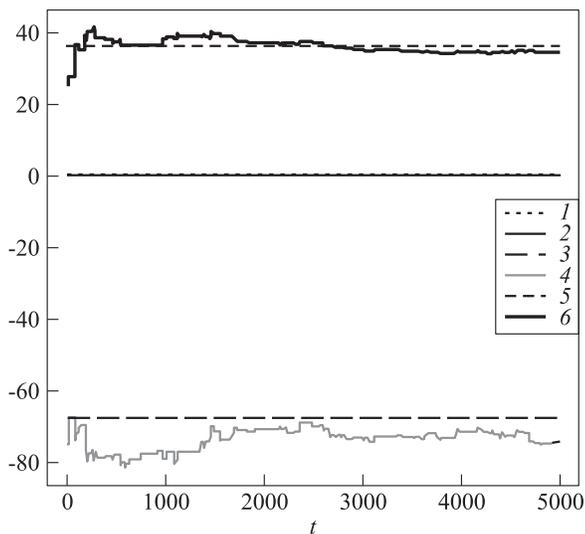
валом  $[-1000,0; 1000,0]$  и 10 000 случайных величин для интегрирования по  $dg$ .

На рис. 8 и 9 приведены результаты идентификации параметров.

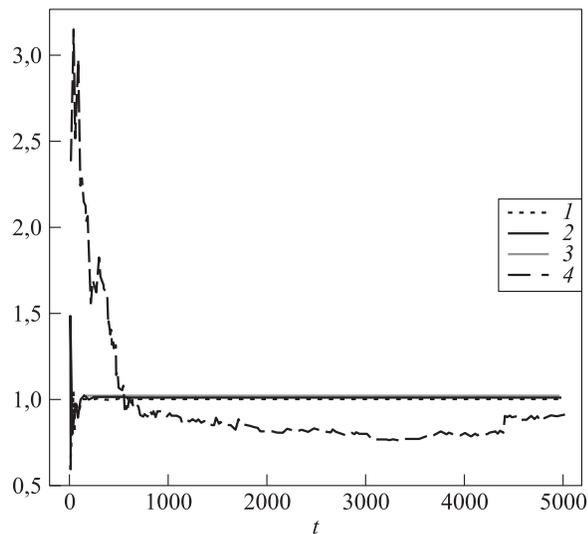
Из полученных результатов видно, что хотя в целом оценки, данные экспертом, достаточно обоснованы, но определенную коррекцию процедура идентификации все-таки вносит. Кроме того, проведенные расчеты свидетельствуют и о практической работоспособности предложенных методов оценивания, и об адекватности используемой математической модели реально наблюдаемым данным.

## 7 Заключение

Представлена модель стохастической динамической системы наблюдения, описывающей процессы изменения активности пользователей, выполняющих запросы к некоторой информационной системе. К ключевым свойствам описываемого



**Рис. 8** Результаты идентификации (вариант 1): 1 —  $a_1$ ; 2 —  $\hat{a}_1(t)$ ; 3 —  $q_1$ ; 4 —  $\hat{q}_1(t)$ ; 5 —  $b_1$ ; 6 —  $\hat{b}_1(t)$



**Рис. 9** Результаты идентификации (вариант 4): 1 —  $c$ ; 2 —  $\hat{c}(t)$ ; 3 —  $\sigma$ ; 4 —  $\hat{\sigma}(t)$

данной моделью процесса предложено отнести следующие предположения:

- наличие нескольких характерных режимов, в рамках которых динамика изучаемого показателя описывается простой линейной моделью;
- скачкообразное изменение характера показателя активности при достижении им заданных границ;
- наличие косвенных линейных наблюдений за показателем.

Случайный процесс, описывающий показатель активности, при практически выполнимых условиях обладает свойством эргодичности, т. е. существует предельное распределение, которое, с одной стороны, отражает свойственный процессу кусочный характер, с другой — гарантирует наличие стационарного режима распределения вероятностей, что очень существенно в практическом смысле.

Применительно к предложенной модели решены две базовые задачи анализа — прогнозирования состояния и идентификации параметров. Работоспособность алгоритмов оценивания проиллюстрирована на численных экспериментах, в том числе с использованием реальных данных, полученных с сайта РАН [www.ras.ru](http://www.ras.ru).

Сами по себе процедуры анализа могут быть использованы лишь в простейших задачах оптимизации работы самой информационной системы, ее программного обеспечения. Однако их основная ценность заключается в возможности применения в полноценных постановках оптимизационных задач. Такие постановки возникают, например, в задачах управления вычислительными ресурсами, используемыми информационной системой при реализации собственной функциональности. Некоторые из решенных задач будут описаны в следующих работах.

## Литература

1. *Микадзе И. С., Хочолава В. В.* Исследование длины очереди в однолинейной СМО с ненадежным прибором // Автоматика и телемеханика, 2005. № 1. С. 72–81.
2. *Печинкин А. В., Соколов И. А., Чаплыгин В. В.* Многолинейная система массового обслуживания с конечным накопителем и ненадежными приборами // Информатика и её применения, 2007. Т. 1. Вып. 1. С. 27–39.
3. *Jacobson V.* Congestion avoidance and control // ACM SIGCOMM'88, 1988.
4. *Allman M., Paxson V., Stevens W.* TCP congestion control // RFC 2581, 1999. <http://www.ietf.org/rfc/rfc2581.txt>.
5. *Kelly F. P., Maulloo A., Tan D.* Rate control in communication networks: Shadow prices, proportional fairness and stability // J. Operational Res. Soc., 1998. Vol. 49. P. 237–252.
6. *Low S. H., Paganini F., Doyle J. C.* Internet congestion control // IEEE Control Systems Magazine, 2002. Vol. 22. No. 1. P. 28–43.
7. *Миллер Б. М., Миллер Г. Б., Семенухин К. В.* Методы синтеза оптимального управления марковским процессом с конечным множеством состояний при наличии ограничений // Автоматика и телемеханика, 2011. № 2. С. 111–130.
8. *Батракова Д. А., Королев В. Ю., Шоргин С. Я.* Новый метод вероятностно-статистического анализа информационных потоков в телекоммуникационных сетях // Информатика и её применения, 2007. Т. 1. Вып. 1. С. 40–53.
9. *Дейт К. Дж.* Введение в системы баз данных. — М.: Вильямс, 2000.
10. *Гордеев А. В., Молчанов А. Ю.* Системное программное обеспечение. — СПб.: Питер, 2001.
11. Информационный веб-портал. Свидетельство об официальной регистрации программы для ЭВМ № 2005612992. Зарегистрировано в Реестре программ для ЭВМ 18.11.2005 г.
12. *Босов А. В.* Порталы в системах органов государственной власти // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 24–34.
13. *Босов А. В., Борисов А. В.* Модели оптимизации функционирования Информационного web-портала // Технологии программирования и хранения данных: Труды ИСА РАН, 2009. Т. 45. С. 107–133.
14. *Босов А. В.* Моделирование и оптимизация процессов функционирования Информационного web-портала // Программирование, 2009. № 6. С. 53–66.
15. Slashdot effect. [http://en.wikipedia.org/wiki/Slashdot\\_effect](http://en.wikipedia.org/wiki/Slashdot_effect).
16. *Tong H.* Non-linear time series: A dynamical approach. — Oxford: Clarendon Press, 1990.
17. *Борисов А. В., Босов А. В., Стефанович А. И.* Оптимальное оценивание показателей функционирования информационного web-портала // Автоматика и телемеханика, 2010. № 3. С. 16–33.
18. *Bhattacharya R. N., Lee C.* Ergodicity of nonlinear first order autoregressive models // J. Theor. Prob., 1995. Vol. 8. No. 1. P. 207–219.
19. *Иванов А. В.* Математические модели базовых процессов функционирования информационного Web-портала // Системы и средства информатики, 2010. Вып. 20. № 1. С. 106–132.
20. *Гихман И. И., Скороход А. В.* Теория случайных процессов. Т. 1. — М.: Наука, 1971.
21. *Ширяев А. Н.* Вероятность. — М.: Наука, 1989.
22. *Aggoun L., Elliott R.* Measure theory and filtering. Introduction and applications. — Cambridge: Cambridge University Press, 2004.

# РЕШЕНИЕ ПРОБЛЕМ ВЗАИМОДЕЙСТВИЯ С СУБД В КРОССПЛАТФОРМЕННОЙ БИБЛИОТЕКЕ EFFIDB

А. В. Янушко<sup>1</sup>, А. В. Бабанин<sup>2</sup>, О. А. Кузнецова<sup>3</sup>, С. В. Петрушенко<sup>4</sup>

**Аннотация:** Статья затрагивает проблемы унифицированного взаимодействия с разнотипными СУБД в различных программных средах. Рассмотрены существующие разработки в данной области, проанализированы их достоинства и недостатки. Перечислены требования к кроссплатформенному инструменту взаимодействия прикладного кода C++ с СУБД. Предложено решение, реализованное в виде динамической библиотеки. Библиотека предоставляет специализированные классы для каждого из понятий реляционных баз данных: собственно база данных, соединение, таблица, средства манипулирования данными и т. д. Проанализированы границы применимости предложенных решений и практика использования библиотеки в реальных проектах. Приведены примеры кода.

**Ключевые слова:** СУБД; C++; библиотека взаимодействия; кроссплатформенность

## 1 Введение

Современные информационные системы автоматизации деятельности предприятий представляют собой многокомпонентные программные комплексы. С прикладной точки зрения в любой информационной системе можно выделить три логических компоненты: слой отображения данных, слой обработки данных и слой хранения данных.

Данная работа посвящена фундаментальной составляющей информационной системы, а именно системам управления базами данных и средствам взаимодействия серверов приложений с различными СУБД.

Несмотря на то что базы данных и СУБД составляют обширную тему исследований и разработок, в этой области остается много нерешенных проблем.

В статье рассматривается проблема организации работы приложений, написанных на C++, с множеством разнотипных СУБД. Предлагаемое решение заключается в создании промежуточного звена доступа к различным СУБД, при этом разработчик не должен задумываться о наличии такого промежуточного слоя и желательно, чтобы взаимодействие со средством доступа происходило стандартным образом, с использованием распространенных технологий.

## 2 Универсальная библиотека взаимодействия с СУБД — EffiDB

C++ библиотека EffiDB представляет собой набор интерфейсов, позволяющий взаимодействовать единым образом с системами управления базами данных. Целью создания библиотеки ставилось предоставление унифицированного доступа к различным СУБД и изоляция прикладного программиста от какого бы то ни было непосредственного низкоуровневого взаимодействия с базой данных.

Библиотека EffiDB доступна для использования в среде Linux (компилятор gcc версий 4.2–4.5), Windows (компилятор MS VC++ 2010), MacOS (компилятор gcc 4.2.1 и выше). Авторам неизвестно никаких препятствий для сборки библиотеки иными компиляторами, поддерживающими стандарт C++ ISO/IEC 14882 1998 г., а также препятствий для использования библиотеки под управлением иных POSIX-совместимых ОС. Однако никаких работ в этом направлении авторами не проводилось.

Специализация работы с каждой конкретной СУБД заключена в отдельных библиотеках, которые далее называются драйверами. Библиотека EffiDB управляет соединениями с базой данных, формирует и исполняет SQL-запросы, интерпретирует полученные результаты. Она также предостав-

<sup>1</sup>АСофт, yan@asoft.ru

<sup>2</sup>Всероссийский научный-исследовательский институт проблем вычислительной техники и информатизации, ababanin@pvti.ru

<sup>3</sup>АСофт, ok@asoft.ru

<sup>4</sup>АСофт, op@asoft.ru

ляет механизмы кэширования результатов запросов и автоматического менеджмента памяти, что повышает надежность и производительность кода.

В настоящий момент доступны драйверы для работы со следующими СУБД: MySQL, MSSQL, Oracle и SQLite.

В случае когда отдельные модули системы выполняют кооперированные действия с одним и тем же объектом базы данных, требуется, чтобы все операции совершались в рамках одной транзакции. Специализированный драйвер Invariant инициирует отдельный процесс исключительно для коммуникации с базой данных, что позволяет объединить все обращения к базе данных в одном соединении. Драйвер Invariant не имеет специфики СУБД и может быть использован с любой из поддерживаемых СУБД.

В зависимости от поставленных задач и степени подготовки разработчика можно выделить 4 уровня использования EffiDB:

1. **Создание программных приложений со статической структурой базы данных.** Простейший вариант использования библиотеки в приложениях с заранее определенной схемой базы данных, которая не изменяется во время работы приложения.
2. **Создание программных приложений с динамической структурой базы данных.** Предусматривается возможность изменения или расширения схемы базы данных в процессе эксплуатации приложения.
3. **Разработка адаптеров для типов данных.** Библиотека EffiDB не использует никаких «внутренних» типов данных для хранения параметров запросов и результатов выборок из базы данных. Вместо этого используется подход «адаптеров», которые позволяют применять в EffiDB любые пользовательские типы данных. Поставка EffiDB включает в себя адаптеры для популярных (в том числе встроенных в C++) типов данных, таких как int, string, double и т. п. Механизм адаптации любого другого типа данных в точности такой же. Работа с пользовательским типом данных, для которого разработан адаптер, не отличается от работы с типами, адаптеры для которых включены в дистрибутив библиотеки.
4. **Разработка драйверов для поддержки дополнительных СУБД.** Библиотека EffiDB не ограничивает пользователя перечисленными ранее СУБД. Открытый API (application programming interface) библиотеки предоставляет техническую возможность разработать собственный

драйвер для произвольной СУБД. Как и в случае с пользовательскими типами, вновь разработанный дополнительный драйвер СУБД предоставляет все те же возможности, что и драйверы, входящие в поставку.

Хотя это и может показаться парадоксальным на первый взгляд, дублирование информации приводит к ее потере. При наличии нескольких реплик данных возникают проблемы синхронизации, определения степени актуальности каждого отдельного экземпляра и, как следствие, появляется риск потери информации. Достаточно распространена ситуация, когда необходимо обрабатывать агрегированную информацию из нескольких баз данных (в общем случае с использованием разных СУБД). Копирование данных в единое хранилище с целью их консолидации и дальнейшей обработки представляется неэффективным в силу вышеупомянутых причин.

Библиотека EffiDB позволяет одновременно обращаться к нескольким базам данных под управлением разных СУБД, и необходимость в копировании информации не возникает.

Возможность одновременной работы с разными СУБД оказывается полезной и в другой хорошо известной практической задаче — задаче миграции данных из одной информационной системы в другую. Благодаря EffiDB процедура миграции может выполняться напрямую без промежуточных шагов.

Библиотека EffiDB предназначена для построения сложных многокомпонентных информационных систем. В таких системах бизнес-логика обычно распределяется между несколькими приложениями (модулями), которые совместно работают с общей базой данных. Хрестоматийным примером, иллюстрирующим такую ситуацию, можно считать перевод денег с одного счета на другой: один модуль снимает деньги со счета-донора и вызывает второй модуль, который, в свою очередь, зачисляет средства на счет-акцептор.

Таким образом, одно действие (перевод средств) может быть представлено несколькими отдельными, но тем не менее зависимыми друг от друга операциями приложений над объектами одной базы данных. Реализация такого действия в информационной системе требует, чтобы каждый участвующий в процессе модуль последовательно совершал необходимую операцию и передавал управление следующему модулю до тех пор, пока все операции не будут выполнены. Обеспечение транзакционной целостности базы данных означает, что все операции (снятие и зачисление средств в вышеприведенном примере) должны производиться в одной транзакции. Это было бы невозможно, ес-

ли бы каждый модуль действовал изолированно и устанавливал собственное соединение с базой данных.

Для устранения этого противоречия в EffiDB используется специализированный драйвер Invariant. Если в индивидуальных процессах приложений требуется взаимодействие с базой данных, драйвер создает специальный процесс для коммуникации с базой данных и устанавливает единое соединение с ней. В результате логически связанные действия разных приложений выполняются в рамках одной транзакции.

### 3 Аналогичные разработки

В настоящий момент разработчики могут выбирать среди достаточно большого количества ин-

струментов доступа к данным, хранящимся в реляционных СУБД. Можно выделить низкоуровневые средства доступа, наиболее популярные из которых ODBC (Open DataBase Connectivity) (для языков C/C++) и JDBC (для Java), а также высокоуровневые — так называемые средства объектно-реляционного отображения, весьма представительный список которых представлен в Wikipedia [1].

Некоторые популярные инструменты сведены в табл. 1, где указаны поддерживаемые операционные системы, СУБД, языки программирования. Также в таблице содержатся наиболее важные с точки зрения разработчика прикладного программного обеспечения (ПО) объективные параметры этих инструментов. Такие важнейшие параметры, как лаконичность и простота использования языка запросов, сложно сравнивать объективно, поэтому они исключены из рассмотрения.

**Таблица 1** Инструменты доступа к данным, хранящимся в РСУБД

Продукт	Операционная система	СУБД	Язык	Уровень изоляции	Контроль синтаксиса запросов	Типы данных
ODBC	Windows, Unix	Все промышленные РСУБД	C	Низкий	В момент исполнения	Простые ODBC специфичные
JDBC	Windows, Unix	Все промышленные РСУБД	Java	Низкий	В момент исполнения	Простые Java специфичные
LINQ	Windows	MS SQL <sup>(*)</sup>	C#, VB.NET и другие MSIL-совместимые .NET языки	Высокий. Требуется регенерация модели данных прикладного ПО	На этапе разработки	Простые .NET специфичные
Hibernate	Windows, Unix	Все промышленные РСУБД	Java	Высокий. Допускает переключение между СУБД «на лету»	В момент исполнения	Простые Java специфичные
SOCI	Unix	Oracle, PostgreSQL, MySQL	C++	Низкий	В момент исполнения	Стандартные C++ типы
ODB	Windows, Unix	MySQL	C++	Высокий <sup>(**)</sup> . Требуется перекомпиляция	На этапе разработки	Любые, определяемые пользователем
EffiDB	Windows, Unix	MSSQL, Oracle, MySQL, SQLite <sup>(***)</sup>	C++	Высокий. Допускает переключение между СУБД «на лету»	На этапе разработки	Любые, определяемые пользователем

<sup>(\*)</sup> Поддержка Oracle, MySQL, PostgreSQL, SQLite возможна при покупке дополнительных инструментов сторонних разработчиков.

<sup>(\*\*)</sup> Поддерживается очень ограниченный набор возможностей SQL. В частности, недоступно объединение таблиц (JOIN).

<sup>(\*\*\*)</sup> Поддержка других РСУБД возможна в режиме ODBC-совместимости.

Под термином «уровень изоляции» авторы понимают степень изоляции прикладного кода от специфики целевой СУБД, в том числе в разрезе поддержки того или иного диалекта SQL (Structured Query Language). Низкий уровень изоляции подразумевает необходимость внесения изменений в код прикладного ПО при смене СУБД, высокий уровень изоляции гарантирует переносимость кода прикладного ПО между СУБД без каких-либо изменений. Однако даже с высоким уровнем изоляции адаптация под новую СУБД может требовать регенерации, перекомпиляции или статической перелинковки кода, а может осуществляться динамически во время работы программы. Контроль синтаксиса запросов в разных системах может осуществляться либо на этапе разработки прикладного ПО (во время компиляции/генерации кода) средствами самого инструмента, либо в момент исполнения SQL-запроса средствами СУБД. Первый подход существенно удешевляет разработку ПО за счет снижения затрат на тестирование (или повышает качество ПО при равных затратах). Набор поддерживаемых типов данных практически не влияет на выбор инструментов доступа при разработке прикладного ПО с нуля, однако в реальной жизни при наличии унаследованного кода или при желании разработчика использовать какие-нибудь дополнительные библиотеки/модули и т. п. вопрос об используемых типах данных может оказаться критически важным.

Анализ присутствующих на рынке инструментов доступа к реляционным СУБД (РСУБД) показывает, что наиболее продвинутым продуктом является LINQ (Language Integrated Query) компании Microsoft — встроенное в .NET средство доступа к данным. Полная изоляция от SQL, проверка синтаксиса на этапе компиляции, лаконичность и удобочитаемость синтаксиса, наличие средств генерации языковых объектов на основании существующей базы данных делают LINQ лучшим на данный момент средством доступа к реляционным данным с точки зрения удобства использования. Однако LINQ доступен только как часть платформы .NET для проектов, работающих под управлением ОС Windows, и нацелен на работу с СУБД MS SQL, что существенно ограничивает область применимости этого инструмента и требует достаточно больших материальных затрат — хотя сам LINQ и является бесплатным, для его использования требуется как минимум покупка лицензий Windows и MS SQL, а также зачастую средств разработки (Microsoft Visual Studio и т. п.).

Другие инструменты, хотя и лишены различных недостатков LINQ, но предоставляют в целом меньшее, иногда существенно меньшее удобство при

разработке прикладного ПО. Кроме того, большинство существующих инструментов ориентированы на интерпретируемые языки (PHP, Python, Ruby и т. п.) или на языки, компилируемые в промежуточный код (Java, .NET), в то время как разработчики, использующие наиболее популярные компилируемые языки C++ и C, вынуждены пользоваться либо очень неудобными низкоуровневыми библиотеками ODBC (существует несколько реализаций), либо такими функционально ограниченными инструментами, как SOCI, ODB и т. п. При этом, по данным крупнейшей Интернет-площадки свободных разработок sourceforge.net, проекты на C/C++ занимают лидирующие позиции — на их долю приходится более четверти всех проектов.

Такое положение вещей и побудило авторов к разработке собственного инструмента — многоплатформенного, поддерживающего разные СУБД, изолирующего прикладного программиста от специфики СУБД и SQL диалектов, способного работать с унаследованными приложениями/библиотеками, разработанными на C++ и C, обладающего компактным и легко читаемым синтаксисом.

## 4 Цели создания EffiDB

- **Переносимость кода между СУБД.** В IT-индустрии часто приходится разрабатывать большое количество приложений для разных клиентов. Как следствие, необходимо поддерживать разнообразные конфигурации клиентского оборудования и иногда переносить унаследованное ПО на новые платформы [2, 3]. Вообще говоря, это ПО должно быть в состоянии работать на разных СУБД в зависимости от требований клиента — представляется нецелесообразным всякий раз, когда новому клиенту требуется очередная СУБД, заниматься ее поддержкой и обучать собственных разработчиков работе с этой СУБД. Библиотека EffiDB обеспечивает независимость кода программного приложения от СУБД. Библиотека покрывает большую часть функциональности, предоставляемой различными СУБД, и гарантирует, что программное приложение будет работать одинаковым образом на всех поддерживаемых СУБД.
- **Проверка запросов на момент компиляции.** Приложения, взаимодействующие с базой данных, генерируют SQL-запросы, которые затем интерпретируются и исполняются на СУБД. Так как SQL и код на C++, вообще говоря, никак не связаны между собой, построение SQL-за-

просов подвержено ошибкам, причем обычно синтаксические ошибки, допущенные при создании запроса, обнаруживаются довольно поздно — только на этапе тестирования приложения. Библиотека EffiDB строит запросы автоматически и производит проверку в процессе компиляции — это гарантирует отсутствие синтаксических ошибок в SQL-запросах.

- **Динамическое подключение базы данных.** В некоторых случаях приложение должно работать с несколькими СУБД — одновременно либо с той или иной СУБД, определяемой конкретной инсталляцией. Так, например, такое требование актуально при интеграции унаследованных информационных систем, разработка которых велась независимо разными подразделениями и производителями [3]. Динамическое подключение базы данных, реализованное в EffiDB, позволяет «на лету» переключаться между СУБД: если изначально предполагалось использовать приложение с MySQL, то нет необходимости перекомпилировать его, например, под Oracle — для того чтобы запустить приложение на конкретной СУБД, достаточно подгрузить соответствующую динамическую библиотеку. Можно также менять СУБД в процессе работы приложения и использовать несколько СУБД одновременно.
- **Лаконичный синтаксис.** Чрезвычайно важным элементом разработки является лаконичность, понятность и эргономичность кода, который приходится писать прикладным разработчикам. Любая недоработка в этой области приводит к снижению поддерживаемости и увеличивает количество ошибок. Благодаря удобному синтаксису EffiDB объем текста, который требуется написать на C++ для создания SQL-запроса, практически не превосходит размер самого SQL-запроса.
- **Простота построения запроса из одной таблицы.** На практике большинство SQL-запросов представляют собой запросы по первичному ключу в одной таблице. Этот простейший и часто встречающийся частный случай имеет смысл рассматривать отдельно. Если не выделять этот важный сценарий и обрабатывать его обычным образом, то такой подход порождает определенные неудобства. Во-первых, таким образом упускается возможность повысить производительность. Во-вторых, часто разработчики поддаются соблазну объединить несколько запросов в один и сконструировать сложный запрос «на все случаи жизни» — такие запросы могут снижать производительность очень сильно и,

что хуже, непредсказуемым образом. Библиотека EffiDB предоставляет специальный синтаксис для запросов из одной таблицы — это очень простой способ составления запроса, позволяющий обходиться минимумом кода. Фактически код C++ в этом случае даже короче, чем конечный SQL-запрос. Программисту не надо составлять сложные универсальные запросы — часто проще писать именно тот запрос, который нужен в данном конкретном случае.

- **Автоматическое управление транзакциями.** Управление транзакциями в СУБД представляет собой нетривиальную задачу и, следовательно, порождает серьезные проблемы в проекте разработки программного продукта [4, 5]. Если фиксировать изменения (Commit changes) слишком часто, данные в БД могут потерять целостность. Если же фиксировать изменения слишком редко, то снижается производительность системы и становятся возможными клинчи (deadlocks). При создании библиотеки EffiDB авторы исходили из того соображения, что прикладной программист не должен заниматься управлением транзакциями, т. е. такими операциями, как управление соединениями, явная фиксация и откат транзакции. Библиотека обеспечивает автоматическое управление транзакциями, и прикладной программист не обязан думать о транзакциях. Тем не менее, если возникает необходимость явным образом управлять транзакциями, разработчик имеет возможность это делать.
- **Автоматическая блокировка.** Довольно часто несколько транзакций пользуются одними и теми же данными, что приводит к хорошо известным проблемам управления параллельным выполнением операций. Необходимо соблюдать разумный баланс между производительностью и изолированностью транзакций. Один из двух крайних подходов, который повышает производительность и пренебрегает изолированностью, приводит к потере данных. Противоположная стратегия, концентрирующаяся на изолированности транзакций, вызывает взаимные блокировки и снижает производительность [4]. Авторы полагают, что целостность данных является жизненно важным свойством информационной системы, и по умолчанию предпочитают изолированность производительности. Для того же, чтобы обеспечить приемлемую производительность и уменьшить вероятность клинчей, в EffiDB реализована поддержка автоматической блокировки. Различаются две категории транзакций:

«безопасные» и быстрые транзакции чтения и «потенциально опасные» транзакции записи. Транзакции записи блокируют все данные, с которыми работают (путем использования SELECT FOR UPDATE), что гарантирует изолированность. В тех случаях, когда такая блокировка является избыточной, ее можно отключить.

- **Поддержка произвольных типов данных.** Как правило, унаследованное ПО использует свои собственные типы данных, которые в общем случае не совпадают с типами данных СУБД. Если проект разработки ПО предполагает интеграцию такого рода приложений с одной или несколькими СУБД, необходимо преобразование типов данных. Библиотека EffiDB обеспечивает поддержку произвольных типов данных и предоставляет возможность разработки «адаптера» для собственных типов данных. Ниже приведен список типов данных, «адаптеры» для которых входят в дистрибутив библиотеки:

- std::string;
- double;
- int32\_t;
- int64\_t;
- Int — синоним для типа boost::optional<int32\_t>;
- Int64 — синоним для типа boost::optional<int64\_t>;
- Str — синоним для типа boost::optional<string>;
- Time — синоним для типа boost::posix\_time::ptime;
- Effi::Blob для больших блоков бинарных данных;
- Effi::GeoData для больших блоков бинарных данных, используемых в геоинформационных приложениях;
- Effi::Decimal.

EffiDB использует типы данных библиотеки Boost Optional [6].

## 5 Концепции EffiDB

Перейдем к изложению идей, положенных в основу библиотеки EffiDB. Для изоляции программиста от специфики взаимодействия с конкретной СУБД EffiDB использует специализированные классы C++ для каждого из понятий реляционных баз данных.

В тех случаях, когда это представляется полезным, текст сопровождается примерами использования библиотеки.

### 5.1 База данных

Понятию «база данных» соответствует класс Effi::DataBase. Этот класс представляет собой фабрику объектов Connection (соединение) — соединение является единственным способом общения с сервером базы данных.

```
DataBase database=
    new DataBase(libName, dbParams);
```

Здесь libName — имя драйвера СУБД, а dbParams — параметры базы данных. Набор параметров зависит от СУБД. Так, для работы с MySQL инициализация может иметь вид:

```
string libName="libardbms_mysql.so";
std::map<string, string> dbParams;
dbParams["DSN"]="<Имя источника данных>";
dbParams["AutoCommit"]="off";
dbParams["UserName"]="<Имя\ пользователя>";
dbParams["Password"]="<Пароль>";
dbParams["Port"]="<Порт>";
dbParams["Host"]="<Адрес сервера базы данных>;
```

### 5.2 Соединение

Соединение с базой данных позволяет прикладному ПО общаться с ПО сервера базы данных. Соединение является необходимым элементом процесса, позволяющим посылать команды СУБД и получать ответы.

Процедура установления соединения является время- и ресурсоемкой операцией, и было бы неэффективно инициировать новое соединение всякий раз, когда приложению нужно связаться с базой данных. Поэтому в EffiDB реализован алгоритм организации связанного пула соединений. Когда приложению требуется связь с базой данных, сначала происходит поиск свободного соединения в пуле уже существующих соединений. Только в том случае, если свободных соединений нет, создается новое соединение. Эта возможность реализована в функции DataBase::GetConnection(). Когда необходимость в соединении пропадает, приложение должно освободить его функцией DataBase::ReleaseConnection(). Таким образом EffiDB формирует пул свободных переиспользуемых соединений.

Соединение представлено классом Effi::Connection.

```
Connection* conn;
conn=database->GetConnection();
...
database->ReleaseConnection(conn);
```

### 5.3 Средства манипулирования данными

Как уже было сказано, код приложения должен быть унифицированным и работать с любой из поддерживаемых СУБД. С другой стороны, каждое SQL-выражение исполняется на вполне конкретной СУБД и, следовательно, должно подчиняться строгим правилам синтаксиса данной СУБД.

Чтобы разрешить это противоречие, EffiDB строит SQL-запрос в несколько этапов. Параметры SQL-запроса аккумулируются одним из четырех классов (каждый класс соответствует одному из видов SQL-выражений): `Effi::Selector`, `Effi::Inserter`, `Effi::Deleter`, `Effi::Updater`. К параметрам относятся все необходимое для формирования запроса: множество выбираемых данных, список таблиц, из которых выбираются данные, условия выборки, условия группировки, ограничения, условия сортировки и т. п. Далее будем называть эти четыре класса «агентами».

Классы-агенты не зависят от специфики СУБД. Их назначение — сбор полной информации, достаточной для построения выражения, при этом они

не формируют собственно SQL-выражения. Библиотека EffiDB позволяет строить разнообразные выражения, в том числе с использованием псевдонимов, инструкций `join`, сортировки и т. п.

Каждый класс-агент имеет функцию `Execute()`, которая, в свою очередь, вызывает соответствующую функцию драйвера СУБД. Эта функция формирует строку SQL-запроса в соответствии с синтаксисом SQL конкретной СУБД.

## 6 Формирование SQL-запросов

Приведем несколько примеров использования библиотеки (предположим, что работаем с MySQL).

Рассмотрим базу данных книжного магазина. Книга имеет название, автора и цену. Известно количество экземпляров каждой книги.

Цель — сформировать типичные SQL-запросы из приложения, написанного на C++.

База данных содержит, в частности, таблицу `Bookstore` — она представляет собой список книг:

ID	Title	Author	Price	Qty
101	Algorithms + Data Structures = Programs	Wirth	120.35	14
187	1001 Nights	NULL	999.99	2
318	The Decameron	Boccaccio	1285.99	1
325	C++	Stroustrup	59.85	10
400	The Art of Computer Programming	Knuth	299.99	0

### 6.1 Установление соединения

В начале работы необходимо определить, драйвер какой СУБД будет использован, задать параметры базы данных, создать объект `DataBase` и установить соединение:

```
DataBase* database=
    new DataBase(libName, dbParams);
Connection* conn;
try {
    conn=database->GetConnection();
}
catch(Exception e) {
    cerr << "ERROR: " << e.What() << endl;
    database->ReleaseConnection(conn);
    throw;
}
```

### 6.2 Выборка из таблицы

Рассмотрим теперь задачу построения запроса, выбирающего дорогие книги и возвращающего

идентификатор книги, ее название и цену с учетом НДС.

Таблица базы данных представлена объектом класса `Effi::SimpleTable`, колонки таблицы — объектами класса `Effi::Column`:

```
// Определение таблицы базы данных
SimpleTable bookStore("Bookstore");

// Определение колонок таблицы, необходимых
// для запроса
Column ID(bookStore, "ID");
Column Title(bookStore, "Title");
Column Price(bookStore, "Price");
```

Для доступа к результатам запроса используются вспомогательные переменные. Соответствие между полями таблиц базы данных и вспомогательными переменными устанавливается посредством технологии связывания (*binding*), как будет показано ниже.

```
// Определение вспомогательных переменных
// для доступа к результатам запроса
int id;
```

```
string title;
double totalPrice;

// Пороговое значение цены
double threshold = 300;
// Ставка НДС, \%
double vat = 0.2;

Собственно формирование запроса состоит из
указания колонок и связанных с ними вспомога-
тельных переменных, а также установки критериев
выбора:

Selector s;
// Определение набора колонок таблицы
// и связывание их
// со вспомогательными переменными
s << ID.Bind(id) << Title.Bind(title) <<
(Price*(1+vat)).As("TotalPrice").
Bind(totalPrice);
// Определение критериев выборки
s.Where(Price > threshold);
```

Выполнение запроса осуществляется посред-
ством вызова метода `Effi::Selector::Execute()`:

```
// Выполнение SQL-запроса
DataSet data=s.Execute(conn);
```

Результатом выполнения SQL-запроса типа SE-
LECT является набор записей. Доступ к резуль-
татам запроса осуществляется посредством класса
`Effi::DataSet`. Функция `DataSet::Fetch()` считывает
очередную запись в соответствующие вспомо-
гательные переменные, связанные с колонка-
ми таблицы результатов функцией `Bind()`. Так,
в рассматриваемом примере колонка ID связана
с переменной `id`, колонка Title — с переменной
`title`, а выражение `(Price*(1+vat))` — с переменной
`totalPrice`. После считывания внутренний указатель
функции `Fetch()` смещается на следующую запись.
Если запись последняя, то функция `Fetch()` воз-
вращает `false`. Если поле таблицы результатов не

связано ни с одной вспомогательной переменной,
то данные этого поля недоступны из C++ кода.

Следующий код выводит результат исполнения
запроса SELECT:

```
while(data.Fetch()) {
    cerr << id << "; " << title << "; "
    << totalPrice << endl;
}
```

Сформированный SQL-запрос имеет вид:

```
SELECT ID, Title, Price*1.2 AS TotalPrice
FROM Bookstore
WHERE Price > 300;
```

и возвращает следующие данные:

ID	Title	TotalPrice
187;	1001 Nights;	1199.99
318;	The Decameron;	1543.19

### 6.3 Классы-обертки реляционных таблиц

Описанный выше метод универсален в том смы-
сле, что позволяет работать с таблицами, структура
которых неизвестна на момент компиляции.

Представляется, однако, что в большинстве ре-
альных ситуаций структура базы данных известна
заранее, и тогда такой подход оказывается избыточ-
ным и неудобным — он требует создавать представ-
ления реляционных сущностей (таблиц и колонок
базы данных) в C++ всякий раз, когда с ними ра-
ботают. Так, если некая таблица участвует в одном
SQL-запросе дважды, то требуется создавать два
объекта класса `SimpleTable`, описывать для каждого
из них одинаковый набор колонок и т. п.

Библиотека `EffiDB` предоставляет возможность
создать универсальный класс-обертку для работы с
известной на момент компиляции таблицей базы
данных. Такой класс должен быть производным от
класса `Effi::Domain`:

```
class Bookstore : public Domain {
public :
    struct BookstoreDomain {
        Column ID;
        Column Title;
        Column Author;
        Column Price;
        BookstoreDomain(const SimpleTable& table)
            : ID(table, "ID");
            , Title(table, "Title");
            , Author(table, "Author");
            , Price(table, "Price"){ }
    } Domain_;
```

```

boost::optional<int> ID;
boost::optional<string> Title;
boost::optional<string> Author;
Decimal Price;
Bookstore() : Domain("Bookstore"), Domain_(*this) {
    // COLUMN_PK означает, что колонка входит в первичный ключ
    columns_.push_back(ColSpec(Domain_.ID.Bind(ID), COLUMN_PK));
    // COLUMN_ORDINARY означает, что колонка не является ключевой
    columns_.push_back(ColSpec(Domain_.Title.Bind(Title), COLUMN_ORDINARY));
    columns_.push_back(ColSpec(Domain_.Author.Bind(Author),
COLUMN_ORDINARY));
    columns_.push_back(ColSpec(Domain_.Price.Bind(Price), COLUMN_ORDINARY));
}
Bookstore(const Bookstore& copy) : Domain("Bookstore"), Domain_(*this)
, ID(copy.ID);
, Name(copy.Title);
, Author(copy.Author);
, Phone(copy.Price) {
    columns_.push_back(ColSpec(Domain_.ID.Bind(ID), COLUMN_SK));
    columns_.push_back(ColSpec(Domain_.Title.Bind(Title), COLUMN_ORDINARY));
    columns_.push_back(ColSpec(Domain_.Author.Bind(Author),
COLUMN_ORDINARY));
    columns_.push_back(ColSpec(Domain_.Price.Bind(Price), COLUMN_ORDINARY));
}
Bookstore& operator= (const Bookstore& copy) {
    ID = copy.ID;
    Title =copy.Title;
    Author=copy.Author;
    Price =copy.Price;
    return *this;
}
virtual ~Bookstore() {}
BookstoreDomain* operator->() { return &Domain_; }
};

```

В классе-обертке определяются колонки таблицы и набор вспомогательных переменных, соответствующий набору колонок, а также устанавливается соответствие между ними.

Имена объектов-колонок и имена вспомогательных переменных задаются совпадающими с именами полей таблиц. Имена объектов-колонок доступны посредством оператора «→», например Bookstore→Price, а значения полей — посредством оператора «.», например Bookstore.Price.

Колонки, входящие в первичный ключ таблицы, помечаются как ключевые.

Таким образом, необходимость всякий раз при использовании экземпляров класса Bookstore связывать колонки таблицы со вспомогательными переменными отпадает.

Только в тех случаях, когда вспомогательные переменные не имеют соответствующей колонки, но требуются в конкретном запросе (totalPrice в рас-

сматриваемом примере), следует по-прежнему явным образом создавать сами переменные и устанавливать их связь с колонками таблицы результатов запроса.

Работа с классами-обертками обеспечивает ряд немаловажных преимуществ.

Во-первых, описание таких сущностей базы данных, как таблица и ее колонки, локализовано в одном месте кода. Это снижает до минимума риск появления синтаксических ошибок и облегчает процесс поддержки и развития приложения.

Во-вторых, код, формирующий рассмотренный выше SQL-запрос, оказывается гораздо короче:

```

Bookstore tbl;
double threshold = 300;
double vat = 0.2;
double totalPrice;
Selector s;

```

```
s << tbl->ID << tbl->Title
  << (tbl->Price*(1+vat))
  .As("TotalPrice").Bind(totalPrice);
s.Where(tbl->Price>threshold);
```

Легко видеть, что создание объекта, представляющего таблицу, сводится теперь к объявлению переменной, а объявление вспомогательных переменных и связывание их с колонками базы данных не требуется.

И, наконец, рассмотрим сценарий, в котором выгода от использования класса-обертки наиболее существенна и очевидна. На практике чаще всего используются SQL-запросы по первичному ключу таблицы. Синтаксис EffiDB в этом случае крайне прост, а размер текста кода оказывается даже меньшим, чем размер итогового SQL-запроса.

Попробуем, к примеру, выполнить все 4 возможных вида SQL-выражений на рассматриваемой базе данных:

```
// Определение уникального кода
// для новой книги
// и проверка его доступности tbl.ID=100;
if (tbl.Select(conn)) throw Exception
("Книга с таким кодом уже существует.");
```

```
// Добавление новой книги
tbl.Title = "Hobbit";
tbl.Author = "Tolkien";
tbl.Insert(conn);
```

```
// Изменение названия книги
tbl.Title = "Hobbit or There and
            Back Again";
tbl.Update(conn);
```

```
// Удаление книги
tbl.Delete(conn);
```

Соответствующие SQL-выражения:

```
SELECT *
FROM Bookstore
WHERE ID = 100;
```

```
INSERT INTO Bookstore (ID, Title,
                      Author, Price, Qty)
VALUES (100, "Hobbit", "Tolkien",
       NULL, NULL);
```

```
UPDATE Bookstore
SET Title = "Hobbit or There and Back
  Again", Author = "Tolkien",
  Price = NULL, Qty = NULL
WHERE ID = 100;
```

```
DELETE FROM Bookstore
WHERE ID = 100;
```

Аналогичным образом составляются запросы по составному первичному ключу.

## 7 Интегрированная среда разработки приложений Effi

Библиотека EffiDB является лишь одним из компонентов интегрированной среды разработки приложений Effi.

Разработчик, использующий среду Effi, описывает приложение на декларативном языке в формате XML (eXtensible Markup Language), оперируя понятиями классов, методов и элементов графического интерфейса. Все объекты C++ кода, представляющие понятия реляционной базы данных, генерируются автоматически. В частности, получается полностью готовый к использованию класс-обертка. Разработчик имеет возможность сосредоточиться на программировании логики приложения и не тратить усилия на описание вспомогательных объектов.

## 8 Заключение

Для решения проблемы унифицированного взаимодействия с разнотипными СУБД создана оригинальная C++ библиотека EffiDB; библиотека доступна в сети Интернет по адресу <http://sourceforge.net/projects/effidb> [7].

Библиотека EffiDB гарантирует полную переносимость программных приложений за счет изоляции прикладного кода от специфики конкретной СУБД. В библиотеке реализован механизм динамического подключения драйверов разных СУБД в процессе работы приложения, а также возможность одновременного подключения нескольких СУБД. Библиотека обеспечивает защиту от синтаксических ошибок в SQL-выражениях за счет разбиения построения SQL-выражений на несколько этапов и проверки на момент компиляции.

Библиотека EffiDB используется в ряде проектов разработки ПО компании ASoft <http://asoft.ru>, среди которых:

- аппаратно-программный комплекс центра хранения электронных копий материалов уголовных дел для Министерства внутренних дел РФ «Невод-Р»;
- объединенная поисковая федеральная система генетической идентификации «Ксенон-2»;

- единая автоматизированная информационная система дежурных частей органов внутренних дел;
- система автоматизации деятельности по линии борьбы с экстремизмом для Департамента по противодействию экстремизму МВД России;
- автоматизированная система управления взаимоотношениями с контрагентами ASoft CRM;
- геоинформационная система хранения, обработки и визуального представления картографической информации;
- конструктор отчетов, позволяющий строить отчеты любой степени сложности и вложенности, а также задавать макет представления данных;
- информационная система для совместной работы над документами ASoft Collaboration.

## Литература

1. Список современных средств объектно-реляционного отображения доступа к данным. [http://en.wikipedia.org/wiki/List\\_of\\_object-relational\\_mapping\\_software](http://en.wikipedia.org/wiki/List_of_object-relational_mapping_software).
2. Янушко А. В. Современные реляционные СУБД // Банки и технологии, 1998. № 2.
3. Ambler S. W., Sadalage P. J.. Refactoring databases: Evolutionary database design. — Addison-Wesley Professional, 2006. 384 p.
4. Gray J., Reuter A. Transaction processing: Concepts and techniques. — Morgan Kaufmann Pubs., 1993.
5. Бегг К., Коннолли Т., Страчан А. Базы данных: Проектирование, реализация и сопровождение: Теория и практика / Пер. с англ. — М.: Вильямс, 2001.
6. Библиотека Boost Optional в сети Интернет. [http://www.boost.org/doc/libs/1\\_41\\_0/libs/optional/doc/html/index.html](http://www.boost.org/doc/libs/1_41_0/libs/optional/doc/html/index.html).
7. Библиотека EffiDB в сети Интернет. <http://sourceforge.net/projects/effidb>.

# ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКАЯ ОЦЕНКА АДЕКВАТНОСТИ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ

Л. А. Кузнецов<sup>1</sup>

**Аннотация:** Приведены математические основы и оригинальная методология разработки систем оценки семантической близости информационных объектов (ИО), представленных на естественном языке. Вводится вероятностно-статистическое представление сопоставляемых ИО. Используется теория информации для оценки уровня семантической близости ИО. Методология доведена до алгоритмов ее реализации в виде соответствующей автоматизированной системы. Представлены результаты практической проверки эффективности методологии.

**Ключевые слова:** информационные объекты; естественный язык; семантическая адекватность; вероятностная модель; теория информации

## 1 Введение. Информация, знания, семантический анализ

Основным мотивом перехода от индустриальной к постиндустриальной модели развития в промышленно развитых странах, начавшегося в конце прошлого века, является стремительное увеличение скорости развития науки и знания. В постиндустриальной модели развития, по мнению ведущих западных ученых в области социального развития и управления, принципиальным является изменение статуса и значения информации, науки и знания, которые становятся важнейшими факторами, определяющими эволюцию общества.

Питер Ф. Друкер, признанный специалист в области организационного управления, пишет: «Изменение значения знания, начавшееся 250 лет тому назад, преобразовало общество и экономику. Знание стало сегодня основным условием производства. Традиционные «факторы производства» — земля (природные ресурсы), рабочая сила и капитал — не исчезли, но приобрели второстепенное значение. Эти ресурсы можно получить, причем без особого труда, если есть необходимые знания» [1].

Информация и знания становятся главной движущей силой экономического развития и переходят из категории бесплатного общественного блага в категорию товара. В промышленно развитых странах разработка и внедрение технологических инноваций — решающий фактор социального и экономического развития, залог экономической безопасности. В США, по оценкам американских специалистов, прирост душевого национального дохода благодаря этому фактору составляет 90%.

Беспрецедентный рост потока информации и знаний, скорости их передачи и возможностей доступа на первый план научных проблем выдвигает разработку технологий их автоматической обработки. Большая часть существующих и вновь формируемых знаний и информации представлена на естественном языке.

Одной из актуальных, фундаментальных проблем в области обработки информации становится обеспечение возможности формального семантического сравнения, оценки семантической близости ИО, представленных на естественном языке. Разработка формализованных технологий оценки семантической близости ИО позволила бы перейти к практической реализации важных задач в сфере обработки информации, распространения знаний и образования.

В настоящее время в литературе задача семантического сравнения двух текстов, в основном, рассматривается в контексте дубликатов в веб-документах и в системах автоматизированного перевода. При поиске дубликатов опираются на число слов, совпавших в двух текстах. Алгоритм сравнения на основе шинглов является наиболее простым и распространенным. Такой подход используется для нахождения копий текстов, полученных копированием и перестановкой слов, но он не позволяет оценить семантическую близость текстов.

При более сложном анализе текстов учитывается структура входящих в них предложений. В предложениях выделяются элементы (слова или группы слов) и сопоставляются определенные шаблоны для этих элементов. Данный подход описан в книге [2] и используется в ряде кандидатских диссертаций.

<sup>1</sup>Липецкий государственный технический университет, kuznetsov@stu.lipetsk.ru

Однако задача оценки информационной близости двух текстов в обнаруженных автором работах не затрагивается. Используемые там концепции не ориентированы на ее решение и не могут быть использованы в качестве основы для ее решения.

В данной статье предлагается оригинальная методология оценки информационной близости текстов на основании вероятностно-статистического подхода и теории информации. Реализация методологии позволит перейти к практическому решению разнообразных задач, в которых требуется определять меру информационной адекватности документов, представленных на естественном языке. В частности, разработанная концепция будет использована для синтеза автоматизированных систем оценки уровня знаний.

## 2 Формализация анализа текстов

Развитие информационных технологий, предоставляющих широкие возможности автоматизированного анализа и обработки вербально представленной информации, существенно повысило интерес к разработке формальных методов исследования и сопоставления текстов. Современные компьютерные системы позволяют хранить и обрабатывать практически неограниченные объемы текстовой информации. Это стимулирует разработку формальных методов для поддержки выполнения постоянно расширяющихся и углубляющихся исследований информации, представленной на естественном языке. В настоящее время интенсивно разрабатываются формальные методы, позволяющие автоматизировать решение задач в области морфологического, синтаксического и семантического анализа текстовой информации.

Из имеющихся публикаций следует, что методологии различных видов анализа базируются на сходных концепциях и достаточно близки по своему содержанию. Формализация морфологического анализа направлена на алгоритмическое представление грамматики русского языка. Части речи русского языка определены, однозначно определены формы, в которых они могут быть, определены правила, следуя которым должно осуществляться изменение слов, принадлежащих к различным частям речи при образовании соответствующих форм. Значительная часть правил изменения частей речи уже отражена в словарях. Все правила могут быть представлены в виде соответствующих процедур, функций, подпрограмм и т. п. В соответствии с имеющимися правилами может быть идентифицировано, какой частью речи является конкретное слово, в какой форме оно находится. Поэтому мо-

жет быть написана программа, обеспечивающая автоматизированное выполнение морфологического анализа, так что на ее вход будет поступать слово предложения, а на выходе она выдаст результат анализа: какой частью речи является данное слово и в какой форме оно находится.

Формализация синтаксического анализа — задача также понятная: существует синтаксис русского языка, представляющий свод правил, следуя которым могут быть достаточно четко определены члены предложения. Раз правила существуют, то их можно представить в виде набора процедур, обеспечивающих выявление состояния (роли) каждого слова в предложении. Правила и процедуры могут быть более или менее сложными, но принципиально то, что правила имеются, а следовательно, и процедуры могут быть синтезированы по ним. На основании этих процедур может быть разработана система синтаксического анализа, которая, получая на свой вход предложение, выполнит его разбор и анализ и на выходе, как примерный ученик выдаст о каждом слове, входящем в предложение, информацию: каким его членом оно является.

Проблема семантического анализа текстов интенсивно исследуется в различных аспектах: разрабатываются правила и алгоритмы анализа предложений, выявления их структуры, установления соответствия между разноязычными текстами, поиска информации и т. д. При этом, однако, формализация семантического анализа какого-либо одноязычного текста или даже одного предложения представляется задачей весьма малопонятной.

Под формализацией обычно понимается однозначное математическое представление существующих правил, которые, возможно, в текстовом, вербальном виде содержат определение способа извлечения нужных сведений из начальных, исходно заданных данных. В контексте формализации семантического анализа математическому оформлению должны подлежать правила, позволяющие извлечь из слова его смысл. Но, в отличие от морфологии и синтаксиса, не существует каких-либо формальных семантических правил, следуя которым можно было бы установить смысл, вложенный в предложение или в каждое отдельное его слово. Поэтому невозможно представить систему семантического анализа в виде упорядоченного набора правил или предписаний, которая (по аналогии с системами морфологического или синтаксического анализа) получала бы на входе предложение или слово, а на выходе выдавала бы его смысл. Ибо слово и есть его смысл. В толковом словаре, конечно, разъясняется смысл отдельных слов, но, в конечном итоге, это разъяснение представляет сопоставление одному слову других слов, близких по

смыслу, и следует из словаря, а не из каких-либо правил, которые можно было бы формализовать.

Следовательно, если морфологический и синтаксический анализ действительно представляют анализ в соответствии со смыслом этого слова, т. е. разбор предложения на составляющие его элементы и выяснение их роли и состояния, то семантический анализ может пониматься только в смысле сравнения и выяснения смысловой близости разных слов и текстов. Только при наличии эталона анализируемого предложения, смысл которого известен, опираясь на словари, в которых отражена семантическая близость отдельных слов и словосочетаний, может быть получен ответ, что анализируемое предложение находится в некотором соответствии с эталоном и, следовательно, имеет определенный смысл. Например, при переводе смысл на языке-оригинале принимается за известный эталон. С помощью словаря, в котором имеется соответствие между словами и словосочетаниями, находится соответствующее выражение на другом языке. Важно понимать, что соответствие при этом следует не из слов, а из словаря.

Таким образом, представляется, что при сопоставлении одноязычных текстов более правильно говорить не об их семантическом анализе, а об установлении уровня их информационной адекватности, об определении взаимного количества информации, общего для сравниваемых текстов, из общего объема информации, содержащегося в одном из них, принимаемом за эталонный текст.

Автоматизированная технология должна обеспечивать реализацию функций определения пересечения, общей части дубля и эталона. Для этого в автоматизированной технологии должны быть разработаны формально-математические инструменты для представления текстов дубля и эталона в виде, позволяющем оценить количество информации, содержащейся в них, определить долю информации в дубле, отражающую содержание эталона, и на этой основе сформировать оценку их семантической близости.

### 3 Ограниченность возможностей детерминированного подхода

Имеется принципиальное базовое отличие семантического анализа от морфологического и синтаксического. Отмеченное кратко выше показывает, что объектом морфологического и синтаксического анализа является фиксированный, полностью однозначно определенный текстовый фрагмент. Определение роли и состояния оборотов

или отдельных слов, составляющих предложения анализируемого фрагмента, производится по четко определенным, детерминированным правилам, которые могут быть представлены в виде более или менее сложных алгоритмов, формирующих морфологические или синтаксические характеристики предложений, оборотов и слов. Процесс и правила формирования морфологических и синтаксических характеристик и сами характеристики определены и закономерны. Поэтому эти виды анализа закономерны или детерминированы.

На первый взгляд, кажется заманчивой идея представить текст эталона и дубля в виде предложений, предложения в виде деревьев или иных детерминированных структур и затем сравнить структурированное таким образом представление эталона и дубля. Однако русский язык, а здесь подразумевается, что именно он используется для вербального представления информации, совершенно игнорирует какие-либо структурные ограничения по расположению членов предложения, по виду предложений, формированию фрагментов предложений из групп слов, изобилует бесконечным многообразием форм управления отдельными словами и группами слов. По этой причине можно ожидать, что в эталоне и дубле не окажется тождественно равных структурных единиц, а число альтернатив, подлежащих сравнению, может быть бесконечным. В такой ситуации становится принципиальной проблема определения альтернатив и логических правил их разрешения. Поэтому представляется, что детерминированный семантический анализ не вполне соответствует содержательному существу проблемы.

По мнению автора, семантический анализ как термин не вполне удачен. Речь может идти об установлении уровня информационного соответствия содержания одного, анализируемого текста, который здесь именуется дублем, содержанию другого, именуемого эталоном, текста. Представляется, что реализация сравнения, выявления уровня соответствия нескольких русскоязычных текстов использованием детерминированного структурирования и детерминированных правил оценки близости не представляется практически возможной. С учетом интеллектуальной специфики одна и та же информация может случайным образом облекаться в различную текстовую оболочку. Основная проблема оценки степени близости информационных объектов, представленных на естественном языке, следует из семантической многозначности слов и наличия синонимов. Эти обстоятельства приводят к неоднозначности лексического представления семантического содержания текстов. Проблемы неоднозначности текстовой информации известны и

активно исследуются специалистами в области русского языка. Детерминированный подход сравнения текстов оказывается нацеленным фактически на формальное описание тонкостей образования синтаксических форм русского языка, многообразия которых представляется бесконечным.

Очевидно, что случай полного совпадения текстов, используемый при формировании регламента доступа к информации, здесь не рассматривается. Должна присутствовать возможность выделения общности информационного содержания сопоставляемых информационных объектов из их случайным образом выбранной формы представления на естественном языке. Решение такой задачи может быть получено только при описании взаимосвязи семантического (информационного) содержания и лексического оформления обоих сравниваемых текстов с вероятностно-статистических позиций.

## 4 Элементы теории информации

Более полувека существует в виде научной дисциплины теория информации. Ее основоположником является американский специалист в области передачи информации в технических линиях связи Клод Шеннон [3]. Значительный вклад в теорию информации, особенно в строгое доказательство ее основных принципов, внесли советские ученые школы А. Н. Колмогорова [4]. В теории информации исследуются проблемы передачи и преобразования информации, при этом вводится количественная оценка информации. Применительно к проблеме сравнения близости ИО, которой посвящена данная статья, является важным, что в теории информации разработаны теоретические основы исследования близости сообщений, переданного передатчиком и принятого приемником на другом конце линии связи. При этом вводится количественная мера информации, которая позволяет осуществить сопоставление информационной емкости переданного и принятого сообщений и на этой основе оценить искажение (потери) информации в линии связи при ее передаче.

Теория информации может быть использована для решения проблемы оценки близости ИО, представленных на естественном языке. Ничто не мешает вместо сообщений, принятого и переданного, рассматривать ИО, трактуя один из них —

аналог переданного сообщения — как эталонный информационный объект (ЭИО), а другой — аналог принятого сообщения — как дубль ЭИО (ДИО). Как будет видно дальше, мера количества информации в одном объекте о другом симметрична, поэтому при количественной оценке их близости не важно, какой объект считать эталоном, а какой — дублем. Понятно, что сравнение ЭИО и ДИО на абсолютное их совпадение неприемлемо. В теории информации исследуется количественная, а не содержательная сторона информации. В связи с наличием случайных помех в системах формирования и линиях передачи информации сообщения, переданное и принятое, интерпретируются случайными величинами  $\xi$ . Шенноном было предложено использовать энтропию<sup>1</sup> как вероятностную меру количества информации [3].

Энтропия исхода определяется в виде логарифма вероятности этого исхода:

$$H(\xi_i) = -\log p(\xi_i), \quad (1)$$

а усредненная энтропия случайной величины  $\xi$  выражается через функцию распределения ее вероятностей в виде:

$$H_\xi = -\sum_{\xi} p(\xi) \log p(\xi), \quad (2)$$

где  $\xi$  — случайная величина;  $p(\xi) \leq 1$  — распределение ее вероятностей.

В рассматриваемом здесь случае анализа текстов случайными величинами могут быть слова или другие конструкции. Под количеством информации в теории информации понимается неопределенность, устраняемая в результате выяснения исхода, т. е. значения, принимаемого случайной величиной.

Простейший содержательный пример в контексте статьи может быть следующим. Слова: *пример*, *образец*, *экспонат* в некотором контексте являются синонимами и могут использоваться для обозначения объекта  $\xi$  с вероятностями:  $p(\xi = \text{пример}) = 0,2$ ;  $p(\xi = \text{образец}) = 0,4$ ;  $(\xi = \text{экспонат}) = 0,6$ . В этом случае количество информации, получаемое при реализации конкретного исхода, допустим, при использовании слова экспонат, т. е.  $\xi = \text{экспонат}$ , будет в соответствии с (1) равно  $\log p(0,6)$ , а усредненная неопределенность объекта  $\xi$  будет по (2)

<sup>1</sup>Энтропия (от греческого *entropia* — превращение) введена в 1865 г. немецким физиком Р. Клаузиусом как функция состояния термодинамической системы, изменение которой  $dS$  в равновесном процессе равно отношению количества теплоты  $dQ$ , подведенного к системе или отведенного от нее, к термодинамической температуре системы  $T$ :  $dS = dQ/T$ . Л. Больцман, один из основателей статистической термодинамики, предложил использовать энтропию как меру вероятности пребывания системы в данном состоянии. Шеннон ввел энтропию в теорию информации в качестве меры количества информации, которое выражается через распределение вероятностей.

равна:  $H_\xi = -p(0,2) \log p(0,2) - p(0,4) \log p(0,4) - p(0,6) \log p(0,6)$ . В теории информации наиболее часто используются логарифмы по основанию 2, в этом случае количество информации определяется в битах.

В реальности чаще интерес представляет сравнение информационной емкости сообщений, оценка имеющейся в них совместной информации. Для этого по распределениям вероятностей сообщений определяется энтропия каждого из них (количество информации в каждом из них), а по совместному распределению вероятностей — совместная энтропия. По энтропиям оценивается количество взаимной информации.

При использовании теории информации для описания закономерностей передачи информации энтропия переданного сообщения определяет количество переданной информации, а энтропия принятого сообщения — количество принятой информации. Общая часть в переданном и принятом сообщениях определяет количество взаимной информации. Обозначив через  $\xi$  переданное сообщение, а через  $\eta$  — принятое, количество взаимной информации  $I(\xi\eta)$ , или количество информации, содержащееся в принятом сообщении  $\eta$  из переданного сообщения  $\xi$ , можно определить, следуя теории информации [3], в виде:

$$I_{\xi\eta} = \int_X \int_Y p_{\xi\eta}(x, y) \log \frac{p_{\xi\eta}(x, y)}{p_\xi(x)p_\eta(y)} dx dy, \quad (3)$$

где  $p_\xi(x)$  — плотность распределения переданного сообщения;  $p_\eta(y)$  — плотность распределения принятого сообщения;  $p_{\xi\eta}(x, y)$  — плотность совместного распределения;  $X$  и  $Y$  — области определения  $x$  и  $y$  соответственно.

В (3) имеет место равноправное симметричное вхождение  $\xi$  и  $\eta$ , поэтому взаимная информация симметрична относительно  $\xi$  и  $\eta$ . Отсюда следует, что при количественной оценке взаимной информации не важно, какое сообщение выступает в роли переданного, а какое — в роли принятого.

Именно взаимная информация может использоваться в качестве меры подобия ИО. Нетрудно видеть, что содержательное существо теории информации, направленное на оценку потерь информации при ее передаче, адекватно содержательной сущности многих задач в области исследования семантической близости ИО на естественном языке. Применение теории информации для решения проблемы оценки близости ИО может основываться на замене сообщений исследуемыми ИО. Один объект (эталон) может интерпретироваться переданным сообщением, а второй (дубль) — принятым сообщением. Вследствие симметрии от перемены

ролей количество взаимной информации не изменится. Формула (3) отражает количество взаимной информации непрерывных сообщений, точнее сообщений, случайный характер которых определяется непрерывными функциями распределения вероятностей вида  $p_\xi(x)$ ,  $x \in X = [x', x'']$ , где  $x'$  и  $x''$  — предельные значения  $x$ .

При анализе текстов в качестве случайных величин будут выступать синтаксические или морфологические компоненты, которые являются дискретными величинами. Их случайными лексическими значениями будут выступать слова или словосочетания, характеризуемые дискретными вероятностями, как это показано в приведенном выше кратком примере. В примере под случайным объектом или компонентом  $\xi$  может пониматься подлежащее при использовании синтаксической структуризации или существительное при использовании морфологической структуризации. Компонент  $\xi$  в обоих случаях может принимать случайные значения *пример*, *образец*, *экспонат* с вероятностями  $p(\xi = \text{пример}) = 0,2$ ;  $p(\xi = \text{образец}) = 0,4$ ;  $p(\xi = \text{экспонат}) = 0,6$ . Дискретные значения вероятностей могут суммироваться, и поэтому вместо интегралов, присутствующих в (3), будут использоваться суммы по всем возможным значениям случайных величин. Количество взаимной информации в дискретном случае будет определяться следующим образом:

$$I_{\xi\eta} = \sum_{\xi \in X} \sum_{\eta \in Y} p(\xi, \eta) \log \frac{p(\xi, \eta)}{p(\xi)p(\eta)}, \quad (4)$$

где  $X = \{x_1, x_2, \dots, x_n\}$ ,  $Y = \{y_1, y_2, \dots, y_m\}$  — множества значений случайных величин  $\xi$  и  $\eta$ ,  $p(\xi)$ ,  $p(\eta)$  и  $p(\xi, \eta)$  — распределения их вероятностей.

Данная статья посвящена изложению общей концепции автоматизированной технологии оценки степени близости ИО, представленных на естественном языке. Поэтому здесь ограничимся этим кратким представлением основного существа теории информации и ее дальнейшее использование объясним «на словах». Достаточно полные и строгие сведения по энтропии и взаимной информации интересующийся читатель сможет найти в оригинальной литературе, например [3–5], а их применение в далекой от передачи информации области управления качеством и технологиями — в работах автора [6, 7] и др.

## 5 Понятие вероятностной модели

Структуризация текста с целью извлечения заключенного в нем смысла представляется не вполне

определенной задачей. Неопределенность следует из сложности представления ее содержательного существа, откуда вытекают и проблемы с определением методов решения. При оценке степени подобия содержания двух ИО смысл каждого из них, вообще говоря, интереса не представляет, так как целью является не выяснение семантики, а оценка степени их содержательного подобия. Для этого необходимо оценить меру совпадения в текстах того, о чем (ком) идет речь, что, как, где, когда и т. п. с ними происходит или они делают.

Синтез формального подхода к оценке близости ИО, представленных на естественном языке, требует формального представления самих сравниваемых объектов, т. е. разработки модели представления ИО. Математические модели объектов являются основой для разработки систем управления этими объектами, а также решения задач анализа объектов, исследования взаимосвязей между компонентами, образующими объект, синтеза суждений о состоянии и эволюции объекта. Поэтому структура и содержание модели должны разрабатываться с учетом четкого представления целей, для достижения которых она будет использоваться. Модель должна адекватно отражать все наиболее важные для правильного решения поставленной задачи содержательные аспекты объекта и игнорировать те, которые, усложняя модель, не способствуют повышению качества решения. Модели одного и того же объекта, предназначенные для решения различных задач, могут значительно различаться глубиной учета отдельных деталей.

Применительно к проблеме формального представления ИО в задачах оценки их семантической близости модель должна обеспечивать возможность сопоставления эквивалентных компонентов объектов, отражающих содержание сопоставляемых ИО, и игнорировать стилистические тонкости, влияющие на форму представления содержания, но не на его смысл.

Обычно первым шагом при построении модели является структуризация объекта, выделение его компонентов, которые в совокупности определяют рассматриваемый объект.

При разработке структуры модели ИО можно было бы, следуя имеющимся в литературе примерам, исходить из структуры простых предложений, в виде совокупности которых тем или иным способом может быть представлен ИО. Простое предложение русская грамматика определяет центральной грамматической единицей. «Это определяется тем, что простое предложение представляет собой элементарную предназначенную для передачи относительно законченной информации единицу. . . » [8, с. 405]. Но далее следуют 154 параграфа, в которых

излагаются типы и формы простых предложений. Их многообразие и присутствие неполной четкости деления по типам и формам делает нереальной задачу формального описания даже простых предложений, не говоря о более сложных типах предложений. Именно по этой причине детерминированный подход, опирающийся на представления русской грамматики, как отмечалось выше, представляется малоприменимым для анализа семантической близости ИО.

Вследствие того, что целью разрабатываемой методологии является не анализ текстов с позиций грамматики русского языка, а сопоставление их семантического содержания, которое может случайным образом облекаться в лексическую оболочку, к определению структуры модели представляется целесообразным подойти с вероятностно-статистических позиций.

В теории вероятностей [9] существует вероятностная модель, которая позволяет дать формальное, максимально полное описание вероятностно-статистического объекта. Она определяется на множестве элементарных событий  $\{\omega_1, \omega_2, \dots, \omega_n\}$ , которое образует пространство элементарных событий, или исходов  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ . Известны вероятности элементарных событий  $p(\omega_i)$ ,  $i = 1, 2, \dots, n$ . На множестве элементарных событий задается алгебра  $\aleph = (A_j | A_j \subseteq \Omega)$  или, иначе, система случайных событий, составленных каким-либо определенным образом из элементарных событий  $\omega_i \in \Omega$ . Для каждого из случайных событий  $A_j = \{\omega_i \in \Omega\}$ , образующих алгебру, по вероятностям элементарных исходов  $p(\omega_i)$ ,  $\omega_i \in A_j$ , определяется его вероятность  $P(A_j)$ .

Набор: множество элементарных событий  $\Omega = \{\omega_1, \dots, \omega_n\}$ , система случайных событий (алгебра)  $\aleph = (A_j | A_j \subseteq \Omega)$  и вероятности случайных событий  $P(A_j)$  — образует вероятностную модель случайного объекта. Она содержит всю информацию, которой может быть охарактеризован случайный объект. Формально вероятностная модель (или вероятностное пространство эксперимента с конечным пространством исходов  $\Omega$  и алгеброй событий  $\aleph$ ) может быть представлена в виде:

$$M_\Omega = \{\Omega, \aleph, P(A)\}, \quad (5)$$

где  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ,  $\aleph = (A_j | A_j \subseteq \Omega)$ ,  $P(A) = (P(A_j) | A_j \in \aleph)$ .

Определить вероятностную модель конкретного случайного объекта значит определить все ее элементы — множество элементарных исходов, систему случайных событий и их вероятности — для этого конкретного объекта.

## 6 Вероятностно-статистическая морфологическая модель информационного объекта

Информационный объект может быть представлен в виде вероятностной модели. В нем множество элементарных исходов  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  представляют слова  $\omega_i$ ,  $i = 1, 2, \dots, n$ , составляющие текст ИО. Существуют системы структуризации лексического материала. Достаточно общими и пригодными для использования при разработке вероятностно-статистической модели ИО являются синтаксическая и морфологическая структуризации русского языка. Морфологическая структуризация задается определением частей речи русского языка, которые разделяют язык на самые крупные грамматические классы слов [8]. Различают десять частей речи, среди которых шесть знаменательных: существительные, прилагательные, числительные, местоимения-существительные, наречия, глаголы и три служебные: предлоги, союзы, частицы. Десятой частью являются междометия. Части речи, к которым относятся отдельные слова, могут трактоваться случайными событиями  $A_j$ ,  $j = 1, 2, \dots, 10$ . Каждое отдельное слово (реализация, элементарный исход)  $\omega_i$ ,  $i = 1, 2, \dots, n$ , входит в текст с определенной вероятностью  $p(\omega_i)$ . В тексте роль вероятности играет относительная частота  $p(\omega_i) = n_i/n$ , где  $n_i$  — число употреблений в ИО слова  $i$ ,  $n$  — общее количество слов в ИО. Относительная частота получается экспериментально и называется в теории вероятностей эмпирической вероятностью. По вероятностям  $p(\omega_i)$  отдельных слов вычисляются вероятности событий  $A_j$  — частей речи. Вероятностно-статистическая модель ИО (5), в которой алгебра (способ структуризации) слов определяется морфологией, может быть названа вероятностно-статистической морфологической моделью (ВСММ) ИО, которая может быть по аналогии с (5) записана в виде:

$$M_M = \{\Omega, \aleph_M, P(A)\}, \quad (6)$$

где индекс  $M$  подчеркивает морфологический характер модели, который отражается через определение алгебры  $\aleph_M$ .

Конкретный ИО представляется в виде соответствующего вероятностно-статистического морфологического образа ИО. Он синтезируется на основании модели (6) введением конкретного множества элементарных исходов  $\Omega_O = (\omega_1, \omega_2, \dots, \omega_n)$  — слов. Обозначение  $\Omega_O$  подчеркивает, что это множество слов конкретного ИО. Множество структурируется в соответствии с введенной алгеброй  $\aleph_M = (A_1, A_2, \dots, A_J)$ , где  $J$  — число случайных событий (частей речи), используемых в образе,  $J \leq 10$ , т. е. некоторые части речи, например междометия, предлоги, могут не использоваться при формировании образа. В результате полностью определяется вероятностно-статистический морфологический образ (ВСМО) ИО ВСММ (5) конкретного ИО, который может быть представлен в виде:

$$O_M = \{\Omega_O, \aleph_M, P_O(A)\}, \quad (7)$$

где обозначения следуют из (5), (6) и текста.

По множеству элементарных исходов  $\Omega_O$  вычисляются количественные характеристики образа: вероятности  $p(\omega_i)$  и вероятности случайных событий  $P(A_j)$ . По вероятностям может быть в соответствии с (2) определена энтропия  $H_O$  образа ИО, характеризующая количество информации в образе.

Для пояснения, возможно, непривычного для исследований в области русского языка подхода и терминологии воспользуемся наглядной иллюстрацией. Вероятностно-статистическая морфологическая модель ИО может быть представлена в виде табл. 1.

В шапке таблицы для сокращения размеров примера указаны в явном виде только 4 части речи из десяти, имеющихся в языке. В модели будет использоваться таблица с полным набором частей речи. Шапка таблицы является атрибутом модели. Она отражает структуру ВСММ и является общей для представления образов всех ИО. Части речи, указанные в шапке, трактуются случайными величинами. В вероятностном смысле шапка таблицы содержит все рассматриваемые при морфологическом подходе случайные события или полное поле событий. Смысл полного поля событий в данном

Таблица 1 Представление ВСММ ИО (фрагмент)

Существительные		Прилагательные		Числительные		...	Глаголы	
Слова	Характеристики	Слова	Характеристики	Слова	Характеристики	...	Слова	Характеристики
1 Дом	$p(\text{дом})$	Серый	$p(\text{сер.})$	Три	$p(\text{три})$	...	Стоит	$P(\text{стоит})$
2 Стол	$p(\text{стол})$	Белый	$p(\text{бел.})$	Два	$p(\text{два})$	...	Идет	$P(\text{идет})$
...	...	...	...	...	...	...	...	...

контексте в том, что любое встреченное в тексте (в ИО) слово относится к одному из них (является какой-либо частью речи).

Все то, что находится в таблице под шапкой, отражает конкретный ИО, т. е. определяет ВСМО ИО (7). Слова *дом*, *стол* являются в примере случайными значениями, которые приняла часть речи «существительное», аналогично *серый*, *белый* — случайные значения «прилагательного», *стоит*, *идет* — «глагола». Кроме собственно значений, которые принимают части речи в ИО, в модели отражаются их случайные характеристики, например относительные частоты.

Вероятностно-статистический морфологический образ (7) может быть сформирован для любого произвольного ИО, представленного на русском языке, да и не на русском тоже. При его формировании могут быть использованы имеющиеся достаточно эффективные инструменты морфологического анализа текстов, которые позволяют автоматизировать процедуры отнесения слов к частям речи.

При решении задачи оценки семантической близости ИО ВСМО синтезируется для обоих сравниваемых объектов. Для определенности один из них называется эталоном (ИОЭ), его морфологический образ обозначается  $O_{МЭ}$ , а второй — дублем (ИОД), его образ —  $O_{МД}$ . Вероятностно-статистический морфологический образ содержит все слова ИО, структурированные по частям речи, вероятности отдельных слов и частей речи в ИО. Вероятности количественно характеризуют ВСМО ИОЭ и ВСМО ИОД. На их основе может быть осуществлена оценка количества информации в каждом из объектов и оценено количество взаимной информации (4) в объектах. Выражение для оценки количества взаимной информации (4) может быть приведено к виду:

$$I_{МЭД} = H(O_{МЭ}) + H(O_{МД}) - H(O_{МЭ}, O_{МД}), \quad (8)$$

где обозначения совпадают с введенными раньше.

Можно утверждать, что полное совпадение ВСМО объектов будет иметь место при полной идентичности ИОД и ИОЭ. Наличие отклонений ВСМО дубля от ВСМО эталона будет указывать на несовпадение содержания ИО, количественной оценкой которого является количество совместной информации.

Формирование ВСМО объектов может опираться на имеющиеся фундаментальные исследования в области морфологии русского языка и разработки мощных инструментов морфологической структуризации. В частности, выполнение морфологической структуризации в данном исследовании

опирается на электронную версию словаря А. А. Зализняка [10], для практического использования которой разработаны оригинальные программные продукты.

## 7 Вероятностно-статистическая синтаксическая модель информационного объекта

С другой стороны, в русском языке классифицированы члены предложения. Определение членов предложения задает синтаксическую структуру русского языка. Важно подчеркнуть, что части речи обладают общностью синтаксических функций, так что эти два способа структуризации лексического состава русского языка взаимосвязаны и дополняют друг друга. Вероятностно-статистическая модель, синтезируемая на синтаксической основе, будет отличаться только системой случайных событий, образующих полное поле событий, т. е. шапкой таблицы, отражающей модель.

Между морфологической и синтаксической структуризацией имеется значительное отличие, следующее из того, что морфологическая структуризация является фиксированной, так как имеется всего 10 частей речи. Синтаксическая структуризация допускает введение более детальной структуры членов предложения. Она определяется разработчиком системы сравнения объектов и допускает определенный волюнтаризм в выборе системы случайных событий. Для возможности отражения семантических оттенков в систему случайных событий могут быть введены разнообразные синтаксические конструкции, связанные со спецификой содержания сравниваемых ИО. Понятно, что увеличение отражаемого в модели разнообразия конструкций, с одной стороны, будет способствовать повышению качества сравнения текстов, а с другой — усложнению модели. Однако если учесть табличное представление модели, стандартное для реляционных баз данных, то увеличение таблиц не приведет к принципиальным затруднениям при реализации систем оценки близости ИО.

Вероятностно-статистическая синтаксическая модель (ВССМ) ИО аналогична ВСММ ИО и может быть представлена в виде таблицы, подобной табл. 1. Шапка таблицы будет отражать принцип структурирования по случайным событиям, которые в ней связываются уже с членами предложения, т. е. с синтаксическими конструкциями языка. Алгебра вероятностной модели в этом случае будет определяться типами членов предложения, которые используются для представления ИО:  $\aleph_C =$

$= \{B_1, B_2, \dots, B_L\}$ , где  $B_1$  — подлежащее;  $B_2$  — сказуемое;  $B_3$  — определение и т. д.;  $L$  — общее число типов членов предложения, используемых в синтаксической модели и образующих в ней полное поле событий;  $B_l$ ,  $l = 1, 2, \dots, L$ , как и  $A_j$ , трактуются как случайные синтаксические события. Таким образом, ВССМ будет иметь вид:

$$M_C = \{\Omega, \aleph_C, P(B)\}, \quad (9)$$

отличающийся от (6) только алгеброй  $\aleph_C$ .

Исследуемый ИО посредством какого-либо синтаксического анализатора разделяется на члены предложения. На основе этого разделения формируются случайные события и синтезируется вероятностно-статистический синтаксический образ (ВССО) ИО:

$$O_C = \{\Omega_O, \aleph_C, P_O(B)\}. \quad (10)$$

Для оценки семантической близости ИО ВССО синтезируется для обоих сравниваемых объектов. На их основе может быть оценено количество взаимной информации в сравниваемых объектах (4). Выражение для оценки количества взаимной информации получается из (8) заменой морфологических образов синтаксическими:

$$I_{CЭД} = H(O_{CЭ}) + H(O_{CД}) - H(O_{CЭ}, O_{CД}), \quad (11)$$

где обозначения совпадают с введенными раньше.

Синтаксический анализ текста представляет отдельную проблему, отличающуюся от рассматриваемой здесь. Поэтому для определения инструментов формирования синтаксических образов были проанализированы имеющиеся в литературе разработки в этом направлении и практически использовался «Синтаксический анализатор Cognitive Dwarf 2.0» [11–13].

## 8 Методология оценки семантической близости информационных объектов

В морфологической и в синтаксической структуре структурные компоненты несут достаточно определенную и близкую семантическую нагрузку. Поэтому могут быть установлены отношения эквивалентности между компонентами двух структур. Вследствие того, что эти структуры охватывают весь лексический состав и грамматический строй

русского языка, они обеспечивают отражение семантического содержания текстов и, следовательно, являются достаточными для сопоставления этого семантического содержания.

Таким образом, ИО может быть представлен в виде морфологического<sup>1</sup> и/или синтаксического образа. Формальное представление ИО в виде математических объектов — вероятностно-статистических образов — позволяет использовать математический аппарат для получения количественных оценок их близости. Можно утверждать, что полное совпадение как ВСМО, так и ВССО сопоставляемых объектов будет соответствовать равенству представляемых ими ИО.

Оценка степени близости ИО, представленных на естественном языке, может быть реализована на основе применения вероятностно-статистической морфологической и/или синтаксической модели.

Мерой степени близости служит энтропия и взаимная информация, количественные значения которых вычисляются по (2), (8) и (11). Выше отмечалось, что под количеством информации в теории информации понимается количество неопределенности случайного объекта, которое исчезает при выяснении этой неопределенности. Неопределенность объекта характеризуется распределением его вероятностей. В использованном выше примере объекта  $\xi$  с синонимами было задано распределение вероятностей:  $p(\xi = \text{пример}) = 0,2$ ;  $p(\xi = \text{образец}) = 0,4$ ;  $p(\xi = \text{экспонат}) = 0,6$ . Так что в этом случае энтропия отдельных исходов будет:  $H(\xi = \text{пример}) = -\log 0,2$ ,  $H(\xi = \text{образец}) = -\log 0,4$ , а  $H(\xi = \text{экспонат}) = -\log 0,6$ , а усредненная энтропия  $H_\xi = -0,2 \log 0,2 - 0,4 \log 0,4 - 0,6 \log 0,6$ . Таким образом, энтропия будет некоторым числом, зависящим от распределения вероятностей случайных событий, но не от их содержания.

Из выражений (8) и (11) для взаимной информации можно видеть, что, во-первых, она тоже является числом, так как выражается через числа — значения соответствующих энтропий. Во-вторых, выражения (8) и (11) отражают смысл взаимной информации как меры неопределенности.

Пусть используются синтаксические образы сопоставляемых объектов, а их взаимная информация оценивается выражением (11). Рассмотрим два предельных случая.

В первом пусть ВССО эталона  $O_{CЭ}$  не имеет ничего общего с ИССО дубля  $O_{CД}$ . Отсутствие общего означает, что вероятность совместного распределе-

<sup>1</sup>Минимальное количество грамматических терминов, используемых в статье, заимствовано из [8] с единственной целью: приблизить изложение терминологически к области русского языка, хотя содержание статьи, как представляется, достаточно далеко от вопросов собственно языка.

ния  $P(O_{CЭ}, O_{CД}) = 0$ . В теории информации принято считать  $0 \log 0 = 0$ , поэтому  $H(O_{CЭ}, O_{CД}) = 0$  и из (11) следует, что количество совместной информации, содержащееся в двух сопоставляемых объектах, равно их общей неопределенности:  $I_{CЭД} = H(O_{CЭ}) + H(O_{CД})$ .

Во втором случае пусть дубль полностью совпадает с эталоном, т. е.  $O_{CЭ} = O_{CД}$ . Тогда совпадут энтропии  $H(O_{CЭ}) = H(O_{CД})$ , более того, и совместная энтропия  $H(O_{CЭ}, O_{CД})$  будет равна энтропии эталона или дубля. Так что количество совместной информации будет равно  $I_{CЭД} = H(O_{CЭ})$ , т. е. неопределенность дубля отсутствует, вся неопределенность связана только с неопределенностью эталона, только с количеством заключенной в нем информации. Отсюда можно заключить, что количество информации (11) изменяется от значения  $I_{CЭД} = H(O_{CЭ})$  до значения  $I_{CЭД} = H(O_{CЭ}) + H(O_{CД})$ . При этом взаимная информация — величина положительная. Это следует из положительности энтропий:  $p(\xi) \leq 1$ ,  $\log p(\xi) \leq 0$  и, следовательно,

$$H(\xi) = -\log p(\xi) \geq 0$$

и факта

$$H(O + CЭ) + H(O_{CД}) \geq H(O_{CЭ}, O_{CД}).$$

Разумеется, такой же результат может быть получен и для (8). Вследствие свойств логарифмической функции количество информации изменяется монотонно в определенных выше пределах, что и позволяет использовать его в качестве меры семантической близости ИО.

Для практического применения абстрактные значения энтропии и взаимной информации необходимо градуировать в некоторых понятных и связанных с содержательным существом задачи мерах оценки семантической близости ИО. Такая градуировка (тарирование) их значений может осуществляться разными способами и, в частности, обеспечивать реализацию функций адаптации системы к различным задачам и типам ИО. Например, в простейшем случае может быть взят реальный эталонный ИО такого типа, для работы с которым предполагается использовать систему. Искажением эталонного ИО случайным образом и в заданных объемах может быть получена серия дублей с известной степенью семантического несовпадения. Для каждой пары «эталон—дубль» находится значение взаимной информации, которое сопоставляется с известной степенью семантического несовпадения. На основании сопоставления определяется линейное преобразование перевода количества информации в удобную меру оценки степени семантического соответствия ИО.

Методология реализуется в виде последовательности следующих этапов:

- выбор вида и формирование вероятностно-статистической модели конкретизацией алгебры (системы случайных событий)  $\aleph_M$  или  $\aleph_C$ ;
- введение содержания (текстов) образов ИО;
- формирование образа эталонного ИО в виде  $O_{MЭ}$  или  $O_{CЭ}$ ;
- формирование образа второго ИО (дубля) в виде  $O_{MД}$  или  $O_{CД}$ ;
- определение энтропии эталонного ИО в виде  $H(O_{MЭ})$  или  $H(O_{CЭ})$ ;
- определение энтропии второго ИО (дубля) в виде  $H(O_{MД})$  или  $H(O_{CД})$ ;
- определение совместной энтропии эталонного ИО и дубля в виде  $H(O_{MЭ}, O_{MД})$  или  $H(O_{CЭ}, O_{CД})$ ;
- определение совместной информации в соответствии с (8) или/и в соответствии с (11);
- перевод совместной информации в выбранную меру информационной близости ИО.

Разработанная методология оценки семантической близости ИО, кажущаяся, на первый взгляд, совершенно от семантики оторванной, имеет глубокую содержательную основу. Если следовать более общему представлению языка, чем детальное грамматическое, то множества элементарных исходов (слов), образующих в вероятностно-статистических образах случайные события (части речи в ВСМО и члены предложения в ВССО) могут трактоваться как соответствующие обобщенные части речи и члены предложения, образующие эти образы. Такое представление позволяет выделить главное содержание в сравниваемых ИО. Содержательным примером, подтверждающим реальность разработанного подхода, является достаточно часто встречающееся в реальности продуктивное общение людей на плохо знакомом им языке. Они не владеют склонениями, спряжениями, формами времени и другими элементами грамматики, но, зная две—три сотни слов, достаточно успешно общаются, вполне понимая друг друга.

Здесь излагается ядро методологии и не рассматриваются возможности привлечения дополнительных инструментов, повышающих адекватность оценки, таких как использование синонимов, введение весовых коэффициентов, детализация и комбинация событий и т. п.

Заметим еще, что такой подход может быть использован и для оценки близости ИО, реализованных на других языках и с использованием иных

алфавитов. Другие языки, например английский или немецкий, отличаются от русского в сторону уменьшения свободы в порядке слов и разнообразия способов управления, что упрощает задачи их структурирования и построения вероятностно-статистических моделей ИО, не требуя изменения методологии.

В ряде случаев ИО могут быть представлены с использованием не естественного языка, а, например, формального математического языка формул. В этом случае изменяется входной алфавит и, возможно, принцип синтеза алгебры случайных событий. Но эти изменения не касаются представленной здесь собственно методологии оценки подобию ИО.

## 9 Оценка знаний

Одной из проблем, для решения которой принята данная разработка, является автоматизированный контроль знаний. Использование для этой цели системы тестов представляется автору неприемлемым по множеству причин. На кафедре АСУ Липецкого государственного технического университета разрабатывается «Автоматизированная система поддержки образовательной программы обучения» (АСПОП) [14], одним из важнейших компонентов которой является подсистема автоматизированного контроля знаний. Концепция подсистемы базируется на изложенной методологии.

Практическая проверка в минимально возможном объеме работоспособности принципиальных положений концепции осуществлена проверкой знаний студентов. При реализации проверки студентам на экране демонстрировался эталонный ответ из АСПОП, который они воспроизводили на память и заносили в компьютер. По эталонным ответам из АСПОП формировались ВСМО  $O_{МЭ}$  и ВССО  $O_{СЭ}$ . По ответам студентов формировались соответствующие ВСМО  $O_{МД}$  и ВССО  $O_{СД}$ . По морфологическим образам  $O_{МЭ}$  и  $O_{МД}$  определялись энтропии  $H(O_{МЭ})$ ,  $H(O_{МД})$  и  $H(O_{МЭ}, O_{МД})$ , а по ним оценивалось количество взаимной информации. Также обрабатывались синтаксические образы эталонных ответов и их дублей — ответов студентов.

Распечатанные эталонные ответы и ответы студентов анонимно сопоставлялись группой преподавателей, которые выставляли оценки студентам по существующей методике по 100-балльной шкале. Оценки преподавателей надлежащим образом усреднялись. По оценкам преподавателей и количествам взаимной информации, определенным автоматизированной системой, определялись параметры масштабного преобразования количества

информации в принятые в университете 100-балльные оценки. После введения коэффициентов масштабного преобразования система, как и преподаватели, выдавала 100-балльные оценки.

Оценки, автоматически сформированные системой, были сопоставлены с оценками, выставленными преподавателями. В итоге было получено, что среднее квадратическое отклонение оценок, вычисленных системой на основании сопоставлений вероятностно-статистических образов эталона и ответа по изложенной методологии, от оценок, выставленных преподавателями, по 100-балльной шкале составило 10%–15%. Результаты сопоставления ВСМО и ВССО эталона и ответа оказались достаточно близкими. Отметим, что это была пробная проверка, предпринятая исключительно для обретения уверенности в практической эффективности оригинальной концепции.

Углубление и детализация вероятностно-статистических моделей ИО на естественном языке, их исследование и применение представляют неограниченное, научно новое и практически полезное поле деятельности, в освоении которого автор может оказать посильную помощь.

## 10 Заключение

Разработана оригинальная методология оценки степени семантической близости информационных объектов. Методология может служить формально-математической основой в сфере современных информационных технологий для решения разнообразных задач сравнения и оценки подобию информационных объектов, представленных на естественном языке. Методология вводит вероятностно-статистическую модель представления русскоязычного текста и определяет способы представления текстов в виде вероятностно-статистических морфологических и синтаксических образов, которые позволяют оценить количественно и объем информации в информационных объектах, и степень их семантического совпадения. Экспериментальная прикладная проверка показала эффективность применения методологии для разработки автоматизированных систем оценки знаний. Практическое применение методологии только в этой сфере может привести к принципиальным изменениям в сфере образования.

## Литература

1. Друкер П. Посткапиталистическое общество. Новая постиндустриальная волна на Западе: Антология / Под ред. В. Л. Иноземцева. — М.: Academia, 1990.

2. Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст». — 2-е изд. — М.: Школа «Языки русской культуры», 1999.
3. Шеннон К. Математическая теория связи. 1948 // Работы по теории информации и кибернетике / Пер. с англ. под ред. Р.Л. Добрушина и О.Б. Лупанова. — М.: ИЛ, 1963.
4. Колмогоров А. Н. Теория информации и теория алгоритмов. — М.: Наука, 1987.
5. Стратонович Р.Л. Теория информации. — М.: Сов. радио, 1975.
6. Кузнецов Л. А. Введение в САПР производства проката. — М.: Металлургия, 1991.
7. Kuznetsov L. A. The entropy and information application to identify fuzzy sets // ICSC Symposium (International) on Fuzzy Logic Proceedings. — Academic Press, 1995. P. A109–A111.
8. Белоусов В. Н., Ковтунова И. И., Кручинина И. Н. и др. Краткая русская грамматика / Под ред. Н. Ю. Шведовой и В. В. Лопатина — М.: Рус. яз., 1989.
9. Гнеденко Б. В. Курс теории вероятностей: Учебник. — 9-е изд., испр. — М.: ЛКИ, 2007.
10. Зализняк А. А. Грамматический словарь русского языка: Словоизменение. — 3-е изд., стер. — М.: Рус. яз., 1987. 880 с.
11. Синтаксический анализатор Cognitive Dwarf 2.0. <http://cs.isa.ru:10000/dwarf/d2/dw2.html>.
12. Антонова А. А., Мисюрев А. В. Реализация синтаксического разбора для русского и английского языков // Системный анализ и информационные технологии (САИТ 2005): Мат-лы I Междунар. конф. — Переславль-Залесский, 2005. — Переславль-Залесский, 2005. С. 245–249.
13. Антонова А. А., Мисюрев А. В. Синтаксический анализатор для русского и английского языков // Сб. трудов ИСА РАН / Под ред. В. Л. Арлазарова и Н. Е. Емельянова. — М.: УРСС, 2007.
14. Кузнецов Л. А., Фарафонов А. С., Тищенко А. Д., Капнин А. В. Автоматизированная система поддержки образовательной программы обучения // Качество. Инновации. Образование, 2010. № 9. С. 12–20.

# УПРАВЛЕНИЕ ИНФОКОММУНИКАЦИОННЫМИ ПРОЕКТАМИ: «СВОЕВРЕМЕННОСТЬ–ПРОИЗВОДИТЕЛЬНОСТЬ– ИНФОРМАЦИЯ»

А. А. Зацаринный<sup>1</sup>, А. П. Шабанов<sup>2</sup>

**Аннотация:** Рассматривается методологический подход к управлению инфокоммуникационными проектами с обоснованием требований к производительности трактов контрольно-технологических систем (КТС) и к числу субъектов функциональных организационных структур, выполняющих работы в соответствии с принимаемыми из этих трактов сообщениями.

**Ключевые слова:** управление проектом; контрольно-технологическая система; организационная структура; своевременность; производительность; информация

## 1 Постановка задач

В комплексе мер по обеспечению модернизации и технологического обновления производственной сферы страны информационно-телекоммуникационные системы (ИТКС) являются материальной основой для обеспечения высокой степени качества управления в организационных структурах любого масштаба. Поэтому одна из важнейших задач заключается в повышении эффективности управления проектами по созданию и сопровождению ИТКС. На этом пути возможны как учет фактического состояния или изменения структуры и параметров системы в реальном времени, так и адаптация к различным факторам неопределенности на основе накопления и использования информации о предметных сущностях деятельности. Разработка материала настоящей статьи осуществлялась на принципах системного подхода [1] с учетом принятой в рассматриваемой области терминологии, результатов авторских исследований, на основе положений и систем классификации, приведенных в документах:

- ГОСТ Р ИСО/МЭК 15288 «ИТ. Системная инженерия. Процессы жизненного цикла систем»;
- практическое руководство по инновационному управлению качеством и рисками в жизненном цикле систем [2];
- практическое руководство по проектированию ИТКС [3].

Ниже приведены основные термины и определения, используемые в статье:

- инфокоммуникационные проекты — проекты по созданию, эксплуатации, модернизации, развитию ИТКС. Сложность проектов подразумевает наличие технических, организационных и ресурсных задач, решение которых предполагает нетривиальные подходы к их решению;
- управление инфокоммуникационным проектом — управление процессами жизненного цикла ИТКС. Жизненный цикл ИТКС — промежуток времени между моментом появления, зарождения проекта и моментом его ликвидации, завершения. Жизненный цикл ИТКС, как правило, включает стадии замысла, разработки, производства, применения, поддержки применения, прекращения применения и списания. Процессами жизненного цикла ИТКС являются: процессы соглашения; процессы предприятия; процессы проекта; технические процессы;
- контрольно-технологические системы — информационные системы, которые образуются на базе ресурсов ИТКС для передачи информации, относящейся к решению задач оперативного управления техническими объектами и их контроля;
- своевременность представления информации — свойство системы обеспечивать представление информации в задаваемые сроки, гарантирующие выполнение соответствующей функции согласно целевому назначению системы;
- функциональная организационная структура — объединение субъектов в штатной организаци-

<sup>1</sup>Институт проблем информатики Российской академии наук, AZatsarinny@ipiran.ru

<sup>2</sup>ООО «ИБС Экспертиза», AShabanov@ibs.ru

онной структуре предприятия, каждый из которых выполняет свои функции для решения общей для них и четко определенной, конкретной задачи;

- технологическая информация — хранящиеся в базе данных конфигурационных единиц (БДКЕ) ИТКС взаимозависимые модули, в своей совокупности описывающие предметные сущности, необходимые для выполнения работ.

В статье исследуются вопросы, относящиеся к проблеме обоснования требований к производительности КТС. В рамках этой проблемы решаются следующие задачи:

- определение требований к производительности трактов КТС в зависимости от заданных требований к своевременности представления информации (первая задача);
- определение требований к числу субъектов функциональных организационных структур, выполняющих работы в соответствии с принимаемыми из трактов КТС сообщениями (вторая задача).

При управлении проектом обе задачи решаются на *стадии замысла* в жизненном цикле ИТКС в ходе реализации *технического процесса по анализу требований*.

Управление процессом анализа требований осуществляется на основе общесистемного подхода к созданию ИТКС, изображенного на рис. 1, который иллюстрирует приоритетность требований бизнеса перед другими компонентами.

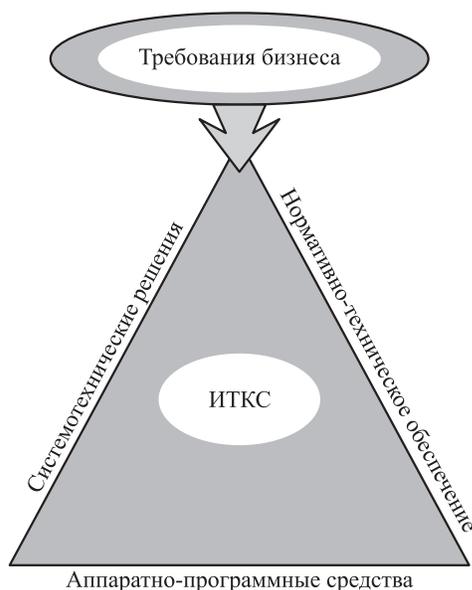


Рис. 1 Подход к созданию ИТКС

## 2 Управление техническим процессом анализа требований

Применяя данный подход к решению поставленных выше задач, проведем анализ:

- зависимости требований бизнеса к своевременности доставки сообщений и требований к производительности используемых для этого трактов ИТКС (рис. 2);
- зависимости требований бизнеса к своевременности выполнения работ в соответствии с доставленным сообщением и требований к числу

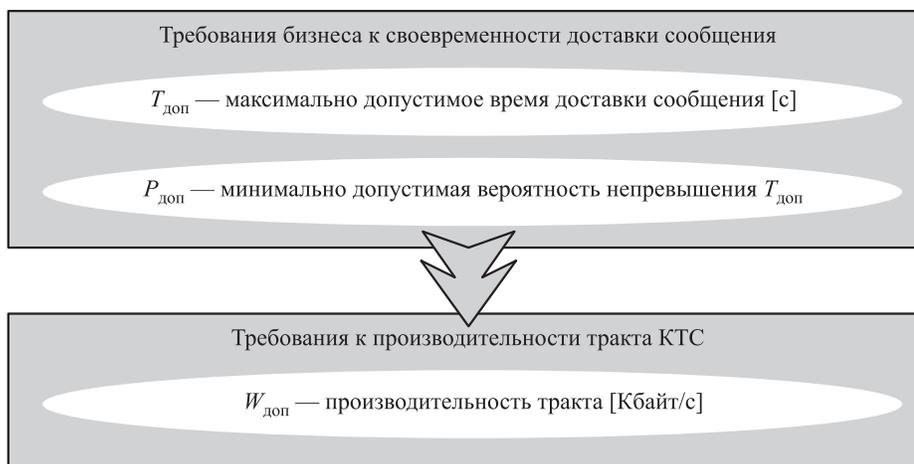


Рис. 2 Обоснование требований к производительности тракта

субъектов функциональной организационной структуры, созданной для выполнения этих работ (рис. 3).

Технический процесс управления анализом требований в рассматриваемой здесь области включает в себя следующие основные функции:

- разработка модели КТС как системы бизнес-класса по отношению к ИТКС;
- разработка методического подхода к обоснованию требований к производительности трактов КТС и к числу субъектов функциональной организационной структуры.

Модель для описания КТС приведена на рис. 4. В состав КТС входят следующие, основные для рассматриваемых задач, части:

- технические источники информации — датчики  $D_{1-1}, \dots, D_{1-N_1}$ , осуществляющие автоматический контроль параметров объектов бизнеса, оснащенные преобразователями сигналов контроля в информационные сообщения; тип групповых трактов ИТКС, используемых для передачи сообщений, поступивших из технических источников информации, назовем техническим трактом КТС;
- субъективные источники информации — субъекты бизнеса, оснащенные персональными компьютерами  $D_{1-1}, \dots, D_{1-N_n}$  для формирования сообщений об объектах; тип групповых трактов ИТКС, используемых для передачи сообщений из субъективных источников информации, назовем информационным трактом КТС;
- функциональная организационная структура — субъекты бизнеса, оснащенные персональными компьютерами и выполняющие работы в соответствии с поступающими из трактов КТС сообщениями.

### 3 Методика обоснования требований к производительности тракта

Методика обоснования требований к производительности  $W_{\text{доп}}$  группового тракта КТС состоит из следующих шагов:

*Шаг 1.* Определение вероятностей  $P_j$  того, что поступившее в тракт сообщение застанет в нем  $j$  требований, где  $j = 0, 1, 2, \dots$

*Шаг 2.* Выполнение последовательного суммирования вероятностей  $P_j$  для значений  $j = 0, 1, 2, \dots$  до тех пор, пока выполняется условие

$$\sum_{j=0}^{J_{\text{доп}}} P_j \leq P_{\text{доп}}, \quad (1)$$

где  $J_{\text{доп}}$  — максимально допустимое число сообщений в тракте, при котором еще выполняется требование к вероятности  $P_{\text{доп}}$  непревышения максимально допустимого времени  $T_{\text{доп}}$  доставки сообщения (см. рис. 2). Условие (1) эквивалентно условию:

$$J_{\text{доп}} T_{\text{обсл}} \leq T_{\text{доп}},$$

где  $T_{\text{обсл}}$  — временной интервал обслуживания (ВИО) одного сообщения.

*Шаг 3.* Определение минимально допустимой производительности тракта КТС:

$$W_{\text{доп}} = \frac{J_{\text{доп}} K_{\text{обсл}}}{T_{\text{доп}}},$$

где  $K_{\text{обсл}}$  — объем передаваемого в тракт сообщения, для технических трактов это постоянная величина, для информационных трактов используется среднее значение.

При определении вероятностей  $P_j$  для информационных трактов КТС применяется аппарат теории массового обслуживания, например при расчетах можно использовать представленные в работах [2, 4] программно-математические и имитационные средства моделирования различных систем массового обслуживания.

Для обоснования требований к производительности  $W_{\text{доп}}$  технического тракта КТС разработана модель тракта, приведенная на рис. 5.

В случаях, когда загрузка  $\rho$  тракта меньше 1, учитывая известный из теории массового обслуживания подход к аппроксимации  $N$  детерминированных потоков суммирующим пуассоновским потоком, вероятности  $P_j$  можно определить с помощью формул [4]:

$$P_j = \begin{cases} 1 - \rho & \text{для } j = 0; \\ (1 - \rho)(e^\rho - 1) & \text{для } j = 1; \\ (1 - \rho) \sum_{k=1}^j \left\{ (-1)^{j-k} e^{k\rho} \times \right. \\ \left. \times \left[ \frac{(k\rho)^{j-k}}{(j-k)!} + \frac{(k\rho)^{j-k-1}}{(j-k-1)!} \right] \right\} & \text{для } j \geq 2. \end{cases} \quad (2)$$

В то же время очевидно, что технический тракт КТС может быть загружен полностью ( $\rho = 1$ ). Этому состоянию тракта соответствует условие:

$$T_{\text{пост}} = N T_{\text{обсл}}.$$

Выделим следующие особенности функционирования технических трактов КТС:

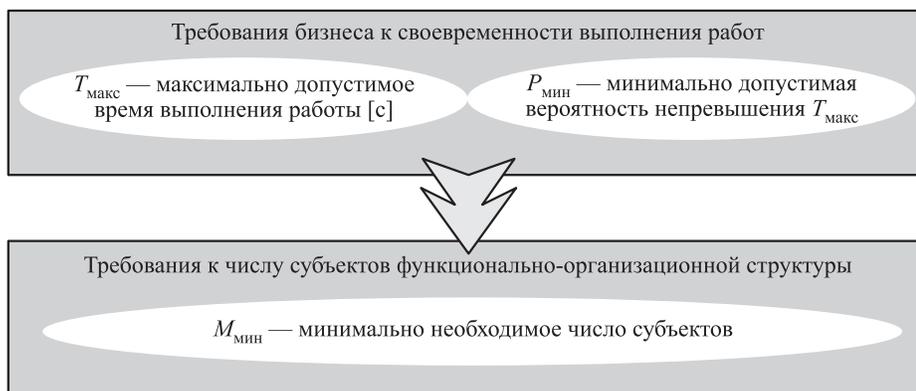


Рис. 3 Обоснование требований к числу субъектов организационной структуры

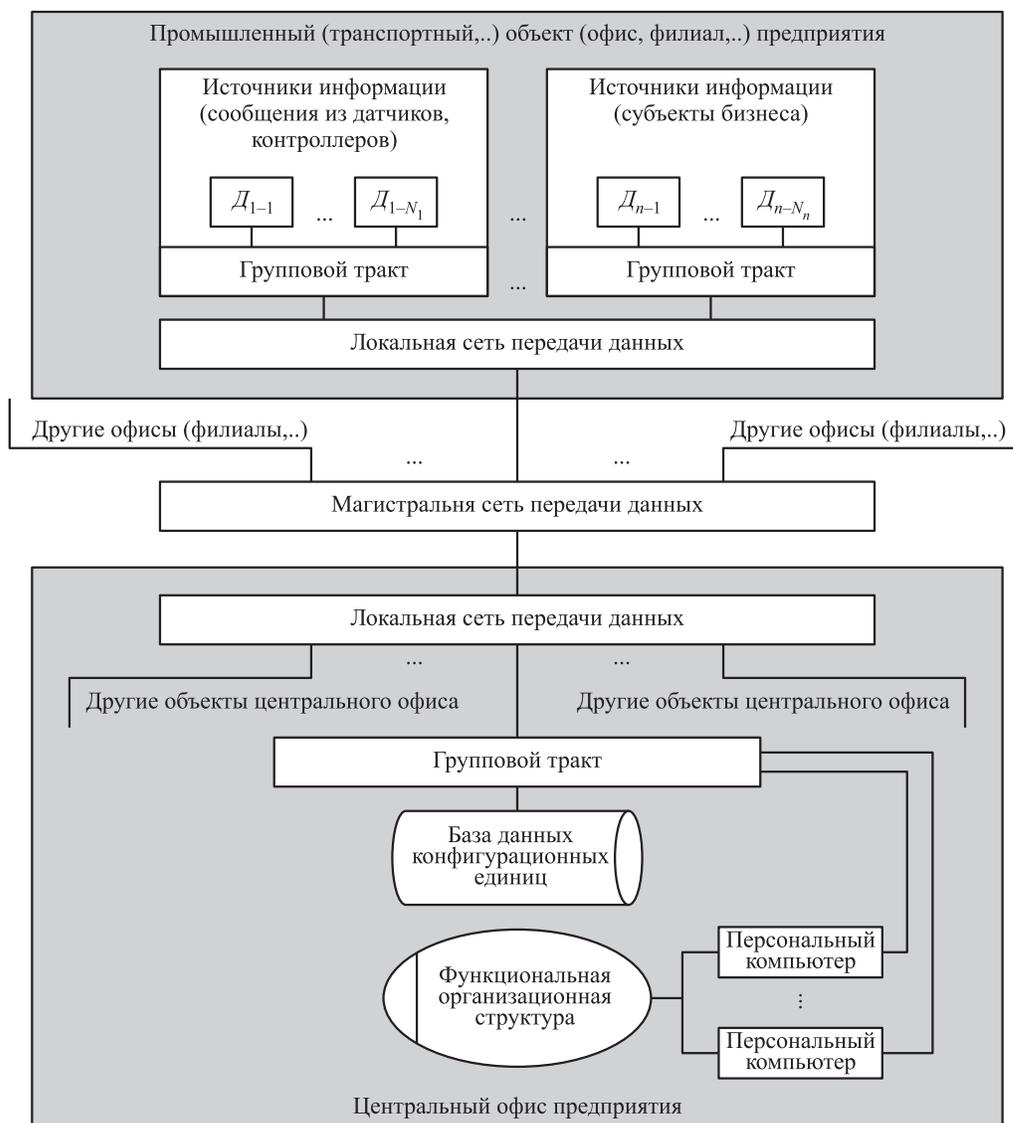


Рис. 4 Модель КТС

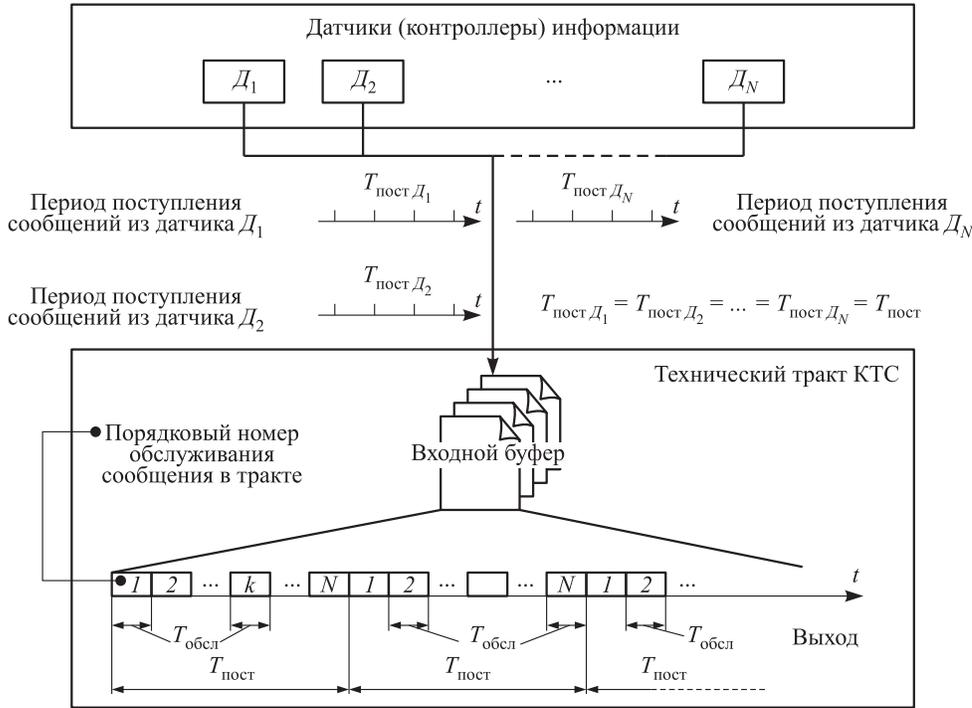


Рис. 5 Модель технического тракта КТС:  $T_{\text{пост}}$  — период поступления сообщения из каждого датчика;  $N$  — число датчиков, подключенных к тракту (число детерминированных потоков)

- поступление сообщения в тракт от каждого из  $N$  датчиков в каждом периоде  $T_{\text{пост}}$  может произойти равновероятно в любом  $k$ -м ВИО, где  $1 \leq k \leq N$ , независимо от моментов поступления сообщений от других датчиков;
- число сообщений  $j$ , которое застаёт в тракте вновь поступившее  $k$ -е сообщение, определяется не только его местом в периоде  $T_{\text{пост}}$ , но и тем, какое число сообщений застаёт в тракте предыдущее  $(k - 1)$ -е сообщение.

Данные особенности отражает граф вероятностей состояний тракта (рис. 6).

На этом графе  $P_N^k(j)$  — вероятность того, что к моменту поступления  $k$ -го ( $1 \leq k \leq N$ ) по порядку в периоде  $T_{\text{пост}}$  сообщения в тракте находится  $j$  ( $0 \leq j \leq k - 1$ ) сообщений при известном числе  $N$  источников информации; при этом соблюдается условие:

$$\sum_{j=0}^{k-1} P_N^k(j) = 1.$$

Тогда вероятности  $P_j$  того, что поступившее в тракт сообщение застанет в нем  $j$  сообщений ( $0 \leq j \leq N - 1$ ) при известном числе  $N$  источников информации будут равны:

$$P_j = P_N(j) = \frac{1}{N} \sum_{k=j+1}^N P_N^k(j), \quad (3)$$

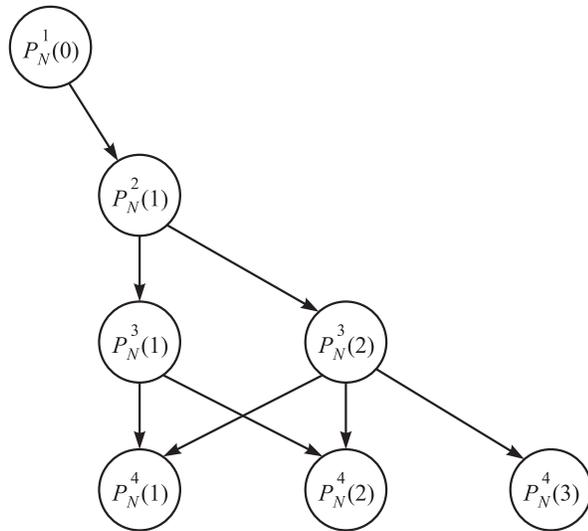
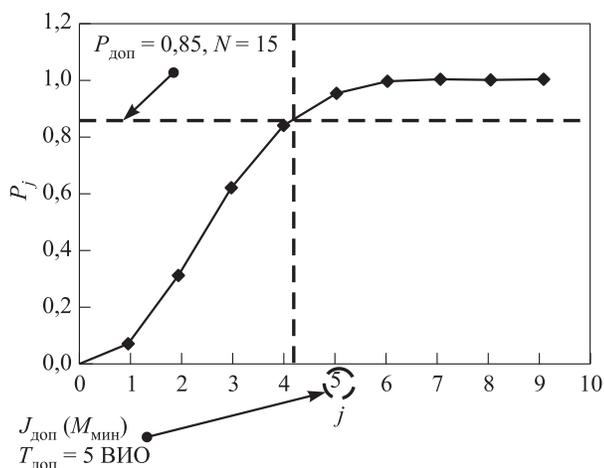


Рис. 6 Граф вероятностей состояний технического тракта КТС

где

$$\sum_{j=0}^{N-1} P_N(j) = 1. \quad (4)$$

С помощью аппарата теории вероятностей и комбинаторного раздела математики, путем выяв-



**Рис. 7** Пример расчета минимально допустимой производительности тракта:  $P_j$  — вероятность нахождения в тракте  $j$  сообщений;  $J_{\text{доп}}(M_{\text{мин}})$  — допустимое число сообщений в тракте (число субъектов функциональной организационной структуры);  $T_{\text{доп}} = 5$  ВИО — максимально допустимое время доставки сообщения;  $N$  — число датчиков

ления и анализа всевозможных комбинаций событий, в своей совокупности определяющих вероятности состояний тракта в различных ВИО периода  $T_{\text{пост}}$  для различных значений  $N$  источников информации, в работе [5] получены следующие формулы для вероятностей  $P_N^k(j)$ :

$$P_N^1(0) = 1; \quad (5)$$

$$P_N^2(1) = 1; \quad (6)$$

$$P_N^k(j) = \frac{(N-1)}{N^{N-2}} \left\{ \frac{(k-j)^{k-j-2}}{(k-j-1)!} \times \left( \frac{(N-k+j+1)^{N-k+j-1}}{(N-k+j)!} - \sum_{m=1}^{j-1} \frac{(N-k+j)^{N-k+j-m-1}}{(m-1)!(N-k+j-m)!} \right) + \sum_{x=1}^{j-1} \left( \sum_{y=0}^x (-1)^y (x-y+1)^y \times \frac{(k-j+x-y)^{k-j+x-y-2}}{y!(k-j+x-y-1)!} \times \left( \frac{(x+1)(N-k+j+1)^{N-k+j-x-1}}{(N-k+j-x)!} - \sum_{z=0}^{j-x-1} \frac{(N-k+j)^{l-k+j-x-z-1}}{z!(N-k+j-x-z)!} (x+z) \right) \right) \right\} \quad (7)$$

для  $N = 3, 4, \dots; k = 3, 4, \dots, N; j = 1, 2, \dots, k-1$ ;

$$P_N^k(0) = 0 \text{ для } k = 2, 3, \dots, N. \quad (8)$$

Подставляя (5)–(8) в (3) и выполняя действия в соответствии с шагами 2 и 3 вышеприведенной методики, получим значения  $W_{\text{доп}}$  минимально допустимой производительности технического тракта КТС для различного числа  $N$  источников информации. На рис. 7 приведен пример расчета допустимого числа  $J_{\text{доп}}$  сообщений, при котором соблюдается вероятность  $P_{\text{доп}}$  непревышения максимально допустимого времени  $T_{\text{доп}}$  доставки сообщения (выбирается большее целое,  $J_{\text{доп}} = 5$ ).

## 4 Функциональная организационная структура. Информация

Методика обоснования требований к числу  $M_{\text{мин}}$  субъектов функциональных организационных структур, выполняющих работы в соответствии с принимаемыми из трактов КТС сообщениями, состоит из следующих шагов:

*Шаг 1.* Определение вероятностей  $P_j$  того, что при поступлении в функциональную организационную структуру нового сообщения в этой структуре уже выполняется  $j$  работ ( $j = 0, 1, 2, \dots$ ). Эти действия выполняются так же, как и при обосновании производительности трактов КТС: с помощью теории массового обслуживания, например с помощью формул (2) либо, если заданы значения максимально допустимого, как правило, нормированного в регламенте организационной структуры, времени  $T_{\text{макс}}$  выполнения работ, с помощью аппарата теории вероятностей, например с помощью формул (3)–(8).

*Шаг 2.* Выполнение последовательного суммирования вероятностей  $P_j$  для значений  $j = 0, 1, 2, \dots$  до тех пор, пока выполняется условие

$$\sum_{j=0}^{M_{\text{мин}}} P_j \leq P_{\text{мин}},$$

где  $M_{\text{мин}}$  — минимально допустимое число субъектов, при котором выполняется требование к вероятности  $P_{\text{мин}}$  непревышения времени  $T_{\text{макс}}$ , параметр  $M_{\text{мин}}$  определяется аналогично параметру  $J_{\text{доп}}$ , например, как показано на рис. 7, при этом

$$M_{\text{мин}} T_{\text{обсл}}^1 \leq T_{\text{макс}},$$

где  $T_{\text{обсл}}^1$  — время выполнения одной работы, приведенное к одному субъекту.

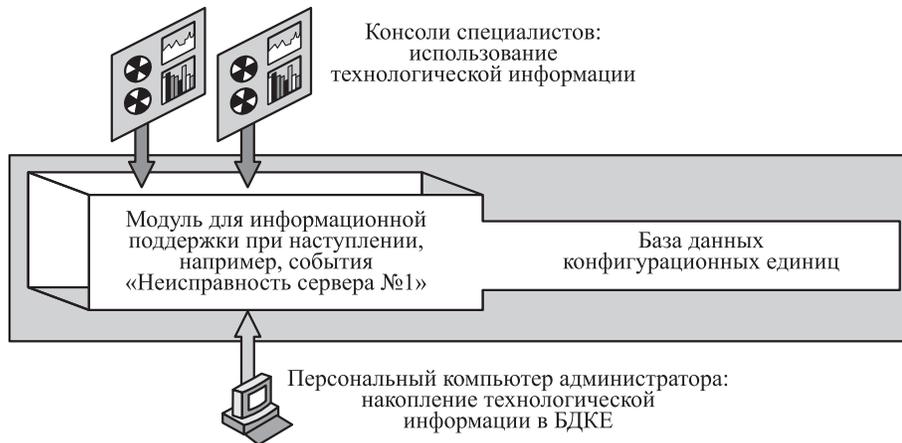


Рис. 8 Пример модели накопления и использования технологической информации

Современным функциональным организационным структурам КТС присущи следующие свойства: формализуемость управленческой деятельности; автоматизация процессов накопления и использования технологической информации о предметных сущностях для выполнения работ; их исчисляемость. На рис. 8 приведен пример модели накопления и использования информации.

Указанные свойства позволяют принять следующие гипотезы:

- существует прямая зависимость объема технологической информации от количества предметных сущностей;
- существует прямая зависимость уровня знаний субъектов от объема накопленной и освоенной ими технологической информации;
- существует обратная зависимость времени выполнения работы от уровня знаний, доступ к

которым имеют субъекты; косвенно данное утверждение подтверждается результатами исследований уровня знаний у работников в области информационных технологий и его влиянием на эффективность их деятельности [6].

Данные гипотезы приводят к утверждению, что нормированные значения времени  $T_{\text{макс}}$  выполнения работ могут со временем пересматриваться в сторону уменьшения в соответствии с определенными законами  $A(t)$ , свойственными различным КТС:

$$T_{\text{макс}} = T_0 - \int_0^{\infty} A(t) dt.$$

Выбор варианта подготовки технологических данных для КТС — это самостоятельная задача, во многом определяемая условиями функционирования конкретного предприятия. Приведенные на

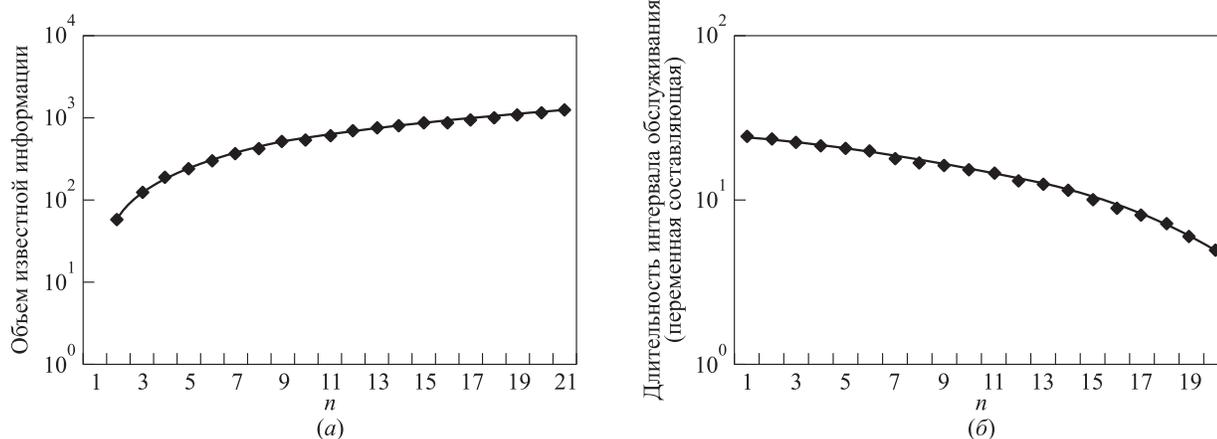


Рис. 9 Примеры оценки влияния объема технологической информации: (а) объем известной информации, (б) длительность интервала обслуживания (переменная составляющая);  $n$  — порядковые номера обновлений

рис. 9 примеры [7] позволяют приблизительно оценить эффект от накопления и использования технологических данных.

## 5 Заключение

Рассмотренный в статье методологический подход к управлению сложными инфокоммуникационными проектами относится к стадии замысла в жизненном цикле ИТКС и к реализации технического процесса по анализу требований. Использование на практике данного подхода к определению производительности трактов КТС и к определению состава их организационных структур для предприятий государственного сектора и коммерческих предприятий позволит исключить или значительно снизить инвестиционные риски при строительстве систем и последовательно повышать эффективность управления при их эксплуатации.

## Литература

1. *Зацаринный А. А.* Основные принципы системного подхода при проектировании, внедрении и развитии современных корпоративных сетей // Системы и средства информатики. — М.: Наука, 2002. Вып. 12. С. 58–66.
2. *Костогрызов А. И., Степанов П. В.* Инновационное управление качеством и рисками в жизненном цикле систем. — М.: ВПК, 2008. 404 с.
3. *Зацаринный А. А., Ионенков Ю. С., Козлов С. В.* Некоторые вопросы проектирования информационно-телекоммуникационных систем. — М.: ИПИ РАН, 2010. 218 с.
4. *Саати Т. Л.* Элементы теории массового обслуживания и ее приложения. — М.: Сов. радио, 1971.
5. *Шабанов А. П.* О распределении времени ожидания внутри интервала занятости входного группового тракта устройства коммутации цифровых каналов // МРС ТТЭ. Сер. О, 1981. Вып. 11. Деп. в ВИМИ.
6. *Гусев А. В., Сурков С. А.* Влияние информационных технологий на эффективность использования результатов бизнес-образования // Применение новых технологий в образовании: XV Междунар. конф. — Троицк: Байтик, 2004. С. 53.
7. *Аракелян М. А., Шабанов А. П.* Технологические данные в ИТ-поддержке бизнеса // Директор информационной службы, 2007. № 1.

# МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ФОРМИРОВАНИЯ ЭКСПЕРТНЫХ ЗНАНИЙ ДЛЯ МОНИТОРИНГА ПРОГРАММНО-ЦЕЛЕВОЙ ДЕЯТЕЛЬНОСТИ\*

И. М. Зацман<sup>1</sup>, А. А. Дурново<sup>2</sup>

**Аннотация:** Рассматривается постановка проблемы компьютерного представления целевых систем знаний (ЦСЗ) об индикаторах мониторинга и предлагается ее решение, состоящее из четырех компонентов: (1) стационарной модели компьютерного представления ЦСЗ об индикаторах; (2) множества точек, на основе которого определяются значения количественных характеристик процессов формирования ЦСЗ; (3) нестационарной модели компьютерного представления ЦСЗ; (4) проективного словаря системы мониторинга для компьютерного представления ЦСЗ об индикаторах в динамике их формирования. Первые три компонента представляют собой теоретическую часть решения этой проблемы, а четвертый — прикладную часть ее решения.

**Ключевые слова:** проблема представления целевых систем знаний об индикаторах; лакуарность систем знаний об индикаторах; семиотические модели представления знаний об индикаторах; концепты индикаторов; денотаты индикаторов

## 1 Введение

В процессе реформирования бюджетного процесса программно-целевая деятельность стала доминирующей во всех сферах, где расходуются бюджетные средства. Согласно Концепции реформирования бюджетного процесса с 2004 г. в РФ осуществляется переход на реализацию долгосрочных программ с явным описанием их целей, задач, ресурсов и ожидаемых результатов, а также с использованием индикаторов для мониторинга и оценивания полученных результатов, эффективности и результативности программно-целевой деятельности [1].

Применение индикаторов, традиционно используемых при статистическом наблюдении бюджетного процесса до его реформирования, выявило неполноту имеющихся наборов индикаторов. Так, они не позволяют оценивать согласованность поставленных целей и имеющихся ресурсов. Необходимость разработки новых индикаторов стимулировала проведение исследований с целью формирования новых концептуальных подходов и методов проектирования индикаторов, так как в бюджетном процессе принятие решений стало существенно зависеть от результатов мониторинга и индикаторного оценивания программно-целевой деятельности.

Целенаправленное формирование знаний, концептуальных подходов и создание новых методов

для разработки программно-ориентированных индикаторов стало актуальной проблемой для многих сфер деятельности, что нашло отражение в официальных документах. Так, задачи разработки новых методов оценки эффективности научной деятельности, создания систем мониторинга, анализа и оценки результатов деятельности юридических и физических субъектов сферы науки включены в тематическое направление «29. Системы автоматизации, CALS-технологии, математические модели и методы исследования сложных управляющих систем и процессов» Программы фундаментальных научных исследований РАН на 2008–2012 годы [2].

Кроме прикладного значения исследования процессов целенаправленного формирования знаний имеют фундаментальный научный аспект. В настоящее время формируется новое научное направление (новая проблематика), объединяющее исследование процессов целенаправленного формирования новых знаний и их компьютерного представления. С прикладной точки зрения это направление связано с широким спектром сфер деятельности.

Становление этого направления во многом было обусловлено 7-й Рамочной программой Европейского Союза (ЕС). В документах этой программы, принятой на период с 2007 по 2013 гг., для сферы информационно-коммуникационных технологий (ИКТ) определен ряд актуальных направлений

\* Работа выполнена при поддержке РФФИ, грант № 09-07-00156.

<sup>1</sup> Институт проблем информатики Российской академии наук, iz\_ipi@al170.ipi.ac.ru

<sup>2</sup> Институт проблем информатики Российской академии наук, duralex49@mail.ru

исследований и разработок, включая исследования возможностей существующих и разработку новых видов ИКТ, обеспечивающих представление в информационных системах формируемых знаний и целенаправленное влияние на процесс их формирования [3–7].

Результаты анализа этих направлений, приведенные в работе [8], позволяют утверждать, что ИКТ играют ключевую роль в исследованиях процессов целенаправленного формирования знаний, в том числе в решении следующих проблем:

- формирование и компьютерное представление в цифровой электронной среде (далее — цифровая среда)<sup>1</sup> личностных и согласованных концептов как структурных элементов ЦСЗ, формируемых экспертами, — *проблема представления ЦСЗ*;
- анализ и оценивание степени релевантности разных вариантов ЦСЗ социальным, экономическим, технологическим и другим общественно значимым потребностям, в интересах которых они были сформированы, — *проблема релевантности ЦСЗ*;
- целенаправленное влияние средствами ИКТ на формирование и эволюцию ЦСЗ, необходимое для получения запланированных результатов, — *проблема направляемого развития ЦСЗ*.

Данная статья относится к первой из трех перечисленных проблем, т. е. к проблеме представления ЦСЗ. Эта проблема рассматривается с точки зрения создания систем информационного мониторинга и индикаторного оценивания программно-целевой деятельности (ПЦД). Предполагается, что в интересах мониторинга и индикаторного оценивания группа экспертов должна сформировать или пополнить ЦСЗ о программно-ориентированных индикаторах.

Модели формирования новых или развития существующих систем знаний по степени их общности предлагается разделить на две основные категории: (1) концептуальные модели, которые не зависят от той предметной области, где происходят процессы формирования или пополнения систем знаний; (2) предметно-ориентированные модели, которые зависят от этой предметной области.

Широко известная в настоящее время спиральная модель формирования знаний, которая описана в работах Нонака и Такеучи [9–11]<sup>2</sup>, и обобщенный вариант этой модели, предложенный в работах

Вежбицки и Накамори [12, 13], не зависят от предметной области и с точки зрения предложенного деления на две категории являются концептуальными.

Опыт применения ИКТ, разработанных в исследовательском институте JAIST (Japan Advanced Institute of Science and Technology) для реализации спиральной модели в интересах поддержки процесса формирования научных знаний [14], позволяет утверждать, что, с одной стороны, спиральная модель и ее обобщение уже используются на практике, с другой стороны, остался ряд нерешенных вопросов, ограничивающих сферу их использования.

Перечислим основные вопросы, которые не были решены в рамках спиральной модели и ее обобщенного варианта:

- не фиксируются изменения состояния формируемых личностных знаний человека в зависимости от времени;
- не определены объекты интерпретации, являющиеся источниками новых знаний человека;
- не выделены структурные элементы формируемых знаний, соответствующие объектам интерпретации;
- не фиксируются моменты времени начала и завершения генерации каждого нового структурного элемента знаний.

В интересах решения вышеперечисленных вопросов на основе спиральной модели были разработаны концептуальные основы [15] и семиотические модели целенаправленного формирования и компьютерного представления формируемых знаний, которые относятся к категории концептуальных моделей. Разработанные стационарная и нестационарная семиотические модели не зависят от предметной области, в которой происходят процессы формирования или пополнения систем знаний, что иллюстрируется примерами из разных предметных областей [16–18].

Основной целью данной работы является описание *предметно-ориентированных* семиотических моделей процессов целенаправленного формирования знаний об индикаторах, формируемых группой экспертов для мониторинга и индикаторного оценивания ПЦД в сфере науки. Еще одной целью является описание концептуальных и технических решений по созданию проективного словаря [18], с помощью которого был проведен эксперимент

<sup>1</sup> Согласно ГОСТ Р 52292-2004, электронная среда — это среда технических устройств (аппаратных средств), функционирующих на основе физических законов и используемых в информационной технологии при обработке, хранении и передаче данных. Цифровая электронная среда — это цифровые технические устройства (аппаратные средства) электронной среды, функционирующие на основе физических законов и используемые в информационной технологии при обработке, хранении и передаче данных.

<sup>2</sup> По данным Google Scholar, на 17.06.2011 общее число цитирований работ [9, 10] равно 22 344.

по реализации компьютерного представления экспертных знаний об индикаторах с отражением динамики их формирования.

## 2 Процессы целенаправленного формирования знаний об индикаторах

В предлагаемой постановке проблемы представления ЦСЗ ключевыми являются процессы концептуализации и интерпретации, т. е. формирование смыслового содержания индикаторов при их разработке экспертами. Понятия «концептуализация» и «интерпретация» являются ключевыми и для решения проблемы представления ЦСЗ.

*Концептуализацию* определим как итерационный процесс формирования экспертом в течение некоторого периода времени концепта (смыслового содержания) нового индикатора. Концепт индикатора как структурный элемент знаний эксперта является результатом анализа алгоритма изменяемой экспертом компьютерной программы и данных, используемых для определения значений этого индикатора, а также его значений, вычисленных этой программой.

Для обозначения объекта содержательного анализа, который состоит из компьютерной программы, данных и вычисленных значений индикатора, в статье используется термин «денотат индикатора», или просто «денотат», когда речь идет о стационарной модели, или термин «состояние денотата», когда речь идет о нестационарной модели.

Предполагается, что на каждой итерации группа экспертов может формировать несколько новых индикаторов одновременно и что каждая итерация концептуализации состоит из двух стадий: стадии изменения денотатов формируемых индикаторов (результат изменения будем называть новым состоянием денотата) и стадии интерпретации. Последовательные моменты времени начала стадий интерпретации обозначим как  $\{t_i - \Delta_i, \Delta_i < t_i - t_{i-1}, t_0 = 0, t_i > t_{i-1}, i = 1, 2, \dots\}$ , а завершения стадий интерпретации обозначим как  $\{t_i, i = 1, 2, \dots\}$ .

Предполагается, что на стадии интерпретации компьютерная программа и данные, используемые этой программой для определения значений индикатора, остаются неизменными. Изменения могут происходить только на стадиях изменения денотатов формируемых индикаторов, т. е. на интервалах времени  $[t_{i-1}, t_i - \Delta_i]$ .

*Интерпретацию* определим как процесс содержательного анализа денотата, выполняемый экспертом на одном из интервалов времени  $[t_i - \Delta_i, t_i]$ ,

или состояния денотата индикатора, если программа и/или данные изменяются в течение концептуализации.

Предполагается, что группа экспертов, разрабатывающая новые индикаторы, руководствуется явно заданными целями генерации ЦСЗ о программно-ориентированных индикаторах. Между любыми двумя последовательными стадиями интерпретации эксперты могут вносить изменения в алгоритмы компьютерных программ и в данные, используемые для определения значений индикаторов.

Если на некоторой стадии интерпретации денотата было сгенерировано и описано смысловое содержание (концепт) нового индикатора, а затем между двумя последовательными стадиями интерпретации был изменен алгоритм и/или данные (было получено новое состояние денотата), то на следующей стадии может быть сгенерирован другой концепт разрабатываемого индикатора как результат интерпретации нового состояния денотата.

Используя вышеопределенные понятия «концептуализация» и «интерпретация», **сформулируем проблему представления ЦСЗ** о новых программно-ориентированных индикаторах следующим образом. Предположим, что перед группой экспертов поставлена задача формирования или пополнения ЦСЗ об индикаторах, предназначенных для мониторинга и оценивания ПЦД, используя некоторую существующую систему классификации индикаторов.

Процесс формирования ЦСЗ об индикаторах представляет собой итерационный процесс концептуализации денотатов, целью которого является формирование новых программно-ориентированных индикаторов. Каждое состояние денотата индикатора (как зафиксированный в цифровой среде объект интерпретации) представляет собой совокупность трех следующих компонентов, которые не изменяются на стадии интерпретации:

- (1) программы определения значений нового индикатора;
- (2) данных, которые обрабатываются программой определения его значений;
- (3) значений нового индикатора, которые получены в результате выполнения программы.

Между стадиями интерпретации эксперты имеют возможность изменять программы и данные, которые обрабатываются программой. Эксперт может написать новую (изменить существующую) программу определения значений индикатора и/или подготовить (изменить) данные, которые обрабатываются этой программой, затем выполнить этот

вариант программы и получить результаты вычислений. На каждой стадии интерпретации эксперт анализирует алгоритм программы, использованные данные и полученные значения с целью формирования на этой стадии его личностного концепта (смыслового содержания) варианта нового индикатора.

Каждая стадия интерпретации денотата индикатора необязательно завершается формированием концепта индикатора (например, эксперт изменил алгоритм и/или данные, но не смог интерпретировать полученное состояние денотата). В случае формирования концепта эксперт выражает свое личностное понимание индикатора сначала в виде позиции (рубрики) в заданной системе классификации индикаторов, а затем дефиницией. Также он может присвоить индикатору имя.

Кроме того, он начинает процесс согласования полученного им личностного концепта, выраженного в виде рубрики и дефиниции, а также имени индикатора с другими экспертами группы. Цель процесса согласования — убедить других экспертов в корректности построения и правильности своей интерпретации индикатора, представленной в виде рубрики в заданной системе классификации и дефиниции, а также обосновать выбор имени индикатора. Если он получит согласование еще хотя бы одного эксперта, то концепт этого индикатора получает статус согласованного концепта.

Каждая стадия интерпретации индикатора может завершиться одним из четырех вариантов следующих результатов, которые регистрируются в лингвистическом и других видах обеспечения системы информационного мониторинга:

- (1) написана новая (изменена существующая) программа вычисления значений варианта нового индикатора и/или подготовлены (изменены) данные, которые обрабатываются этой программой, и получены результаты вычислений, но личностный концепт не сформирован (получено новое неинтерпретированное состояние денотата, иначе говоря, понимание этого состояния денотата отсутствует);
- (2) сформирован личностный концепт и определена его позиция (рубрика) в заданной системе классификации, но отсутствуют дефиниция как вербальное описание концепта нового индикатора и имя для полученного варианта нового индикатора;
- (3) дополнительно экспертом сформулирована дефиниция как вербальное описание концепта нового индикатора и выбрано для него имя;

- (4) позиция (рубрика) в заданной системе классификации, дефиниция и имя нового индикатора согласованы с одним или несколькими экспертами группы.

На следующей стадии интерпретации эксперт может сформировать новый денотат (изменить существующий денотат, получив новое его состояние) и повторить перечисленные операции или продолжить процесс согласования ранее созданных индикаторов с другими экспертами группы. Если в последнем случае согласование состоялось, то результатом будет изменение степени согласованности индикаторов между экспертами. Предполагается, что задан некоторый пороговый уровень степени согласованности, при достижении которого индикатор получает статус *сформированного*.

При перечисленных исходных условиях и вышеописанных вариантах формирования индикаторов требуется:

- определить пространство компьютерных кодов состояний денотатов индикаторов, их концептов и имен, включающее ось времени, для отражения в этом пространстве результатов концептуализации денотатов индикаторов и при этом зафиксировать моменты времени начала и завершения каждой стадии интерпретации;
- разработать модели компьютерного представления экспертных знаний для описания процессов формирования индикаторов в пространстве компьютерных кодов состояний денотатов, концептов и имен индикаторов.

Существенной особенностью сформулированной проблемы является представление результатов концептуализации денотатов индикаторов в пространстве компьютерных кодов состояний денотатов, концептов и имен индикаторов. Сама постановка этой проблемы говорит о том, что в ней охватываются сущности различной природы из трех разных сред: ментальной (концепты индикаторов), социально-коммуникационной (имена индикаторов, присвоенные экспертами) и цифровой (три компонента денотатов, а также компьютерные коды концептов, имен и состояний денотатов, присвоенные системой информационного мониторинга).

В приведенной постановке решением сформулированной проблемы представления ЦСЗ будем считать создание совокупности таких компонентов, как:

- (1) стационарная модель компьютерного представления знаний о *неизменяемых* программно-ориентированных индикаторах, которые

являются составной частью ЦСЗ, и модель компьютерного представления в некоторый момент времени формируемых знаний об *изменяемых* индикаторах;

- (2) множество точек пространства действительных чисел  $R_2$ , на основе которого определяются значения количественных характеристик процесса формирования ЦСЗ (это множество точек будем называть пространством Фреге);
- (3) нестационарная модель компьютерного представления ЦСЗ об *изменяемых* программно-ориентированных индикаторах, формируемых группой экспертов;
- (4) компонент лингвистического обеспечения системы информационного мониторинга для компьютерного представления ЦСЗ в динамике формирования индикаторов, который будем называть *проективным словарем*. Этот словарь представляет собой один из возможных вариантов технической реализации моделей компьютерного представления ЦСЗ.

Первые три из четырех компонентов решения сформулированной проблемы представления ЦСЗ, которым посвящен следующий раздел статьи, представляют собой теоретическую часть решения этой проблемы. Четвертый компонент, которому посвящен раздел 4, представляет собой один из вариантов прикладного ее решения.

### 3 Предметно-ориентированные семиотические модели

#### 3.1 Первый этап решения проблемы

На первом этапе решения этой проблемы определим стационарную семиотическую модель, которая предназначена для компьютерного кодирования в системе информационного мониторинга ПЦД экспертных знаний о неизменяемых индикаторах, уже сформированных группой экспертов. Затем определим нестационарную семиотическую модель, которая предназначена для компьютерного кодирования экспертных знаний об изменяемых индикаторах, формируемых группой экспертов.

Сформулируем три исходных положения для построения обеих моделей согласно вышеприведенной постановке проблемы представления ЦСЗ, используя рис. 1.

**Положение 1.** В процессе целенаправленного формирования знаний об индикаторах для каждого индикатора должны кодироваться три его компонента: концепт (смысловое содержание), имя и денотат индикатора, если кодируемый индикатор является стабильным, или состояние денотата изменяемого индикатора в некоторый момент времени, если кодируемый индикатор является формируемым, т. е. нестабильным.

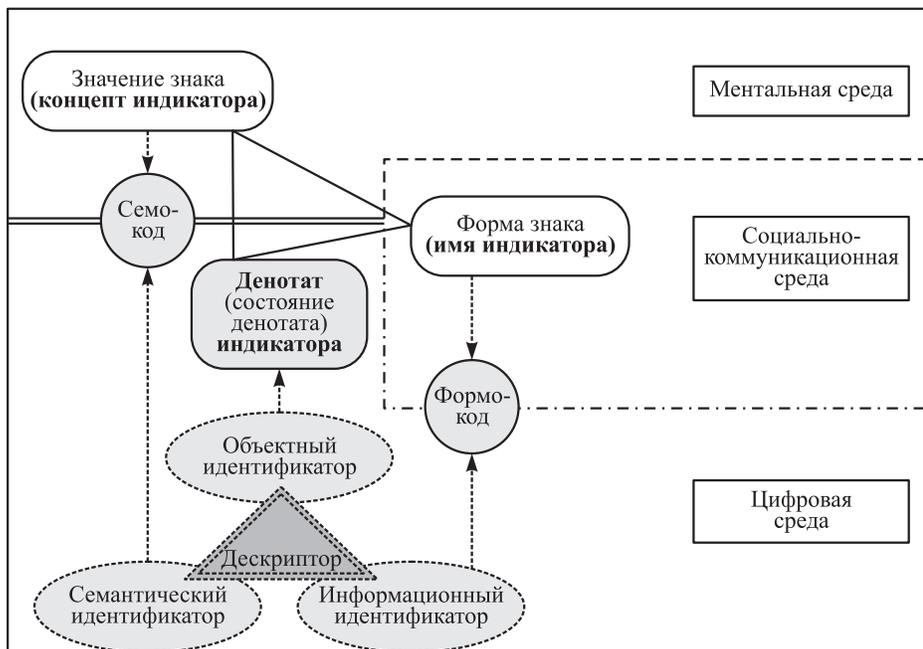


Рис. 1 Стационарная семиотическая модель индикатора, его семокод и формокод

Концепт, имя и денотат (состояние денотата) являются вершинами *треугольника Фреге* соответствующего индикатора<sup>1</sup> [19–21].

**Положение 2.** Концепт, имя и денотат индикатора имеют различную природу. Поэтому для построения стационарной семиотической модели будем использовать следующие три среды: ментальную, социально-коммуникационную и цифровую (см. рис. 1).

**Положение 3.** Каждый дескриптор, который строится экспертом в целях описания формируемого им индикатора, включает три уникальных (в рамках проективного словаря) идентификатора:

- (1) *семантический идентификатор*, предназначенный для компьютерного представления в системе информационного мониторинга концепта индикатора;
- (2) *информационный идентификатор*, предназначенный для компьютерного кодирования имени индикатора;
- (3) *объектный идентификатор*, предназначенный для кодирования денотата индикатора, если кодируемый индикатор является стабильным, или состояния денотата изменяемого индикатора, если кодируемый индикатор является формируемым, т. е. нестабильным.

Совокупность числовых представлений семантического, информационного и объектного идентификаторов дескриптора будем называть цифровым семиотическим треугольником по аналогии с треугольником Фреге.

**Определение 1.** Пусть для некоторого стабильного индикатора известны его концепт, имя и денотат, которые не изменяются во времени, относятся к ментальной, социально-коммуникационной и цифровой средам соответственно и образуют треугольник Фреге. Тогда **стационарной семиотической моделью** компьютерного кодирования этого индикатора называется его треугольник Фреге, для которого построен дескриптор и трем вершинам которого назначены семантический идентификатор для концепта индикатора, информационный идентификатор для имени и объектный идентификатор для денотата как совокупности трех компонентов: программы, данных и значений этого индикатора.

Рисунок 1 включает ментальную, социально-коммуникационную и цифровую среды, содержа-

щие соответственно концепт индикатора, его имя и денотат, которые и образуют треугольник Фреге (составлен тремя полужирными отрезками). Построенный дескриптор с тремя вершинами, которым согласно определению 1 назначены семантический идентификатор для концепта, информационный идентификатор для имени и объектный идентификатор для денотата этого индикатора, обозначен равнобедренным треугольником с двойными точечными сторонами. На этом же рисунке двумя кругами условно обозначены два понятия: «*семокод*» и «*формокод*», которые введем как следствие определения 1.

**Следствие 1.** Пусть для некоторого стабильного индикатора заданы семантический идентификатор для концепта индикатора и информационный идентификатор для его имени. Тогда **семокодом** этого индикатора называется совокупность концепта индикатора и уникального семантического идентификатора дескриптора индикатора, а **формокодом** этого индикатора называется совокупность имени индикатора и уникального информационного идентификатора дескриптора индикатора [15].

**Определение 2.** Пусть для некоторого нестабильного индикатора известны в некоторый момент времени состояние его денотата, концепт как структурный элемент экспертных знаний, полученный в результате интерпретации этого состояния, и имя этого индикатора. Тогда семиотической моделью компьютерного кодирования индикатора в этот момент времени называется соответствующий ему треугольник Фреге, для которого построен дескриптор и трем вершинам которого (концепту, имени и состоянию денотата) назначены семантический идентификатор для концепта, информационный идентификатор для имени и объектный идентификатор для состояния денотата как совокупности трех компонентов: программы, данных и значений индикатора в этот момент времени.

**Следствие 2.** Пусть для некоторого нестабильного индикатора заданы в некоторый момент времени семантический идентификатор для концепта индикатора и информационный идентификатор для его имени. Тогда **семокодом** этого индикатора называется совокупность концепта и уникального семантического идентификатора дескриптора индикатора, а

<sup>1</sup>Чтобы построить треугольник Фреге для изменяемого индикатора в некоторый момент времени, необходимо использовать понятие авторского знака индикатора, две стороны которого: форма знака (имя индикатора) и значение знака (концепт индикатора) — могут находиться в отношении временной связи. Эта связь является опосредованной сознанием автора этого знака и представляет собой нестабильное единство, которое посредством сенсорно воспринимаемой автором формы знака репрезентирует персонально приданное этому знаку значение. Значение авторского знака индикатора является результатом личностного анализа автором объекта интерпретации как состояния рассматриваемого денотата в момент времени, указанный автором этого знака при его регистрации в системе информационного мониторинга [8, 15].

формокодом этого индикатора называется совокупность имени индикатора и уникального информационного идентификатора дескриптора индикатора [15].

В отличие от концепта индикатора, его имени и денотата, понятия «семокод» и «формокод» являются двуедиными сущностями, которые относятся одновременно к двум средам. Поэтому семокод находится на границе ментальной и цифровой сред, а формокод — на границе социально-коммуникационной и цифровой сред (см. рис. 1).

Предложенные определения двух моделей и их следствия являются первым из трех компонентов теоретической части решения проблемы представления ЦСЗ об индикаторах. Этот компонент, полученный на первом этапе, представляет собой модели компьютерного кодирования вершин треугольника Фреге, принадлежащих разным средам, а также понятия «семокод» и «формокод», используемые при описании технических решений компьютерного кодирования концептов и имен индикаторов.

Модель компьютерного представления в некоторый момент времени формируемых знаний об изменяемых индикаторах понадобится далее для определения пространства Фреге и построения нестационарной модели. Отметим, что в определении 1 говорится о денотатах, так как стационарная модель используется для случая неизменяемого индикатора, а в определении 2 — о состояниях денотатов. При этом предполагается, что фиксируется момент времени завершения стадии интерпретации изменяемого индикатора, рассматривается соответствующее состояние денотата этого индикатора и семиотическая модель применяется для кодирования вершин треугольника Фреге только в этот момент времени.

### 3.2 Второй этап решения проблемы

На втором этапе решения проблемы представления ЦСЗ об индикаторах определим пространство Фреге для компьютерного представления процессов формирования ЦСЗ об индикаторах, формируемых экспертами. Кроме вышеперечисленных положений 1–3 дополнительно понадобятся следующие исходные положения, которые будем нумеровать начиная с номера 4.

**Положение 4.** На каждой итерации концептуализации группа экспертов может формировать несколько новых индикаторов одновременно, и каждая итерация состоит из двух стадий: стадии изменения денотатов формируемых индикаторов (результат изменения денотата будем называть новым его состоянием) и стадии интерпретации. Каждая стадия интерпретации может завершиться одним из

четырёх возможных вариантов окончания, которые были описаны в предыдущем разделе.

**Положение 5.** Результаты каждого из четырёх вариантов окончания стадии интерпретации должны регистрироваться в лингвистическом и других видах обеспечения системы информационного мониторинга по завершении каждой итерации концептуализации индикатора, в том числе регистрируются моменты времени начала и завершения каждой стадии интерпретации. Последовательные моменты времени начала стадий интерпретации были ранее обозначены как  $\{t_i - \Delta_i, \Delta_i < t_i - t_{i-1}, t_0 = 0, t_i > t_{i-1}, i = 1, 2, \dots\}$ , а моменты времени завершения стадий интерпретации — как  $\{t_i, i = 1, 2, \dots\}$ .

**Положение 6.** Для первого варианта окончания стадии интерпретации должны регистрироваться момент времени окончания стадии, используемый вариант программы вычисления значений варианта нового индикатора, данные, которые обрабатываются этим вариантом программы, и полученные результаты вычисления значений индикатора, т. е. фиксируется каждый момент времени формирования состояния денотата индикатора и само это состояние как совокупность трех перечисленных компонентов. В проективном словаре экспертом-разработчиком программы строится дескриптор, содержащий ссылки на каждый из трех компонентов, а также в нем запоминается персональный идентификатор этого эксперта.

**Положение 7.** Для второго варианта окончания стадии интерпретации дополнительно к требованиям положения 6 должна быть выбрана позиция (*рубрика для концепта индикатора*) в существующей системе классификации и указан персональный идентификатор эксперта, который выбрал эту рубрику.

**Положение 8.** Для третьего варианта дополнительно к требованиям положений 6 и 7 должны регистрироваться *дефиниция* и выбранное *имя* для нового индикатора, а также персональный идентификатор эксперта, который разработал эту дефиницию и выбрал имя.

**Положение 9.** Для четвертого варианта дополнительно к требованиям положений 6–8 должны регистрироваться результаты согласования в группе экспертов рубрики заданной системы классификации, дефиниции и имени нового индикатора, а также персональные идентификаторы соответствующих экспертов.

**Положение 10.** Ссылки на три компонента каждого нового состояния денотата индикатора, его рубрика в заданной системе классификации, дефиниция и имя, а также моменты времени начала и завершения каждой стадии интерпретации этого

индикатора должны храниться в системе информационного мониторинга в виде *атрибутов дескриптора* проективного словаря лингвистического обеспечения этой системы, построенного на данной стадии интерпретации индикатора.

**Положение 11.** Кроме ссылки на три компонента каждого нового состояния денотата индикатора дескриптор также должен включать атрибуты, фиксирующие его положение внутри проективного словаря, которые являются ссылками на другие дескрипторы (эти ссылки строятся на основе заданной классификации индикаторов [22]).

Используя вышеперечисленные положения, определим вариант пространства Фреге, предназначенный для количественного описания процессов целенаправленного формирования ЦСЗ об индикаторах, используя семантические, информационные и объектные идентификаторы дескрипторов состояний этих индикаторов, зарегистрированных в системе информационного мониторинга. Определенный ниже вариант пространства Фреге представляет собой предметно-ориентированную версию пространства Фреге, построенного в работе [17].

Состояния формируемых индикаторов будем фиксировать в последовательные моменты времени, которые обозначены как  $\{t_i, i = 1, 2, \dots\}$ , где  $t_i$  — момент времени *завершения*  $i$ -й стадии интерпретации индикаторов, в том числе построения дескрипторов новых состояний индикаторов. В каждый из этих моментов времени будем фиксировать описание концептов формируемых индикаторов (их рубрики и дефиниции); соответствующие им информационные объекты (имена индикаторов); три компонента состояния каждого из тех денотатов, которые были сгенерированы или изменены на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ , т. е. к моменту времени *начала*  $i$ -й стадии интерпретации.

За  $t_1$  принимается тот момент времени, когда эксперты начали фиксировать в системе информационного мониторинга процесс формирования ЦСЗ об индикаторах. Сам процесс формирования ЦСЗ мог начаться и до момента времени  $t_1$ , но момент времени завершения первой стадии интерпретации и построения дескрипторов обозначим именно как  $t_1$ .

Предполагается, что в каждый из моментов времени  $\{t_i, i = 1, 2, \dots\}$ , когда завершаются стадии интерпретации индикаторов и построения дескрипторов новых состояний индикаторов, система информационного мониторинга генерирует

три уникальных (в рамках проективного словаря) идентификатора и ими кодируются вершины семиотических треугольников Фреге согласно семиотической модели. При этом в каждый момент времени  $t_i$  может быть закодировано одновременно несколько формируемых индикаторов с использованием уникальных идентификаторов дескрипторов, если эксперты их построят.

**Определение 3.** Пусть заданы моменты времени начала стадий интерпретации  $\{t_i - \Delta_i, \Delta_i < t_i - t_{i-1}, t_0 = 0, t_i > t_{i-1}, i = 1, 2, \dots\}$ <sup>1</sup>, моменты времени завершения этих стадий  $\{t_i, i = 1, 2, \dots\}$  и пусть для каждого момента времени  $t_i$  известно число тех индикаторов, денотаты которых были сгенерированы или изменены на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ , которое обозначим как  $S_i$ . В общем случае число  $S_i$  может зависеть от длины интервала времени  $[t_{i-1}, t_i - \Delta_i]$ . Предположим, что для каждого из  $S_i$  денотатов экспертам известны их состояния, эти состояния стали объектами интерпретации и экспертами было построено  $S_i$  дескрипторов на интервале времени  $[t_i - \Delta_i, t_i]$ .

Тогда *пространство Фреге* в целях регистрации характеристик процесса формирования ЦСЗ об индикаторах в системах информационного мониторинга определим как четырехмерное пространство действительных чисел  $\mathbf{R}_2$ , в котором задано множество точек  $\{(t_i, n_{ij}, m_{ij}, k_{ij}), j = 1, \dots, S_i, i = 1, 2, \dots\}$ , где

- (1)  $t_i$  — момент времени завершения  $i$ -й стадии интерпретации, в течение которой было построено  $S_i$  дескрипторов на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ , и генерации для каждого дескриптора трех уникальных идентификаторов в момент времени  $t_i$ ;
- (2)  $n_{ij}$  — числовое представление семантического идентификатора  $j$ -го дескриптора, сгенерированного в момент времени  $t_i$  ( $n_{ij} = 0$ , если концепта нет, т. е. у экспертов отсутствует понимание совокупности программы, данных и значений, зафиксированных с помощью этого дескриптора; эту совокупность эксперты пытались интерпретировать на интервале времени  $[t_i - \Delta_i, t_i]$ );
- (3)  $m_{ij}$  — числовое представление информационного идентификатора имени  $j$ -го дескриптора, сгенерированного в момент времени  $t_i$  ( $m_{ij} = 0$ , если имени нет);
- (4)  $k_{ij}$  — числовое представление объектного идентификатора состояния денотата  $j$ -го де-

<sup>1</sup>В общем случае в процессе формирования или пополнения ЦСЗ об индикаторах задаются не моменты времени начала и завершения стадий интерпретации, а алгоритм, с помощью которого в момент времени  $t_{i-1}$  можно определить моменты времени  $(t_i - \Delta_i)$  и  $t_i$ .

скриптора, сгенерированного в момент времени  $t_i$ ; каждое состояние денотата представляет собой совокупность программы вычисления, обрабатываемых данных и значений этого индикатора, вычисленных на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ .

По определению пространство Фреге, которое является вторым компонентом решения проблемы, включает ось времени и три оси с числовыми значениями идентификаторов концептов, имен и состояний денотатов индикаторов. Для дескрипторов, построенных на интервалах времени  $\{[t_i - \Delta_i, t_i], i = 1, 2, \dots\}$ , пространство Фреге не показывает связей идентификаторов дескрипторов с концептами, именами и состояниями денотатов соответствующих индикаторов. Однако в любой момент времени  $t_i$  эти связи фиксируются в описании тех дескрипторов, которые были построены на интервале времени  $[t_i - \Delta_i, t_i]$ .

### 3.3 Третий этап решения проблемы

На третьем этапе теоретической части решения проблемы представления ЦСЗ об индикаторах определим нестационарную семиотическую модель, которая является третьим компонентом решения этой проблемы.

**Определение 4.** Пусть заданы моменты времени начала стадий интерпретации  $\{t_i - \Delta_i, \Delta_i < t_i - t_{i-1}, t_0 = 0, t_i > t_{i-1}, i = 1, 2, \dots\}$ , моменты времени завершения этих стадий  $\{t_i, i = 1, 2, \dots\}$  и пусть для каждого момента времени  $t_i$  известно число тех индикаторов, денотаты которых были сгенерированы или изменены на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ , которое обозначим как  $S_i$ .

Предположим, что для каждого из  $S_i$  денотатов экспертам известны их состояния, эти состояния стали объектами интерпретации и экспертами было построено  $S_i$  дескрипторов на интервале времени  $[t_i - \Delta_i, t_i]$ .

Тогда *нестационарной семиотической моделью* компьютерного представления концептов ЦСЗ о формируемых индикаторах, кодирования состояний их денотатов и имен в дискретные моменты времени  $\{t_i, i = 1, 2, \dots\}$  называется совокупность следующих трех составляющих:

- (1) множество семиотических треугольников Фреге для индикаторов, описание которых в виде дескрипторов получено экспертами на интервалах времени  $\{[t_i - \Delta_i, t_i], i = 1, 2, \dots\}$  (это множество обозначим как  $\{T_{ij}, j = 1, \dots, S_i, i = 1, 2, \dots\}$ );
- (2) множество цифровых семиотических треугольников, соответствующих  $\{T_{ij}\}$ , кото-

рые состоят из числовых представлений семантического, информационного и объектного идентификаторов, генерируемых системой информационного мониторинга на интервалах времени  $\{[t_i - \Delta_i, t_i], i = 1, 2, \dots\}$  согласно некоторому заданному правилу назначения идентификаторов (множество цифровых семиотических треугольников обозначим как  $\{D_{ij}, j = 1, \dots, S_i, i = 1, 2, \dots\}$ );

- (3) пространство Фреге, определенное как множество точек  $\{(t_i, n_{ij}, m_{ij}, k_{ij}), j = 1, \dots, S_i, i = 1, 2, \dots\}$ , полученных в результате отображения в  $\mathbf{R}_2$  цифровых семиотических треугольников  $\{D_{ij}\}$ .

Первая из трех составляющих модели описывает содержательные характеристики процесса формирования индикаторов в виде дескрипторов, в том числе состояния их денотатов, концепты и имена, вторая составляющая представляет собой числовые представления семантического, информационного и объектного идентификаторов, а третья составляющая представляет собой множество точек, которое планируется использовать для построения различных функциональных характеристик процесса формирования индикаторов (пример одной из характеристик будет рассмотрен далее в этом разделе).

Отметим взаимную связанность трех перечисленных составляющих нестационарной семиотической модели. Зная некоторую точку  $(t_i, n_{ij}, m_{ij}, k_{ij})$  в пространстве Фреге, можно определить момент времени  $t_i$ , когда был сформирован соответствующий этой точке дескриптор, и его уникальный объектный идентификатор  $k_{ij}$ , который является одной из вершин цифрового семиотического треугольника  $D_{ij}$ . Именно по такому коду можно найти этот дескриптор (построенный для семиотического треугольника Фреге  $T_{ij}$ ), который содержит:

- дефиницию концепта  $j$ -го индикатора, сформированного на интервале времени  $[t_i - \Delta_i, t_i]$  (если  $n_{ij} = 0$ , то рубрика и дефиниция отсутствуют);
- имя (название)  $j$ -го индикатора, выбранного на интервале времени  $[t_i - \Delta_i, t_i]$  (если  $m_{ij} = 0$ , то имя отсутствует);
- ссылки на описание программы вычисления значений  $j$ -го индикатора и обрабатываемых этой программой данных, использованных для вычисления значений  $j$ -го индикатора на интервале времени  $[t_i - \Delta_i, t_i]$ .

И наоборот, зная момент времени  $t_i$ , можно найти все дескрипторы, построенные для треуголь-

ников  $T_{ij}$ , число которых равно  $S_i$ . Для каждого из этих дескрипторов можно определить его идентификаторы, числовые значения которых  $n_{ij}$ ,  $m_{ij}$ ,  $k_{ij}$  вместе с  $t_i$  дадут точку  $(t_i, n_{ij}, m_{ij}, k_{ij})$  в пространстве Фреге.

Определенная нестационарная семиотическая модель является третьим компонентом решения проблемы представления ЦСЗ об индикаторах. В этом определении множество точек  $\{(t_i, n_{ij}, m_{ij}, k_{ij})\}$  является зависимым от используемого правила назначения информационного, семантического и объектного идентификаторов для дескрипторов проективного словаря.

В качестве примера одной их характеристик процесса формирования индикаторов построим функцию степени согласованности индикаторов между экспертами, отражающую динамику их согласования.

В процессе построения будут использоваться данные проведенного эксперимента по разработке нескольких вариантов индикатора, характеризующих распределение публикационной активности научного коллектива по возрастным группам (далее — ЦСЗ об индикаторах возрастного распределения публикаций (ВРП)) [8].

Сначала определим множество целочисленных параметров  $\{L_{ij}^p \geq 0, j = 1, \dots, S_i, i = 1, 2, \dots, p = 0, 1, 2, \dots\}$ , значение каждого из которых равно числу экспертов, согласованно интерпретирующих в момент времени  $t_{i+p}$  концепт  $j$ -го индикатора, сгенерированного на интервале времени  $[t_i - \Delta_i, t_i]$ , согласных с выбором его имени и состоянием его денотата, созданным на интервале времени  $[t_{i-1}, t_i - \Delta_i]$ .

Предположим также, что задано число  $N_C$  как граница между категориями формируемых и сформированных вариантов индикатора ВРП. В проведенном эксперименте вариант индикатора считался сформированным, если число экспертов, согласованно интерпретирующих его концепт и согласных с выбором его имени, было больше или равно  $N_C$ . Было также определено понятие неактуального дескриптора варианта индикатора ВРП: если в момент времени  $t_i$  все эксперты, участвующие в формировании вариантов, приняли решение о неперспективности дальнейшего использования некоторого уже построенного дескриптора, то в момент времени  $t_i$  он помечается как неактуальный.

В процессе проведения эксперимента функция степени согласованности, которую обозначим как  $F_{\text{con}}$ , была определена следующим образом: для любого  $i$  при условии, что  $S_i \neq 0$ ,

$$F_{\text{con}}(i, p, n_{ij}, m_{ij}, k_{ij}) = \begin{cases} 0, & \text{если дескриптор с кодом } n_{ij} \text{ является} \\ & \text{неактуальным в момент } t_{i+p}; \\ 1, & \text{если дескриптор с кодом } n_{ij} \text{ является} \\ & \text{авторским в момент } t_{i+p}; \\ L_{ij}^p, & \text{если дескриптор с кодом } n_{ij} \text{ является} \\ & \text{коллективным, а число экспертов,} \\ & \text{его согласовавших в момент } t_{i+p}, \\ & \text{меньше } N_C; \\ N_C, & \text{если дескриптор с кодом } n_{ij} \text{ является} \\ & \text{сформированным в момент } t_{i+p}, \\ & \text{т. е. число экспертов, его согласовавших} \\ & \text{в момент } t_{i+p}, \text{ равно} \\ & \text{или больше } N_C, \end{cases}$$

где  $p = 0, 1, 2, \dots$  является пятым измерением области определения  $F_{\text{con}}$  кроме четырех измерений  $i, n, m$  и  $k$  (пятое измерение введено для отражения изменения степени согласованности дескрипторов индикаторов, начиная с момента создания каждого из этих дескрипторов).

Согласно этому определению,  $F_{\text{con}} = 1$  для всех тех семиотических треугольников  $T_{ij}$ , которым в момент времени  $t_{i+p}$  соответствует авторский дескриптор с личностным концептом, который не согласован с другими экспертами. Если в некоторый момент времени эксперт-автор этого дескриптора и все остальные эксперты, участвующие в формировании ЦСЗ, отмечали его как неактуальный для процесса построения ЦСЗ, то тогда  $F_{\text{con}} = 0$ .

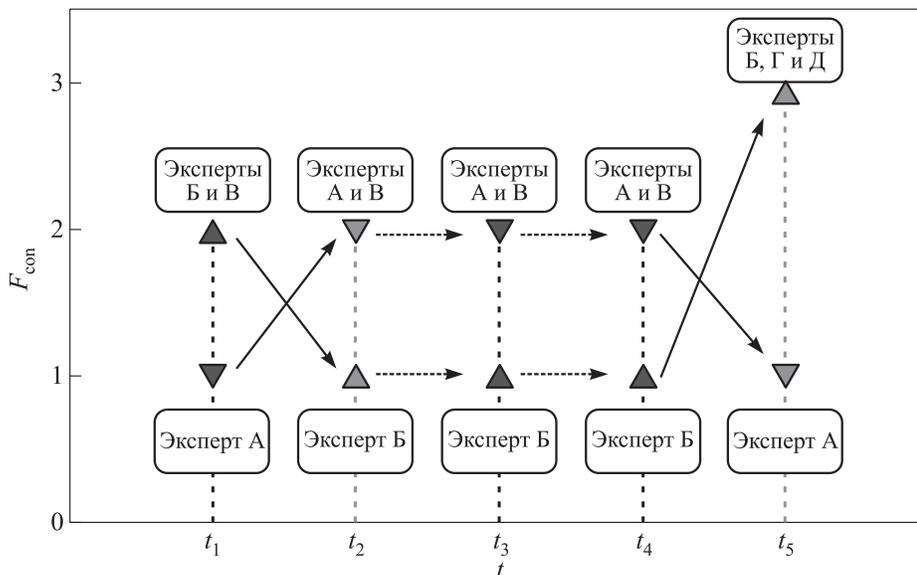
Значения определенной функции  $F_{\text{con}}$  не зависят от числовых представлений семантических, информационных и объектных идентификаторов  $n_{ij}$ ,  $m_{ij}$  и  $k_{ij}$ , а зависят только от хода процесса согласования экспертами смысла вариантов индикатора, т. е. их рубрик и дефиниций, а также имен вариантов индикатора ВРП.

Рассмотрим первые пять итераций проведенного эксперимента. В процессе разработки вариантов индикатора ВРП участвовало пять экспертов (обозначим их как А, Б, В, Г и Д).

*Первая итерация.* Эксперт А создал первый вариант при следующих условиях:

- для вычисления значений первого варианта индикатора ВРП использовались статьи сотрудников одного из подразделений ИПИ РАН, напечатанные в журналах и сборниках в 2009 г. и введенные в базы данных системы информационного мониторинга;
- возрастная группа<sup>1</sup> каждого соавтора статьи получала за одну статью 1 балл;

<sup>1</sup>В эксперименте учитывались 14 возрастных групп: 20–24, 25–29 и далее до группы 85–89 лет.



**Рис. 2** Десять из 20 значений  $F_{con}$  на пяти итерациях в следующих точках  $t_{i+p}$ : ( $i = 1, p = 0$ ); ( $i = 1, p = 1$ ); ( $i = 3, p = 0$ ); ( $i = 4, p = 0$ ); ( $i = 4, p = 1$ ). Сплошные стрелки соединяют те пары значений функции, которым соответствуют идентичные концепты; точечными стрелками обозначены отношения наследования

– отсутствовала нормализация относительно численности возрастных групп.

В это же время (на этой же итерации) эксперты Б и В создали совместно и согласованно второй вариант индикатора ВРП при тех же первом и третьем условиях, но второе из трех вышеперечисленных условий отличалось: возрастная группа каждого соавтора получала  $1/N$  балла за статью, у которой  $N$  соавторов.

На первой итерации порождается два треугольника Фреге (следовательно,  $S_i = 2$  для  $i = 1$ ), для них строятся два дескриптора и вычисляются два значения  $F_{con}$  в точке ( $i = 1, p = 0$ ):

- (1) первое значение  $F_{con}$  равно 1, так как дескриптор эксперта А на первой итерации является авторским; это значение обозначено треугольником с вершиной ниже его основания (см. рис. 2, на котором все значения  $F_{con}$  для дескрипторов первого варианта индикатора, порожденных на последующих итерациях, обозначены таким же треугольником);
- (2) второе значение равно 2, так как дескриптор экспертов Б и В на первой итерации является коллективным; это значение обозначено треугольником с вершиной выше его основания (см. рис. 2, на котором все значения  $F_{con}$  для дескрипторов второго варианта индикатора, порожденных на последующих итерациях, обозначены таким же треугольником).

**Вторая итерация.** На второй итерации эксперт В решил изменить свою точку зрения и принять точ-

ку зрения эксперта А. Иначе говоря, эксперт В на второй итерации отказывается в согласовании второго варианту индикатора ВРП, так как считает правильным добавлять всем соавторам по одному баллу, и согласовывает первый вариант индикатора ВРП. Первый вариант, сгенерированный экспертом А становится коллективным, а второй — авторским. Эти варианты индикатора ВРП идентичны вариантам первой итерации, что обозначено двумя серыми треугольниками и двумя пересекающимися сплошными стрелками (см. рис. 2).

Так как имеющиеся концепты, имена и состояния денотатов на второй итерации не изменялись, а новые не формировались, то новые дескрипторы не строились и число порожденных треугольников Фреге на второй итерации равно нулю ( $S_i = 0$  для  $i = 2$ ), значения функции  $F_{con}$  не определены в точке ( $i = 2, p = 0$ ) при  $S_2 = 0$ , но два значения этой функции определены в точке ( $i = 1, p = 1$ ), так как  $S_1 = 2$ . Эти значения равны 2 и 1 (см. рис. 2).

**Третья итерация.** Эксперты А, Б и В одновременно принимают решение учесть численности возрастных групп при вычислении значений своих вариантов индикатора ВРП, что находит отражение в изменении соответствующих алгоритмов программ вычисления их значений с целью нормализации. При этом на третьей итерации эксперты связывают новые порожденные концепты с концептами, созданными на первой итерации и изменившими степень согласованности на второй итерации, отношениями наследования, что обозна-

чено двумя точечными стрелками. Так как формируются два новых концепта, то на третьей итерации для них строятся два дескриптора проективного словаря ( $S_i = 2$  для  $i = 3$ ). При этом вычисляются четыре значения функции  $F_{\text{con}}$ : по два значения в точках ( $i = 3, p = 0$ ) и ( $i = 1, p = 2$ ).

*Четвертая итерация.* Эксперты А, Б и В одновременно приняли решение учитывать только те статьи, которые опубликованы в журналах из перечня ВАК. На этой итерации эксперты связывают новые порожденные концепты с концептами, созданными на третьей итерации, отношениями наследования, что обозначено еще двумя точечными стрелками. Так как формируются два новых концепта, то на четвертой итерации для них строятся два дескриптора проективного словаря ( $S_i = 2$  для  $i = 4$ ). При этом вычисляются шесть значений функции  $F_{\text{con}}$ : по два значения в точках ( $i = 4, p = 0$ ), ( $i = 3, p = 1$ ) и ( $i = 1, p = 3$ ).

*Пятая итерация.* На этой итерации разработчики эксперты оставляют неизменными состояния денотатов, концепты и имена двух вариантов индикатора ВРП, учитывающие численность групп и перечень журналов ВАК, но точки зрения экспертов изменяются следующим образом. Эксперт В отказывается от варианта эксперта А. Следовательно, этот концепт становится личностным концептом эксперта А. Другой вариант приобретает двух новых сторонников — экспертов Г и Д, что приводит к изменению степени его согласованности. У этих двух концептов изменилась только степень согласованности, но они остались идентичными концептам, сгенерированным на четвертой итерации, что условно обозначено двумя серыми треугольниками на пятой итерации и двумя пересекающимися сплошными стрелками между четвертой и пятой итерациями (см. рис. 2).

Так как новые концепты не формировались, а существующие не изменялись, то на пятой итерации новые дескрипторы не строились и число порожденных треугольников Фреге равно нулю ( $S_i = 0$  для  $i = 5$ ). Значение функции  $F_{\text{con}}$  не определено для пары ( $i = 5, p = 0$ ). При этом вычисляются шесть значений функции  $F_{\text{con}}$  в других точках: по два значения в точках ( $i = 4, p = 1$ ), ( $i = 3, p = 2$ ) и ( $i = 1, p = 4$ ) (см. рис. 2).

Все значения функции  $F_{\text{con}}$  для новых или измененных концептов обозначены черными треугольниками, а им идентичные — серыми треугольниками (у этих концептов изменилась только степень их согласованности экспертами, формирующими ЦСЗ).

Рисунок 2 содержит не все 20 значений функции степени согласованности  $F_{\text{con}}$ , вычисленных на

первых пяти итерациях, а только те ее 10 значений, соответствующие дескрипторы которых либо связаны отношениями наследования, либо изменили степень согласованности.

Рассмотренный эксперимент по компьютерному представлению экспертных знаний и вычислению значений функции степени согласованности  $F_{\text{con}}$  иллюстрирует потенциал использования пространства Фреге для построения функциональных характеристик процесса формирования индикаторов и областей определения соответствующих функций.

Таким образом, получены три из четырех компонентов решения сформулированной проблемы представления ЦСЗ, которые представляют собой теоретическую часть решения этой проблемы.

## 4 Проективный словарь лингвистического обеспечения

Данный раздел посвящен описанию выбранного подхода к технической реализации положений 6–9, которые были перечислены в разд. 3. Рассматривается реализация именно этих четырех положений, поскольку они формулируют непосредственные требования к созданию проективного словаря лингвистического обеспечения системы информационного мониторинга (как четвертого компонента решения сформулированной проблемы).

В процессе создания проективного словаря с помощью реляционной СУБД Microsoft SQL Server разработано шесть таблиц реляционной базы данных (РБД), названия которых можно увидеть в первой строке табл. 1. Столбцы, кроме первого, соответствуют таблицам РБД. Первый столбец содержит номера тех четырех положений (6, 7, 8 и 9), на основе которых разрабатывались шесть таблиц РБД.

В ячейке этой таблицы может стоять символ «+», либо ячейка может быть не заполнена. Незаполненность ячейки говорит о том, что для реализации положения соответствующая таблица РБД не используется. Символ «+» в ячейке строки с номером  $N$  говорит о том, что для технической реализации положения с номером  $N$  всегда используется таблица РБД из соответствующего этой ячейке столбца.

Таблица *Денотат* предназначена для описания состояний денотатов индикаторов. В столбце *Денотат* проставлен символ «+» для всех четырех положений. Это означает, что для любого из четырех возможных вариантов окончания стадии ин-

**Таблица 1** Взаимосвязь положений из разд. 3 с таблицами проективного словаря

Номер положения	Денотат	Рубрика	Дефиниция	Коллектив	Дескриптор	Эксперт
6	+				+	+
7	+	+			+	+
8	+	+	+		+	+
9	+	+	+	+	+	+

терпретации в таблице *Денотат* регистрируются: идентификатор денотата индикатора, идентификатор состояния денотата, ссылки на предыдущее и последующее состояния денотата, момент времени окончания стадии интерпретации, ссылки на описание используемого варианта программы вычисления значений нового индикатора, данные, которые обрабатываются этим вариантом программы, и полученные результаты вычисления значений индикатора, а также персональный идентификатор того эксперта, который является автором созданного состояния денотата.

Таблица *Рубрика* предназначена для указания одной или нескольких рубрик используемых систем классификации индикаторов с помощью ссылок на эти системы (а также используемых версий систем классификации), отобранных в результате интерпретации состояний денотатов индикаторов. В столбце *Рубрика* проставлен символ «+» для положений 7, 8 и 9. Это означает, что для любого из этих трех возможных вариантов окончания стадии интерпретации в таблице *Рубрика* регистрируются: идентификатор денотата индикатора, идентификатор состояния денотата, идентификаторы используемых систем классификации индикаторов, версий систем классификации, рубрики как минимум из одной системы, ссылки на ключевые слова из терминологических портретов этих рубрик, отобранных экспертом, а также персональный идентификатор этого эксперта.

Таблица *Дефиниция* предназначена для описания структурированных и параметризованных дефиниций, полученных в результате интерпретации состояний денотатов индикаторов. Подобная дефиниция представляет собой совокупность текстовых фрагментов, в которую включены параметры используемых вариантов программ и параметры отбора тех данных, которые включены в это состояние денотата [23]. В столбце *Дефиниция* проставлен символ «+» для положений 8 и 9. Это означает, что для любого из этих двух возможных вариантов окончания стадии интерпретации в таблице *Дефиниция* регистрируются: идентификатор денотата индикатора, идентификатор состояния денотата, ссылка на структурированную и параметризованную де-

финицию, значения параметров используемых вариантов программ и параметров отбора данных, выбранные экспертом, а также персональный идентификатор этого эксперта.

Таблица *Коллектив* предназначена для хранения списка тех экспертов, с которыми были согласованы построенный денотат индикатора, описание его интерпретируемого состояния, в том числе рубрики систем классификации, дефиниция и имя. В столбце *Коллектив* поставлен символ «+» для положения 9. Это означает, что для четвертого варианта окончания стадии интерпретации в таблице *Коллектив* регистрируются: идентификатор денотата индикатора, идентификатор состояния денотата, персональный идентификатор эксперта-автора описания этого состояния, а также идентификаторы тех экспертов, с которыми были согласованы построенный денотат и описание его интерпретируемого состояния.

Для последних двух столбцов символ «+» проставлен для всех четырех положений. Это означает, что в таблицах *Дескриптор* и *Эксперт* для любого из четырех возможных вариантов окончания стадии интерпретации в этих таблицах регистрируются описание каждого дескриптора и данные об экспертах.

Необходимо заметить, что строки в таблицах РБД никогда не редактируются и не удаляются. Любое изменение формируемого индикатора порождает одну или несколько новых строк, соответствующих новому состоянию денотата, концепту или имени этого индикатора. Потерявшие актуальность строки остаются в таблице и помечаются как «закрытые». Это позволяет в любой момент времени восстановить любое предыдущее состояние формируемого индикатора, а также проследить его эволюцию во времени.

Первый вариант технической реализации выбранного подхода к построению таблиц РБД использовался для проведения эксперимента по реализации компьютерного представления экспертных знаний об индикаторах в динамике их формирования. Этот вариант технической реализации охватывал не все предусмотренные возможности таблиц РБД. В частности, использовались только

три из четырех возможных вариантов окончания стадии интерпретации, кроме второго, т. е. в описании каждого состояния формируемого индикатора отсутствовали рубрики систем классификации индикаторов и ссылки на ключевые слова из терминологических портретов этих рубрик.

## 5 Заключение

Основные полученные результаты, рассмотренные в статье, заключаются в следующем. Во-первых, определены две основные категории моделей формирования новых или пополнения существующих систем знаний (концептуальные и предметно-ориентированные), которые отличаются по степени общности. Во-вторых, на основе концептуальных семиотических моделей представления процессов формирования знаний из работ [16, 17] дано описание предметно-ориентированных моделей для случая формирования или пополнения системы экспертных знаний об индикаторах мониторинга. В-третьих, предложено решение проблемы представления ЦСЗ об индикаторах мониторинга, включающее теоретическую и прикладную части ее решения.

В теоретической части решения определена функция степени согласованности формируемых индикаторов между экспертами в динамике их формирования. В прикладной части решения предложен подход к технической реализации разработанных моделей.

Проведенный эксперимент позволяет сделать следующий вывод: создание семиотических моделей компьютерного представления экспертных знаний об индикаторах, разработка на их основе проективного словаря и предлагаемый подход к его технической реализации дают возможность группе экспертов совместно разрабатывать новые алгоритмы для вычисления значений программно-ориентированных индикаторов и развивать систему экспертных знаний об индикаторах мониторинга. Проективный словарь предназначен для уменьшения лакуарности этой системы, и процесс его формирования отражает целенаправленную деятельность экспертов по заполнению лакун в системе экспертных знаний об индикаторах теми структурными элементами знаний, которые оказались необходимы для мониторинга и индикаторного оценивания ПЦД. При этом у экспертов имеется возможность фиксировать в проективном словаре этапы формирования смыслового содержания индикаторов, а также различия в рубриках, дефинициях, именах формируемых индикаторов и степень их согласованности между экспертами.

## Литература

1. Концепция реформирования бюджетного процесса в Российской Федерации в 2004–2006 годах. Одобрена постановлением Правительства РФ от 22 мая 2004 года № 249 «О мерах по повышению результативности бюджетных расходов». <http://government.consultant.ru/page.aspx?787610> (дата обращения: 09.07.2011).
2. Программа фундаментальных научных исследований государственных академий наук на 2008–2012 годы. — М.: Наука, 2008.
3. CORDIS ICT Programme Home. [http://cordis.europa.eu/fp7/ict/programme/home\\_en.html](http://cordis.europa.eu/fp7/ict/programme/home_en.html) (дата обращения: 23.05.2011).
4. ICT FP7 Work Programme 2007-08. [ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2007-08\\_en.pdf](ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2007-08_en.pdf) (дата обращения: 23.05.2011).
5. ICT FP7 Work Programme 2009-10. [ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2009-10\\_en.pdf](ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2009-10_en.pdf) (дата обращения: 23.05.2011).
6. ICT FP7 Work Programme 2011-12. [ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2011-12\\_en.pdf](ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2011-12_en.pdf) (дата обращения: 23.05.2011).
7. FP7 Exploratory Workshop 4 «Knowledge Anywhere Anytime». [http://cordis.europa.eu/ist/directorate\\_f/f\\_ws4.htm](http://cordis.europa.eu/ist/directorate_f/f_ws4.htm) (дата обращения: 23.05.2011).
8. Зацман И. М., Косарик В. В., Курчавова О. А. Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
9. Nonaka I. The knowledge-creating company // Harvard Business Rev., 1991. Vol. 69. No. 6. P. 96–104.
10. Nonaka I., Takeuchi H. The knowledge-creating company. — Oxford; N.Y.: Oxford University Press, 1995. (Пер. на русск. яз.: Нонака И., Такеучи Х. Компания — создатель знания. — М.: Олимп-бизнес, 2003.)
11. Knowledge emergence / Eds. I. Nonaka, T. Nishiguchi. — Oxford; N.Y.: Oxford University Press, 2001.
12. Wierzbicki A. P., Nakamori Y. Basic dimensions of creative space // Creative space: Models of creative processes for knowledge civilization age / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin—Heidelberg: Springer Verlag, 2006. P. 59–90.
13. Wierzbicki A. P., Nakamori Y. Knowledge sciences: Some new developments // Zeitschrift für Betriebswirtschaft, 2007. Vol. 77. No. 3. P. 271–295.
14. Ren H., Tian J., Nakamori Y., Wierzbicki A. P. Electronic support for knowledge creation in a research institute // J. Syst. Sci. Syst. Eng. 2007. Vol. 16. No. 2. P. 235–253.
15. Зацман И. М. Концептуальный поиск и качество информации. — М.: Наука, 2003.

16. Зацман И. М. Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // Информатика и её применения, 2009. Т. 3. Вып. 2. С. 65–81.
17. Зацман И. М. Нестационарная семиотическая модель компьютерного кодирования концептов, информационных объектов и денотатов // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 87–101.
18. Zatsman I., Durnovo A. Incompleteness problem for indicators system of research programme // 11th Conference (International) on Science and Technology Indicators (STI'2010): Book of Abstracts. — Leiden: Universiteit Leiden, 2010. P. 309–311.
19. Успенский В. А. К публикации статьи Г. Фреге «Смысл и денотат» // Семиотика и информатика. — М.: Языки русской культуры, 1997. Вып. 35. С. 351–352.
20. Фреге Г. Смысл и денотат // Семиотика и информатика. — М.: Языки русской культуры, 1997. Вып. 35. С. 352–379.
21. Фреге Г. Понятие и вещь // Семиотика и информатика. — М.: Языки русской культуры, 1997. Вып. 35. С. 380–396.
22. Зацман И. М. Категоризация результатов и индикаторов программ научных исследований в информационных системах мониторинга // Системы и средства информатики. — М.: ИПИ РАН, 2009. Доп. вып. С. 200–219.
23. Кожунова О. С. Технология разработки семантического словаря системы информационного мониторинга: Автореф. дисс. . . . канд. техн. наук. — М.: ИПИ РАН, 2009. 23 с.

# ТРАНСФОРМАЦИОННЫЕ МОДЕЛИ ЯЗЫКОВЫХ СТРУКТУР ДЛЯ ФРАНЦУЗСКО-РУССКОГО МАШИННОГО ПЕРЕВОДА

Ю. И. Морозова<sup>1</sup>

**Аннотация:** Данная работа посвящена актуальным проблемам исследования трансформационных свойств языковых объектов при переводе предикативных структур с французского языка на русский. Основное внимание уделено изменению категориальной принадлежности и изменению грамматических характеристик предикатных слов при переводе. Материалом исследования послужили патентные тексты на французском языке и их переводы на русский язык, выполненные специалистами-переводчиками.

**Ключевые слова:** французско-русский автоматический перевод; функциональная семантика; языковые трансформации; вершинные грамматики

## 1 Введение

Данное исследование направлено на исследование предикатных фразовых структур на основе вершинных грамматик применительно к задачам моделирования машинного перевода и извлечения знаний из текста для французско-русского направления. Основной задачей ставилось создание унифицированной модели функциональных значений синтаксем, в которой бы учитывались сдвиги значений, производимые переводческими трансформациями. Предикатные слова являются вершинами синтаксической структуры предложения, а также вершинами внутреннего представления знаний в структурах баз знаний, поэтому описание их дистрибутивных и трансформационных свойств имеет первостепенное значение.

Исследования ведутся в рамках проекта по созданию многоязычного лингвистического процессора для задач машинного перевода и извлечения знаний из текстов, разрабатываемого на основе функционально-семантического подхода [1]. В качестве материала исследования были использованы фрагменты параллельных текстов патентов, содержащие предикатные выражения. Модель перевода с учетом трансформаций для русско-французской языковой пары основана на многовариантной когнитивной трансферной грамматике (МКТГ), разработанной Е. Б. Козеренко [1–4]. Данный формализм имеет определенные черты грамматики составляющих и вершинной грамматики HPSG (Head-driven phrase structure grammar) [5]. Преимущество данного формализма заключается в том, что он позволяет описывать как отношения линейного порядка, так и отношения зависимости в рамках одной и той же фразовой структуры. Формализ-

мы, основанные на порождающей грамматике Хомского и на вершинной грамматике HPSG, широко применяются для создания систем автоматической обработки текстов на английском языке и других европейских языках (французском, испанском, немецком, чешском). Однако возможности применения данных формализмов для автоматической обработки русского языка изучены недостаточно.

## 2 Грамматические формализмы, используемые для создания лингвистических процессоров

Для формального описания синтаксиса естественных языков применительно к задачам автоматической обработки языка наиболее часто используются следующие виды формализмов: регулярные грамматики, контекстно-свободные грамматики, мягко контекстно-зависимые грамматики. Регулярные грамматики не могут быть использованы для полноценного описания синтаксиса. Данный формализм используется только для частичного синтаксического анализа предложений (shallow parsing). С помощью контекстно-свободных грамматик можно описать большинство предложений естественного языка, однако грамматики данного класса не позволяют описывать предложения с разрывными структурами. Наконец, мягко контекстно-зависимые грамматики являются наиболее мощным формализмом и позволяют описывать любые виды предложений естественных языков, однако их применение в системах автоматической обработки естественного языка связано с большими

<sup>1</sup>Институт проблем информатики Российской академии наук, yulia-ipi@yandex.ru

вычислительными затратами. Для описания явлений естественных языков применяются, в основном, контекстно-свободные грамматики с некоторыми расширениями. Однако вопрос о выборе наиболее адекватного формализма для описания синтаксиса естественных языков и создания лингвистических процессоров остается открытым. Во многих системах автоматической обработки текстов на английском языке используются модернизированные грамматики Н. Хомского [6].

Для создания систем автоматической обработки текстов на языках с богатой морфологией и относительно свободным порядком слов часто используется вершинная грамматика HPSG, разработанная Карлом Поллардом и Иваном Сагом [5]. Согласно данной теории описание грамматики языка должно состоять из очень подробного словаря и очень небольшого количества грамматических правил, носящих универсальный характер. Словарь имеет хорошо проработанную иерархическую структуру, которая характеризуется наследованием свойств по умолчанию. В словарном описании слов, которые могут являться вершинами синтаксических групп (существительных, глаголов, предлогов, прилагательных) есть поле HEAD, в котором описываются такие важные с точки зрения синтаксического поведения слова свойства, как часть речи, признаки согласования, форма, предикативность и др. Данные свойства передаются группам, порождаемым данными вершинами, в соответствии с правилами грамматики. Таким образом, процесс порождения правильно построенных предложений определяется свойствами вершин (отсюда название *Head-driven phrase structure grammar*). Одним из основных понятий HPSG является структура свойств (*feature structure*). Это набор атрибутов с их значениями, например словарное описание лексемы задается следующей структурой свойств: [PHON... SYN... SEM... ARG-ST...].

Значение каждого из свойств может представлять собой как единый элемент, так и структуру свойств, например свойство AGR (согласование) имеет следующую структуру: AGR [PER..., NUM..., GEND...].

Унификационный механизм, использующийся в грамматике HPSG, позволяет объединить два описания структур свойств. Результатом данной операции является структура свойств, содержащая информацию из обеих структур. Механизм унификации используется при проверке согласования морфологических характеристик слов, необходимой для включения их в одну синтаксическую группу. Вершинная грамматика HPSG была с успехом применена при создании систем автоматической обработки текстов на разных языках (английском,

французском, чешском). Один из примеров такой системы — грамматика английского языка LinGO English Resource Grammar (ERG) и синтаксически аннотированный корпус Redwoods, размеченный автоматически с использованием грамматики ERG [7].

В работе [8] предлагается формализм, имеющий некоторые черты универсальной грамматики Хомского, вершинной грамматики HPSG и лексико-функциональной грамматики LFG (*Lexical functional grammar*). С использованием данного формализма была реализована программа синтаксического анализа русского языка, которая строит структуру предложения в двух аспектах: как структуру составляющих и как функциональную структуру. Для описания согласования морфологических характеристик применяется аппарат унификации, используемый в вершинной грамматике HPSG. В данной работе обосновывается возможность применения грамматик составляющих с различными модификациями для автоматической обработки текстов на русском языке (в частности, для перевода с русского языка на другой язык).

### 3 Современные подходы к проблеме машинного перевода

В области машинного перевода существуют два основных направления исследований — подход на основе правил и статистический подход. Системы, созданные в рамках подхода на основе правил, включают в себя компоненты, отвечающие за последовательный морфологический, синтаксический и семантический анализ предложений исходного языка и синтез предложений целевого языка (с прохождением тех же уровней). Создание таких систем требует многолетней кропотливой работы лингвистов, так как для функционирования системы необходим словарь с подробными синтаксико-семантическими описаниями словарных единиц и правила анализа и синтеза предложений (морфологического, синтаксического и семантического уровней). Достоинствами данного подхода являются высокое качество перевода, соответствие теоретическим концепциям и возможность удобного внесения изменений. Недостатками являются большие трудозатраты для создания словарей и систем правил перевода.

Статистический подход заключается в выявлении закономерностей перевода путем автоматического анализа параллельных текстов с использованием методов математической статистики и без

использования лингвистических знаний. Достоинством статистического подхода является быстрота создания подобных систем. Для того чтобы система начала работать, необходим лишь текстовый корпус, переводческий словарь (возможно, неполный) и словарь основ. По данным из [9], требуется всего несколько часов для того, чтобы система начала работать, и 1–2 недели, чтобы настроить ее и получать приемлемые результаты. Недостатком данного подхода является необходимость использования больших параллельных корпусов (от 1 млн слов) для получения удовлетворительных результатов перевода. Не для всех языковых пар существуют такие большие текстовые коллекции. Если же использовать не только статистические методы, но добавить и частичную лингвистическую разметку, размер корпуса можно существенно уменьшить (с 1 млн до 300 тыс. слов) [10].

Современный период развития исследований и разработок в области машинного перевода и систем извлечения знаний из текстов характеризуется интенсивным процессом «гибридизации» подходов и моделей. Создатели систем, основанных на правилах, вводят в правила различные стохастические модели, которые позволяют отобразить динамику и разнообразие языковых форм и значений, порождаемых в процессе речевой деятельности, а сторонники статистических методов построения лингвистических моделей все чаще обращаются к подходам, основанным на лингвистических знаниях, рассматривая их как средства «интеллектуализации» систем. В настоящее время появляется все больше исследований в рамках синергетического подхода, использующего лингвистические знания, статистические методы и механизмы машинного обучения [4]. Как пишут авторы [11], «наше убеждение состоит в том, что в долгосрочной перспективе самые эффективные технологии машинного перевода объединят в себе преимущества обоих подходов». По мнению авторов [11], подход на основе правил следует применять для анализа тех уровней языка, для которых существуют детальные лингвистические теории, описывающие подавляющее большинство случаев, в то время как статистический подход следует применять для извлечения лексической и предметно-ориентированной лингвистической информации, для которой пока что не существует разработанной теории.

Наиболее перспективными направлениями в области статистического машинного перевода являются перевод цепочек слов (*phrase-based translation*) и синтаксический перевод (*syntax-based translation*). При использовании метода перевода цепочек слов сопоставлению и переводу подверга-

ются цепочки слов (обычно не длиннее трех слов), выделенные путем применения статистических методов. Они не всегда совпадают со словосочетаниями в традиционном лингвистическом понимании (группа слов, взаимосвязанных синтаксически и семантически). Например, группа слов «*in accordance with the*» является цепочкой слов, подлежащей переводу, в рамках статистического машинного перевода, но не является словосочетанием в лингвистическом смысле. При синтаксическом переводе сопоставлению и переводу подвергаются синтаксические поддеревья, а не конкретные слова или словосочетания.

#### 4 Создание системы правил для русско-французского машинного перевода

В работе [12] описывается система перевода с английского на французский язык, сочетающая в себе традиционный правилковый подход и статистический подход — перевод цепочек слов с использованием соответствий, извлеченных из параллельного текстового корпуса. В качестве цепочек слов авторы предлагают использовать не любые последовательности слов, а только синтаксически мотивированные, другими словами, из текста извлекаются именные, глагольные группы, группы прилагательных и наречий. При выборе наилучшего варианта перевода цепочки слов предпочтение отдается цепочке слов, имеющей ту же самую синтаксическую категорию, т. е. в качестве перевода для именных групп используются именные группы и т. д. Французско-русское направление машинного перевода в нашей стране развивается с самого начала исследований по машинному переводу. Первыми появились экспериментальные системы французско-русского автоматического перевода ФРАП (1976–1986 гг.) [13] и ЭТАП-1 (1985 г.) [14]. Эти системы были основаны на последовательном морфологическом, синтаксическом и семантическом анализе предложений исходного языка с последующим синтезом предложений целевого языка (с прохождением тех же уровней). В системах использовались словари с подробными синтактико-семантическими описаниями слов и системы правил анализа и синтеза предложений естественного языка.

В 1990-е гг. появилась первая коммерческая система автоматического перевода французско-русского направления ПРОМТ. В основу архитектуры систем было положено представление процесса перевода как процесса с объектно-ориентированной организацией, основанной на иерархии обработки

ваемых компонентов предложения. В системах работают сетевые грамматики, близкие по типу к расширенным сетям переходов, а также процедурные алгоритмы заполнения и трансформаций фреймовых структур для анализа сложных предикатов [15].

Качество перевода в современных системах машинного перевода французско-русского направления достигло высокого уровня, однако многие особенности синтаксиса русского языка, а также многие типы регулярных трансформаций, происходящих при переводе с русского языка на французский, остаются неучтенными в этих системах.

В рамках описываемых проектов разрабатывается система правил трансфера синтаксических структур, учитывающая возможность синтаксических трансформаций при переводе и многовариантность перевода. Козеренко была разработана многоязычная семантическая грамматика русского и английского языков для задач автоматической обработки текстов — МКТГ [1–4]. Данная грамматика является разновидностью унификационно-порождающей грамматики. Многовариантные правила функционально-семантического переноса фразовых структур задают алгоритм перевода с одного языка на другой, причем учитывается вероятность каждого из вариантов перевода. Функциональные значения языковых единиц отражены в расширенной системе категориально-функциональных атрибутов. Структуры атрибутов и значений и правила их преобразования задаются в виде контекстно-свободных и мягко контекстно-зависимых продукционных правил. Отношения зависимости реализуются через механизм головных вершин фразовых структур, а сами фразовые структуры задают линейные последовательности языковых объектов. Лингвистический процессор сегментирует входные предложения на фразовые структуры и осуществляет трансфер этих структур в соответствующие им структуры целевого языка. Сегментация фразовых структур входного предложения проводится с учетом смысла структур, который при переводе должен быть передан средствами целевого языка. Задачей проведенных исследований ставилось создание системы правил многовариантного трансфера для перевода с русского языка на французский.

С точки зрения синтаксической структуры предложения русский и французский языки очень сильно отличаются друг от друга. Во французском языке большинство предложений двусоставны, т. е. и подлежащие, и сказуемые выражены на поверхностном уровне, причем сказуемое всегда выражается личной формой глагола. В русском языке кроме канонической структуры «Подлежащее (выраженное существительным в именительном падеже) + сказуемое (выраженное личным глаголом)»

возможны также другие синтаксические структуры:

- В предложении отсутствует сказуемое, выраженное глаголом в личной форме. Сказуемое выражено кратким причастием, кратким или полным прилагательным, существительным, предложной группой, инфинитивом и др.

#### Примеры:

*Дом красив* (краткое прилагательное). *Дом построен* (краткое причастие). *Пьер — учащийся* (существительное).

Также к данному классу относятся случаи назывных предложений, состоящих из одного подлежащего, выраженного существительным в именительном падеже (например, заголовки) и случаи эллипсиса, когда глагол в личной форме «подразумевается», но не выражен в поверхностной структуре предложения. Приведем пример эллипсиса из текста научного патента:

*Рисунки 1–4 ИЗОБРАЖАЮТ продольный разрез половины детали различных вариантов, соответствующих выполнению зубного штифта согласно первому варианту осуществления изобретения* (полная структура).

*Рисунки 5 и 6 — продольный разрез половины детали моноблочного компонента протеза* (структура с эллипсисом).

Подобные структуры являются трудными для синтаксического анализа и перевода на французский язык, так как при переводе требуется вставить пропущенный глагол в личной форме (глагол *быть* или другой глагол). Многие из структур данного вида в существующих системах автоматического перевода с русского языка на французский язык обрабатываются некорректно.

- В предложении отсутствует подлежащее, выраженное существительным в именительном падеже. К данному типу предложений относятся безличные предложения, неопределенно-личные предложения, определено-личные предложения, инфинитивные предложения.

#### Примеры:

*Мне нравится работать* (безличное предложение).

*Маше подарили книгу* (неопределенно-личное предложение).

*Еду в кино* (определенно-личное предложение).

*Нам бы сессию сдать* (инфинитивное предложение).

Также к данному классу относятся предложения, в которых подлежащее выражено инфинитивом.

#### Пример:

*Курить — здоровью вредить.*

Предложения с подобной синтаксической структурой также являются источником значительных трудностей при автоматическом анализе и переводе, так как французский язык требует, чтобы в каждом предложении было подлежащее, выраженное существительным или местоимением без предлогов (функционально соответствует существительному в именительном падеже в русском языке). Следовательно, при переводе предложения русского языка, в котором нет подлежащего, выраженного существительным в именительном падеже, требуется восстановить подлежащее, используя «формальное» подлежащее (безличное местоимение *il*) или личное местоимение.

Чтобы обосновать описание всех типов предложений в создаваемой системе правил многовариантного перевода, было проведено исследование частоты встречаемости предложений различных типов в текстах научных патентов. Предложения были разделены на 3 класса.

**Класс 1.** Подлежащее (выраженное существительным в именительном падеже) + сказуемое (выраженное глаголом в личной форме).

**Класс 2.** Сказуемое, выраженное глаголом в личной форме, отсутствует.

**Класс 3.** Подлежащее, выраженное существительным в именительном падеже, отсутствует.

Некоторые предложения относятся одновременно и к классу 2, и к классу 3. Такие предложения были отнесены к классу 2.

В результате распределения предложений по группам и подсчета относительной частоты встречаемости предложений каждого вида в текстах научных патентов были получены следующие результаты: класс 1 — 48%; класс 2 — 38%; класс 3 — 14%.

Как видно, предложения, относящиеся к каждому из трех классов, встречаются в текстах с достаточно большой частотой, и существует необходимость включения правил для всех трех классов в систему многовариантного трансфера.

Кроме различия структурных типов предложений французский и русский языки также очень существенно различаются между собой в аспекте порядка слов в предложении. Во французском языке большинство предложений имеют канонический порядок слов:

*Подлежащее — сказуемое — прямое дополнение — косвенные дополнения.*

В русском языке данный порядок регулярно нарушается как в устной, так и в письменной речи. Несоответствие порядка слов в предложениях на

русском и французском языке создает необходимость изменения порядка слов при переводе.

При создании системы правил перевода будем использовать классификации предложений русского языка, изложенные в учебниках по русскому синтаксису [16, 17], а также типичные переводческие трансформации, описанные в учебниках по переводу с французского языка на русский [18].

Рассмотрим наиболее частотные типы предложений русского языка, которые создают трудности при переводе, на материале патентных текстов.

- **Безличные предложения.** Будем понимать под безличным предложением такое предложение, которое содержит глагол в личной форме, но не содержит существительного в именительном падеже, которое выполняло бы роль подлежащего. Безличным предложениям русского языка соответствуют предложения с безличным местоимением *il* французского языка. Пример перевода:

*Однако оказалось, что эта прочность может в конечном счете привести к нарушению надежности соединения. → Toutefois, il est apparu que cette robustesse pouvait finalement porter atteinte à la fiabilité de la liaison.*

Правило переноса выглядит следующим образом:

$$V[\text{Person } 3, \text{NumberSG}, \text{GenderNEUTR}] \& \& \text{NONP}[\text{CaseNOM}] \rightarrow \Pi + V[\text{Person}3, \text{NumberSG}, \text{GenderMASC}] .$$

- **Неопределенно-личные предложения.**

**Пример:**

*Указанное раструбное соединение осуществляют [ . . . ].*

При переводе на французский язык чаще всего используется пассивная конструкция:

*Cet emboîtement est effectué [ . . . ].*

Возможен и другой вариант перевода, с использованием безличного местоимения *on*:

*On effectue cet emboîtement [ . . . ].*

Правило переноса выглядит следующим образом:

$$\text{NP}[\text{CaseAcc}] + V[\text{Person}3, \text{NumberPL}] \rightarrow \rightarrow \{ \text{NP}^* + V(\text{be})^* + V[\text{FORMPART}, \text{TENSEPAST}]^* \} \text{OR} \{ \text{On} + V[\text{Person}3, \text{NumberSG}] + \text{NP} \} .$$

Знак \* означает согласование морфологических признаков.

## 5 Трансфер пропозиционального ядра в русско-французской языковой паре

Основу семантико-синтаксической структуры предложения составляет пропозициональное ядро, прежде всего языковые средства предикации. Были изучены структуры когнитивного переноса в рамках поля функционального переноса (ПФП) первичной и вторичной предикации для русско-французской языковой пары по аналогии с русско-английской языковой парой. Были выделены базовые правила когнитивного переноса для различных функциональных значений (частотные характеристики были выделены на основании анализа патентных текстов). Материалом анализа послужили параллельные тексты патентов и/или рефератов патентов на русском и французском языках, взятые из базы данных Роспатента.

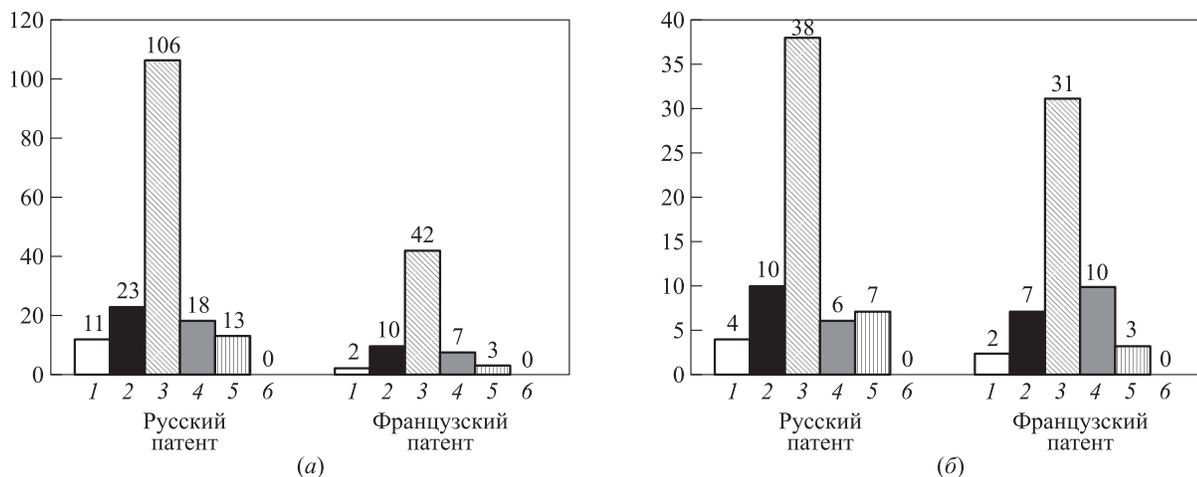
Сравнение русских и французских текстов рефератов научных патентов показало, что доля действительно параллельных текстов в них составляет примерно 30%. Остальные тексты можно назвать когнитивно-сопоставимыми, причем объем русского текста может превышать объем французского

на две трети. Однако распределение по частям речи в русских и французских научных текстах патентных рефератов (и самих патентов) очень близко по составу и объему, что отражено на рис. 1. Русский текст в целом на 30%–35% более номинативен, чем французский, в котором в поле вторичной предикации предпочтение отдается инфинитиву (в русском — отглагольным существительным).

В первом примере тексты рефератов не параллельные, а когнитивно-сопоставимые, во втором тексты русского и французского реферата параллельны: перевод выполнен точно, почти дословно. В любом случае, как видно из примеров, и в русских, и во французских патентных текстах очень высока доля именных групп, что вообще всегда характерно для научно-технических текстов.

Правила когнитивного переноса для функциональных значений цели и назначения представлены на рис. 2.

Таким образом, набор структур, используемых для выражения цели действия, одинаков для русского и французского языка, однако французский тяготеет к инфинитивной структуре, а русский — к именной (*Для увеличения способности сети к обобщению... / Afin d'augmenter la capacité du réseau de généraliser...* ).



**Рис. 1** Распределение по частям речи в русских и французских научных текстах патентных рефератов: (а) реферат патента WO2004009333; (б) реферат патента WO2004017987; 1 — предложения; 2 — строки; 3 — существительные; 4 — глаголы; 5 — причастия; 6 — дееспричастия (герундии)

### Цель, назначение

Русский язык

**Существительное (98)**

Инфинитив (2)

Придаточное предложение (0)

Французский язык

**Инфинитив (72)**

← → Существительное (26)

Придаточное предложение (2)

**Рис. 2** Правила когнитивного переноса для функциональных значений цели и назначения

**Примеры.**

1. [Cat : VerbNoun] {для распознавания} {pour la reconnaissance} — предложная группа: предлог + существительное.
2. [Cat : VerbInf] {чтобы распознать} {afin de reconnaître} — союз + инфинитив.
3. [Cat : Sentence] {чтобы распознавание было эффективным} {pour que la reconnaissance soit efficace} — придаточное предложение, присоединяемое подчинительной связью (союзом цели). При трансформации русского отглагольного существительного во французский инфинитив необходимо сделать выбор между его активной и пассивной формой. Видимо, в рамках системы автоматического перевода данный выбор лучше всего осуществляется с применением статистических данных (активный инфинитив встречается в текстах намного чаще пассивного; в анализируемых текстах французский пассивный инфинитив в качестве перевода русского отглагольного существительного встретился в 13% случаев).

**6 Заключение**

Были описаны предикативные синтаксические структуры русского языка, принадлежащие к функционально-семантическому полю первичной и вторичной предикации и соответствующие им синтаксические структуры французского языка. Была составлена подробная классификация типов предложений русского языка с точки зрения синтаксиса и соответствующих им синтаксических типов во французском языке, которая может быть использована при написании правил переноса синтаксических структур, происходящего при переводе с русского языка на французский. В классификации учтено синтаксическое многообразие русского языка: назывные предложения, безглагольная предикация (в случае невыраженного глагола «быть» в настоящем времени), безличные предложения, определенно-личные предложения, неопределенно-личные предложения, двусоставные предложения с различным типом сказуемых (глагольные, именные и пр.).

Были изучены категориальные трансформации предикативных структур, происходящие при переводе с русского языка на французский и в обратном направлении. Моделирование трансформаций предикативных структур для задачи машинного перевода является актуальной задачей, так как это явление мало исследовано с точки зрения компьютерной реализации и недостаточно учтено в дей-

ствующих системах машинного перевода. Кроме того, правила, задающие функциональную синонимию языковых конструкций, могут использоваться также при машинном обучении на корпусе параллельных текстов, позволяя избежать формирования избыточных правил и «шумов».

Дальнейшие исследования будут направлены на уточнение системы синтаксических соответствий с помощью параллельного корпуса текстов научных патентов, а также на расширение числа типов трансформаций при русско-французском машинном переводе, дальнейшее изучение дистрибутивно-трансформационных характеристик языковых структур и сбор статистической информации.

**Литература**

1. Козеренко Е. Б. Моделирование переноса функциональных значений для англо-русского машинного перевода // Компьютерная лингвистика и интеллектуальные технологии: Труды Междунар. конф. Диалог'2004. — М.: Наука, 2004.
2. Kozerenko E. B. Cognitive approach to language structure segmentation for machine translation algorithms // Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA, 2003. P. 49–55.
3. Козеренко Е. Б. Функционально-семантические инварианты для алгоритмов синтаксического анализа и разметки полнотекстового научного документа // Системы и средства информатики. — М.: Наука, 2003. Вып. 13. С. 298–312.
4. Козеренко Е. Б. Лингвистическое моделирование для систем машинного перевода и обработки знаний // Информатика и её применения, 2007. Т. 1. Вып. 1. С. 54–66.
5. Sag I., Wasow Th., Bender E. M. Syntactic theory: A formal introduction. — Stanford: CSLI Publications, 2003.
6. Chomsky N., Lasnik H. The theory of principles and parameters // The minimalist program. — Cambridge: MIT Press, 1995.
7. Oepen S., Toutanova K., Shieber S., Manning C., Flickinger D., Brants T. The LinGO Redwoods Treebank: Motivation and preliminary applications // 19th Conference (International) on Computational Linguistics Proceedings. — Taipei, Taiwan, 2002. P. 1253–1257.
8. Перекрестенко А. Разработка и программная реализация системы автоматического выделения синтаксических групп для естественных языков // Системы и средства информатики. — М.: Наука, 2007. Вып. 17. С. 273–291.
9. Brown R. D. Example-based machine translation in the Pangloss system // 16th Conference (International) on Computational Linguistics (COLING-96) Proceedings. — Copenhagen, Denmark, 1996. P. 169–174.

10. *Brown R. D.* Adding linguistic knowledge to a lexical example-based translation system // 8th Conference (International) on Theoretical and Methodological Issues in Machine Translation Proceedings. — Chester, UK, 1999. P. 22–32.
11. *Grishman R., Kosaka M.* Combining rationalist and empiricist approaches to machine translation // 4th Conference (International) on Theoretical and Methodological Issues in Machine Translation Proceedings. — Montreal, Canada, 1992. P. 263–274.
12. *Dugast L., Senellart J., Koehn P.* Selective addition of corpus-extracted phrasal lexical rules to a rule-based machine translation system // 12th Machine Translation Summit Proceedings. — Ottawa, ON, Canada, 2009. P. 222–229.
13. *Леонтьева Н. Н., Никогосов С. Л.* Система ФРАП как информационная система // Актуальные вопросы практической реализации систем автоматического перевода. — М.: МГУ, 1982. С. 134–166.
14. *Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л. и др.* Лингвистическое обеспечение системы французско-русского автоматического перевода ЭТАП-1. 1. Общая характеристика системы // Теория и модели знаний (Теория и практика создания систем искусственного интеллекта): Труды по искусственному интеллекту. Ученые записки Тартуского гос. ун-та. — Тарту, 1985. Вып. 714. С. 20–39.
15. *Соколова С.* Как переводит компьютер. [http://www.translationmemory.ru/technology/articles/article\\_Sokolova.php](http://www.translationmemory.ru/technology/articles/article_Sokolova.php).
16. *Валгина Н. С.* Синтаксис современного русского языка. — М.: Агар, 2000. 416 с.
17. *Шелякин М. А.* Справочник по русской грамматике. — М.: Дрофа, 2006. 355 с.
18. *Гак В. Г., Григорьев Б. Б.* Теория и практика перевода: Французский язык. — СПб.: Интердиалект+, 2000. 456 с.

# THE STRATEGIES OF SYNTACTIC ANALYSIS BASED ON HEAD-DRIVEN GRAMMARS AND THE METHODS OF THEIR IMPLEMENTATION IN INFORMATION SYSTEMS\*

E. B. Kozerenko<sup>1</sup> and P. V. Ermakov<sup>2</sup>

**Abstract:** The paper deals with the problems of design and development of syntactic parses in multilingual natural language processing systems, machine translation, and knowledge extraction from texts. The grammar formalisms and approaches to parses construction are considered that take into account such challenges of translation as language transformations. The approach based on the hybrid grammar catching the functional parameters of language structures is proposed.

**Keywords:** formal grammars; machine translation; syntactic analysis; statistical models; functional approach

## 1 Introduction

The paper is focused on discovering the ways of combining the solutions from the two research paradigms for design of an optimum parsing technique, i. e., a syntactic analyzer featuring both rule-based and statistical methods with the view of enhancing the existing language processing technologies. The present state-of-the-art requires capturing human language intuition in statistical translation models and multilingual knowledge extraction systems<sup>3</sup>.

In statistical machine translation (SMT), the task of translating from one natural language into another is treated as a machine learning problem. This means that via training on a very large number of hand-made translation samples, the SMT algorithms master the rules of translation automatically. The application of statistical models has considerably advanced the area of machine translation since the last decade of the previous century; however, now new ideas and methods appear aimed at creating systems that efficiently combine symbolic and statistical approaches comprising different models. Both the paradigms move towards each other: more and more linguistics is being introduced into stochastic models of machine translation, and the rule-based systems include statistics into their linguistic rule systems for disambiguation of language structures. The main method for solving numerous problems, including the part of speech establishment and tagging, is the Bayesian approach. The architecture of stochastic systems is based on the dynamic programming algorithm.

Machine learning is rooted in the stochastic research paradigm. The object of machine learning is the automatic inference of the model for some subject area based on the data from this area. Thus, a system learning, for example, syntactic rules should be supplied with a basic set of phrase structure rules. The widely used methods lately have been the N-grams of variable length, in particular, introduction of semantic information into N-grams. In [1], a detailed description is given of the approach to creating a statistical machine translation based on N-grams of bilingual units called “tuples” and the four special attribute functions.

The statistical models are built on the data obtained from the parallel corpora in different languages. Usually, the texts are compared within language pairs. The text in the language from which the translation should be done is called the source text, and the text which is its translation is called the target text.

## 2 Techniques for Handling Syntactic Structures

Segmentation and unification of utterances in the course of translation is a major task for human professional interpreters. They would even say that syntax is “interpreter’s enemy.” The selectivity of languages as to the choice of specific characteristics of description of one and the same situation results in numerous distinctions, and one of the most crucial of them is the degree of particularity in conveying a referential situation. There-

\*This research is partially funded by the Russian Foundation for Basic Research, Grant No. 11-06-00476-a.

<sup>1</sup>Institute of Informatics Problems, Russian Academy of Sciences, kozerenko@mail.ru

<sup>2</sup>Institute of Informatics Problems, Russian Academy of Sciences, petcazay@mail.ru

<sup>3</sup>The research presented in the paper is further developed within the framework of the MRASP International Project for creating the multilingual statistical parser.

fore, a situation which in one language is described by means of one specific feature, in another language may require two or more characteristics. Thus, in many cases the English language is more economical (about 30%, according to the reports of simultaneous interpreters) [2, 3] in expressing a thought than Russian. In practice, the technique applied to overcome this problem is *utterance segmentation* which consists in sectioning a source Russian sentence into two or more utterances in the resulting English sentence.

Another important rule is the least possible change of word order. But this inflicts other unavoidable transformations, and not all of them can be implemented within the framework of machine translation. For example, the general rule for interpreters: a Russian noun which appears at the very beginning of a sentence and has the form of an oblique case, i. e., indirect object standing at the beginning of a Russian sentence, should be transformed into the subject of an English sentence notwithstanding its initial syntactic role. For example,

*Na vstreche dogovorilis'. . . (At the meeting agreed. . .)*

should be translated as

*The meeting reached an agreement. . .*

The requirement of denotational equivalence involves numerous lexical grammatical shifts which cause transformations of the semantic structure of an utterance [2, 3].

Thus, it is obvious that the search for equivalence should be carried out starting from the establishment of semantic equivalence of patterns notwithstanding their structural dissimilarity.

### 3 Establishment of Cross-Language Matches and Structural Synonymy

Translation activity involves the search for equivalence between structures of different languages. However, to establish whether the structures and units are equal or not, one needs some general equivalent against which the language phenomena would be matched. The authors' approach based on the principle "from the meaning to the form" focusing on Functional Syntax would yield the necessary basis for equivalence search. The following types of structural semantic matches have been observed:

*word* → *word*, *phrase structure* → *phrase structure*,  
*word* → *phrase structure*, *morpheme* → *word*,  
*morpheme* → *phrase structure*.

Syntactically, languages are most different in the basic word order of verbs, subjects, and objects in declar-

ative clauses. English is an SVO (i. e., *Subject–Verb–Object*) language, while Russian has a comparatively flexible word order. The syntactic distinction is connected with a semantic distinction in the way languages map underlying cognitive structures onto language patterns, which should be envisaged in MT implementations.

The basis of Cognitive Transfer Grammar (CTG) is composed of the prototypical structures of the languages (in the initial model Russian and English) being investigated, their most probable positions in a sentence, statistical data about the distributive characteristics of structures (the information about the contextual conditions of the use of the investigated objects, i. e., the information about the structural contexts), the schemes of the complete parse of sentences.

The authors' studies are based on the concepts of the functional approach, which is used for the multilingual situation. The basic design unit of the spaces of cognitive transfer is a *transfeme*, i. e., a unit of cognitive transfer which implies a semantic element embodied in a translatable semantically relevant language segment taken in the unity of its categorial and functional characteristics. It establishes the semantic correspondence between the language structures which belong to different language levels and systems [4].

## 4 Some Particulars of the English–Russian Transfer

The machine translation technique employed presupposes three stages: analysis, transfer, and generation. The stage of analysis results in parse representing the structure of the input sentences. Transfer is a bridge between the parse structure of the source language and the input to the generation procedure for the target language. At this stage, the transformation is performed of one parse tree (applicable for the source language presentation) into another tree (presenting the target language). Thus, syntactic transformations imply the mapping of one tree structure to another.

It is very important that *a parse for MT differs from parses required for other purposes*. Thus, the grammar formalisms developed for a unilingual situation (phrase structure rules systems for the English language) [5] would give an untransferable parse in many crucial situations. The segmentation of phrase patterns used for the input language parse was carried out with the consideration of semantics to be reproduced via the target language means. Both the most important universals such as enumeration, comparison, modality patterns, etc. and less general structures were singled out and assigned corresponding target language equivalents.

## 5 Generalized Cognitive Structures Underlying Transfemes

Actually, the process of transfer goes across the functional-categorial values of language units. A language structure which can be subjected to transfer has to be semantically complete from the point of view of its function. The cases of categorial shifts, in particular, when the technique of conversion is employed, require special treatment: the categorial shift of a syntax unit is determined by the functional role of this unit in a sentence (e.g., noun as a modifier  $\rightarrow$  adjective). *Only by creating the centaur concepts. . . 'constituency-dependency,' 'linearity-nonlinearity,' 'form-function,' etc. can we get a reasonably clear picture of linguistic reality* [6].

The starting idea for the language structures segmentation strategy was the notion of functional semantic fields. The system of grammar units, classes, and categories with generalized content supplementary to the content of lexical units, together with the rules of their functioning, is a system which in the end serves for transmission of generalized categories and structures of mental content which make the foundation of utterance sense and constitute the basis of language grammar formation [7].

As it was exhibited in [8], language coding technique is to a great extent determined by the deep semantic structure, and of considerable advantage is such a presentation method that takes semantic level for the starting point and particular semantic units are confronted with the coding devices expressing them. The approach of functional semantics concords in many aspects with the categorial grammar. The system of sentence members (functional roles) is being modified, but its essence is preserved in the new facts qualification via the traditional categories [9].

The formalism developed employs feature-based parse, and head-feature inheritance for phrase structures which are singled out on the basis of functional identity in the source and target languages. To implement the feature-valued inheritance, sometimes broader contexts are taken.

## 6 The Existing Formalisms Influence

At present, basic instruments of formal and natural languages analysis are context-free grammars. Analytical engine for context-free grammars is a one-sided nondeterministic automaton with stack outer memory. In the most trivial case of such automaton's algorithm implementation, it is characterized by an exponential complexity. But iterative upgrade of the algorithm may

lead to its polynomial time value depending on the length of the input set of symbols and necessary for its analysis. Among the existing context-free languages, one can distinguish a class of deterministic context-free languages, which are interpreted by deterministic automaton with stack outer memory. The principal feature of these languages is their unambiguity: it is proved that one can always build an unambiguous grammar for any deterministic context-free language. Since the languages are unambiguous, they are most useful when it comes to building compilers. Moreover, among all the deterministic context-free languages, there exist such classes of languages which allow building a linear recognizer for them. This is the recognizer which time value related to the time for decision-making about a set of symbols belonging to a language has a linear dependency on the chain length. Syntactic constructions in the majority of the existing programming languages might be classified as ones of the class mentioned. This aspect is very important when developing up-to-date high-speed compilers.

As a rule, the more complex from the mathematical perspective an analytical engine is, the more challenging is its technical implementation. Case with finite state automata (both with memory and without it) is not an exception. For instance, rules of natural language include a plenty of features of natural language structures which are attributive by their nature. Processing of such transformation rules implemented using finite-state automaton may lead to the increase of the automaton states and to the growth of amount of the transformation rules when changing the states. Moreover, when implementing the above mentioned transformation rules, one needs to create a problem-oriented analytical engine to handle processing of the input set of symbols of the natural language structures and to apply interpreting rules to them.

The present implementation formalism was developed taking into account the apparatus of phrase structure and unification grammars: Head-Driven Phrase Structure Grammars (HPSG) [10], Generalized Phrase Structure Grammars (GPSG) [11], and Revised Generalized Phrase Structure Grammars (RGPSG) [12]. Categorial and Dependency [13] grammars were also considered. Important was the strict lexicalism principle of the HPSG, i.e., word and phrase structures are governed by independent principles. Roles are determined by verbal valences; utterances are a blend of categorial meanings and role meanings and their structural projections which are specific for every particular language.

The technique of categorization, i.e., generation of a backbone grammar of atomic categories from distinct sets of feature bindings was first suggested in [14]. It was used for building a shift-reduce table for the Alvey grammar [5], but there it was necessary to subsume

each category into its most general unifying category, so reducing the overall number of categories.

Generally, for the Russian language, dependency grammars have been applied. And phrase structure approach seemed to be less applicable here. Hence, of particular interest were the study and comparison of both the formal approaches, so that practical algorithmic solutions could be worked out.

A certain key was suggested in the coexisting systems of Immediate Dominance (ID) rules and phrase structure (PS) rules in ANLT (Alvey Natural Language Tools) [5] based on a variant of GPSG. The ID rules encode unordered dependency relations and further are subjected to linearization to be applied for the parse. Generalized phrase structure grammar may be thought of as a grammar for generating a context-free grammar. The generation process begins with ID rules which are context-free productions with unordered right-hand sides. An important feature of ID rules is that nonterminals in the rules are not atomic symbols (e. g., NP — Not Present). Rather, GPSG nonterminals are sets of [feature, feature-value] pairs. For example, [N+] is a [feature, feature-value] pair, and the set {[N+], [V-], [BAR 2]} is the GPSG representation of a noun phrase. Next, metarules are applied to the ID rules that results in an enlarged set of ID rules. In the RGPSG, the finite closure problem is used to determine the cost of metarule application. Principles of universal feature instantiation (UFI) are applied to the resulting enlarged set of ID rules, defining a set of phrase structure trees of depth one (local trees). One principle of UFI is the head feature convention which ensures that phrases are projected from lexical heads. Finally, linear precedence statements are applied to the instantiated local trees. The final result is a set of ordered local trees, and these are equivalent to the context-free productions in a context-free grammar. The process of assigning structural descriptions to utterances consists of two steps in GPSG: the projection of ID rules to local trees and the derivation of utterances from nonterminals, using the local trees.

In GPSG, there are three category-valued features: SLASH which marks the path between a gap and its filler with the category of the filler; AGR which marks the path between an argument and a functor that syntactically agrees with it (between the subject and the matrix verb, for example); and WH which marks the path between a wh-word and the minimal clause that contains it with the morphological type of the wh-word. In RGPSG, the revision is unit feature closure: to limit category-valued features to containing only 0-level categories, i. e., 0-level categories do not contain any category-valued features. The GPSG's ID/LP format models the head parameter and some free word order information. The HPSG formalism is based on phrase

structure rules, but dominance relations are implemented via head elements. Phrasal types are also treated in terms of multiple inheritance hierarchies that allow generalizations about diverse construction types to be factored into various cross-cutting dimensions.

In fact, each nonlinear dependency rule is an encoded potential for actualization of a set of possible linear phrase structures. Therefore, the authors assumed a more computationally practical approach (to our knowledge, never used before in a bilingual situation), that of feature-valued head-driven phrase structures for both English and Russian.

## 7 Parse Engines: Viterbi and Early Algorithms

The *Viterbi algorithm* is a dynamic programming algorithm for finding the most likely sequence of hidden states — called the *Viterbi path* — that results in a sequence of observed events, especially in the context of Markov information sources, and more generally, hidden Markov models. The algorithm makes a number of assumptions:

- both the observed events and hidden events must be in a sequence, this sequence often corresponds to time;
- these two sequences need to be aligned, and an instance of an observed event needs to correspond to exactly one instance of a hidden event; and
- computing the most likely hidden sequence up to a certain point  $t$  must depend only on the observed event at point  $t$ , and the most likely sequence at point  $t - 1$ .

These assumptions are all satisfied in a first-order hidden Markov model. The terms “Viterbi path” and “Viterbi algorithm” are also applied to related dynamic programming algorithms that discover the single most likely explanation for an observation. For example, in statistical parsing, a dynamic programming algorithm can be used to discover the single most likely context-free derivation (parse) of a string, which is sometimes called the “Viterbi parse.” The Viterbi algorithm operates on a state machine assumption. That is, at any time, the system being modeled is in one of a finite number of states. While multiple sequences of states (paths) can lead to a given state, at least one of them is a most likely path to that state, called the “survivor path.” This is a fundamental assumption of the algorithm because the algorithm will examine all possible paths leading to a state and only keep the one most likely. This way the algorithm does not have to keep track of all possible paths, only one per state.

Another basic assumption is that a transition from a previous state to a new state is marked by an incremental metric, usually a number. This transition is computed from the event. The third fundamental assumption is that the events are cumulative over a path in some sense, usually additive. When an event occurs, the algorithm examines moving forward to a new set of states by combining the metric of a possible previous state with the incremental metric of the transition due to the event and chooses the best. The incremental metric associated with an event depends on the transition possibility from the old state to the new state. For example, in data communications, it may be possible to only transmit half the symbols from an odd numbered state and the other half from an even numbered state.

*Earley's parsing algorithm* is a general algorithm, able to handle any context-free grammar [15]. Earley parsers operate by constructing a sequence of sets, sometimes called Earley sets. Given an input  $x_1x_2 \dots x_n$ , the parser builds  $n + 1$  sets: an initial set  $S_0$  and one set  $S_i$  for each input symbol  $x_i$ . Elements of these sets are referred to as (Earley) items, which consist of three parts: a grammar rule, a position in the right-hand side of the rule indicating how much of that rule has been seen, and a pointer to an earlier Earley set. Typically, Earley items are written as  $[A \rightarrow \alpha \bullet \beta, j]$  where the position in the rule's right-hand side is denoted by a dot ( $\bullet$ ) and  $j$  is a pointer to set  $S_j$ . An Earley set  $S_i$  is computed from an initial set of Earley items in  $S_i$ , and  $S_{i+1}$  is initialized, by applying the following three steps to the items in  $S_i$  until no more can be added:

SCANNER . If  $[A \rightarrow \dots \bullet a \dots, j]$  is in  $S_i$   
 and  $a = x_i + 1$ , add  $[A \rightarrow \dots a \bullet \dots, j]$  to  $S_{i+1}$ .  
 PREDICTOR . If  $[A \rightarrow \dots \bullet B \dots, j]$  is in  $S_i$ ,  
 add  $[B \rightarrow \bullet \alpha, i]$  to  $S_i$  for all rules  $B \rightarrow \alpha$ .  
 COMPLETER . If  $[A \rightarrow \dots \bullet, j]$  is in  $S_i$ ,  
 add  $[B \rightarrow \dots A \bullet \dots, k]$  to  $S_i$   
 for all items  $[B \rightarrow \dots \bullet A \dots, k]$  in  $S_j$ .

An item is added to a set only if it is not in the set already. The initial set  $S_0$  contains the item  $[S_- \rightarrow \bullet S, 0]$  to begin with -. We assume the grammar is augmented with a new start rule  $S_- \rightarrow S$ , and the final set must contain  $[S_- \rightarrow S \bullet, 0]$  for the input to be accepted.

Many different metrics exist for evaluating parsing results, including Viterbi, Crossing Brackets Rate, Zero Crossing Brackets Rate, and several others. However, most parsing algorithms, including the Viterbi algorithm, attempt to optimize the same metric, namely, the probability of getting the correct labeled tree. By choosing a parsing algorithm appropriate for the evaluation metric, better performance can be achieved.

## 8 Experiments with Functional Grammars

Representation of grammar rules, or transformation rules, as mathematical functions has certain advantages in comparison with formal grammar apparatus:

- usage of functional programming tools to build systems of transfer immediately; and
- possibility of higher-order function applications.

The approach denoted by the concept *system of transfer* is employed for experimental software implementation of an analytic engine for processing of transformations expressed as functions [16].

Using *n-tuples* as a form of representation of natural language structures enables the system to generalize attributive characteristics of words and use pattern alternations of such structures in prospect, using *n-tuples* for storage of attributive features of language structures allows to extract functions according to language structures of various levels of abstraction (for instance, a word, a phrase, a sentence, etc.), any of such language structures has its own set of attributes; hence, there should be functions which have words, phrases, and so on, as their arguments.

The second distinctive feature of the functional approach is a technical one. It is essential to design and implement a relevant analytic engine for every information system the tasks of which correlate with natural language structures analysis based on mathematical apparatus of formal grammars. In the case of functional representation of grammar rules as an analytical machine, the environment of the functional programming might be used (for instance, Erlang, Haskell) [17].

## 9 Implementation Techniques

The primary purpose in introducing feature structures and unification was to provide a way to express syntactic constraints that would be difficult to express using the mechanisms of context-free grammars alone. The next step was to design a way to integrate feature structures and unification operations into the specification of a grammar.

A constraint-based formalism comprising some features of the HPSG was developed. The formalism provides representation mechanisms for the fine-grained information about number and person, agreement, subcategorization, as well as semantics for syntactic representations. The system of rules based on this formalism can be called the Cognitive Transfer Grammar [18] and consists of transferable phrase structures together with

the transfer rules which are combined within the same pattern. Such patterns, or Cognitive Transfer Structures (CTS), are constitutional components of the declarative syntactical processor module and encode both linear precedence and dependency relations within phrase structures.

In the present approach, the direct encoding of possible subcategorization features is made via a verbal CTS. Since the verbs can subcategorize for quite complex frames composed of many different phrasal types, the authors first established a list of possible phrasal types that can make up these frames, e. g., VPto “*I want to know*,” VPing “*He contemplates using them*,” Sto “*feel themselves to be relatively happy*.” Each verb allows many different subcategorization frames.

If compared with the existing phrasal subcategorization frames [19], in the present authors’ system, the emphasis is laid on functional motivation.

The above stated methods are being employed for design and development of a linguistic knowledge base Intertext. It is a linguistic resource with semantic grouping of phrase structure patterns provided with the links to synonymous structures at all language levels for the languages included into the linguistic base.

The authors’ focus on configurations provides high portability to the language processing software designed under these principles: we can operate with a lexicon which has only standard linguistic information including morphological characteristics, part of speech information, and the indication of transitivity for verbs. The Intertext linguistic knowledge base comprises the following components:

- the inventory of phrasal configurations (CTS);
- the database of semantically aligned parallel texts; and
- structural parse editor (at present, under development).

## 10 Concluding Remarks

The functionally motivated approach to natural language parsers design and development provides a sound and extensible platform for simulation of cross-lingual syntactic-semantic transfer and can be applied to a greater number of languages (especially, with similar categorial feature-value structures). However, the problems of discontinuity, reference resolution, and ambiguity, though partially treated, still remain. Further research is connected with introducing special feature-value augmentations to the existing presentations for tracing the discontinuous structures, specifying the semantic values of particular head features and verbal subcategorization frames, and numerous phrasal units adjustment.

## References

1. Marino J. B., Banchs R. E., Crego J. M., de Gispert A., Lambert P., Fonollosa J. A. R., Costa-Jussa M. R. N-gram-based machine translation // *Computational Linguistics*, 2006. Vol. 32. No. 4. P. 527–549.
2. Visson L. Syntactical problems for the Russian–English interpreter. No uncertain terms // *FBIS*, 1989. Vol. 4. No. 2. P. 2–8.
3. Visson L. From Russian into English: An introduction to simultaneous interpretation. — Ann Arbor, Michigan: Ardis, 1991.
4. Comrie B. Language universals and linguistic typology. — 2nd ed. — Oxford: Basil Blackwell, 1989.
5. Grover C., Carroll J., Briscoe T. The Alvey natural language tools grammar (4th Release). Technical Report. Computer Laboratory, University of Cambridge, 1993.
6. Shaumyan S. A semiotic theory of language. — Indiana University Press, 1987.
7. Bondarko A. V. Functional grammar principles and aspectology questions. — Moscow, 2001. [In Russian.]
8. Kibrik A. E. Studies in general and applied linguistics issues. — 2nd ed. — Moscow, 2001. [In Russian.]
9. Zolotova G. A. Communicative principles of the Russian syntax. — Moscow, 2001. [In Russian.]
10. Pollard C., Sag I. A. Head-driven phrase structure grammar. — Chicago: University of Chicago Press, 1994.
11. Gazdar G., Klein E., Pullum G., Sag I. Generalized Phrase Structure Grammar. — Oxford: Basil Blackwell, 1985.
12. Ristard E. S. Computational complexity of current GPSG theory // 24th Annual Meeting of the Association for Computational Linguistics Proceedings. — Columbia University, New York: Association for Computational Linguistics, 1986. P. 30–39.
13. Mel’cuk I. A. Dependency syntax: Theory and practice. — State University of New York Press, 1988.
14. Gazdar G., Mellish C. Natural language processing in Prolog. — Wokingam, UK: Addison-Wesley, 1989.
15. Earley J. An efficient context-free parsing algorithm // *Communications of the ACM*, 1970. Vol. 13. P. 94–102.
16. Ermakov P., Kozhunova O. Application of the functional programming tools in the tasks of language and interlanguage structures representation // SYRCoSE 2011. — Moscow–Yekaterinburg, 2011. P. 48–54.
17. Web sites for Erlang <http://erlang.org>; Website for Haskell <http://haskell.org>.
18. Kozerenko E. Features and categories design for the English–Russian transfer model // *Advances in Natural Language Processing and Applications Research in Computing Science*, 2008. Vol. 33. P. 123–138.
19. Baker C. F., Fillmore C. J., Lowe J. B. The Berkeley FrameNet project // *COLING/ACL-98*, 1998. P. 86–90.

## СТРАТЕГИИ СИНТАКСИЧЕСКОГО АНАЛИЗА НА ОСНОВЕ ВЕРШИННЫХ ГРАММАТИК И МЕТОДЫ ИХ РЕАЛИЗАЦИИ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Е. Б. Козеренко<sup>1</sup>, П. В. Ермаков<sup>2</sup>

<sup>1</sup>Институт проблем информатики Российской академии наук, kozerenko@mail.ru

<sup>2</sup>Институт проблем информатики Российской академии наук, petcazay@mail.ru

**Аннотация:** Рассмотрены задачи проектирования и разработки синтаксических анализаторов (парсеров) в многоязычных системах обработки естественного языка, машинного перевода и извлечения знаний из текстов. Рассмотрены грамматические формализмы и подходы к созданию парсеров, в которых учитываются реальные трудности перевода, такие как языковые трансформации. Предложен подход, учитывающий функциональные параметры языковых структур на основе гибридной грамматики.

**Ключевые слова:** формальные грамматики; машинный перевод; синтаксический анализ; статистические модели; функциональный подход

### STOCHASTIC INFORMATIONAL TECHNOLOGIES FOR NONLINEAR CIRCULAR STOCHASTIC SYSTEMS ANALYSIS

I. N. Sinitsyn

IPI RAN, sinitsin@dol.ru

The paper is devoted to stochastic informational technologies (StIT) for analysis, analytical and statistical modeling of circular processes in circular nonlinear stochastic systems (StS) based on statistical linearization by wrapped normal distribution. For StIT, software tools “CStS-ANALYSIS” in MATLAB are developed and tested.

**Keywords:** analytical modeling; circular stochastic process; circular stochastic system; circular statistical linearization; computer aided support of statistical scientific research; MATLAB; correlation equations; spectral and correlation equations; stochastic informational technologies; statistical modeling

### DISCRETE TIME QUEUEING SYSTEM WITH UNRELIABLE SERVER

A. V. Pechinkin<sup>1</sup> and I. A. Sokolov<sup>2</sup>

<sup>1</sup>IPI RAN, apechinkin@ipiran.ru

<sup>2</sup>IPI RAN, isokolov@ipiran.ru

Consideration is given to the discrete time queueing system  $\text{Geo}/G/1/\infty$  with server subject to two types of breakdowns. The server can break down either when it is busy or when it is idle, which happens with different probabilities. Repair time distribution depends on the type of the breakdown. Expressions for the stationary probability distribution and other main stationary characteristics are given. It is shown how the obtained results can be used to find some stationary characteristics of  $\text{Geo}_2/\bar{G}_2/1/\infty$  with two types of customers and preemptive priority.

**Keywords:** queueing system; discrete time; unreliable server; breakdowns

### ON A CLASS OF MARKOVIAN QUEUES

Ya. A. Satin<sup>1</sup>, A. I. Zeifman<sup>2</sup>, A. V. Korotysheva<sup>3</sup>, and S. Ya. Shorgin<sup>4</sup>

<sup>1</sup>Vologda State Pedagogical University, yacovi@mail.ru

<sup>2</sup>Vologda State Pedagogical University; IPI RAN; VSCC CEMI RAS, a\_zeifman@mail.ru

<sup>3</sup>Vologda State Pedagogical University, a\_korotysheva@mail.ru

<sup>4</sup>IPI RAN, SShorgin@ipiran.ru

The nonstationary continuous-time Markovian queueing models are considered. Arrival and service rates are supposed to be independent on the length of the queue. The bounds of the rate of convergence and stability for some characteristics of such systems are obtained.

**Keywords:** nonstationary Markovian queues; rate of convergence; stability; bounds

## ACTIVITY MAXIMA IN FREE-SCALE RANDOM NETWORKS WITH HEAVY TAILS

A. V. Lebedev

M. V. Lomonosov Moscow State University, avlebed@yandex.ru

The oriented power-law random graphs are considered as the models of information networks, where each node has a random information activity whose distribution has heavy (regularly varying) tail. The model of a random graph, in which incoming degrees of vertices are independent and have distribution with power tail, is used. Sufficient conditions have been got under which the maximum total activity (of the node and its incoming neighbors) increases asymptotically as well as the maximum of individual activities, and therefore, for them, Frechet limit law is hold.

**Keywords:** maxima; random sums; free-scale networks; power law; random graph; heavy tail; regular variation; Frechet distribution

## ANALYTICAL MODEL FOR CALCULATING THE PERFORMANCE PLAN OF DISTRIBUTING MULTIPROCESSOR SYSTEM RESOURCES IN SOLVING THE PROBLEMS OF SPECIAL CLASS

M. Ya. Agalarov

IPI RAN, murad-agalarov@yandex.ru

A model of multiprocessor system designed to solve the tasks that are parallelized on weakly dependent calculations is considered. As a model, multiservice queuing system with apparent losses, Poisson incoming flow, and general service time distribution functions of tasks is used. Recursion formulas for calculating the stationary probability distribution of states and the explicit expressions for the probability of system failure for different types of tasks are obtained. In this model, a method for assessing the capacity of a multiprocessor system for a given resource allocation static plan is suggested.

**Keywords:** multiprocessor system; queuing system; multiservice system; distribution of computing resources

## DECONVOLUTION UNDER PARTIALLY KNOWN ERROR DISTRIBUTION

V. G. Ushakov<sup>1</sup> and N. G. Ushakov<sup>2</sup>

<sup>1</sup>Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vgushakov@mail.ru

<sup>2</sup>Institute of Microelectronics Technology and High Purity Materials, Russian Academy of Sciences, ushakov@math.ntnu.no

The problem of nonparametric estimation of a probability distribution is considered for a case when the sample is contaminated by a random noise. It is supposed that the distribution of the error is known only partially. Identifiability and consistent estimation are investigated.

**Keywords:** nonparametric estimation; deconvolution

## ANALYSIS AND OPTIMIZATION PROBLEMS FOR SOME USERS ACTIVITY MODEL. PART 1. ANALYSIS AND PREDICTION

A. V. Bosov

IPI RAN, AVBosov@ipiran.ru

A mathematical model describing the activity of users, forming a query to some information system, is suggested. The properties of the model are investigated, the procedures of prediction and identification of equations parameters are described. The results of numerical experiments, including real data processing, are presented.

**Keywords:** information systems; modeling; ergodic random process; parametric identification; prediction

## SOLUTION OF DBMS INTERACTION PROBLEM IN THE CROSS-PLATFORM LIBRARY EFFIDB

A. V. Yanushko<sup>1</sup>, A. V. Babanin<sup>2</sup>, O. A. Kuznetsova<sup>3</sup>, and S. V. Petrushenko<sup>4</sup><sup>1</sup>ASoft, yan@asoft.ru<sup>2</sup>Russian Research Institute of Computer Science and Information, ababanin@pvti.ru<sup>3</sup>ASoft, ok@asoft.ru<sup>4</sup>ASoft, op@asoft.ru

The article covers the problems of unified interaction with different database management systems in various software environments. It examines present-day solutions in this field and analyzes their advantages and drawbacks. The requirements to cross platform instrument for interaction between C++ application code and DBMS are listed and solution implemented as a dynamic library is suggested. The library provides specialized classes for each of the concepts of relational databases: the actual database connection, table, tools for data manipulation, etc. The limits of applicability of the proposed solutions and the practice of using the library in real projects are also analyzed. Code case studies are provided as well.

**Keywords:** DBMS, C++, connectivity library, cross-platform

## PROBABILISTIC STATISTICAL EVALUATION OF THE INFORMATION OBJECTS ADEQUACY

L. A. Kuznetsov

Lipetsk State Technical University, kuznetsov@stu.lipetsk.ru

Mathematical basis and original methodology for developing evaluation systems of semantic proximity of information objects (IO) in natural language are presented. A probabilistic statistical representation of the compared IO is introduced. The information theory is used to estimate the semantic proximity of IO. The methodology can be used for synthesis of computer-based systems. The results of the practical testing of the methodology effectiveness are presented.

**Keywords:** information objects; natural language; semantic adequacy; probabilistic model; information theory

INFORMATION AND TELECOMMUNICATION PROJECTS MANAGEMENT:  
“TIMELINESS–PERFORMANCE–INFORMATION”A. A. Zatsarinny<sup>1</sup> and A. P. Shabanov<sup>2</sup><sup>1</sup>IPI RAN, AZatsarinny@ipiran.ru<sup>2</sup>IBS Expertiza Company, AShabanov@ibs.ru

Methodological approach to the management of information and telecommunication projects is discussed. Justification of performance requirements to the paths of the technological systems and to the number of subjects of the functional organizational structures, performing work in accordance with the messages taken from these tracts is examined.

**Keywords:** project management; technology system; organizational structure; timeliness; performance; information

## MODELING OF PROCESSES FOR CREATION OF EXPERT KNOWLEDGE FOR MONITORING OF GOAL-ORIENTED PROGRAMME ACTIVITIES

I. M. Zatsman<sup>1</sup> and A. A. Durnovo<sup>2</sup><sup>1</sup>IPI RAN, iz\_ipi@a170.ipi.ac.ru<sup>2</sup>IPI RAN, duralex49@mail.ru

Statement of a problem of goal-oriented knowledge representation about indicators of monitoring is considered and its decision consisting of four components is offered: (i) stationary model of computer representation of goal-oriented knowledge about indicators; (ii) Frege space for computer representation; (iii) nonstationary model of computer representation; and (iv) the proactive dictionary of a lingware of an evaluation system. The first three of the four components represent a theoretical part of the decision of this problem, and the fourth component represents an applied part of its decision.

**Keywords:** problem of goal-oriented knowledge representation about indicators; semiotic models of computer representation of knowledge about indicators; indicator concepts; indicator denotata

## TRANSFORMATIONAL MODELS OF LANGUAGE STRUCTURES FOR MACHINE TRANSLATION FROM FRENCH INTO RUSSIAN

Yu. I. Morozova

IPI RAN, yulia-ipi@yandex.ru

The paper focuses on the problems of studying transformational properties of language objects in the process of translation of predicative structures from French into Russian. The paper studies the cases when in the process of translation, predicative words change their syntactic category or their grammatical features. The texts of patents in French and their translations into Russian performed by professional translators were used as the material for the research.

**Keywords:** machine translation from French into Russian; functional semantics; language transformations; head-driven grammars

## STRATEGIES OF SYNTACTIC ANALYSIS BASED ON HEAD-DRIVEN GRAMMARS AND METHODS OF THEIR IMPLEMENTATION IN INFORMATION SYSTEMS

E. B. Kozerenko<sup>1</sup> and P. V. Ermakov<sup>2</sup><sup>1</sup>IPI RAN, kozerenko@mail.ru<sup>2</sup>IPI RAN, petcazay@mail.ru

The problems of design and development of syntactic parsers in multilingual natural language processing systems, machine translation, and knowledge extraction from texts are considered. The grammar formalisms and approaches to parsers construction are considered that take into account such challenges of translation as language transformations. An approach based on the hybrid grammar catching the functional parameters of language structures is proposed.

**Keywords:** formal grammars; machine translation; syntactic analysis; statistical models; functional approach

## Об авторах

**Агаларов Мурад Яверович** (р. 1987) — аспирант ИПИ РАН

**Бабанин Алексей Васильевич** (р. 1966) — заместитель директора Всероссийского научно-исследовательского института проблем вычислительной техники и информатизации

**Босов Алексей Вячеславович** (р. 1969) — доктор технических наук, заведующий сектором ИПИ РАН

**Дурново Александр Адрианович** (р. 1949) — ведущий программист ИПИ РАН

**Ермаков Петр Вячеславович** (р. 1985) — младший научный сотрудник ИПИ РАН

**Зацаринный Александр Алексеевич** (р. 1951) — доктор технических наук, профессор, заместитель директора ИПИ РАН

**Зацман Игорь Моисеевич** (р. 1952) — кандидат технических наук, заведующий отделом ИПИ РАН

**Зейфман Александр Израилевич** (р. 1954) — доктор физико-математических наук, профессор, декан факультета прикладной математики и компьютерных технологий Вологодского государственного педагогического университета, старший научный сотрудник ИПИ РАН, ведущий научный сотрудник ВНКЦ ЦЭМИ РАН

**Козеренко Елена Борисовна** (р. 1959) — кандидат филологических наук, заведующая лабораторией ИПИ РАН

**Коротышева Анна Владимировна** (р. 1988) — аспирантка факультета прикладной математики и компьютерных технологий Вологодского государственного педагогического университета

**Кузнецов Леонид Александрович** (р. 1942) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий кафедрой Липецкого государственного технического университета

**Кузнецова Ольга Анатольевна** (р. 1974) — руководитель проектов внедрения компании АСофт

**Лебедев Алексей Викторович** (р. 1971) — кандидат физико-математических наук, доцент Московского государственного университета им. М. В. Ломоносова

**Морозова Юлия Игоревна** (р. 1984) — научный сотрудник ИПИ РАН

**Петрушенко Сергей Владимирович** (р. 1965) — заместитель генерального директора компании АСофт

**Печинкин Александр Владимирович** (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник ИПИ РАН; профессор Российского университета дружбы народов

**Сатин Яков Александрович** (р. 1978) — кандидат физико-математических наук, старший преподаватель кафедры математического анализа Вологодского государственного педагогического университета

**Синицын Игорь Николаевич** (р. 1940) — доктор технических наук, профессор, заведующий отделом ИПИ РАН

**Соколов Игорь Анатольевич** (р. 1954) — академик (действительный член) Российской академии наук, доктор технических наук, директор ИПИ РАН

**Ушаков Владимир Георгиевич** (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

**Ушаков Николай Георгиевич** (р. 1954) — доктор физико-математических наук, ведущий научный сотрудник Института проблем технологии микроэлектроники и особочистых материалов РАН

**Шабанов Александр Петрович** (р. 1949) — кандидат технических наук, главный эксперт ООО «ИБС Экспертиза»

**Шоргин Сергей Яковлевич** (р. 1952) — доктор физико-математических наук, профессор, заместитель директора ИПИ РАН

**Янушко Александр Владимирович** (р. 1972) — технический директор компании АСофт

## *About Authors*

**Agalarov Murad Ya.** (b. 1987) — PhD student, Institute of Informatics Problems, Russian Academy of Sciences

**Babanin Alexey V.** (b. 1966) — Deputy Director, Russian Research Institute of Computer Science and Information

**Bosov Alexey V.** (b. 1969) — Doctor of Science in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

**Durnovo Alexander A.** (b. 1949) — leading programmer, Institute of Informatics Problems, Russian Academy of Sciences

**Ermakov Piotr V.** (b. 1985) — junior scientist, Institute for Informatics Problems, Russian Academy of Sciences

**Korotysheva Anna V.** (b. 1988) — PhD student, Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University

**Kozerenko Elena B.** (b. 1959) — Candidate of Science (PhD) in linguistics, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

**Kuznetsov Leonid A.** (b. 1942) — Doctor of Science in technology, professor, Head of Department, Lipetsk State Technical University

**Kuznetsova Olga A.** (b. 1974) — Head of Implementation, ASoft

**Lebedev Alexey V.** (b. 1971) — Candidate of Science (PhD) in physics and mathematics, associate professor, M. V. Lomonosov Moscow State University

**Morozova Yulia I.** (b. 1984) — scientist, Institute of Informatics Problems, Russian Academy of Sciences

**Pechinkin Alexander V.** (b. 1946) — Doctor of Science in physics and mathematics; principal scientist, Institute of Informatics Problems, Russian Academy of Sciences; professor, Peoples' Friendship University of Russia

**Petrushenko Sergey V.** (b. 1965) — Chief Operations Officer, ASoft

**Satin Yakov A.** (b. 1978) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Vologda State Pedagogical University

**Shabanov Alexander P.** (b. 1949) — Candidate of Science (PhD) in technology, principal expert, IBS Expertiza Company

**Shorgin Sergey Ya.** (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

**Sinityn Igor N.** (b. 1940) — Doctor of Science in technology, professor, honored scientist of RF, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

**Sokolov Igor A.** (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, director, Institute of Informatics Problems, Russian Academy of Sciences

**Ushakov Nikolai G.** (b. 1954) — Doctor of Science in physics and mathematics, leading scientist, Institute of Microelectronics Technology and High Purity Materials, Russian Academy of Sciences

**Ushakov Vladimir G.** (b. 1952) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

**Yanushko Alexander V.** (b. 1972) — Chief Technology Officer, ASoft

**Zatsarinny Alexander A.** (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

**Zatsman Igor M.** (b. 1952) — Candidate of Science (PhD) in informatics, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

**Zeifman Alexander I.** (b. 1954) — Doctor of Science in physics and mathematics, professor, Dean of the Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; leading scientist, VSCC CEMI RAS

## АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2011 г.

	<b>Выпуск</b>	<b>Стр.</b>
<b>Агаларов М. Я.</b> Аналитическая модель расчета эффективности плана распределения вычислительных ресурсов многопроцессорной системы при решении специального класса задач.....	4	29
<b>Бабанин А. В.</b> см. Янушко А. В.		
<b>Баша Н. С., Шульга Л. А.</b> Алгоритм автоматического выделения лица на термографических изображениях.....	1	73
<b>Бенинг В. Е., Горшенин А. К., Королев В. Ю.</b> Асимптотически оптимальный критерий проверки гипотез о числе компонент смеси вероятностных распределений.....	3	4
<b>Босов А. В.</b> Задачи анализа и оптимизации для модели пользовательской активности. Часть 1. Анализ и прогнозирование.....	4	40
<b>Гавриленко С. В.</b> Уточнение неравномерной оценки скорости сходимости распределений пуассоновских случайных сумм к нормальному закону.....	1	12
<b>Горшенин А. К.</b> см. Бенинг В. Е.		
<b>Демин П. В.</b> О двух моделях распределения ресурсов при организации инвестиционных процессов.....	1	65
<b>Дурново А. А.</b> см. Зацман И. М.		
<b>Ермаков П. В.</b> см. Козеренко Е. Б.		
<b>Зацаринный А. А., Шабанов А. П.</b> Управление инфокоммуникационными проектами: «своевременность—производительность—информация».....	4	76
<b>Зацман И. М., Дурново А. А.</b> Моделирование процессов формирования экспертных знаний для мониторинга программно-целевой деятельности.....	4	84
<b>Зейфман А. И., Коротышева А. В., Панфилова Т. Л., Шоргин С. Я.</b> Оценки устойчивости для некоторых систем обслуживания с катастрофами.....	3	27
<b>Зейфман А. И.</b> см. Сатин Я. А.		
<b>Игнатъев М. Б.</b> Теория сложных систем и кибернетическая картина мира.....	2	58
<b>Козеренко Е. Б., Ермаков П. В.</b> Стратегии синтаксического анализа на основе вершинных грамматик и методы их реализации в информационных системах.....	4	107
<b>Козмидиади В. А.</b> Дерево работ и массово-параллельная обработка.....	2	90
<b>Кольцов П. П.</b> Использование метрик при сравнительном исследовании качества работы алгоритмов сегментации изображений.....	3	53
<b>Корепанов Э. Р.</b> Стохастические информационные технологии на основе фильтров Пугачёва.....	2	36
<b>Королев В. Ю., Крылов В. А., Кузьмин В. Ю.</b> Устойчивость конечных смесей обобщенных гамма-распределений относительно возмущений параметров.....	1	31
<b>Королев В. Ю., Шевцова И. Г., Шоргин С. Я.</b> О неравенствах типа Берри—Эссеена для пуассоновских случайных сумм.....	3	64
<b>Королев В. Ю.</b> см. Бенинг В. Е.		
<b>Королев В. Ю.</b> см. Соколов И. А.		
<b>Коротышева А. В.</b> см. Зейфман А. И.	3	27
<b>Коротышева А. В.</b> см. Сатин Я. А.		
<b>Кристоф Г., Ульянов В. В.</b> О точности приближений нормированных хи-квадрат распределений асимптотическими разложениями Эджворта—Чебышева.....	1	25
<b>Крылов А. С.</b> см. Павельева Е. А.		
<b>Крылов В. А.</b> см. Королев В. Ю.		

	Выпуск	Стр.
<b>Кудрявцев А. А., Шоргин С. Я.</b> Об уточнении некоторых результатов для одной байесовской модели массового обслуживания .....	1	78
<b>Кудрявцева И. А., Пантелеев А. В.</b> Приближенные методы решения задачи диагностики плоским зондом сильноионизованной плазмы с учетом Кулоновских столкновений .....	3	46
<b>Кузнецов Л. А.</b> Вероятностно-статистическая оценка адекватности информационных объектов .....	4	64
<b>Кузнецова О. А.</b> см. Янушко А. В.		
<b>Кузьмин В. Ю.</b> см. Королев В. Ю.		
<b>Лебедев А. В.</b> Максимумы активности в безмасштабных случайных сетях с тяжелыми хвостами .....	4	25
<b>Лери М. М., Челпюкова И. А.</b> Об одной статистической задаче для случайных графов Интернет-типа .....	3	34
<b>Марков Ю. Г., Перепелкин В. В., Сеницын И. Н., Семендяев Н. Н.</b> Информационные модели неравномерности вращения Земли .....	2	17
<b>Морозова Ю. И.</b> Трансформационные модели языковых структур для французско-русского машинного перевода .....	4	99
<b>Нефедова Ю. С., Шевцова И. Г.</b> О точности нормальной аппроксимации для распределений пуассоновских случайных сумм .....	1	39
<b>Павельева Е. А., Крылов А. С.</b> Алгоритм сравнения изображений радужной оболочки глаза на основе ключевых точек .....	1	68
<b>Пантелеев А. В., Рыбаков К. А.</b> Синтез оптимальных нелинейных стохастических систем управления спектральным методом .....	2	69
<b>Пантелеев А. В.</b> см. Кудрявцева И. А.		
<b>Панфилова Т. Л.</b> см. Зейфман А. И.		
<b>Перепелкин В. В.</b> см. Марков Ю. Г.		
<b>Петрушенко С. В.</b> см. Янушко А. В.		
<b>Печинкин А. В.</b> Двухприоритетная система с резервированием каналов и марковским входящим потоком .....	1	2
<b>Печинкин А. В., Соколов И. А.</b> Система массового обслуживания с ненадежным прибором в дискретном времени .....	4	6
<b>Преображенский Н. Б., Файзулхаков Я. Р.</b> Проблема компенсации Рэлеевских замираний в радиоканалах подвижных систем голосовой связи .....	2	82
<b>Пугачёв В. С., Сеницын И. Н.</b> Структурная теория сложных стохастических систем ...	2	4
<b>Разумчик Р. В.</b> Система массового обслуживания с отрицательными заявками, бункером для вытесненных заявок и различными интенсивностями обслуживания ...	3	41
<b>Рыбаков К. А.</b> см. Пантелеев А. В.		
<b>Сатин Я. А., Зейфман А. И., Коротышева А. В., Шоргин С. Я.</b> Об одном классе марковских систем обслуживания .....	4	18
<b>Семендяев Н. Н.</b> см. Марков Ю. Г.		
<b>Сеницын И. Н.</b> Стохастические информационные технологии для исследования нелинейных круговых стохастических систем .....	4	2
<b>Сеницын И. Н.</b> см. Марков Ю. Г.		
<b>Сеницын И. Н.</b> см. Пугачёв В. С.		
<b>Сеницын И. Н.</b> см. Соколов И. А.		
<b>Соколов И. А., Королев В. Ю.</b> Предисловие .....	3	2
<b>Соколов И. А., Сеницын И. Н.</b> Предисловие .....	2	2
<b>Соколов И. А., Шоргин С. Я.</b> К 60-летию члена редколлегии журнала «Информатика и её применения», заместителя директора ИПИ РАН по научной работе, доктора технических наук А. А. Зацаринного .....	3	86
<b>Соколов И. А.</b> см. Печинкин А. В.		
<b>Ульянов В. В.</b> см. Кристоф Г.		

	<b>Выпуск</b>	<b>Стр.</b>
<b>Ушаков В. Г., Ушаков Н. Г.</b> Декомпозиция при частично известном распределении ошибки .....	4	36
<b>Ушаков В. Г., Ушаков Н. Г.</b> Об одной ядерной оценке плотности .....	3	67
<b>Ушаков В. Г., Шестаков О. В.</b> Реконструкция распределений случайных функций в задачах однофотонной эмиссионной томографии при помощи аппроксимации экспоненциального множителя тригонометрическими многочленами .....	3	17
<b>Ушаков Н. Г.</b> см. Ушаков В. Г.		
<b>Ушаков Н. Г.</b> см. Ушаков В. Г.		
<b>Фазекаш И.</b> см. Чупрунов А. Н.		
<b>Файзулхаков Я. Р.</b> см. Преображенский Н. Б.		
<b>Целищев М. А.</b> см. Яковенко Д. О.		
<b>Чеплюкова И. А.</b> см. Лери М. М.		
<b>Чиркунов К. С.</b> Агентное моделирование развития территориальной системы .....	1	58
<b>Чубич В. М.</b> Информационная технология активной параметрической идентификации стохастических квазилинейных дискретных систем .....	1	46
<b>Чупрунов А. Н., Фазекаш И.</b> Усиленные законы больших чисел для числа безошибочных блоков при помехоустойчивом кодировании .....	3	80
<b>Шабанов А. П.</b> см. Зацаринный А. А.		
<b>Шевцова И. Г.</b> см. Королев В. Ю.		
<b>Шевцова И. Г.</b> см. Нефедова Ю. С.		
<b>Шестаков О. В.</b> О скорости сходимости распределения выборочного абсолютного медианного отклонения к нормальному закону .....	3	74
<b>Шестаков О. В.</b> см. Ушаков В. Г.		
<b>Шоргин С. Я.</b> см. Зейфман А. И.		
<b>Шоргин С. Я.</b> см. Королев В. Ю.		
<b>Шоргин С. Я.</b> см. Кудрявцев А. А.		
<b>Шоргин С. Я.</b> см. Сатин Я. А.		
<b>Шоргин С. Я.</b> см. Соколов И. А.		
<b>Шульга Л. А.</b> см. Баша Н. С.		
<b>Яковенко Д. О., Целищев М. А.</b> Диверсификация и ее связь с мерами риска .....	3	21
<b>Янушко А. В., Бабанин А. В., Кузнецова О. А., Петрушенко С. В.</b> Решение проблем взаимодействия с СУБД в кроссплатформенной библиотеке EFFIDB .....	4	53

## 2011 AUTHOR INDEX

	Issue	Page
<b>Agalarov M. Ya.</b> Analytical Model for Calculating the Performance Plan of Distributing Multi-processor System Resources in Solving the Problems of Special Class.....	4	29
<b>Babanin A. V.</b> see Yanushko A. V.		
<b>Basha N. and Shulga L.</b> Algorithm of Automatic Face Detection in Thermal Images.....	1	73
<b>Bening V. E., Gorshenin A. K., and Korolev V. Yu.</b> An Asymptotically Optimal Test for the Number of Components of a Mixture of Probability Distributions.....	3	4
<b>Bosov A. V.</b> Analysis and Optimization Problems for Some Users Activity Model. Part I. Analysis and Prediction.....	4	40
<b>Cheplyukova I. A.</b> see Leri M. M.		
<b>Chirkunov K. S.</b> Agent Modeling of Territorial System Development.....	1	58
<b>Chubich V. M.</b> Information Technology of Active Parametric Identification of Stochastic Quasi-Linear Discrete Systems.....	1	46
<b>Chuprunov A. N. and Fazekas I.</b> Strong Laws of Large Numbers for a Number of Error-Free Blocks Under Error-Corrected Coding.....	3	80
<b>Cristoph G. and Ulyanov V. V.</b> On Accuracy of Approximations for Standardized Chi-Squared Distributions by Edgeworth–Chebyshev Expansions.....	1	25
<b>Demin P. V.</b> Two Models of Resource Allocation Under the Organization of Investment Processes	1	65
<b>Durnovo A. A.</b> see Zatsman I. M.		
<b>Ermakov P. V.</b> see Kozerenko E. B.		
<b>Fayzulkhakov Ya. R.</b> see Preobrajenskiy N. B.		
<b>Fazekas I.</b> see Chuprunov A. N.		
<b>Gavrilenko S. V.</b> Improvements of the Nonuniform Estimate for Convergence of Distributions of Poisson Random Sums to the Normal Distribution.....	1	12
<b>Gorshenin A. K.</b> see Bening V. E.		
<b>Ignatyev M. B.</b> Theory of Complex Systems and Cybernetic Picture of the World.....	2	58
<b>Jakovenko D. O. and Tselishchev M. A.</b> Diversification and Its Links with Risk Measures.....	3	21
<b>Koltsov P. P.</b> Comparative Study of Image Segmentation Algorithms Processing Quality on Metric Base.....	3	53
<b>Korepanov E. R.</b> Stochastic Informational Technologies Based on Pugachev Filtering.....	2	36
<b>Korolev V. Yu., Krylov V. A., and Kuz'min V. Yu.</b> Stability of Finite Mixtures of Generalized Gamma-Distributions with Respect to Disturbance of Parameters.....	1	31
<b>Korolev V. Yu., Shevtsova I. G., and Shorgin S. Ya.</b> On the Berry–Esseen Type Inequalities for Poisson Random Sums.....	3	64
<b>Korolev V. Yu.</b> see Bening V. E.		
<b>Korotysheva A. V.</b> see Satin Ya. A.		
<b>Korotysheva A. V.</b> see Zeifman A. I.		
<b>Kozerenko E. B. and Ermakov P. V.</b> Strategies of Syntactic Analysis Based on Head-Driven Grammars and Methods of Their Implementation in Information Systems.....	4	107
<b>Kozmidiady V.</b> Work Tree and Massively Parallel Processing.....	2	90
<b>Krylov A. S.</b> see Pavelyeva E. A.		
<b>Krylov V. A.</b> see Korolev V. Yu.		
<b>Kudriavtsev A. A. and Shorgin S. Ya.</b> On a Refinement of Certain Results for a Bayesian Queuing Model.....	1	78

	Issue	Page
<b>Kudryavtseva I. A. and Panteleyev A. V.</b> Application of the Statistical Method and Finite-Difference Method for Strongly Ionized Collisional Plasma Diagnostics Problem Solution by the Flat Probe .....	3	46
<b>Kuz'min V. Yu.</b> see Korolev V. Yu.		
<b>Kuznetsov L. A.</b> Probabilistic Statistical Evaluation of the Information Objects Adequacy .....	4	64
<b>Kuznetsova O. A.</b> see Yanushko A. V.		
<b>Lebedev A. V.</b> Activity Maxima in Free-Scale Random Networks with Heavy Tails .....	4	25
<b>Leri M. M. and Cheplyukova I. A.</b> On a Statistical Problem for Random Internet-Type Graphs .....	3	34
<b>Markov Yu. G., Perepelkin V. V., Sinitsyn I. N., and Semendyaev N. N.</b> Informational Models for the Earth Irregular Rotation .....	2	17
<b>Morozova Yu. I.</b> Transformational Models of Language Structures for Machine Translation from French into Russian .....	4	99
<b>Nefedova Yu. S. and Shevtsova I. G.</b> On the Accuracy of the Normal Approximation to Distributions of Poisson Random Sums .....	1	39
<b>Panfilova T. L.</b> see Zeifman A. I.		
<b>Panteleev A. V. and Rybakov K. A.</b> Synthesis of Optimal Nonlinear Stochastic Control Systems by the Spectral Method .....	2	69
<b>Panteleyev A. V.</b> see Kudryavtseva I. A.		
<b>Pavelyeva E. A. and Krylov A. S.</b> Iris Images Comparison Algorithm Based on Iris Key Points ...	1	68
<b>Pechinkin A. V.</b> Two-Priority Queueing System with Reserving Several Channels and with Markovian Input Flow .....	1	2
<b>Pechinkin A. V. and Sokolov I. A.</b> Discrete Time Queueing System with Unreliable Server .....	4	6
<b>Perepelkin V. V.</b> see Markov Yu. G.		
<b>Petrushenko S. V.</b> see Yanushko A. V.		
<b>Preobrajenskiy N. B. and Fayzulkhakov Ya. R.</b> Problem of Compensation of Rayleigh Fading in Radio Channels of Voice Mobile Communication Systems .....	2	82
<b>Pugachev V. S. and Sinitsyn I. N.</b> Theory of Complex Stochastic Systems .....	2	4
<b>Razumchik R. V.</b> Queueing System with Negative Customers, Bunker for Ousted Customers, and Different Service Rates .....	3	41
<b>Rybakov K. A.</b> see Panteleev A. V.		
<b>Satin Ya. A., Zeifman A. I., Korotysheva A. V., and Shorgin S. Ya.</b> On a Class of Markovian Queues .....	4	18
<b>Semendyaev N. N.</b> see Markov Yu. G.		
<b>Shabanov A. P.</b> see Zatsarinny A. A.		
<b>Shestakov O. V.</b> On the Rate of Convergence of Sample Median Absolute Deviation Distribution to the Normal Law .....	3	74
<b>Shestakov O. V.</b> see Ushakov V. G.		
<b>Shevtsova I. G.</b> see Korolev V. Yu.		
<b>Shevtsova I. G.</b> see Nefedova Yu. S.		
<b>Shorgin S. Ya.</b> see Korolev V. Yu.		
<b>Shorgin S. Ya.</b> see Kudryavtsev A. A.		
<b>Shorgin S. Ya.</b> see Satin Ya. A.		
<b>Shorgin S. Ya.</b> see Zeifman A. I.		
<b>Shulga L.</b> see Basha N.		
<b>Sinitsyn I. N.</b> Stochastic Informational Technologies for Nonlinear Circular Stochastic Systems Analysis .....	4	2
<b>Sinitsyn I. N.</b> see Markov Yu. G.		
<b>Sinitsyn I. N.</b> see Pugachev V. S.		
<b>Sokolov I. A.</b> see Pechinkin A. V.		
<b>Tselishchev M. A.</b> see Jakovenko D. O.		
<b>Ulyanov V. V.</b> see Cristoph G.		
<b>Ushakov N. G.</b> see Ushakov V. G.		
<b>Ushakov N. G.</b> see Ushakov V. G.		

	Issue	Page
<b>Ushakov V. G. and Shestakov O. V.</b> Reconstruction of Random Function Distributions in Single Photon Emission Tomography Problems Using Trigonometric Polynomial Approximation of Exponential Multiplier .....	3	17
<b>Ushakov V. G. and Ushakov N. G.</b> Deconvolution Under Partially Known Error Distribution ...	4	36
<b>Ushakov V. G. and Ushakov N. G.</b> On One Kernel Density Estimator .....	3	67
<b>Yanushko A. V., Babanin A. V., Kuznetsova O. A., and Petrushenko S. V.</b> Solution of DBMS Interaction Problem in the Cross-Platform Library EFFIDB .....	4	53
<b>Zatsarinny A. A. and Shabanov A. P.</b> Information and Telecommunication Projects Management: “Timeliness—Performance—Information” .....	4	76
<b>Zatsman I. M. and Durnovo A. A.</b> Modeling of Processes for Creation of Expert Knowledge for Monitoring of Goal-Oriented Programme Activities .....	4	84
<b>Zeifman A. I., Korotysheva A. V., Panfilova T. L., and Shorgin S. Ya.</b> Stability Bounds for Some Queueing Systems with Catastrophes .....	3	27
<b>Zeifman A. I.</b> see Satin Ya. A.		

# Правила подготовки рукописей статей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений. Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи, в виде исключения, могут печататься на английском языке. Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях. Публикация не должна нарушать закон об авторских правах. Направляя свою рукопись в редакцию, авторы автоматически передают учредителям и редколлегии неисключительные права на издание данной статьи на русском языке и на ее распространение в России и за рубежом. При этом за авторами сохраняются все права как собственников данной рукописи. В связи с этим авторами должно быть представлено в редакцию письмо в следующей форме: Соглашение о передаче права на публикацию:

*«Мы, нижеподписавшиеся, авторы рукописи « \_\_\_\_\_ », передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций. Подписи авторов: (ф. и. о., дата, адрес)».*

Указанное соглашение может быть представлено как в бумажном виде, так и в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности опубликования представленной статьи в открытой печати.

2. Статья подписывается всеми авторами. На отдельном листе представляются данные автора (или всех авторов): фамилия, полное имя и отчество, телефон, факс, e-mail, почтовый адрес. Если работа выполнена несколькими авторами, указывается фамилия одного из них, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет самостоятельную экспертизу присланных статей. Возвращение рукописи на доработку не означает, что статья уже принята к печати. Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редакционной коллегии о принятии статьи к печати или ее отклонении сообщается авторам. Редколлегия не обязуется направлять рецензию авторам отклоненной статьи.

5. Корректурa статей высылается авторам для просмотра. Редакция просит авторов присылать свои замечания в кратчайшие сроки.

6. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки. Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3. Основной текст: стиль — «Обычный»: шрифт Times New Roman, размер 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине. Рекомендуемый объем рукописи — не свыше 25 страниц указанного формата. Ознакомиться с шаблонами, содержащими примеры оформления, можно по адресу в Интернете: <http://www.ipiran.ru/journal/template.doc>.

7. К рукописи, предоставляемой в 2-х экземплярах, обязательно прилагается электронная версия статьи (как правило, в форматах MS WORD (.doc) или L<sup>A</sup>T<sub>E</sub>X (.tex), а также — дополнительно — в формате .pdf) на дискете, лазерном диске или по электронной почте. Сокращения слов, кроме стандартных, не применяются. Все страницы рукописи должны быть пронумерованы.

8. Статья должна содержать следующую информацию на русском и английском языках: название, Ф.И.О. авторов, места работы авторов и их электронные адреса, подробные сведения об авторах, оформленные в соответствии с форматом, определяемым файлами [http://www.ipiran.ru/journal/issues/2011\\_05\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2011_05_01/authors.asp) и [http://www.ipiran.ru/journal/issues/2011\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2011_01_eng/authors.asp), аннотация (не более 100 слов), ключевые слова. Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в порядке их первого упоминания. В списке литературы не должно быть позиций, на которые нет ссылки в тексте статьи. Все фамилии авторов, заглавия статей, названия книг, конференций и т. п. даются на языке оригинала, если этот язык использует кириллический или латинский алфавит.

9. Присланные в редакцию материалы авторам не возвращаются.

10. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в случае больших объемов информации возможно использование общеизвестных архиваторов (ZIP, RAR);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

11. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию с авторов не взимается, гонорар авторам не выплачивается.

**Адрес редакции:** Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05 E-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru)