СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 22 № 1 Год 2012

СОДЕРЖАНИЕ

Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем
И. Н. Синицын, Э. Р. Корепанов, В. В. Белоусов, Т. Д. Конашенкова **3**

Персонифицированное преобразование представлений цветных изображений на мониторе ПЭВМ
О. П. Архипов, З. П. Зыкова **22**

Система характеризации самосинхронных элементов
Ю. Г. Дьяченко, Н. В. Морозов, Д. Ю. Степченков, Ю. А. Степченков **38**

Фиксация исключительных ситуаций в рекуррентном операционном устройстве
Р. А. Зеленов, А. А. Прокофьев, Ю. А. Степченков, В. Н. Волчек **49**

Иерархический метод анализа самосинхронных электронных схем
Л. П. Плеханов **62**

Повышение отказоустойчивости данных в кэш-памяти путем их обновления
Б. З. Шмейлин **74**

Моделирование лексической семантики в задачах компьютерной лингвистики
О. С. Кожунова **86**

Program-oriented indicators: Production and application in science
I. Zatsman and A. Durnovo **110**

Предметные словари: назначение, особенности и перспективы
Н. В. Сомин, И. П. Кузнецов, М. М. Шарнин, В. Г. Николаев **121**

Оценки скорости сходимости распределений случайных сумм к несимметричному распределению Стьюдента
В. Е. Бенинг, Л. М. Закс, В. Ю. Королев **132**

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 22 № 1 Год 2012

СОДЕРЖАНИЕ

О точности приближения распределения оценки риска пороговой обработки вейвлет-коэффициентов сигнала нормальным законом при неизвестном уровне шума
О. В. Шестаков **142**

Использование вейвлет-анализа в климатических исследованиях
М. Ш. Хазиахметов, Т. В. Захарова **153**

Устойчивость масштабных смесей нормальных законов относительно изменений смешивающего распределения
А. К. Горшенин **167**

О неравномерных оценках скорости сходимости в центральной предельной теореме
М. Е. Григорьева, С. В. Попов **180**

Abstracts **205**

Об авторах **211**

About Authors **213**

РАЗВИТИЕ МАТЕМАТИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА НЕЛИНЕЙНЫХ МНОГОКАНАЛЬНЫХ КРУГОВЫХ СТОХАСТИЧЕСКИХ СИСТЕМ*

И. Н. Синицын[1], *Э. Р. Корепанов*[2], *В. В. Белоусов*[3], *Т. Д. Конашенкова*[4]

Аннотация: Статья посвящена развитию математического обеспечения (МО) для вероятностного анализа нелинейных многоканальных круговых стохастических информационно-измерительных систем. В основу МО положены методы параметризации круговых распределений. Особое внимание уделено МО на основе методов нормальной аппроксимации и моментов. Описаны модули инструментального программного обеспечения «CStS-ANALYSIS». Приведены тестовые примеры.

Ключевые слова: аналитическое моделирование; коэффициенты кругового ортогонального разложения; круговая случайная величина; круговой стохастических процесс; «намотанная» нормальная плотность; нелинейная многоканальная стохастическая система; одно- и многомерные плотности распределения; ортогональное разложение плотности; эталонная плотность; MATLAB; «CStS-ANALYSIS»

1 Введение

Математическое обеспечение для спектрально-корреляционного анализа процессов в нелинейных круговых стохастических системах (КСтС), основанное на методе круговой статистической линеаризации посредством «намотанного» (wrapped) нормального распределения, рассмотрено в [1–4].

Для анализа одно- и многомерных распределений в нелинейных КСтС невысокой размерности прямые методы решения интегродифференциальных уравнений Пугачёва для характеристических функций приводят к алгоритмам, требующим применения только средств суперкомпьютерной техники [5].

Как известно [6–8], для анализа стохастических процессов в многомерных нелинейных СтС в евклидовом пространстве широкое применение получили

*Работа выполнена при финансовой поддержке РФФИ (проект № 10-07-00021).

[1]Институт проблем информатики Российской академии наук, sinitsin@dol.ru

[2]Институт проблем информатики Российской академии наук, ekorepanov@ipiran.ru

[3]Институт проблем информатики Российской академии наук, vbelousov@ipiran.ru

[4]Институт проблем информатики Российской академии наук, tkonashenkova@ipiran.ru

методы параметризации распределений (нормальной аппроксимации, начальных и центральных моментов, квазимоментов, а также ортогональных разложений и их модификаций). В настоящее время создано инструментальное программное обеспечение в среде MATLAB [3, 7–10].

Применительно к нелинейным гауссовым КСтС такое МО развито в [11]. Рассмотрим развитие МО для офлайн-анализа процессов в многоканальных КСтС [2, 11] на случай многомерных распределений.

2 Ортогональное разложение плотности круговой случайной величины

В задачах стохастической информатики широкое распространение получили модели распределений круговых случайных величин (КСВ), основанные на следующих основных подходах:

- «наматывание» (wrapping) распределений линейных СВ X на круг единичного радиуса $\Theta = X(\mathrm{mod} 2\pi)$;
- преобразования к полярным координатам двумерного линейного распределения (так называемые offset distributions) или использование стереографической проекции;
- характеризации (параметризации) кругового распределения круговыми моментами, семиинвариантами и другими характеристиками, имеющими важное предметное значение, на основе ортогонального разложения плотности по некоторой биортонормальной системе функций с весом, задаваемым некоторой эталонной плотностью распределения.

Рассмотрим ортогональное разложение плотности $f_\theta(\theta)$ для КСВ Θ, обладающей конечными круговыми моментами, по некоторой известной биортогональной системе функций $\{p_\nu(\theta), q_\nu(\theta)\}$ с весом $w_\theta(\theta)$ таким, что

$$\int_{-\infty}^{\infty} w_\theta(\theta) p_\nu(\theta) q_\nu(\theta)\, d\theta = \delta_{\nu\mu} = \begin{cases} 0 & \text{при } \nu \neq \mu\,; \\ 1 & \text{при } \nu = \mu\,, \end{cases}$$

следующего вида [6–8]:

$$f_\theta(\theta) = w_\theta(\theta) \sum_\nu c_\nu p_\nu(\theta)\,, \tag{1}$$

где

$$c_\nu = \int_{-\infty}^{\infty} f_\theta(\theta) q_\nu(\theta) d\theta = \left[q_\nu\left(\frac{\partial}{i\partial\lambda}\right) g(\lambda) \right]_{\lambda=0} \quad \left(i = \sqrt{-1}\right)\,. \tag{2}$$

Здесь величины c_ν называются коэффициентами кругового ортогонального разложения (КОР); $g_\theta(\theta)$ — характеристическая функция, соответствующая плотности $f_\theta(\theta)$; $w_\theta(\theta)$ — плотность эталонного распределения.

Если в (1) потребовать совпадения круговых моментов первого и второго порядка плотностей $f_\theta(\theta)$ и $w_\theta(\theta)$, то КОР примет следующий вид:

$$f_\theta(\theta) = w_\theta(\theta)\left[1 + \sum_{\nu=3}^{\infty} c_\nu p_\nu(\theta)\right]. \tag{3}$$

Функции $p_\nu(\theta)$ и $q_\nu(\theta)$ необязательно должны быть полиномами. Они могут быть любыми функциями, удовлетворяющими условию биортонормальности и условиям существования интегралов (2). Все сказанное о разложении (3) справедливо и в более общем случае. Однако если функции $q_\nu(\theta)$ не являются полиномами, то, несмотря на совпадение моментов первого и второго порядка распределений $f_\theta(\theta)$ и $w_\theta(\theta)$, круговые коэффициенты c_ν не будут равны нулю при $\nu = 1, 2$, вследствие чего суммирование по ν будет начинаться с 1.

Иногда применяют разложение по производным некоторой плотности $w_\theta(\theta)$, имеющей производные и моменты всех порядков [6–8]:

$$f_\theta(\theta) = \sum_{\nu=0}^{\infty} c_\nu w_\theta^{(\nu)}(\theta).$$

В этом случае $p_\nu(\theta) = w_\theta^{(\nu)}(\theta)/w_\theta(\theta)$, а функции $q_\nu(\theta)$ представляют собой полиномы.

В общем случае в зависимости от того, какие величины приняты за параметры конечного отрезка КОР, различают моментные КОР, семиинвариантные КОР, квазимоментные КОР и др. Поэтому, обозначая эти параметры через u, будем записывать КОР (3) для N членов в виде:

$$f_\theta(\theta; u) = w_\theta(\theta; u)\left[1 + \sum_{\nu=3}^{N} c_\nu p_\nu(\theta)\right]; \quad c_\nu = \left[q_\nu\left(\frac{\partial}{i\partial\lambda}\right) g(\lambda)\right]_{\lambda=0},$$

где $c_\nu = c_\nu(u)$; $p_\nu(\theta) = p_\nu(\theta, u)$; $q_\nu = q_\nu(\theta, u)$.

3 Ортогональные разложения одно- и многомерных плотностей круговых стохастических процессов

Для действительного кругового стохастического процесса (КСтП) одномерная плотность для момента времени t в силу (2) и (3) определяется следующим КОР:

$$f_1(\theta_t; t, u) = w_1(\theta_t, t, u)\left[1 + \sum_{\nu=3}^{\infty} c_{\nu t} p_\nu(\theta_t, u)\right], \tag{4}$$

где

$$c_{\nu t} = c_{\nu t}(t,u) = \int\limits_{-\infty}^{\infty} f_1(\theta_t;t,u)q_\nu(\theta_t) = \left[q_\nu\left(\frac{\partial}{i\partial\lambda}\right)g_\theta(\lambda;t,u)\right].$$

Для действительных КСтП n-мерные плотности для моментов времени $t_1, \ldots, t_n$ определяются как совокупность согласованных КОР плотностей КСВ $\Theta_{t_1}, \ldots, \Theta_{t_n}$:

$$f_n(\theta_{t_1}, \ldots, \theta_{t_n};t_1, \ldots, t_n,u) = w_n(\theta_{t_1}, \ldots, \theta_{t_n};t_1, \ldots, t_n,u)\,[1 + \\ + \sum_{k=3}^{\infty}\sum_{\nu_1+\cdots+\nu_n=k} c_{\nu_1,\ldots,\nu_n}(t_1, \ldots, t_n,u)p_{\nu_1,\ldots,\nu_n}(\theta_{t_1}, \ldots, \theta_{t_n};t_1, \ldots, t_n,u)\Bigg], \quad (5)$$

где

$$c_{\nu_1,\ldots,\nu_n}(t_1, \ldots, t_n,u) = \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} f_n(\theta_{t_1}, \ldots, \theta_{t_n};t_1, \ldots, t_n,u) \times \\ \times\, q_{\nu_1,\ldots,\nu_n}(\theta_{t_1}, \ldots, \theta_{t_n};t_1, \ldots, t_n,u)\,d\theta_{t_1}, \ldots, d\theta_{t_n} = \\ = \Bigg[q_{\nu_1,\ldots,\nu_n}\left(\frac{\partial}{i\partial\lambda_1}, \ldots, \frac{\partial}{i\partial\lambda_n};t_1, \ldots, t_n,u\right) \times \\ \times\, g_n(\lambda_1, \ldots, \lambda_n;t_1, \ldots, t_n,u)\Bigg]_{\lambda_1=\cdots=\lambda_n=0}.$$

Здесь $\{p_{\nu_1,\ldots,\nu_n}(\theta_{t_1}, \ldots, \theta_{t_n},u),\ q_{\nu_1,\ldots,\nu_n}(\theta_{t_1}, \ldots, \theta_{t_n},u)\}$ — согласованные биортонормальные системы полиномов, соответствующие согласованным многомерным плотностям $w_n(\theta_{t_1}, \ldots, \theta_{t_n},u)$, имеющим те же круговые моменты первого и второго порядка, что и КСтП $\Theta(t) = \Theta_t$.

Ограничиваясь в (4) и (5) полиномами не выше N-й степени, получим согласованное приближенное представление всех многомерных плотностей КСтП Θ_t. Этим приближенным представлением можно практически пользоваться, если КСтП имеет конечные круговые моменты до N-го порядка включительно независимо от того, существуют или не существуют его моменты высших порядков.

При рассмотрении круговых марковских СтП соответствующие КОР используют для двух плотностей (одномерной и переходной).

В рамках спектрально-корреляционной теории принимают $n = 1, 2$ и ограничиваются рассмотрением круговых дисперсий и ковариационных функций [2].

4 Многоканальные нелинейные круговые стохастические системы

Введем векторный КСтП $Y(t) = Y_t$, составленный из КСтП $\Theta_j(t) = \Theta_{jt}$ $(j = 1, \ldots, r)$, $Y_t = [\Theta_{1t} \cdots \Theta_{rt}]^{\mathrm{T}}$, где $r_y = r$ — число каналов в КСтС. Пусть векторное нелинейное стохастическое дифференциальное уравнение (понимаемое в смысле Ито), описывающее эволюцию Y_t, имеет вид [6–8]:

$$\dot{Y}_t = \varphi(Y_t, t) + \psi(Y_t, t)V_t\,; \quad Y(t_0) = Y_0\,, \tag{6}$$

где $\varphi(Y_t, t)$ и $\psi(Y_t, t)$ — $(r_y \times 1)$- и $(r_y \times r_V)$-мерные в общем случае известные нелинейные функции; V_t — $(r_V \times 1)$-мерный круговой белый шум в строгом смысле, представляющий собой среднюю квадратическую производную по времени от кругового процесса с независимыми приращениями $W_t, V_t = \dot{W}_t$. Начальное значение Y_0 будем считать независимым от приращений W_t.

Обозначим через $\chi(\mu; t)$ логарифмическую производную по времени от одномерной характеристической функции $h_1(\mu, t)$ кругового белого шума V_t, $\chi(\mu; t) = (\partial/\partial t)h_1(\mu; t)$. Тогда при известных условиях [6, 7] одно- и n-мерные плотности будут определяться следующей системой интегродифференциальных уравнений:

$$\frac{\partial}{\partial t_n} f_n(y_{t_1}, \ldots, y_{t_n}) = \; = \frac{1}{(2\pi)^{nr}} \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} \Big[i\lambda_n^{\mathrm{T}} \varphi(\eta_n, t_n) + \\ + \chi(\psi(\eta_n, t_n)^{\mathrm{T}} \lambda_n; t_n) \Big] \exp\left\{ i \sum_{k=1}^{n} \lambda_k^{\mathrm{T}} (\eta_k - y_{t_k}) \right\} \times \\ \times f_n(\eta_1, \ldots, \eta_n; t_1, \ldots, t_n)\, d\eta_1 \cdots d\eta_n d\lambda_1 \cdots d\lambda_n$$

$(n = 1, 2, \ldots)$ при условиях

$$f_n(y_{t_1}, \ldots, y_{t_{n-1}}; t_1, \ldots, t_{n-1}, t_{n-1}) = \\ = f_{n-1}(y_{t_1}, \ldots, y_{t_{n-1}}; t_1, \ldots, t_{n-1}) \delta(y_{t_n} - y_{t_{n-1}})\,; \\ f_1(y_t; t_0) = f_0(y_t)\,.$$

Для дискретных КСтС соответствующие интегроразностные уравнения получаются путем замены оператора дифференцирования по времени на оператор сдвига по времени [6–8].

5 Математическое обеспечение, основанное на ортогональном разложении одномерной плотности

При аппроксимации круговой одномерной плотности конечным отрезком КОР естественно за параметры u принять: m_t — вектор круговых средних; K_t —

круговую ковариационную матрицу (от которых зависят эталонная плотность, а следовательно, и полиномы $\{p_\nu(y_t), q_\nu(y_t)\}$) и коэффициенты $c_{\kappa,t}$. Поэтому из (4) для отрезка КОР имеем:

$$\tilde{f}_1(y_t, t, u) \approx w_1(y_t, t, u) \left[1 + \sum_{l=3}^{N} \sum_{|\kappa|=l} c_{\kappa t} p_\kappa(y_t, u) \right] . \tag{7}$$

При этом, согласно [6, 7], m_t, K_t и $c_{\nu t}$ будут определяться следующей системой обыкновенных дифференциальных уравнений:

$$\dot{m}_t = A_0(m_t, K_t, t) + \sum_{l=3}^{N} \sum_{|\nu|=l} A_{1\nu}(m_t, K_t, t) c_{\nu t}, \;\; m_0 = m(t_0)\,; \tag{8}$$

$$\dot{K}_t = B_0(m_t, K_t, t) + \sum_{l=3}^{N} \sum_{|\nu|=l} B_{1\nu}(m_t, K_t, t) c_{\nu t}, \;\; K_0 = K(t_0)\,; \tag{9}$$

$$\begin{aligned} \dot{c}_{\kappa t} = {} & C_{\kappa 0}(m_t, K_t, t) + A_0(m_t, K_t, t)^{\mathrm{T}} q_\kappa^m(\alpha) + \operatorname{tr}\left[B_0(m_t, K_t, t) q_\kappa^K(\alpha) \right] + \\ & + \sum_{l=3}^{N} \sum_{|\nu|=l} \Big\{ C_{\kappa\nu}(m_t, K_t, t) + A_{1\nu}(m_t, K_t, t)^{\mathrm{T}} q_\kappa^m(\alpha) + \\ & + \operatorname{tr}\left[B_{1\nu}(m_t, K_t, t) q_\kappa^K(\alpha) \right] \Big\} c_{\nu t}\,, \;\; c_{\kappa 0} = c_\kappa(t_0)\,. \end{aligned} \tag{10}$$

Здесь введены следующие обозначения:

$$A_0(m_t, K_t, t) = \int_{-\infty}^{\infty} \varphi(y_t, t) w_1(y_t, u)\, dy_t\,;$$

$$A_{1\nu}(m_t, K_t, t) = \int_{-\infty}^{\infty} \varphi(y_t, t) p_\nu(y_t, u)\, dy_t\,;$$

$$\begin{aligned} B_0(m_t, K_t, t) = {} & B_{01}(m_t, K_t, t) + B_{01}(m_t, K_t, t)^{\mathrm{T}} + B_{02}(m_t, K_t, t) = \\ = {} & \int_{-\infty}^{\infty} \Big[\varphi(y_t, t)(y_t - m_t)^{\mathrm{T}} + (y_t - m_t)\varphi(y_t, t)^{\mathrm{T}} + \\ & + \psi(y_t, t) G_t \psi_(y_t, t)^{\mathrm{T}} \Big] w_1(y_t, u)\, dy_t\,; \end{aligned}$$

$$B_{1\nu}(m_t, K_t, t) = \int\limits_{-\infty}^{\infty} \Big[\varphi(y_t, t)(y_t - m_t)^{\mathrm{T}} + \\ + (y_t - m_t)\varphi(y_t, t)^{\mathrm{T}}\psi(y_t, t)G_t\psi(y_t, t)^{\mathrm{T}}\Big]\, w_1(y_t, u)\, dy_t\,; \quad (11)$$

$$C_{\kappa 0}(m_t, K_t, t) = \int\limits_{-\infty}^{\infty} \left\{ q_\kappa\left(\frac{\partial}{i\partial\lambda}\right) \Big[i\lambda^{\mathrm{T}}\varphi(y_t, t) + \right. \\ \left. + \chi(\psi(y_t, t)^{\mathrm{T}}\lambda; t)\Big] \exp\Big[i\lambda^{\mathrm{T}} y_t\Big] \right\}_{\lambda=0} w_1(y_t, u)\, dy_t\,;$$

$$C_{\kappa\nu}(m_t, K_t, t) = \int\limits_{-\infty}^{\infty} \left\{ q_\kappa\left(\frac{\partial}{i\partial\lambda}\right) \Big[i\lambda^{\mathrm{T}}\varphi(y_t, t) + \right. \\ \left. + \chi(\psi(y_t, t)^{\mathrm{T}}\lambda; t)\Big] \exp\Big[i\lambda^{\mathrm{T}} y_t\Big]\, p_\nu(y_t, u) \right\}_{\lambda=0} w_1(y_t, u)\, dy_t\,,$$

где $q_\kappa^m(y_t)$ — матрица-столбец производных полинома $q_\kappa(y_t)$ по компонентам вектора m_t; $q_\kappa^K(y_t)$ — квадратная матрица производных полинома $q_\kappa(y_t)$ по элементам матрицы K_t; $G_t = [G_{lj}(t)]$ — матрица интенсивностей векторного кругового белого шума V, причем

$$G_{lj}(t) = \left[\frac{\partial^2\chi(\mu; t)}{\partial(i\mu_l)\partial(i\mu_j)}\right]_{\mu=0}\,;$$

$q_\kappa^m(\alpha)$ и $q_\kappa^K(\alpha)$ — результат замены одночленов вида $y_{1t}^{\rho_1}\ldots y_{rt}^{\rho_r}$ соответствующими начальными моментами $\alpha_{\rho_1,\ldots,\rho_r}$, которые, как известно, зависят от $c_{\kappa t}$.

Уравнения (10) нелинейны относительно $c_{\kappa t}$. Если отказаться от требования совпадения первых моментов и задать эти моменты для эталонной плотности априори, то для $c_{\kappa t}$ получатся линейные уравнения. Эти уравнения проще, чем (10), однако при этом придется взять большее N в (7).

Таким образом, при использовании полиномиальной биортонормальной системы $\{p_\nu(y_t, u), q_\mu(y_t, u)\}$ в основе МО кругового ортогонального (7) разложения одномерной плотности КСтП Y_t в КСтС (6) лежат уравнения (8)–(10) при условиях (11).

В состав МО «CStS-ANALYSIS» входят:

- уравнения для различных типов систем;
- типовые системы биортонормальных систем функций для одномерных плотностей;

– ортогональные разложения плотностей одномерных распределений;
– обыкновенные дифференциальные (разностные) уравнения для параметров одномерных распределений с соответствующими начальными условиями;
– выражения для вычисления типовых функционалов, определяющих задачи вероятностного кругового анализа многоканальной КСтС;
– набор тестовых задач.

6 Математическое обеспечение, основанное на ортогональных разложениях n-мерных плотностей

Для КСтС (6) n-мерное ($n \geq 2$) распределение КСтП можно описать в виде следующих конечных отрезков согласованных КОР многомерных плотностей:

$$\tilde{f}_n = \tilde{f}_n(y_{t_1}, \ldots, y_{t_n}, t_1, \ldots, t_n, u) \approx w_n(y_{t_1}, \ldots, y_{t_n}, t_1, \ldots, t_n, u)\Bigg[1 + \\ + \sum_{l=3}^{N} \sum_{|\kappa_1|+\cdots+|\kappa_n|=l} c_{\kappa_1,\ldots,\kappa_n,t} p_{\kappa_1,\ldots,\kappa_n}(y_{t_1}, \ldots, y_{t_n})\Bigg] \ (n \geq 2)\,. \quad (12)$$

Здесь $w_n(y_{t_1}, \ldots, y_{t_n}, t_1, \ldots, t_n, u)$ — согласованная последовательность плотностей, имеющих те же круговые моменты первого и второго порядка, что и соответствующие плотности $\tilde{f}_n$, вследствие чего каждая из плотностей зависит, как от параметров, от значений математического ожидания m_t и ковариационной функции $K(t, t')$ при $t, t' = t_1, \ldots, t_n$.

Уравнение для ковариационной функции $K(t_1, t_2)$, если заменить двумерную плотность $\tilde{f}_2(y_{t_1}, y_{t_2}; t_1, t_2, u)$ аппроксимирующим ее отрезком КОР (12), имеет вид:

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = B_{20}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2), t_1, t_2) + \\ + \sum_{l=3}^{N} \sum_{|\kappa|+|\kappa_2|=l} B_{2\kappa_1,\kappa_2}(m_t, K_t, t_1, t_2) c_{\kappa_1\kappa_2}(t_1, t_2), \quad K(t_1, t_1) = K_{t_1}\,. \quad (13)$$

Здесь введены следующие обозначения:

$$B_{20} = B_{20}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2), t_1, t_2) = \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y_{t_1} - m_{t_1}) \varphi(y_{t_2}, t_2)^{\mathrm{T}} w_2(y_{t_1}, y_{t_2})\, dy_{t_1} dy_{t_2}\,;$$

$$B_{2\kappa_1\kappa_2} = B_{2\kappa_1,\kappa_2}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1,t_2), t_1, t_2) = \\ = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (y_{t_1} - m_{t_1})\varphi(y_{t_2}, t_2)^{\mathrm{T}} p_{\kappa_1\kappa_2}(y_{t_1}, y_{t_2}) w_2(y_{t_1}, y_{y_2})\, dy_{t_1} dy_{t_2}\,.$$

Обобщая [6, 7], приведем уравнения круговых коэффициентов согласованных КОР в (12) и соответствующие начальные условия к виду

$$\frac{\partial c_{\kappa_1,\ldots,\kappa_n}(t_1,\ \ldots\ ,\ t_n)}{\partial t_n} = \\ = C_{\kappa_1,\ldots,\kappa_n,0}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1,t_2), t_1,\ \ldots\ ,\ t_n) + \\ + \left[A_{10}(m_{t_n}, K_{t_n}, t_n)^{\mathrm{T}} + \sum_{l=3}^{N}\sum_{|\nu|=l} c_\nu(t_n) A_{1\nu}(m_{t_n}, K_{t_n}, t_n)^{\mathrm{T}}\right] q^m_{\kappa_1,\ldots,\kappa_n}(\alpha) + \\ + \mathrm{tr}\left[\left\{B_{20}(m_{t_n}, K_{t_n}, t_n) + \sum_{l=3}^{N}\sum_{|\nu|=l} c_{\nu(t_n)} B_{2\nu}(m_{t_n}, K_{t_n}, t_n)\right\} + q^{K_n}_{\kappa_1,\ldots,\kappa_n}(\alpha)\right] + \\ + \sum_{n=1}^{n-1} \mathrm{tr}\left[\left\{B_{20}(m_{t_n}, K_{t_n}, t_n)^{\mathrm{T}} + \right.\right. \\ \left.\left. + \sum_{l=3}^{N}\sum_{|\nu_1|+|\nu_2|=l} c_{\nu_1\nu_2}(t_n,t_n) B_{2\nu_1\nu_2}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1,t_2), t_1, t_2)\right\} \times \right. \\ \left. \times\, q^{K_n}_{\kappa_1,\ldots,\kappa_n}(\alpha)\right] + \sum_{l=3}^{N}\sum_{|\nu_1|+\cdots+|\nu_n|=l} c_{\nu_1,\ldots,\nu_2}(t_1,\ \ldots\ ,\ t_n) \times \\ \times\, B_{\kappa_1,\ldots,\kappa_2,\nu_1,\ldots,\nu_n}(m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1,t_2), t_1,\ \ldots\ ,\ t_n)\,; \quad (14)$$

$$c_{\kappa_1,\ldots,\kappa_n}(t_1,\ \ldots\ ,\ t_{n-1}, t_{n-1}) = c_{\kappa_1,\ldots,\kappa_{n-1}+\kappa_n}(t_1,\ \ldots\ ,\ t_{n-1})\,. \quad (15)$$

Здесь

$$B_{\kappa_1,\ldots,\kappa_n,0}(m_{t_h}, m_{t_l}, K_{t_h}, K_{t_l}, K(t_h,t_l), t_1,\ \ldots\ ,\ t_n) = \\ = \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}\left\{q_{\kappa_1,\ldots,\kappa_n}\left(y_{t_1},\ \ldots\ ,\ y_{t_{n-1}}, \frac{\partial}{i\partial\lambda_n}\right)\times\right.$$

$$\times \left[i\lambda_n^{\mathrm{T}} \varphi(y_{t_n}, t_n) + \chi(\psi(y_{t_n}, t_n)^{\mathrm{T}} \lambda_n; t_n) \right] \times$$

$$\times \exp\left[i\lambda_n^{\mathrm{T}} y_{t_n} \right] \Big\}_{\lambda_n = 0} w_n(y_{t_1}, \ldots, y_{t_n})\, dy_{t_1} \cdots dy_{t_n} \quad (h, l = 1, 2, \ldots);$$

$$B_{\kappa_1, \ldots, \kappa_n, \nu_1, \ldots, \nu_n}(m_{t_h}, m_{t_l}, K_{t_h}, K_{t_l}, K(t_h, t_l), t_1, \ldots, t_n) =$$

$$= \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} \left\{ q_{\kappa_1, \ldots, \kappa_n} \left(y_{t_1}, \ldots, y_{t_{n-1}}, \frac{\partial}{i\partial\lambda_n} \right) \times \right.$$

$$\times \left[i\lambda_n^{\mathrm{T}} \varphi(y_{t_n}, t_n) + \chi(\psi(y_{t_n}, t_n)^{\mathrm{T}} \lambda_n; t_n) \right] \times$$

$$\left. \times \exp\left[i\lambda_n^{\mathrm{T}} y_{t_n} \right] \right\}_{\lambda_n = 0} p_{\nu_1, \ldots, \nu_n}(y_{t_1}, \ldots, y_{t_n}) \times$$

$$\times\, w_n(y_{t_1}, \ldots, y_{t_n})\, dy_{t_1} \cdots dy_{t_n} \quad (h, l = 1, 2, \ldots).$$

Таким образом, для КСтС (6) согласно круговому методу ОР сначала следует проинтегрировать уравнения (8) и (10) для приближенного определения m_t, K_t и c_ν ($|\nu| = 3, \ldots, N$) отрезка КОР (7) одномерной плотности $\tilde{f}_1(y_t; t)$ как функции t. После этого, проинтегрировав уравнения (14) и (15) при $n = 2$, $t_2 > t_1$ совместно с уравнением и начальным условием (13), определим приближенно ковариационную функцию $K(t_1, t_2)$ и все оставшиеся коэффициенты $c_{\nu_1 \nu_2}$ отрезка КОР двумерной плотности $\tilde{f}_2(y_{t_1}, y_{t_2}; t_1, t_2)$. Интегрируя далее уравнения (14) и (15) для $t_1 < \cdots < t_n$ последовательно при $n = 3, \ldots, N$, приближенно определим коэффициенты $c_{\nu_1, \ldots, \nu_n}$ отрезков КОР.

В состав МО «CStS-ANALYSIS» (см. разд. 5) дополнительно входят:

- типовые системы биортонормальных функций для n-мерных плотностей ($n \geq \geq 2$);
- обыкновенные дифференциальные (разностные) уравнения и начальные условия для коэффициентов КОР n-мерных плотностей;
- набор тестовых задач.

7 Математическое обеспечение на основе метода «намотанной» нормальной аппроксимации и статистической линеаризации

Если коэффициент при круговом белом шуме в уравнении (6) зависит от состояния, основные уравнения кругового метода «намотанной» нормальной аппроксимации (МННА) принимают следующий вид [11]:

$$\dot{m}_t = A_0(m_t, K_t, t)\,, \;\; m_0 = m(t_0)\,; \tag{16}$$

$$\dot{K}_t = B_0(m_t, K_t, t)\,, \;\; K_0 = K(t_0)\,; \tag{17}$$

$$\frac{\partial K(t_1,t_2)}{\partial t_2} = \\ = K(t_1,t_2)K(t_2)^{-1}B_{01}(m_{t_2}, K_{t_2}, t_2)\,, \;\; t_1 < t_2\,, \;\; K(t_1,t_1) = K_{t_1}\,. \tag{18}$$

Здесь $A_0(m_t, K_t, t)$, $B_0(m_t, K_t, t)$ и $B_{01}(m_t, K_t, t)$ определены в (11) для «намотанного» нормального распределения.

В состав МННА включено математическое обеспечение [2] для системы (6) при $\psi(y_t, t) = I_r$. Коэффициенты круговой статистической линеаризации для типовых нелинейных функций приведены в приложении.

8 Математическое обеспечение на основе методов круговых начальных и центральных моментов

Из уравнений разд. 5 для круговых «намотанных» начальных моментов $\alpha_\rho = \alpha_\rho(t)$ порядка ρ, обобщая [6, 7], имеем следующие уравнения:

$$\dot{\alpha}_\rho = A_{0,\rho} + \sum_{k=3}^{N} \sum_{|\nu|=k} A_{\nu,\rho} c_\nu(\alpha) \;\; (c_\nu = q_\nu(\alpha))\,, \tag{19}$$

где

$$A_{0,\rho} = \int_{-\infty}^{\infty} \left\{ \frac{\partial^{|\rho|}}{\partial(i\lambda_1)^{\rho_1} \cdots \partial(i\lambda_r)^{\rho_r}} \times \right. \\ \left. \times \left[i\lambda^{\mathrm{T}} \varphi(y_t, t) + \chi(\psi(y_t, t)^{\mathrm{T}} \lambda; t) \right] \exp\left[i\lambda^{\mathrm{T}} y_t \right] \right\}_{\lambda=0} w_1(y_t, u)\, dy_t\,;$$

$$A_{\nu,\rho} = \int_{-\infty}^{\infty} \left\{ \frac{\partial^{|\rho|}}{\partial(i\lambda_1)^{\rho_1} \cdots \partial(i\lambda_r)^{\rho_r}} \times \right. \\ \left. \times \left[i\lambda^{\mathrm{T}} \varphi(y_t, t) + \chi(\psi(y_t, t)^{\mathrm{T}} \lambda; t) \right] \exp\left[i\lambda^{\mathrm{T}} y_t \right] \right\}_{\lambda=0} p_\nu(y_t, u) w_1(y_t, u)\, dy_t$$

$$(\,|\rho| = \rho_1 + \cdots + \rho_r\,, \quad \rho_1, \ldots, \rho_r = 0, 1, \ldots, N)\,.$$

Напомним, что $q_\nu(\alpha)$ представляет собой результат замены всех одночленов $y_{1t}^{\rho_1} \cdots y_{rt}^{\rho_r}$ в выражении полинома $q_\nu(\alpha)$ соответствующими моментами $\alpha_{\rho_1, \ldots, \rho_r}$.

При составлении уравнений (7) в конкретных задачах следует иметь в виду, что число N_r^ρ моментов ρ-го порядка r-мерного вектора определяется формулой:

$$N_r^\rho = C_{r+\rho-1}^\rho = \frac{(r+\rho-1)!}{\rho!(r-1)!}\,,$$

а полное число моментов, не превосходящих N, для r-мерного вектора равно

$$P_r^N = \sum_{l=1}^{N} N_r^l = \frac{(N+r)!}{N!r!} - 1\,.$$

Уравнения (19) линейны относительно α_ρ ($|\rho| = 3, \ldots, N$) и нелинейны относительно моментов первого и второго порядка, поскольку эталонная плотность и полиномы $p_\nu(y_t, u)$ и $q_\nu(y_t, u)$ зависят от моментов первого и второго порядка.

Обобщая [6, 7], представим уравнения для математических ожиданий m_h ($h = 1, \ldots, r$) и центральных моментов μ_ρ порядка ρ в виде:

$$\dot m_h = A_{0,h} + \sum_{k=3}^{\infty}\sum_{|\nu|=k} A_{\nu r} c_\nu, \quad c_\nu = q_\nu(\alpha)\,; \tag{20}$$

$$\dot\mu_\rho = B_{0,\rho} - \sum_{h=1}^{r} \rho_h B_{0,h}\mu_{\rho-e_h} + \sum_{k=3}^{N}\sum_{|\nu|=k}\left[B_{\nu,\rho} - \sum_{h=1}^{r}\rho_h B_{\nu,h}\mu_{\rho-e_h}\right] c_\nu$$
$$(\rho_1, \ldots, \rho_r = 0, 1, \ldots, N\,; \quad |\rho| = 2, \ldots, N)\,. \tag{21}$$

Здесь введены следующие обозначения:

$$e_h = \left[0\cdots 0\underset{h}{1}0\ldots 0\right]^{\mathrm T}\,;$$

$$A_{0,h} = \int\limits_{-\infty}^{\infty} \varphi_h(y_t, t) w_1(y_t, u)\, dy_t\,;$$

$$A_{\nu,h} = \int\limits_{-\infty}^{\infty} \varphi_h(y_t, t) p_\nu(y_t, u) w_1(y_t, u)\, dy_t\,;$$

$$A_{0,\rho} = \int\limits_{-\infty}^{\infty} \left\{ \frac{\partial^{|\rho|}}{\partial(i\lambda_1)^{\rho_1}\cdots\partial(i\lambda_r)^{\rho_r}} \left[i\lambda^{\mathrm T}\varphi(y_t, t) + \chi(\psi(y_t, t)^{\mathrm T}\lambda; t)\right] \times \right.$$
$$\left. \times \exp\left[i\lambda^{\mathrm T}(y_t - m_t)\right]\right\}_{\lambda=0} w_1(y_t, u)\, dy_t\,;$$

$$B_{\nu,\rho} = \int\limits_{-\infty}^{\infty} \left\{ \frac{\partial^{|\rho|}}{\partial(i\lambda_1)^{\rho_1} \cdots \partial(i\lambda_r)^{\rho_r}} \left[i\lambda^{\mathrm{T}} \varphi(y_t, t) + \chi(\psi(y_t, t)^{\mathrm{T}} \lambda; t) \right] \times \right.$$

$$\left. \times \exp\left[i\lambda^{\mathrm{T}} (y_t - m_t) \right] \right\}_{\lambda=0} p_\nu(y_t, u) w_1(y_t, u)\, dy_t \,,$$

а $q_\nu(\alpha)$ должны быть выражены через центральные моменты.

Уравнения (20) и (21) всегда нелинейны из-за наличия слагаемых вида $\mu_{\rho-e_h} q_\nu(\alpha)$.

В практических задачах для полиномиальных функций $\varphi(y_t, t)$, $\psi(y_t, t)$ и $p_\nu(y_t), q_\nu(y_t)$ непосредственно используются символьные вычисления [7–9].

9 Тестовые примеры

Пример 9.1. Разброс характеристик $L(\Theta)$ кругового ограничительного упора размаха l с зоной чувствительности d.

Для характеристики упора $L(\Theta) \approx k_0(\zeta)\mu + k_1(\zeta)\Theta^0$ имеем при $D \neq 0$:

$$\mu = M\Theta\,; \quad \Theta^0 = \Theta - \mu\,;$$

$$\sigma^2 = D\Theta\,; \quad \zeta = \frac{\mu}{\sigma}\,; \quad m_L = k_0(\zeta)\,; \quad \mu = \psi\,;$$

$$\sigma_L = k_1(\zeta)\sigma = \sqrt{-2\ln r}\,; \quad d_1 = \frac{d}{\sigma}\,;$$

$$r(l,d) = \big((\cos l + (1-\cos l)(\Phi(d_1+\zeta) + \Phi(d_1-\zeta)))^2 + \sin^2 l(\Phi(d_1+\zeta) - \Phi(d_1-\zeta))^2\big)^{1/2}\,;$$

$$\psi(l,d) = \arctan\left(\frac{\sin l(\Phi(d_1+\zeta) - \Phi(d_1-\zeta))}{\cos l + (1-\cos l)(\Phi(d_1+\zeta) + \Phi(d_1-\zeta))} \right).$$

В частности, при $d = 0$ отсюда получаем:

$$r(l,0) = \sqrt{\cos^2 l + 4\sin^2 l\, \Phi^2(\zeta)}\,; \qquad \psi(l,0) = \arctan\left(\frac{2\sin l\, \Phi(\zeta)}{\cos l} \right).$$

На рис. 1 и 2 представлены результаты аналитического и статистического моделирования при различных значениях параметра d_1.

Пример 9.2. Для одномерной нелинейной круговой системы

$$\dot{Y} + \varphi(Y) = V$$

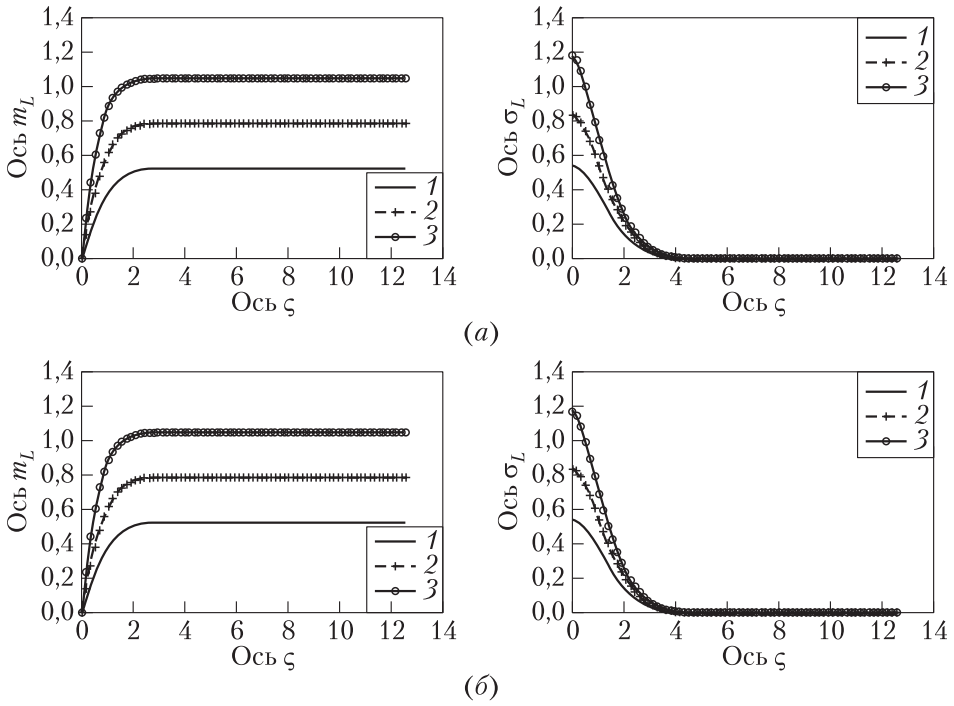


Рис. 1 Результаты аналитического и статистического моделирования при $d = d_1 = 0$ (*а*) и $d \neq 0$, $d_1 = 0{,}01$ (*б*): *1* — $l = \pi/6$; *2* — $\pi/4$; *3* — $l = \pi/3$

($\varphi(Y)$ — скалярная нелинейная функция; V — круговой белый шум интенсивности G_t) уравнения МННА (16)–(18) имеют вид:

$$\dot{m}_t = -A_0(m_t, D_t)\,, \quad \dot{D}_t = -2k_1(m_t, D_t) + G_t\,;$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = -k_1(m_t, D_t)K(t_1, t_2)\,,$$

где $A_0(m_t, D_t)$ и $k_1(m_t, D_t)$ — коэффициенты статистической линеаризации нелинейной функции $\varphi(Y)$ для «намотанного» нормального распределения с параметрами m_t и D_t.

При $N = 4$ уравнения (20) и (21) имеют вид

$$\dot{m}_t = A_0(m_t, D_t) + A_{13}(m_t, D_t)\mu_{3t} + A_{14}(m_t, D_t)\mu_{4t}\,;$$

$$\dot{D}_t = B_0(m_t, D_t) + B_{13}(m_t, D_t)\mu_{3t} + B_{14}(m_t, D_t)\mu_{4t}\,;$$

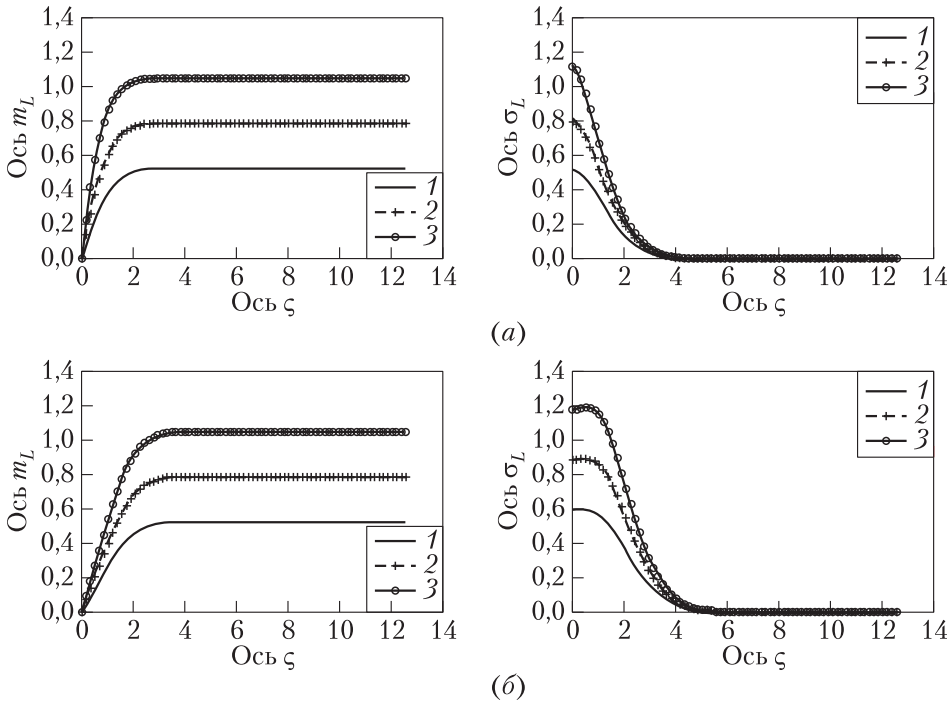


Рис. 2 Результаты аналитического и статистического моделирования при $d_1 = 0{,}1$ (*а*) и $d_1 = 1$ (*б*): *1* — $l = \pi/6$; *2* — $\pi/4$; *3* — $l = \pi/3$

$$\dot{\mu}_{3t} = C_{30}(m_t, D_t) + C_{33}(m_t, D_t)\mu_{3t} + C_{34}(m_t, D_t)\mu_{4t}\,;$$
$$\dot{\mu}_{4t} = C_{40}(m_t, D_t) + C_{41}(m_t, D_t)\mu_{3t} + C_{44}(m_t, D_t)\mu_{4t} - 4A_{13}(m_t, D_t)\mu_{3t}^2 - A_{14}(m_t, D_t)\mu_{3t}\mu_{4t}\,.$$

Для различных нелинейных функций $\varphi(y)$ полученные уравнения позволяют оценить точность МННА.

10 Заключение

Для нелинейных многоканальных круговых стохастических информационно-измерительных систем, описываемых нелинейными уравнениями, содержащими стохастические возмущения, зависящие от состояния, разработано МО для офлайн-анализа «CStS-ANALYSIS».

В основу положены методы параметризации одно- и многомерных круговых плотностей на базе «намотанного» нормального распределения [2, 11].

В настоящее время в ИПИ РАН ведутся работы по созданию МО для других эталонных плотностей, а также для многоканальных сферических стохастических информационно-измерительных систем.

Литература

1. *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 1-й междунар. конф. — СПб., 2010.
2. *Синицын И. Н.* Стохастические информационные технологии для исследования нелинейных круговых стохастических систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 2–5.
3. *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. 7–37.
4. *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 on Abstracts. — M.: IPI RAS, 2011. P. 86–87.
5. *Босов А. В., Будзко В. И., Захаров В. Н., Козмидиади 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.
6. *Пугачев В. С., Синицын И. Н.* Стохастические дифференциальные системы. Анализ и фильтрация. — 2-е изд., доп. — М.: Наука, 1990.
7. *Пугачев В. С., Синицын И. Н.* Теория стохастических систем. — 2-е изд. — М.: ЛОГОС, 2004.
8. *Синицын И. Н.* Канонические представления случайных функций и их применения в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009.
9. *Синицын И. Н.* Стохастические системы. Теория и стохастические информационные технологии (I) // История науки и техники, 2008. — М.: Научтехлитиздат. № 7. С. 9–12.
10. *Синицын И. Н.* Стохастические системы. Теория и стохастические информационные технологии (II) // Системы высокой доступности, 2008. Т. 4. № 1–2. С. 25–35.
11. *Синицын И. Н.* Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 12–18.

ПРИЛОЖЕНИЕ

Коэффициенты круговой статистической линеаризации для типовых нелинейных функций $Y = \varphi(X_N)$,

$$Y \approx U = k_0 + k_1(X - \mu)\,. \qquad (\text{П}.1)$$

Здесь $k_0(\mu,\sigma)$ и $k_1(\mu,\sigma)$ определяются формулами:

$$\mu k_0(\mu,\sigma) = \varphi(\mu,\sigma)\,; \tag{П.2}$$

$$k_1(\mu,\sigma) = \frac{\sqrt{-2\ln r(\mu,\sigma)}}{\sigma}\,, \tag{П.3}$$

где $re^{i\psi} = \mathrm{M}_{WN}\exp\{i\varphi(X_N)\}$; M_{WN} — символ математического ожидания для намотанного нормального распределения $X = X_N$.

1. $Y = l\,\mathrm{sgn}(X_N)\,;$ $\quad k_0 = \dfrac{1}{\mu}\arctan\left(\dfrac{2\sin l\,\Phi\,(\mu/\sigma)}{\cos l}\right)\,;$

$$k_1 = \frac{1}{\sigma}\sqrt{-2\ln\sqrt{\cos^2 l + 4\sin^2 l\,\Phi^2\left(\frac{\mu}{\sigma}\right)}}\,;$$

$$\Phi(\zeta) = \frac{1}{\sqrt{2\pi}}\int\limits_0^{\zeta} e^{-t^2/2}\,dt\,; \quad \zeta = \frac{\mu}{\sigma}\,.$$

Функция $\arctan(S/C)$ определена в [3]:

$$\arctan\left(\frac{S}{C}\right) = \begin{cases} \mathrm{arctg}\left(\dfrac{S}{C}\right), & \text{если } C>0\,,\ S>0\,;\\ \dfrac{\pi}{2}, & \text{если } C=0\,,\ S>0\,;\\ \mathrm{arctg}\left(\dfrac{S}{C}\right)+\pi\,, & \text{если } C<0\,;\\ \mathrm{arctg}\left(\dfrac{S}{C}\right)+2\pi\,, & \text{если } C\geq 0\,,\ S<0\,;\\ \text{не определено}\,, & \text{если } C=0\,,\ S=0. \end{cases}$$

2. $Y = l\,\mathbb{1}\,(X_N)\,;$ $\quad k_0 = \dfrac{1}{\mu}\,\arctan\left(\dfrac{\sin l\ (0{,}5+\Phi\,(\mu/\sigma))}{0{,}5-\Phi\,(\mu/\sigma)+\cos l\ (0{,}5+\Phi\,(\mu/\sigma))}\right)\,;$

$$k_1 = \frac{1}{\sigma}\sqrt{-2\ln\sqrt{\frac{1}{2}(1+\cos l)+2(1-\cos l)\Phi^2\left(\frac{\mu}{\sigma}\right)}}\,.$$

3. $Y = lX_N\,\mathbb{1}(X_N)\,;$ $\quad k_0 = \dfrac{1}{\mu}\,\arctan\left(\dfrac{r_2}{r_1}\right);$

$$k_1 = \frac{1}{\sigma}\sqrt{-2\ln\sqrt{r_1^2+r_2^2}}\,;$$

$$r_2 = \frac{1}{2}\, e^{-l^2\sigma^2/2} \left(\sin(l\mu) - \frac{2}{\sqrt{\pi}} \operatorname{IE}(l\sigma) \cos(l\mu) \right) ;$$

$$r_1 = \left(\frac{1}{2} - \Phi\left(\frac{\mu}{\sigma}\right) \right) + \frac{1}{2}\, e^{-l^2\sigma^2/2} \left(\cos(l\mu) + \frac{2}{\sqrt{\pi}} \operatorname{IE}(l\sigma) \sin(l\mu) \right) ;$$

$$\operatorname{IE}(l\sigma) = \int\limits_{-l\sigma/\sqrt{2}}^{0} e^{y^2}\, dy\,.$$

4. $Y = lX_N^2 \operatorname{sgn}(X_N)\,;$ $\qquad k_0 = \dfrac{1}{\mu} \arctan\left(\dfrac{r_2}{r_1}\right) ;$

$$k_1 = \frac{1}{\sigma} \sqrt{-2 \ln \sqrt{r_1^2 + r_2^2}}\,;$$

$$r_1 = p_1(\cos p_2(1 - \operatorname{Re}\operatorname{erf}(p_3) + \operatorname{Re}\operatorname{erf}(p_4)) + \sin p_2(\operatorname{Im}\operatorname{erf}(p_3) - \operatorname{Im}\operatorname{erf}(p_4)))\,;$$

$$r_2 = p_1(\cos p_2(\operatorname{Im}\operatorname{erf}(p_4) - \operatorname{Im}\operatorname{erf}(p_3)) + \sin p_2(3 - \operatorname{Re}\operatorname{erf}(p_3) + \operatorname{Re}\operatorname{erf}(p_4)))\,;$$

$$p_1 = \frac{\sigma}{2(1 + 4\sigma^4 l^2)^{1/4}}\, e^{-4\mu^2 l^2 \sigma^2/(1+4\sigma^4 l^2)}\,;$$

$$p_2 = \frac{l}{1 + 4\sigma^4 l^2} + \frac{1}{2} \operatorname{arctg}\left(2l\sigma^2\right) ;$$

$$p_3 = \frac{\mu}{\sqrt{2}\sigma} \cos\left(\frac{1}{2} \operatorname{arctg}\left(2l\sigma^2\right)\right) + \imath \frac{\mu}{\sqrt{2}\sigma} \sin\left(\frac{1}{2} \operatorname{arctg}\left(2l\sigma^2\right)\right) ;$$

$$p_4 = \frac{\mu}{\sqrt{2}\sigma} \cos\left(\frac{1}{2} \operatorname{arctg}\left(2l\sigma^2\right)\right) - \imath \frac{\mu}{\sqrt{2}\sigma} \sin\left(\frac{1}{2} \operatorname{arctg}\left(2l\sigma^2\right)\right) .$$

Здесь

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int\limits_0^z e^{-t^2}\, dt\,;$$

$$\operatorname{Re}\operatorname{erf}(z) = \int\limits_0^{Re\, z} e^{\left(k_l^2 - 1\right)x^2} \left(\cos\left(2k_1 x^2\right) + k_1 \sin\left(2k_1 x^2\right) \right) dx\,;$$

$$\operatorname{Im}\operatorname{erf}(z) = \int\limits_0^{Re\, z} e^{\left(k_l^2 - 1\right)x^2} \left(\cos\left(2k_1 x^2\right) - k_1 \sin\left(2k_1 x^2\right) \right) dx\,;$$

$$k_1 = \frac{\operatorname{Im} z}{\operatorname{Re} z}\,.$$

5. $Y = \begin{cases} -l\,, & X_N < -d\,; \\ \dfrac{l}{d}\,, & |X_N| \le d\,, \\ l\,, & X_N \ge -d\,; \end{cases} \quad k_0 = \dfrac{1}{\mu}\arctan\left(\dfrac{r_2}{r_1}\right), \quad k_1 = \dfrac{1}{\sigma}\sqrt{-2\ln\sqrt{r_1^2 + r_2^2}}\,;$

$$r_1 = \cos l\left(1 - \Phi\left(\frac{d+\mu}{\sigma}\right) - \Phi\left(\frac{d-\mu}{\sigma}\right)\right) + \\ + \frac{1}{2}\,e^{-l^2\sigma^2/(2d^2)}\left(\cos\left(\frac{l\mu}{d}\right)\left(\operatorname{Re}\operatorname{erf}\left(\frac{d-\mu}{\sqrt{2\sigma}} - \imath\frac{l\sigma}{\sqrt{2}d}\right) - \right.\right. \\ \left. - \operatorname{Re}\operatorname{erf}\left(\frac{d+\mu}{\sqrt{2\sigma}} + \imath\frac{l\sigma}{\sqrt{2}d}\right)\right) - \sin\left(\frac{l\mu}{d}\right)\left(\operatorname{Im}\operatorname{erf}\left(\frac{d-\mu}{\sqrt{2\sigma}} - \imath\frac{l\sigma}{\sqrt{2}d}\right) - \right. \\ \left.\left. \left. - \operatorname{Im}\operatorname{erf}\left(\frac{d+\mu}{\sqrt{2\sigma}} + \imath\frac{l\sigma}{\sqrt{2}d}\right)\right)\right)\,;$$

$$r_2 = \sin l\left(\Phi\left(\frac{d+\mu}{\sigma}\right) - \Phi\left(\frac{d-\mu}{\sigma}\right)\right) + \\ + \frac{1}{2}\,e^{-l^2\sigma^2/(2d^2)}\left(\cos\left(\frac{l\mu}{d}\right)\left(\operatorname{Im}\operatorname{erf}\left(\frac{d-\mu}{\sqrt{2\sigma}} - \imath\frac{l\sigma}{\sqrt{2}d}\right) - \right.\right. \\ \left. - \operatorname{Im}\operatorname{erf}\left(\frac{d+\mu}{\sqrt{2\sigma}} + \imath\frac{l\sigma}{\sqrt{2}d}\right)\right) - \sin\left(\frac{l\mu}{d}\right)\left(\operatorname{Re}\operatorname{erf}\left(\frac{d-\mu}{\sqrt{2\sigma}} - \imath\frac{l\sigma}{\sqrt{2}d}\right) - \right. \\ \left.\left. \left. - \operatorname{Re}\operatorname{erf}\left(\frac{d+\mu}{\sqrt{2\sigma}} + \imath\frac{l\sigma}{\sqrt{2}d}\right)\right)\right)\,.$$

ПЕРСОНИФИЦИРОВАННОЕ ПРЕОБРАЗОВАНИЕ ПРЕДСТАВЛЕНИЙ ЦВЕТНЫХ ИЗОБРАЖЕНИЙ НА МОНИТОРЕ ПЭВМ

***О. П. Архипов*[1], *З. П. Зыкова*[1]**

Аннотация: Метод, применяемый для повышения точности тоновоспроизведения в полиграфии, основан на применении равноконтрастного градационного преобразования тоновых шкал в цветовом пространстве «стандартного» наблюдателя. Рассмотрена аналогичная задача в цветовом пространстве произвольного пользователя ПЭВМ при цветовоспроизведении на мониторе. Предложен метод, основанный на применении персонифицированного равноконтрастного градационного преобразования тоновых шкал. Приведены примеры, иллюстрирующие эффективность метода.

Ключевые слова: тоновоспроизведение; контраст; градация; тоновые шкалы; равноконтрастные градационные преобразования

1 Введение

При массовом репродуцировании в полиграфии для обеспечения точности тоновоспроизведения широко используется метод предварительного преобразования исходных изображений в соответствии с равноконтрастным градационным преобразованием тоновых шкал, которое зависит от характеристик цветовоспроизведения на применяемом полиграфическом оборудовании [1] и модели цветовосприятия «стандартного» наблюдателя [2]. При этом за счет увеличения степени подобия оригиналу по детализации изображения существенно улучшается восприятие полученных репродукций большинством наблюдателей, цветовосприятие которых близко к стандарту.

Чтобы добиться аналогичного результата в общем случае, необходимо учесть, что значительная часть наблюдателей имеет различные аномалии цветового зрения [2–4]. Поскольку особенности цветового зрения таких наблюдателей не учитываются при стандартном описании, в общем случае требуется персонифицированное моделирование цветовосприятия, соответствующее конкретному устройству вывода и конкретному наблюдателю.

В рамках данной работы для обеспечения максимально точного тоновоспроизведения при выводе цветных изображений на монитор пользовательской

[1]Институт проблем информатики Российской академии наук, ofran@orel.ru

компьютерной системы предлагается метод их предварительного персонифицированного преобразования.

Метод основан на применении RGB-характеризации цветовосприятия [5–11].

2 RGB-характеризация цветовосприятия

Пусть наблюдателю (произвольному пользователю ПЭВМ) присвоен порядковый номер k. Обозначим через F_k функцию цветовосприятия наблюдателем представления RGB-пикселов на мониторе ПЭВМ, аргументами которой являются пикселы из RGB-пространства, а значениями — пикселы из цветового пространства наблюдателя.

В качестве приближенного цифрового описания цветовосприятия k-го наблюдателя в соответствии с [5–11] может использоваться функция Ψ_k, аргументами и значениями которой являются RGB-пикселы.

Функция Ψ_k вычисляется по результатам тестирования различения наблюдателем представления цветных пикселов на мониторе ПЭВМ. Используется табличное задание Ψ_k

$$y = \Psi_k(x)\,, \quad x \in M\,,$$

где M — множество RGB-пикселов, координаты которых кратны 17. В этом случае множества M и $\Psi_k(M)$ могут быть представлены в виде цветных изображений, которые сохраняются на диске ПЭВМ в 24-битовом BMP-формате.

По определению линия уровня функции Ψ_k, построенная для произвольного пиксела, приближенно совпадает с зоной толерантности в цветовом пространстве наблюдателя — линией уровня функции F_k, соответствующей тому же пикселу.

Если для пикселов x' и x'' выполнено соотношение

$$F_k\left(x'\right) = F_k\left(x''\right)\,,$$

то

$$\Psi_k\left(x'\right) \approx \Psi_k\left(x''\right)\,.$$

Если для пикселов x' и x'' выполнено соотношение

$$\Psi_k\left(x'\right) = \Psi_k\left(x''\right)\,,$$

то

$$F_k\left(x'\right) \approx F_k\left(x''\right)\,.$$

Таким образом, функция Ψ_k является персонифицированной RGB-характеризацией цветовосприятия наблюдателя и может использоваться для предсказания восприятия наблюдателем представления цветных изображений на мониторе.

В качестве примера рассматриваются функции RGB-характеризации цветовосприятия:

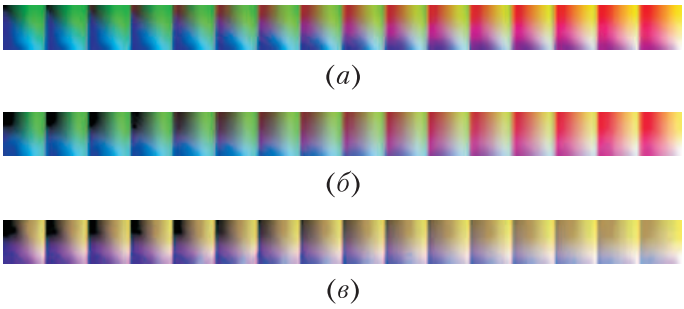


Рис. 1 Изображение множеств: (*а*) M; (*б*) $\Psi_0(M)$; (*в*) $\Psi_1(M)$

– функция Ψ_0, построенная по результатам визуального тестирования в цветовом пространстве одного реального наблюдателя, восприятие которого близко к стандарту;
– функция Ψ_1, построенная по результатам визуального тестирования в цветовом пространстве одного виртуального наблюдателя — дейтеранопа, восприятие которого соответствует модели, использованной в программе *Color Oracle* [3].

Отображение множества M в RGB-пространство, соответствующее указанным функциям, представлено на рис. 1.

3 RGB-характеризация восприятия изображений

Чтобы охарактеризовать восприятие изображения, достаточно преобразовать его с помощью функции Ψ_k:

$$\{x\} \Rightarrow \Psi_k\left(\{x\}\right) = \{\Psi_k(x)\}\ .$$

В общем случае значение Ψ_k от произвольного пиксела x можно вычислить при интерполяции по ближайшим точкам. Такими точками являются вершины наименьшего из RGB-кубиков, которому принадлежит пиксел x и вершинами которого являются пикселы из множества M: (R', G', B'); (R'', G', B'); (R', G'', B'); (R', G', B''); (R'', G'', B'); (R'', G', B''); (R', G'', B''); (R'', G'', B''), где R', R'', G', G'', B', B'' — числа, кратные 17, причем

$$R' \le R \le R'' = R' + 17\,;\ G' \le G \le G'' = G' + 17\,;\ B' \le B \le B'' = B' + 17\,.$$

Теперь для произвольного пиксела x в соответствии, например, с интерполяционной формулой ω из [6] имеем:

$$\Psi_k\left(x\right) \approx \omega\left(x, \Psi_k, M\right)\ .$$

(*а*)

(*б*)

(*в*)

Рис. 2 Вид изображения: (*а*) Img; (*б*) $\Psi_0(\mathrm{Img})$; (*в*) $\Psi_1(\mathrm{Img})$

В качестве примера рассматриваются ранее определенные функции RGB-характеризации цветовосприятия Ψ_0 и Ψ_1 и тестовое изображение Img (рис. 2, *а*), которое имеет очевидную недостаточную детализацию некоторых фрагментов «в тенях».

Восприятие представления изображения Img на мониторе можно предсказать с помощью изображений $\Psi_0(\mathrm{Img})$ и $\Psi_1(\mathrm{Img})$, приведенных на рис. 2, *б* и *в*.

Заметим, что недостатки исходного изображения, связанные с потерей деталей в области «теней», усугубляются в восприятии наблюдателя при их представлении на мониторе. Кроме того, при этом теряются и некоторые детали в области «светов». Чтобы улучшить представление изображений в подобных случаях, применяют различные предварительные преобразования.

Пусть известно табличное задание функции предварительного преобразования RGB-изображения $f_{k,l}$

$$y = f_{k,l}(x)\,, \quad x \in M\,.$$

(*а*)

(*б*)

Рис. 3 Вид изображений: (*а*) $\Psi'_{0,0}(M)$; (*б*) $\Psi'_{1,0}(M)$

Значение $f_{k,l}$ от произвольного пиксела x произвольного RGB-изображения можно вычислить аналогично предыдущему путем интерполяции

$$f_{k,l}(x) \approx \omega\left(x, f_{k,l}, M\right) .$$

Для вычисления изображения $\{\Psi'_{k,l}(x)\}$, предсказывающего восприятие произвольного изображения $\{x\}$ после преобразования, достаточно преобразовать изображение $\{f_{k,l}(x)\}$ с помощью функции Ψ_k:

$$\{x\} \Rightarrow \{f_{k,l}(x)\} \approx \{\omega\left(x, f_{k,l}, M\right)\} \Rightarrow \{\Psi_k\left(\omega\left(x, f_{k,l}, M\right)\right)\} \approx$$
$$\approx \{\omega\left(\omega\left(x, f_{k,l}, M\right), \Psi_k, M\right)\} = \left\{\Psi'_{k,l}(x)\right\} .$$

В качестве примера рассматриваются ранее определенные функции RGB-характеризации цветовосприятия Ψ_0 и Ψ_1, тестовое изображение Img и функция предварительного преобразования $f_{k,0}$, построенная по методу [12].

В соответствии с определением функций Ψ_k и $f_{k,0}$ было вычислено табличное задание функций $\Psi'_{k,0}$, что позволило определить изображения $\Psi'_{0,0}(M)$, $\Psi'_{1,0}(M)$ в соответствии с рис. 3.

По определению $f_{k,0}$ использование соответствующего этой функции преобразования позволяет максимально (насколько позволяет цветовое пространство конкретного наблюдателя) увеличить степень подобия представления изображения его оригиналу в части цветового исполнения. При этом часто улучшается точность тоновоспроизведения, а следовательно, и детализация изображения, что и подтверждается видом изображений $\Psi'_{0,0}(\mathrm{Img})$ и $\Psi'_{1,0}(\mathrm{Img})$ на рис. 4.

Если наиболее важным при представлении изображений является увеличение степени подобия представления изображения его оригиналу в части детализации, то следует применять, как это делается в полиграфии, предварительные преобразования изображений на основе результатов равноконтрастных градационных преобразований ступенчатых тоновых шкал.

(*а*) (*б*)

Рис. 4 Вид изображений: (*а*) $\Psi'_{0,0}(\mathrm{Img})$; (*б*) $\Psi'_{1,0}(\mathrm{Img})$

4 Равноконтрастное градационное преобразование ступенчатых тоновых шкал

Рассмотрим ступенчатые тоновые шкалы

$$T_i \subset M\,, \quad i = 0, 1, \ldots, 6\,,$$

с компонентами вида:

$$\begin{aligned}
&T_0 = \{t_{0,j}\}\,, && t_{0,j} = \{(j\cdot 17, j\cdot 17, j\cdot 17);\\
&T_1 = \{t_{1,j}\}\cup\{t_{1,15+j}\}\,, && t_{1,j} = (j\cdot 17, 0, 0)\,;\ \ t_{1,15+j} = (255, j\cdot 17, j\cdot 17)\,;\\
&T_2 = \{t_{2,j}\}\cup\{t_{2,1+j}\}\,, && t_{2,j} = (j\cdot 17, j\cdot 17, 0)\,;\ \ t_{2,15+j} = (255, 255, j\cdot 17)\,;\\
&T_3 = \{t_{3,j}\}\cup\{t_{3,15+j}\}\,, && t_{3,j} = (0, j\cdot 17, 0)\,;\ \ t_{3,15+j} = (j\cdot 17, 255, j\cdot 17)\,;\\
&T_4 = \{t_{4,j}\}\cup\{t_{4,15+j}\}\,, && t_{4,j} = (j\cdot 17, j\cdot 17, 0)\,;\ \ t_{4,15+j} = (255, 255, j\cdot 17)\,;\\
&T_5 = \{t_{5,j}\}\cup\{t_{5,15+j}\}\,, && t_{5,j} = (0, 0, j\cdot 17)\,;\ \ t_{5,15+j} = (j\cdot 17, j\cdot 17, 255)\,;\\
&T_6 = \{t_{6,j}\}\cup\{t_{6,15+j}\}\,, && t_{6,j} = (j\cdot 17, 0, j\cdot 17)\,;\ \ t_{6,15+j} = (255, j\cdot 17, 255)\,,\\
& && \qquad\qquad j = 0, 1, \ldots, 15\,.
\end{aligned}$$

Заметим, что серая шкала T_0 является равномерной в RGB-пространстве, а каждая из остальных шкал состоит из двух равномерных фрагментов (рис. 5).

Как известно, в общем случае тоновые шкалы не являются равноконтрастными в восприятии k-го наблюдателя, т. е.

$$c_{k,1}\left(\Psi_k(t_{i,j}), \Psi_k(t_{i,j-1})\right) \neq c_{k,1}\left(\Psi_k(t_{i,j}), \Psi_k(t_{i,j+1})\right)\,,$$

где $c_{k,1}$ — визуально определяемый контраст двух пикселов в цветовом пространстве k-го наблюдателя.

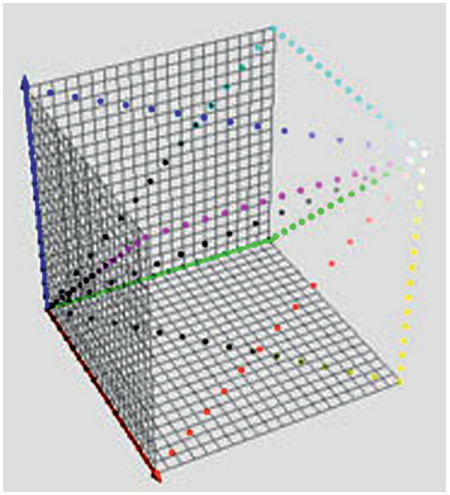


Рис. 5 Вид $\{t_{i,j}\}$ в RGB-пространстве

Также в общем случае тоновые шкалы не являются равноконтрастными в восприятии стандартного наблюдателя, т. е.

$$c_{k,2}\left(\Psi_k(t_{i,j}), \Psi_k(t_{i,j-1})\right) \neq \neq c_{k,2}\left(\Psi_k(t_{i,j}), \Psi_k(t_{i,j+1})\right) ,$$

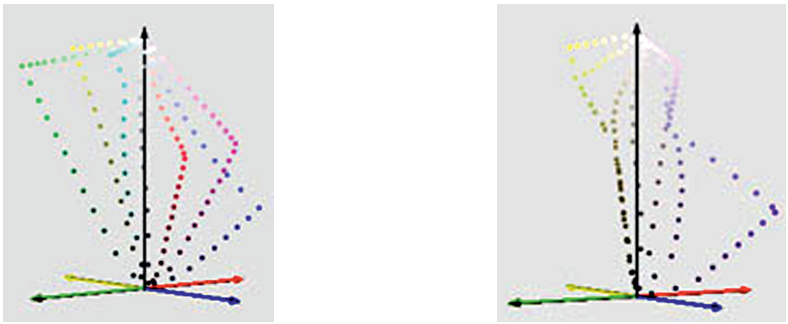
где $c_{k,2}$ — контраст двух пикселов в цветовом пространстве k-го наблюдателя, значение которого автоматически определяется как расстояние ΔE между образами пикселов в Lab-пространстве, когда пикселы считаются различаемыми, если выполнено соотношение [1]:

$$\Delta E > 5 .$$

Для рассматриваемых случаев Ψ_0 и Ψ_1 пикселы соответствующих последовательностей также распределены в Lab-пространстве неравномерно, что отражено на рис. 6.

В [13] предложен метод определения функции $\gamma_{k,l}$, $l \in \{1, 2\}$, которая преобразует равномерные шкалы $\{t_{i,j}\}$ в шкалы $\{t'_{k,l,i,j}\}$:

$$t'_{k,l,i,j} = \gamma_{k,l}(t_{i,j}) ,$$



(*а*) (*б*)

Рис. 6 Вид $\{\Psi_k(t_{i,j})\}$ в Lab-пространстве: (*а*) $k = 0$; (*б*) $k = 1$

для компонентов которых выполняется соотношение:

$$c_{k,l}\left(\Psi_k(t'_{k,l,i,j}),\Psi_k(t'_{k,l,i,j-1})\right)\approx c_{k,l}\left(\Psi_k(t'_{k,l,i,j}),\Psi_k(t'_{k,l,i,j+1})\right)\,, \qquad (1)$$

т. е. шкалы $\{t'_{k,l,i,j}\}$ являются равноконтрастными в соответствии с l-м способом определения контраста.

Поскольку расстояния между соседними членами равномерных шкал равны

$$\rho(t_{i,j},t_{i,j-1})=\rho(t_{i,j},t_{i,j+1})\,,$$

то (1) можно записать в общем виде через нормированные значения контраста:

$$\sigma'_{k,l,i,j+1}\approx\sigma'_{k,l,i,j}\,, \qquad (2)$$

где

$$\sigma'_{k,l,i,j}=\frac{c_{k,l}\left(\Psi_k(t'_{k,l,i,j}),\Psi_k(t'_{k,l,i,j-1})\right)}{\rho(t_{i,j},t_{i,j-1})}\,.$$

Алгоритм реализации метода [13] нуждается в доработке, чтобы (2) выполнялось и в случае неравномерных шкал. Для этого достаточно изменить алгоритм определения нового местоположения градаций.

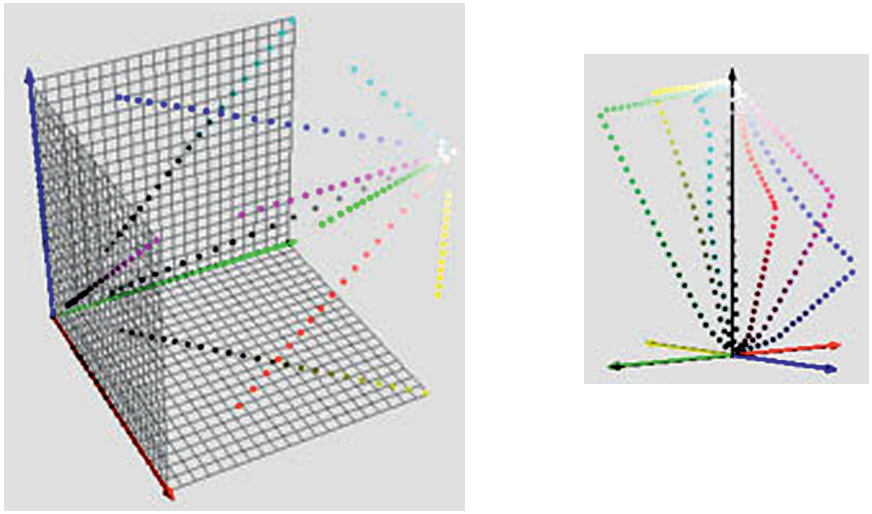
Обозначим через S_i непрерывные шкалы, соответствующие ступенчатым шкалам T_i, а $G_{k,l,i}=\{g_{k,l,i,j}\}$, $j=1,2,\ldots,J_{k,l,i}$, — последовательности градаций, являющиеся подпоследовательностями S_i. Состав $G_{k,l,i}$ зависит от цветовосприятия k-го наблюдателя и l-го способа определения контраста. Видно, что в рассмотренных примерах данные последовательности имеют различное число компонентов $J_{k,l,i}$:

i	$J_{0,1,i}$	$J_{1,1,i}$	$J_{0,2,i}$	$J_{1,2,i}$
0	54	71	19	21
1	58	69	37	24
2	53	61	37	37
3	56	61	43	32
4	59	54	29	24
5	75	63	40	45
6	67	68	36	28

Шкалы $\{S_i\}$, $i\geq 1$, состоят из двух равномерных фрагментов. Обозначим их длины через $L_{i,1}$ и $L_{i,2}$, а расстояния между пикселами — соответственно через $h_{i,1}$ и $h_{i,2}$.

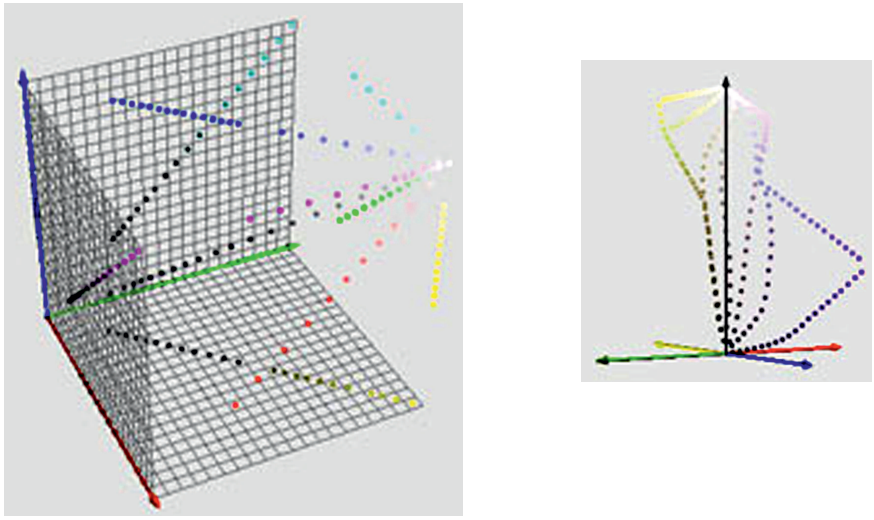
Чтобы градации были распределены равномерно, расстояние между ними должно быть равно

$$H_{k,l,i}=\frac{L_{i,1}+L_{i,2}}{J_{k,l,i}}\,.$$



(*а*) (*б*)

Рис. 7 Вид изображений: (*а*) $\{g_{0,2,i,j}\}$ в RGB-пространстве; (*б*) $\{\Psi_0(g_{0,2,i,j})\}$ в Lab-пространстве



(*а*) (*б*)

Рис. 8 Вид изображений: (*а*) $\{g_{1,2,i,j}\}$ в RGB-пространстве; (*б*) $\{\Psi_1(g_{1,2,i,j})\}$ в Lab-пространстве

В связи с этим на первом фрагменте шкал $\{S_i\}$ должно быть распределено $(1 + [L_{i,1}/H_{k,l,i}])$ градаций, а на втором — $(J_{k,l,i} - [L_{i,1}/H_{k,l,i}])$.

После определения нового местоположения градаций приближенные значения функции равноконтрастного градационного преобразования $\gamma_{k,l}$, обеспечивающей выполнение (2), вычисляются при линейной интерполяции обычным путем.

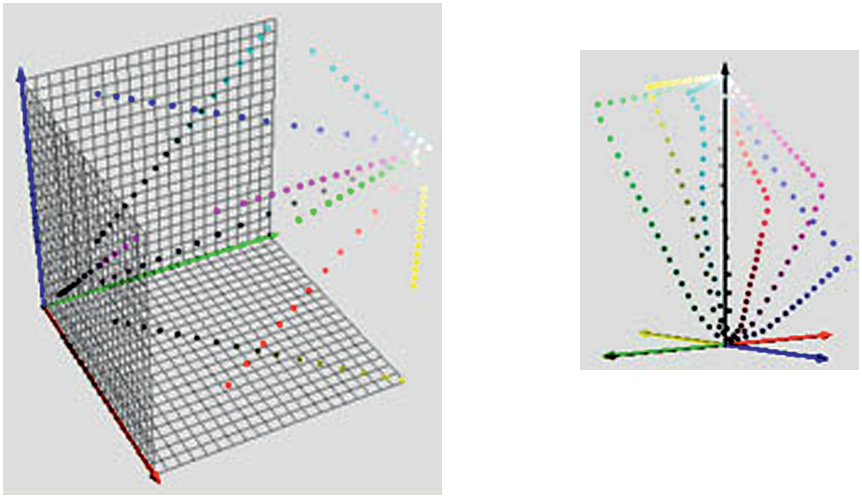
Для двух рассматриваемых случаев при $l = 2$ были найдены различные по количественному и качественному составу последовательности градаций, отраженные на рис. 7 и 8.

В результате равноконтрастных градационных преобразований кусочно-равномерные последовательности $\{T_i\}$ были преобразованы в неравномерные, но приближенно равноконтрастные последовательности $\{T'_{k,2,i}\}$, представленные на рис. 9 и 10.

Введем обозначения:

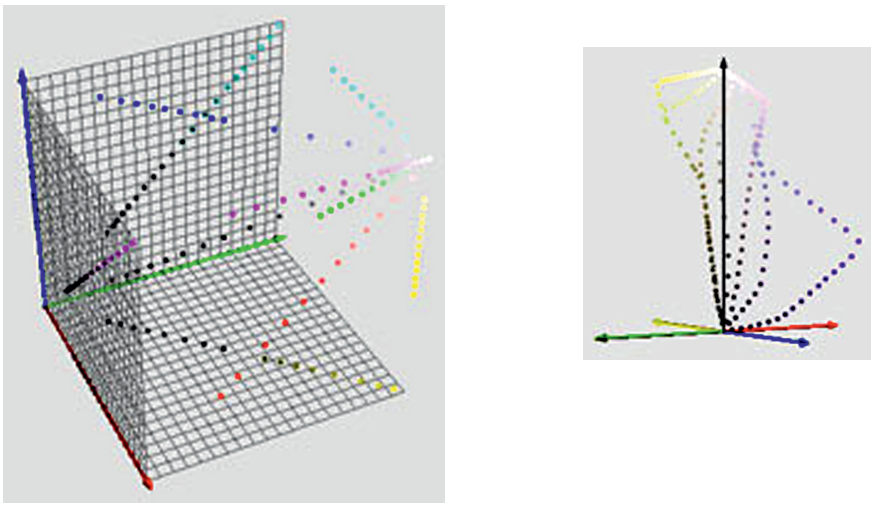
$$\sigma_{k,l,i,j} = \frac{c_{k,l}\left(\Psi_k(t_{i,j}), \Psi_k(t_{i,j-1})\right)}{\rho(t_{i,j}, t_{i,j-1})}\,; \quad \sigma''_{k,l,i,j} = \frac{c_{k,l}\left(t'_{k,l,i,j}, t'_{k,l,i,j-1}\right)}{\rho(t_{i,j}, t_{i,j-1})}\,.$$

Сравнительный анализ последовательностей значений $\{t_{1,j}\}$, $\{t'_{k,l,1,j}\}$, $\{\Psi_k(t_{1,j})\}$, $\{\Psi_k(t'_{k,l,1,j})\}$, $\{\sigma_{k,l,1,j}\}$, $\{\sigma'_{k,l,1,j}\}$ и $\{\sigma''_{k,l,1,j}\}$, проиллюстрированный рис. 11–13, показывает существенные изменения. Так, после применения



(*а*) (*б*)

Рис. 9 Вид изображений: (*а*) $\{t'_{0,2,i,j}\}$ в RGB-пространстве; (*б*) $\{\Psi_0(t'_{0,2,i,j})\}$ в Lab-пространстве



(*а*) (*б*)

Рис. 10 Вид изображений: (*а*) $\{t'_{1,2,i,j}\}$ в RGB-пространстве; (*б*) $\{\Psi_1(t'_{1,2,i,j})\}$ в Lab-пространстве

предложенного метода диапазон изменения контраста для рассмотренных случаев уменьшился более чем в три раза.

Далее займемся определением функций предварительных преобразований изображений $f_{k,l}$ в соответствии с функциями $\gamma_{k,l}$ равноконтрастных градационных преобразований ступенчатых тоновых шкал.

5 Определение функции $f_{k,l}$

Для определения на множестве M табличного задания функции $f_{k,l}$, согласованной с функцией $\gamma_{k,l}$, достаточно приравнять ее значения значениям $\gamma_{k,l}$ для тех пикселов, которые принадлежат совокупности тоновых шкал $\{T_i\}$, и корректно доопределить ее значения для оставшихся пикселов множества M.

Пусть пиксел x принадлежит отрезку $[x_0, x_1]$, концы которого являются соседними пикселами какой-либо тоновой шкалы. Поскольку значения $\gamma_{k,l}(x_0)$ и $\gamma_{k,l}(x_1)$ известны, то полагаем

$$f_{k,l}(p) = \frac{\gamma_{k,l}(p_0)\rho(p,p_1) + \gamma_{k,l}(p_1)\rho(p_0,p)}{\rho\,(p_0,p_1)}\,,\quad x \in [x_0,\, x_1]\,, \tag{3}$$

где ρ — расстояние между двумя пикселами в RGB-пространстве.

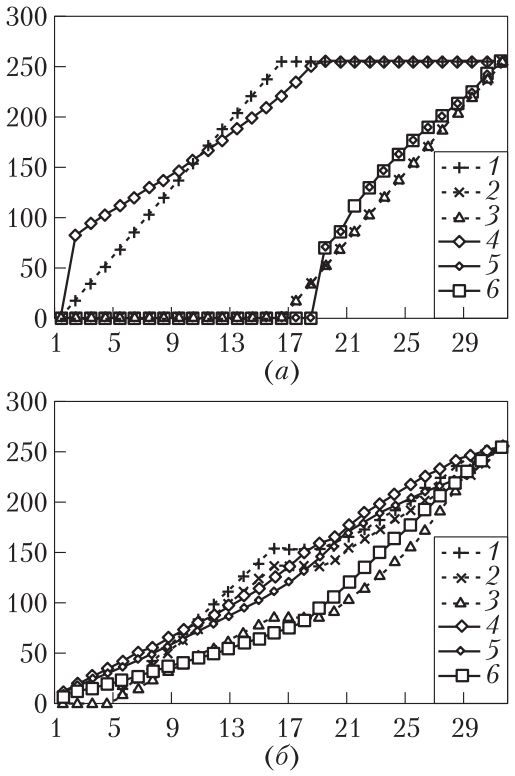


Рис. 11 Результаты преобразований $\gamma_{1,2}$: (*а*) $(R,G,B)_j = t_{1,j}$, $(R',G',B')_j = t'_{1,2,1,j}$; (*б*) $(R,G,B)_j = \Psi_1(t_{1,j})$, $(R',G',B')_j = \Psi_1(t'_{1,2,1,j})$; *1* — R; *2* — G; *3* — B; *4* — R'; *4* — G'; *6* — B'

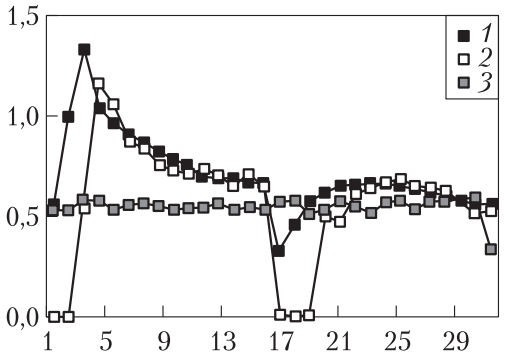


Рис. 12 Показатели контраста до и после преобразования $\gamma_{0,2}$: *1* — $\{\sigma''_{0,2,1,j}\}$; *2* — $\{\sigma_{0,2,1,j}\}$; *3* — $\{\sigma'_{0,2,1,j}\}$

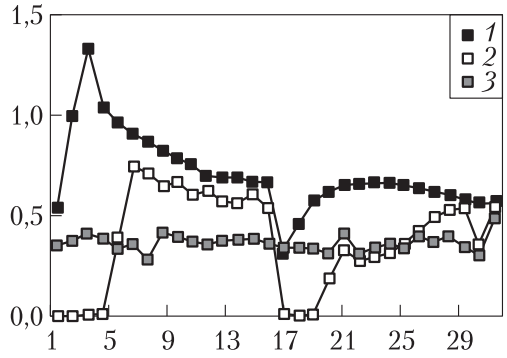


Рис. 13 Показатели контраста до и после преобразования $\gamma_{1,2}$: *1* — $\{\sigma''_{1,2,1,j}\}$; *2* — $\{\sigma_{1,2,1,j}\}$; *3* — $\{\sigma'_{1,2,1,j}\}$

Во всех других случаях вычислим точки пересечения плоскости, проходящей через данную точку x_0 перпендикулярно серой оси RGB-куба, и ломаными линиями, соответствующими серой и двум разным ближайшим цветным тоновым шкалам. Обозначим полученные пикселы x_0, x_1 и x_2 соответственно.

Для каждой из этих точек найдется пара пикселов соответствующей тоновой шкалы, которые являются концами отрезка, содержащего эту точку. Можем определить в них значения $\gamma_{k,l}(x_0)$, $\gamma_{k,l}(x_1)$ и $\gamma_{k,l}(x_2)$, воспользовавшись (3).

Введем в рассмотрение матрицу

$$A_{k,l}(x) = \{a_{k,l,i,j}\}\,, \quad i,j \in \{0,1,2\}\,.$$

Определим ее компоненты из системы уравнений:

$$A_{k,l}(x)x_n = \gamma_{k,l}(x_n)\,, \quad n \in \{0,1,2\}\,.$$

Поскольку точки x_0, x_1 и x_2 заведомо не лежат на одной прямой, то решение системы уравнений существует, является единственным и может быть вычислено с помощью известных методов.

Теперь функция $f_{k,l}$ может быть следующим образом определена для всех пикселов множества M:

$$f_{k,l}(x) = A_{k,l}(x)x\,.$$

Соответствующая функция $\Psi'_{k,l}$ имеет вид:

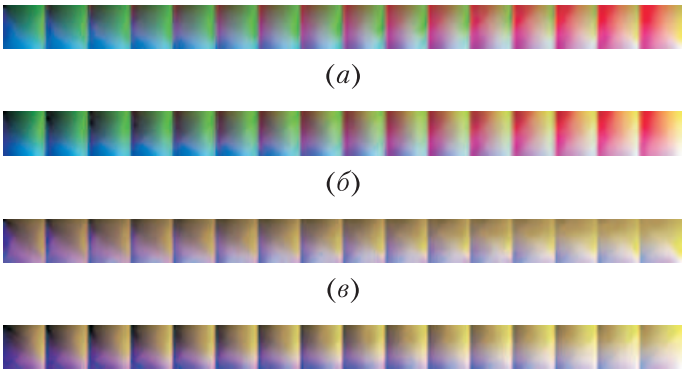
$$\Psi'_{k,l}(x) = \omega\left(\omega\left(x, A_{k,l}(x), M\right), \Psi_k, M\right)\,, \quad x \in M\,.$$

6 Примеры применения функций $f_{k,l}$

В соответствии с определением функций Ψ_k и $f_{k,l}$ было вычислено табличное задание функций $\Psi'_{k,l}$, $k \in \{0, 1\}$, $l \in \{1, 2\}$ (рис. 14).

Это позволило построить изображения $\Psi'_{k,l}(\mathrm{Img})$ (рис. 15 и 16), предсказывающие восприятие тестового изображения после того, как оно было преобразовано с помощью функций $f_{k,l}$, $k \in \{0, 1\}$, $l \in \{1, 2\}$.

Заметим, что, как и следовало ожидать, в рассмотренных примерах после предварительных преобразований на основе равноконтрастного градационного преобразования ступенчатых тоновых шкал улучшилась детализация представления изображений. Особенно заметные улучшения по детализации представле-



(*а*) (*б*) (*в*) (*г*)

Рис. 14 Вид изображений: (*а*) $\Psi'_{0,1}(M)$; (*б*) $\Psi'_{0,2}(M)$; (*в*) $\Psi'_{1,1}(M)$; (*г*) $\Psi'_{1,2}(M)$

(*а*) (*б*)

Рис. 15 Вид изображений: (*а*) $\Psi'_{0,1}(\mathrm{Img})$; (*б*) $\Psi'_{0,2}(\mathrm{Img})$

(*а*) (*б*)

Рис. 16 Вид изображений: (*а*) $\Psi'_{1,1}(\mathrm{Img})$, (*б*) $\Psi'_{1,2}(\mathrm{Img})$

ния изображения обеспечили персонифицированные преобразования с помощью функций $f_{k,1}$, $k \in \{0, 1\}$.

7 Заключение

Рассмотрена задача улучшения качества представления цветных изображений на мониторе в части повышения точности тоновоспроизведения. Предлагается численный метод ее решения, состоящий в определении персонифицированной функции предварительного преобразования изображений, согласованной с функцией равноконтрастного градационного преобразования ступенчатых тоновых шкал.

В связи с неизбежными погрешностями измерений и вычисления функции равноконтрастного градационного преобразования, с неточностью тестирования цветовосприятия, с дискретностью связанных цветовых пространств и с необходимостью применения численных методов возможно только приближенное определение функций предварительного преобразования изображений.

Тем не менее, результаты практической реализации показывают эффективность применения метода. Применение вычисленных функций предварительного преобразования позволяет улучшить качество представления изображений на мониторе в части их детализации.

Детализация изображений улучшалась, даже если не проводилось реальное градационное тестирование наблюдателя, и градации определялись при автоматическом виртуальном Lab-тестировании в цветовом пространстве наблюдателя. Однако, как и следовало ожидать, наилучших результатов удается добиться, если при вычислении функции предварительного преобразования используются данные реального тестирования.

Результаты работы имеют важное практическое значение, поскольку применимы при решении задач повышения качества цветовоспроизведения на периферийных устройствах ПЭВМ не только в интересах пользователей, чье цветовосприятие близко к цветовосприятию «стандартного» наблюдателя, но и в интересах произвольных пользователей, возможно, имеющих аномалии цветового зрения.

Авторы выражают благодарность коллеге Е. В. Рябинкину за создание программного обеспечения, с помощью которого были созданы иллюстрации, приведенные в данной работе.

Литература

1. *Александров Д.* Равноконтрастное градационное преобразование полиграфических изображений // Полиграфия, 1999. № 1. С. 25–26.
2. *Джадд Д., Вышецки Г.* Цвет в науке и технике. — М.: Мир, 1978.
3. *Jenny B., Vaughn Kelso N.* Color Oracle: Design for the color impaired. http://colororacle.cartography.ch.
4. Vischeck. http://www.vischeck.com.
5. *Архипов О. П., Зыкова З. П.* Некоторые формы представления цветового пространства RGB // Информационные технологии, 1999. № 3. С. 17–23.
6. *Архипов О. П., Зыкова З. П.* Интеграция гетерогенной информации о цветных пикселях и их цветовосприятии // Информатика и её применения, 2010. Т. 4. Вып. 4. С. 14–25.
7. *Архипов О. П., Зыкова З. П.* Функциональное описание индивидуального цветовосприятия // Информационные системы и технологии, 2010. № 5. С. 5–12.
8. *Архипов О. П., Зыкова З. П.* RGB-характеризация пространства цветовосприятия // Системы и средства информатики. — М.: ИПИ РАН, 2010. Вып. 20. № 1. С. 73–90.
9. *Архипов О. П., Зыкова З. П.* Программная система многокритериального выбора и тестирования множества тест-пикселей для исследования цветовосприятия (IP_9_3): Свидетельство о государственной регистрации программы для ЭВМ № 2010613793 от 09.06.2010.
10. *Архипов О. П., Зыкова З. П.* Многокритериальный выбор тестового множества при исследовании цветовосприятия // Информационные технологии, 2011. № 2. С. 67–73.
11. *Архипов О. П., Зыкова З. П.* Программная система визуального тестирования и предсказания восприятия цветных изображений (T_10): Свидетельство о государственной регистрации программы для ЭВМ № 2011613764 от 13.05.2011.
12. *Соколов И. А., Архипов О. П., Захаров В. Н., Зыкова З. П., Архипов П. О.* Способ компьютерного распознавания и визуального воспроизведения цветных изображений: Патент РФ № 2005130683 от 04.10.05 // Бюл. № 8 от 20.03.07.
13. *Архипов О. П., Зыкова З. П.* Равноконтрастные градационные преобразования ступенчатых тоновых шкал // Информационные системы и технологии, 2011. № 4. С. 39–46.

СИСТЕМА ХАРАКТЕРИЗАЦИИ САМОСИНХРОННЫХ ЭЛЕМЕНТОВ

Ю. Г. Дьяченко[1], *Н. В. Морозов*[2], *Д. Ю. Степченков*[3], *Ю. А. Степченков*[4]

Аннотация: Представлена методика характеризации (извлечения временны́х и электрических параметров для формирования функционально-логической модели) элементов библиотеки для проектирования самосинхронных (СС) схем. Показано, что специфика поведения СС-элементов накладывает дополнительные ограничения на процесс характеризации и приводит к изменению смысла некоторых параметров модели элемента. Описана программная система характеризации СС-элементов. Представлены результаты ее опытной эксплуатации.

Ключевые слова: самосинхронные элементы; характеризация; функционально-логическая модель; программные средства

1 Введение

Разработка современных больших интегральных схем (БИС) невозможна без использования готовых библиотек стандартных элементов разного уровня — от простейших логических элементов до фрагментов памяти и сложных функциональных блоков. Потребительские характеристики схемы — количество транзисторов, площадь реализации, быстродействие, энергопотребление — не в последнюю очередь зависят от функционального состава используемой библиотеки стандартных элементов. Особенно сильно это проявляется при проектировании СС-схем. Обусловлено это необходимостью индицировать (определять момент окончания переключения) каждый элемент в составе СС-схемы. Под элементом здесь понимается однокаскадная принципиальная схема, выполняющая некоторую логическую функцию.

Современные библиотеки стандартных элементов для проектирования заказных БИС, как правило, не учитывают указанной специфики СС-схем. Это заставляет разработчиков СС-БИС вводить в состав готовой сертифицированной библиотеки новые элементы, повышающие эффективность проектирования СС-схем и улучшающие их потребительские характеристики.

[1]Институт проблем информатики Российской академии наук, diaura@mail.ru
[2]Институт проблем информатики Российской академии наук, NMorozov@ipiran.ru
[3]Институт проблем информатики Российской академии наук, Stepchenkov@mail.ru
[4]Институт проблем информатики Российской академии наук, YStepchenkov@ipiran.ru

Но в этом случае разработчикам приходится сталкиваться с необходимостью характеризации добавляемых в библиотеку элементов, чтобы иметь возможность использовать их в традиционном маршруте проектирования БИС с применением имеющихся систем автоматизированного проектирования (САПР). В настоящее время известны различные программные средства характеризации элементов стандартных библиотек, например [1–3]. Однако они не учитывают специфики функционирования СС-элементов и не позволяют из-за этого получить для них адекватные функционально-логические модели.

Данная работа посвящена проблемам характеризации элементов, составляющих схемотехнический базис разработки СС-схем, и создания программного комплекса автоматизированной характеризации таких элементов.

2 Состав библиотеки самосинхронных элементов

Эффективность проектирования БИС с использованием библиотеки стандартных элементов определяется количественным и качественным составом библиотеки, оптимальностью использованных схемотехнических и топологических решений, а также достоверностью характеристик элемента, заложенных в его модель.

К настоящему времени в ИПИ РАН разработана библиотека СС-элементов IPI18 [4] для проектирования КМОП (комплементарный металл–оксид–полупроводник) БИС по технологии с проектными нормами 180 нм и ниже. Она включает 100 элементов различной сложности:

- однокаскадные комбинационные элементы;
- мультиплексоры с парафазными входами и выходами;
- сумматор и полусумматор с парафазными входами и выходами;
- самосинхронные счетные триггеры с различными вариантами установки;
- однотактные RS-триггеры с бифазными и парафазными входами;
- двухтактные RS-триггеры с бифазными и парафазными входами;
- триггеры с унарным информационным входом;
- индикаторные элементы (гистерезисные триггеры и другие).

Использование данной библиотеки для проектирования заказных БИС затруднено в связи с отсутствием стандартизованных файлов описания моделей элементов, получаемых с помощью процедуры характеризации. Ручная же характеризация по трудоемкости составляет 3–4 человеко-дня для каждого элемента. Поэтому разработка средств, позволяющих автоматизировать процедуру характеризации СС-элементов, является актуальной задачей.

Эти программные средства также существенно сокращают трудозатраты как на настройку готовой библиотеки элементов на новую технологию при изменении

проектных норм или технологических параметров, так и на ввод новых элементов в состав имеющейся библиотеки.

3 Параметры моделей самосинхронных элементов

Функциональная модель элемента, реализованного по КМОП-технологии, включает в себя следующие параметры:

- задержки переключения элемента от каждого входа к каждому выходу;
- временны́е ограничения, накладываемые на моменты изменения входных переменных;
- входные и выходные емкости;
- мощность переключения.

Задержка переключения выхода элемента определяется несколькими факторами. Переключение элемента инициируется соответствующим изменением значений сигналов на его входах. При этом длительность перехода элемента из одного состояния в другое зависит не только от текущей и последующей комбинаций значений входов, но и от порядка изменения входов, длительности процесса переключения входного сигнала и текущего состояния внутренних узлов элемента.

Временны́е ограничения, накладываемые на моменты изменения входных переменных, учитывают взаимоотношения между входными сигналами определенного функционального назначения. Например, в триггерных элементах с входом разрешения записи большое значение имеет временной интервал между моментом изменения информационного входа и последующего переключения сигнала разрешения записи. Информационный сигнал должен измениться раньше сигнала управления и оставаться неизменным некоторое время после окончания переключения последнего.

Энергия, потребляемая в процессе переключения схемы, расходуется на заряд емкостей в узлах схемы и нагревание резистивных компонентов активных транзисторных и пассивных структур. В отдельно взятом элементе основная составляющая энергии потребления определяется зарядом входных, выходных и внутренних емкостей.

В моделях СС-элементов используются следующие параметры:

(1) задержка распространения сигнала от входов к выходам элемента при переключении выхода из 0 в 1;

(2) задержка распространения сигнала от входов к выходам элемента при переключении выхода из 1 в 0;

(3) длительность перепада от 0 до 1 на выходе для каждой зависимой пары «вход–выход» при всех возможных комбинациях остальных входов;

(4) длительность перепада от 1 до 0 на выходе для каждой зависимой пары «вход–выход» при всех возможных комбинациях остальных входов;

(5) задержка входа управления после изменения информационного входа при переключении последнего из 0 в 1;

(6) задержка входа управления после изменения информационного входа при переключении последнего из 1 в 0;

(7) опережение входом управления изменения информационного входа при переключении последнего из 0 в 1;

(8) опережение входом управления изменения информационного входа при переключении последнего из 1 в 0;

(9) статическая мощность потребления;

(10) динамическая мощность переключения выхода из 0 в 1;

(11) динамическая мощность переключения выхода из 1 в 0;

(12) емкость входа;

(13) максимальная допустимая емкость нагрузки на выходе.

Если путей распространения сигнала от данного входа к данному выходу несколько, параметры 1–4, 10 и 11 рассчитываются для каждого из них.

В синхронных схемах используются также параметры «минимальная длительность высокого уровня импульса на входе» и «минимальная длительность низкого уровня импульса на входе». Они указываются для входов синхронизации и предустановки в триггерах и используются для контроля длительности активных уровней соответствующих сигналов. В самосинхронных же элементах вход управления, являющийся аналогом входа синхронизации в синхронной схемотехнике, зависит от индикаторного выхода элемента. Его изменение разрешается только после переключения индикаторного выхода на новое значение. Поэтому нет необходимости в определении минимальной длительности входа управления.

Входы предустановки в СС-схемах также не нуждаются в данном параметре. Асинхронная предустановка, используемая, главным образом, для инициализации схемы при включении питания или общем сбросе, традиционно задается достаточной длительности, чтобы все зависящие от нее элементы и блоки схемы успели установиться в определенное состояние. Самосинхронная предустановка так же связана с переключением индикаторного выхода, как и вход управления.

Таким образом, в СС-элементах данный параметр неактуален и в настоящей версии системы характеризации не реализован.

Энергопотребление при отсутствии переключения выхода элемента, используемое в синхронных схемах, относится ко входу элемента и рассчитывается только в том случае, когда при этом изменяется состояние схемы элемента, а внешние выходы остаются неизменными. Однако в СС-библиотеке нет таких

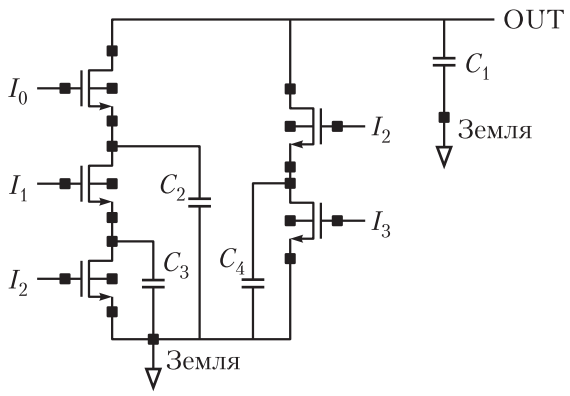


Рис. 1 Фрагмент принципиальной схемы

элементов, внутренние выходы которых переключались бы, а внешние выходы при этом не менялись. Это означало бы отсутствие индицируемости соответствующих внутренних выходов. Если в элементе нет индикаторного выхода, то все его выходы — внешние и переключение внутри элемента обязательно приведет к изменению состояния хотя бы одного из них. Поэтому энергопотребление при отсутствии переключения выхода элемента не включается в модель элементов СС-библиотеки.

Значения задержек и энергопотребления относительно какого-либо входа определяются для всех возможных путей распространения сигнала от данного входа до выхода. Для каждого пути запускается свой сеанс моделирования. Начальное состояние входов и выходов элемента в каждом сеансе выбирается так, чтобы активировалась именно выбранная цепочка транзисторов и перезаряжалось наибольшее количество емкостей на соответствующем пути.

На рис. 1 приведена n-часть принципиальной схемы логического элемента, выполняющего функцию

$$\mathrm{OUT} = \hat{}I_0 * I_1 * I_2 | I_2 * I_3 \,.$$

Начальное напряжение на выходе $\mathrm{OUT} = 1$. Переключение выхода OUT из 1 в 0 в зависимости от входа I_2 логически разрешается при двух комбинациях входных сигналов: $\{I_0 = I_1 = 1,\ I_3 = 0\}$ и $\{I_0 = I_1 = *,\ I_3 = 1\}$, — и может происходить тремя способами:

(1) по цепочке из трех транзисторов n-типа при $I_1 = I_0 = 1$, $I_3 = 0$;

(2) по цепочке из двух транзисторов при $I_0 = 0$, $I_1 = *$, $I_3 = 1$;

(3) по цепочке из двух транзисторов при $I_0 = 1$, $I_1 = 0$, $I_3 = 1$.

В первом случае будут перезаряжаться все паразитные емкости C_1, C_2, C_3 и C_4. Во втором случае будут перезаряжаться только емкости C_1 и C_4. В третьем случае будут перезаряжаться емкости C_1, C_2 и C_4. Эти три варианта переключения выхода OUT из 1 в 0 при подаче перепада $0 \rightarrow 1$ на вход I_2 характеризуются разными значениями задержки распространения, длительности спада выходного сигнала и энергопотребления. Поэтому при характеризации элемента их все необходимо учесть и специфицировать в модели элемента для максимально точного описания поведения элемента в реальных условиях работы. При необходимости число вариантов может быть сведено к одному — наихудшему с точки зрения задержки и энергопотребления (в данном случае — первому).

Вся совокупность параметров функциональной модели элемента рассчитывается на основе электрического моделирования принципиальной схемы элемента с помощью программы Ngspice [5]. Графики изменения уровня входных, выходных и внутренних узлов схемы элемента позволяют выявить все временны́е и электрические характеристики элемента путем анализа отклика схемы элемента на специально подобранные (сформированные) последовательности переключения входных переменных с варьируемой формой импульсов. Эта процедура и называется *характеризацией* элемента. Для ее автоматизации был разработан программный комплекс САХИБ (Система автоматизированной характеризации интегральных библиотек).

4 Структура программного комплекса САХИБ

Комплекс САХИБ разработан в рамках НИР «СТЕРХ» [6] и обеспечивает для стандартных САПР выполнение характеризации отдельных элементов ССБИС (сверхбольших интегральных схем). Комплекс САХИБ состоит из следующих основных частей (рис. 2):

- *библиотеки*, включающей списки цепей элементов, модели транзисторов, учитывающие вариабельность размеров транзисторов в схеме и технологический и физический разброс их параметров, и функционально-логические модели элементов;
- *подсистемы расчета параметров* модели элемента, анализирующей структуру принципиальной схемы элемента, формирующей на основе этого анализа задания для электрического моделирования схемы элемента и извлекающей параметры модели элемента из результатов моделирования;
- *подсистемы трансляции результатов характеризации*, обрабатывающей результаты электрического моделирования элемента и формирующей файлы функционально-логического описания библиотеки элементов в различных форматах;

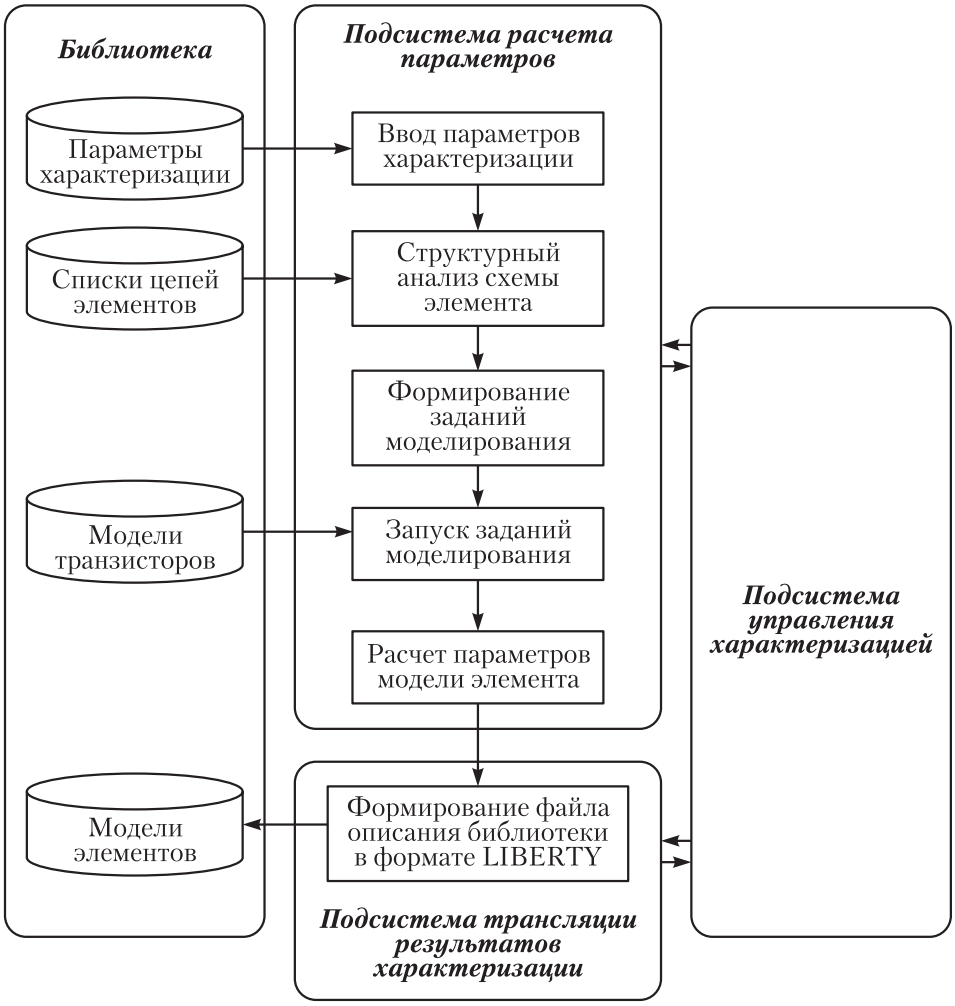


Рис. 2 Структурная схема комплекса САХИБ

– *подсистемы управления характеризацией*, управляющей подсистемами расчета параметров и трансляции результатов и включающей графический интерфейс пользователя, сопровождающий и упрощающий процесс характеризации.

Комплекс САХИБ выполняет следующие действия:

1. Проводит входной контроль: синтаксический и семантический анализ исходного описания характеризуемого элемента.

2. Анализирует транзисторную структуру элемента и определяет набор векторов входных воздействий, обеспечивающих выявление электрических и временны́х параметров элемента для всех условий распространения сигнала от входов к выходам.
3. Задает окружающую среду: напряжение питания, температуру, наклон входных импульсов, нагрузку выходов и настройки процесса моделирования — в соответствии с заданными пользователем ограничениями и характеристиками процесса моделирования.
4. Запускает и выполняет SPICE-подобное моделирование, суммирует результаты: задержки от входов к выходам, временны́е ограничения для входов, значения входных и выходных емкостей, мощности потребления.
5. Генерирует функционально-логическую модель (спецификацию) элемента, включающую задержки от входов к выходам, временные ограничения для входов, параметры выводов.
6. Генерирует файл результатов, который затем средствами САХИБ преобразуется в файл описания библиотеки элементов в стандартном формате Liberty [7].

Ввод исходных данных, запуск моделирования и трансляция полученных результатов поддерживаются графическим пользовательским интерфейсом в ОС Windows версии XP и старше. Для выполнения каждого этапа характеризации подключается соответствующая программа, интегрированная в программный комплекс САХИБ.

Входной контроль заключается в синтаксическом и семантическом анализе исходного описания характеризуемого элемента. В качестве исходного описания элемента используется следующая информация:

– спецификация элемента как «черного ящика»;
– принципиальная схема, включающая КМОП-транзисторы, паразитные резисторы и емкости, в виде списка цепей в SPICE-формате с указанием перечня внешних выводов.

Спецификация элемента, в свою очередь, содержит следующую информацию:

– имя библиотеки;
– имя элемента;
– тип логики (комбинационная, триггерная, псевдодинамическая, с тремя состояниями);
– тип вывода (вход, выход, вход–выход) и дисциплина (унарный, парафазный со спейсером, парафазный без спейсера, бифазный, инфазный);
– функциональное назначение входов и выходов (информационный, управляющий, индикаторный, предустановка).

Данная информация подготавливается пользователем системы САХИБ на основе функциональной характеристики элемента в составе библиотеки и его топологической реализации. Паразитные параметры (емкости и резисторы) принципиальной схемы элемента извлекаются из его топологической реализации с помощью стандартных средств имеющейся САПР.

Тестовые векторы используются для моделирования принципиальной схемы элемента. Каждый вектор описывает начальное состояние выходных узлов схемы и поведение всех ее входов в процессе одного сеанса моделирования. При формировании тестовых векторов учитывается транзисторная структура элемента: статическое состояние неактивных входов выбирается таким образом, чтобы обеспечить прохождение сигнала от анализируемого входа к выходу по критическому, с точки зрения задержки или энергопотребления, пути. В КМОП-логике критический путь определяется максимальным числом транзисторов, через которые протекает ток, приводящий к перезаряду внутренних и выходных (нагрузочных) паразитных емкостей в процессе переключения выхода элемента.

5 Результаты процедуры характеризации

Использование САПР при проектировании БИС предъявляет дополнительные требования как к полноте описания элементов библиотеки, так и к сопровождающим ее файлам технического описания элементов библиотеки. Эти файлы имеют стандартизованные форматы и используются системами логического моделирования и функционального и топологического синтеза разрабатываемой схемы. Они содержат всю необходимую информацию об элементах библиотеки от функционально-логических моделей с набором временны́х и электрических параметров до габаритных прямоугольников топологических реализаций элементов, используемых системами топологического синтеза.

Промышленным стандартом формата описания функционально-логических моделей библиотек стандартных элементов признан формат Liberty [7]. Он используется практически всеми автоматизированными системами функционального моделирования и топологического синтеза БИС. Поэтому автоматическое формирование файла описания моделей элементов в формате Liberty по результатам характеризации элементов библиотеки является также важной задачей.

Все характеризуемые модели элементов в формате Liberty объединяются в библиотеки. Конструкции формата Liberty позволяют описывать достаточно сложные и разнообразные элементы стандартных библиотек, в том числе и СС-элементы. Каждая библиотека состоит из двух частей — заголовка и ссылок на элементы библиотеки.

Заголовок библиотеки содержит информацию об условиях характеризации, единицах измерения параметров моделей, способах измерения сигналов и описание шаблонов таблиц, используемых при описании элементов. Информация

об условиях характеризации содержит данные о температуре и напряжении питания, при которых происходило снятие значений. Под способами измерения сигналов подразумеваются уровни напряжения, при которых снимаются те или иные временны́е и энергетические характеристики. Шаблоны таблиц, входящие в заголовок, описывают тип информации, содержащейся в таблице, размерность таблицы, а также характер значений, по которым происходит характеризация.

Описание элемента содержит сведения о площади его топологической реализации, о входных, внутренних и выходных портах, емкостные и временны́е характеристики портов, представленные в разд. 3. Завершает описание элемента указание значения статической мощности потребления.

Библиотека моделей создается для каждого варианта условий функционирования элементов и моделей КМОП-транзисторов, зависящих от геометрических размеров топологической реализации транзисторов и разброса технологических параметров процесса изготовления БИС.

С помощью разработанного программного комплекса САХИБ была проведена характеризация библиотеки IPI18 СС-элементов КМОП-базиса, созданной в проектных нормах 0,18 мкм, а также библиотеки IPI65 в проектных нормах 65 нм. Опытная эксплуатация системы САХИБ показала ее эффективность при характеризации библиотек СС-элементов.

6 Выводы

1. Комплекс САХИБ решает проблему построения корректных функционально-логических моделей СС-элементов.

2. Разработаны методики построения начального состояния элементов с памятью (триггеров, псевдодинамических элементов), имеющих несколько внутренних и внешних выходов. Это обеспечило реализуемость сеансов электрического моделирования, используемых для вычисления временны́х и электрических параметров характеризуемых элементов.

3. На основе алгоритмов и методик формализации описания принципиальной схемы элемента и формирования задания на электрическое моделирование, выявляющее временны́е и электрические параметры схемы, разработаны программные средства процедуры характеризации с использованием среды разработки Borland Delphi. Они обеспечивают синтаксический и семантический контроль входных данных, формирование задания на моделирование в формате SPICE, запуск задания на моделирование с использованием программы электрического моделирования Ngspice, расчет и аккумуляцию временных и электрических параметров элемента по его принципиальной схеме с паразитными компонентами.

4. Разработана подсистема трансляции результатов моделирования элемента в описание его функционально-логической модели в формате Liberty, используемом большинством современных промышленных САПР для синтеза схемотехники, топологии БИС и оценочного расчета ее временны́х и энергетических характеристик.
5. Разработана подсистема управления характеризацией, обеспечивающая интеграцию программных средств САХИБ в единую систему характеризации элементов КМОП-базиса и предоставляющая пользователю удобный графический интерфейс. Он в интуитивно понятной форме помогает пользователю сформировать задание на характеризацию библиотечного элемента, учитывающее условия его функционирования (наихудшие, типовые, наилучшие) с точки зрения напряжения питания, температуры, технологических параметров активных компонентов, индицирует процесс выполнения характеризации, информируя пользователя о степени завершенности отдельных этапов процесса характеризации и всего процесса в целом.
6. С помощью САХИБ была проведена характеризация библиотеки СС-элементов КМОП-базиса в проектных нормах 0,18 мкм и 65 нм.

Литература

1. Encounter Library Characterizer, Cadence. http://www.cadence.com/rl/Resources/datasheets/library_characterizer_ds.pdf.
2. CHARISMA: система характеризации библиотек стандартных ячеек // Radix Tools. http://www.radixtools.ru/files/Charisma_Presentation_RU.ppt.
3. *Антонов С. В., Чуйко Д. В.* Программа характеризации библиотеки цифровых КМОП-элементов. http://library.mephi.ru/data/scientific-sessions/2003/1/148.html.
4. Библиотека самосинхронных элементов и средств отладки рекуррентного компьютера: Отчет по НИР, шифр «БОРК». — М.: ИПИ РАН, 2010.
5. Ngspice Circuit Simulator. http://ngspice.sourceforge.net.
6. Разработка методики и программных средств характеризации элементов библиотеки САПР самосинхронных БИС: Отчет по НИР, шифр «СТЕРХ». — М.: ИПИ РАН, 2011.
7. Synopsys Inc., Liberty User Guides and Reference Manual Suite Version 2007.12. ftp://ftp.synopsys.com/pub/LUGRM07/Liberty_User_Guides_Reference_Manual_2007.12.pdf.

ФИКСАЦИЯ ИСКЛЮЧИТЕЛЬНЫХ СИТУАЦИЙ В РЕКУРРЕНТНОМ ОПЕРАЦИОННОМ УСТРОЙСТВЕ*

***Р. А. Зеленов*[1], *А. А. Прокофьев*[2], *Ю. А. Степченков*[3], *В. Н. Волчек*[4]**

Аннотация: Обозначены проблемы, связанные с фиксацией исключительных ситуаций (ИС) в многоядерной потоковой вычислительной системе. Анализируются возможные варианты обнаружения ИС и формирования отладочной информации для ее передачи на обработку. Сформулированы принципы создания логики фиксации и обработки ИС, применимые для любого вида вычислительных систем.

Ключевые слова: исключительные ситуации; прерывания; потоковая архитектура; рекуррентность

1 Введение

В Институте проблем информатики Российской академии наук ведется разработка многоядерной потоковой рекуррентной вычислительной системы (МПРВС) с нетрадиционной архитектурой, являющейся развитием потокового подхода. В основе парадигмы вычислений лежит графодинамическое представление алгоритмов, рекуррентно свернутых до момента инициации исполнения и саморазворачивающихся в ходе выполнения задач. В вычислительном процессе задействован единый поток самодостаточных данных, который хранит в себе рекуррентно сжатый алгоритм решения конкретной задачи [1].

Особенность МПРВС заключается в том, что в ней осуществляется взаимодействие двух вычислительных систем с различной архитектурой: управляющего уровня, базирующегося на стандартных фон-неймановских принципах, и операционного уровня, реализованного в виде рекуррентного операционного устройства (РОУ). Следовательно, необходимо решить следующие проблемы взаимодействия этих уровней:

(1) обмен данными (решена путем введения буферной памяти (БП) [2]);

(2) обработка ИС — реакция аппаратуры на ошибки и отладочные события.

*Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований ОНИТ РАН на 2011 г., проект 1.5.

[1]Институт проблем информатики Российской академии наук, graf.developer@gmail.com

[2]Институт проблем информатики Российской академии наук, a.a.prokofyev@mail.ru

[3]Институт проблем информатики Российской академии наук, YStepchenkov@ipiran.ru

[4]Институт проблем информатики Российской академии наук, v_volchek@inbox.ru

Обработка ИС является сложной и чрезвычайно актуальной задачей, которая включает в себя обнаружение ошибок, сбор отладочной информации, выполнение функций обработки возникших ситуаций, а также внедрение поддержки отладочных процедур в вычислительную систему. В данной статье представлен анализ проблем реализации функций фиксации и обработки ИС в РОУ, а также возможных вариантов их решения.

2 Особенности реализации многоядерной потоковой рекуррентной вычислительной системы

Как упоминалось выше, архитектура МПРВС является двухуровневой (рис. 1): в качестве управляющего выступает фон-неймановский процессор, в его задачи входит подготовка данных для операционного уровня и связь с периферийными устройствами; РОУ предназначено для эффективного исполнения параллельных алгоритмов.

В качестве программного обеспечения (ПО) для РОУ выступают капсулы [2]. Это программы, представляющие собой набор элементов самодостаточных данных (операндов), — инструкции по настройке определенных блоков, данные и команды для их обработки.

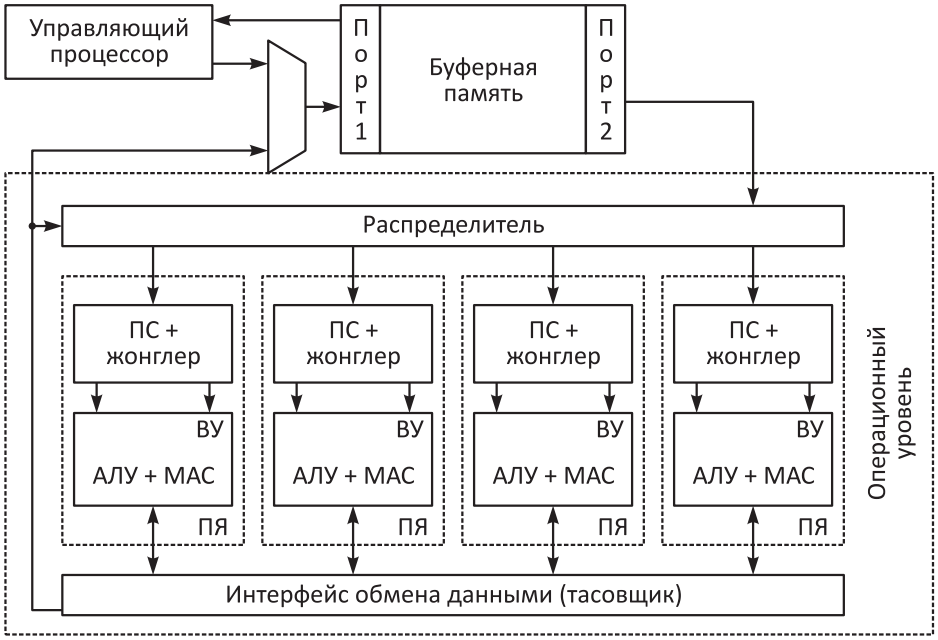


Рис. 1 Структура МПРВС

Взаимодействие двух уровней осуществляется через БП. С целью сокращения объема данных, поступающих со стороны управляющего уровня, перед инициализацией вычислительного процесса в БП записываются шаблоны капсул, содержащие только функциональные поля операндов, в процессе же вычислений идет заполнение лишь содержательных частей (непосредственно данных). Таким образом, в процессе многократного выполнения капсул вся информация о ходе вычислений остается неизменной, меняются только обрабатываемые данные [3].

В текущей реализации операционное устройство (ОУ) имеет в своем составе четыре процессорных ядра (ПЯ). Распределение входного потока самодостаточных данных по секциям осуществляется функциональным блоком (ФБ) — распределителем. Поступающие на вход ПЯ данные сначала находят свою пару в памяти совпадений (ПС), затем в зависимости от режима работы и кодов операций жонглер определяет, на какое плечо вычислительного устройства (ВУ) подать операнды для их дальнейшей обработки [4].

3 Фиксация исключительных ситуаций в рекуррентном операционном устройстве

Как и в любой вычислительной системе, в МПРВС возможно возникновение ошибок, связанных с различного рода ошибками и сбоями, которые невозможно предусмотреть на этапе программирования [5, 6]. Подобные события неизбежно приводят к нарушению нормального хода вычислительного процесса, поэтому встает задача их обнаружения, обработки, восстановления корректности вычислений. Представим структуру функции обработки ИС в виде дерева (рис. 2).

Двухуровневая организация МПРВС накладывает специфику на последовательность действий при работе с ошибками и служебными процедурами. Требуется разделение функций, представленных на рис. 2, между управляющим уровнем (УУ) и РОУ.

Внедрение логики фиксации исключительных ситуаций требует применения аналитического подхода с оценкой различных возможных вариантов исполнения, так как необходимо удовлетворять критериям:

- полноты покрытия ИС: должны обнаруживаться все возможные ИС, которые могут возникнуть в ходе работы РОУ, в том числе и вследствие сбоев в аппаратуре;
- полноты отладочной информации, передаваемой на УУ: в зависимости от места и времени возникновения ИС управляющий уровень будет производить обработку разными способами;
- минимизации аппаратных затрат на реализацию в устройстве;
- раннего обнаружения ИС: если ошибка, приводящая к ИС, может быть обнаружена в нескольких местах, то ее необходимо обнаруживать на как

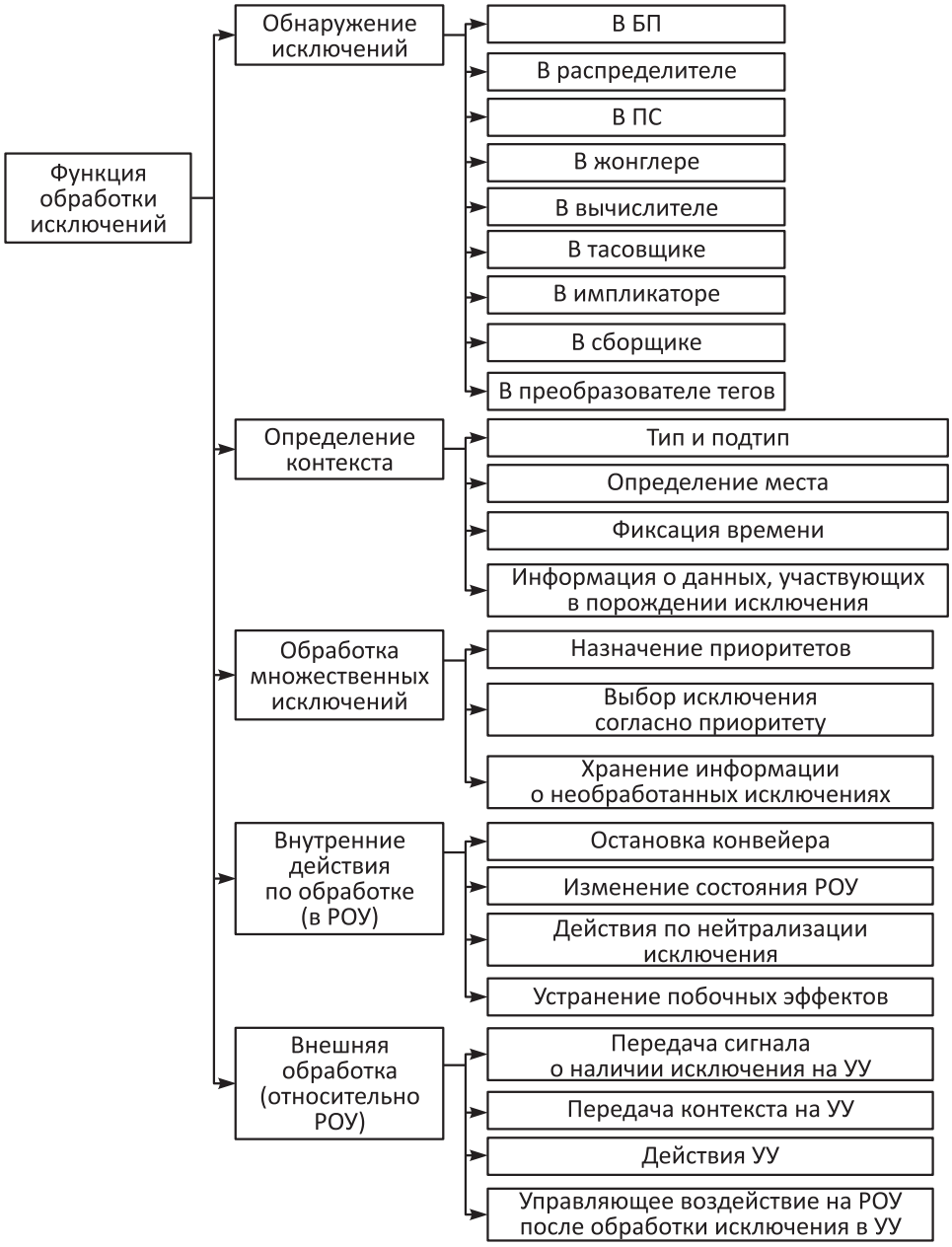


Рис. 2 Структура функции обработки ИС

можно более ранних этапах вычислительного процесса. Это позволит точнее определить причину ее возникновения.

Для реализации фиксации ИС требуется найти ответы на следующие вопросы:

1. Каков перечень ИС, их классификация и приоритеты?
2. Какую информацию о возникшей ИС требуется сохранить?
3. Что должно произойти с вычислительным процессом в РОУ при возникновении ИС?
4. Каким образом УУ получит информацию об ИС?
5. Каковы возможные варианты обработки ИС?

3.1 Перечень исключительных ситуаций

Для составления перечня ИС проанализируем работу каждого ФБ в отдельности с точки зрения спецификации и установим, какие ошибки возможны в ходе вычислительного процесса. Каждый из ФБ работает под управлением данных, поступающих на его вход. В этих данных могут передаваться коды настройки; следовательно, если код настройки не специфицирован, то поведение ФБ не определено и такая ситуация должна вызывать ИС. Подобного рода ИС можно отнести к типу «ошибка по допустимости». Однако будем последовательны и начнем анализ возможных ошибок в функциональных блоках с интерфейса взаимодействия РОУ с управляющим уровнем — буферной памяти.

Анализ структуры и функционирования БП выявил следующие возможные нарушения в корректности работы:

– попытка обращения по несуществующему адресу БП (тип «ошибка по допустимости», так как адрес превышает допустимое адресное пространство):
  - со стороны УУ — неверное значение на шине адреса;
  - через индексные регистры БП — выход за границы адресного пространства при считывании и записи данных;
– запись данных со стороны УУ в непредназначенное для них место: попытка записать функциональную часть операнда в область памяти, предназначенную для хранения содержательной части, и наоборот («ошибка по назначению»);
– отсутствие обязательного конфигурационного операнда по адресу, соответствующему началу капсулы («ошибка по отсутствию»).

Такие события могут происходить, если возник сбой в работе при записи операндов капсулы в БП управляющим уровнем.

После БП операнды попадают в распределитель. Для обеспечения раннего обнаружения логично именно на входе распределителя проверять на соответствие спецификации значения функциональных полей операндов («ошибка по допустимости»). Но данный анализ не может гарантировать полного отсутствия ошибок в функциональных полях, так как вследствие определенных причин значения полей будут специфицированными, но отличными от задуманных программистом, что приведет к ошибкам в ходе вычислительного процесса.

Кроме того, в распределителе могут возникать еще два типа нарушений корректности работы:

(1) по причине неправильного программирования или сбоя возможны ситуации, когда несколько операндов будут направлены на один и тот же путь передачи данных одновременно («ошибка по коллизии»);

(2) вследствие неверных настроек рассылки операндов они не могут быть корректно направлены в места своего назначения («ошибка по репликации»).

Других типов ИС в распределителе не предусмотрено, так как остальные варианты его работы специфицированы и могут корректно обрабатываться.

При определенной настройке жонглера нужно фиксировать ИС, а именно при обнаружении несовместимости в значениях отдельных функциональных полей сцепившихся операндов, когда:

– невозможно определить, нужно ли направлять результирующий операнд на шину обмена данными с распределителем и БП («ошибка по экспозиции»);

– коды операций несовместимы («ошибка по операции»);

– невозможно определить, нужна пересылка результирующего операнда вообще, а если да, то в свою или соседнюю секцию («ошибка по пересылке»).

В ПС может быть зафиксирована ИС, связанная с обращением к несуществующему адресу («ошибка по допустимости»).

Вычислитель выполняет операции над данными и, подобно большинству процессоров, отслеживает сопутствующие операциям события и характеристики получаемых результатов (нулевое значение результата, переносы, переполнения, вхождения в циклы и т. д.). Проведя их анализ, можно прийти к выводу, что только факт переполнения может негативно повлиять на ход вычислительного процесса, а это означает, что при его наступлении должна вырабатываться ИС — «ошибка по переполнению». Переполнения возможны в трех местах вычислителя: арифметико-логическом устройстве (АЛУ), регистре B, регистре C.

Результат вычислений должен через интерфейс обмена данными попадать на межсекционные шины; следовательно, здесь возможна «ошибка по коллизии».

В остальных узлах РОУ фиксации ИС не требуется, так как их функционирование не допускает появления ИС.

3.2 Классификация исключений

Получив перечень ИС, необходимо провести их классификацию, так как это позволит разработать стратегию по их фиксации и дальнейшей обработке. На рис. 3 приведена схема классификации ИС по трем основным признакам:

1. **Критичность** означает возможность или невозможность дальнейшего продолжения вычислительного процесса. Критичные ИС должны немедленно генерировать прерывания на управляющий уровень.

   В процессе отладки может потребоваться переход в пошаговое исполнение при возникновении определенных условий, в то время как в процессе вычислений достаточно только зафиксировать факт возникновения этих условий без генерации прерываний, поэтому некритичные ИС могут иметь возможность настройки своего поведения:

   – логирование информации об ИС;

   – генерация прерываний.

2. **Приоритет** позволит определить очередность и возможность обработки ИС в случае, когда в различных узлах устройства одновременно возникают ошибки. При этом необходимо сохранить информацию о наиболее важных из них. Приоритет может обозначаться числом (увеличение числа соответствует уменьшению приоритета, табл. 1), и при одновременном возникновении обработке будут подлежать ИС с большим приоритетом.

3. **По назначению** ИС делятся на два вида:

   (а) использующиеся постоянно для фиксации ошибок;

   (б) определяемые пользователем для целей отладки.

Перечень типов ИС и их классификация представлены в табл. 1.

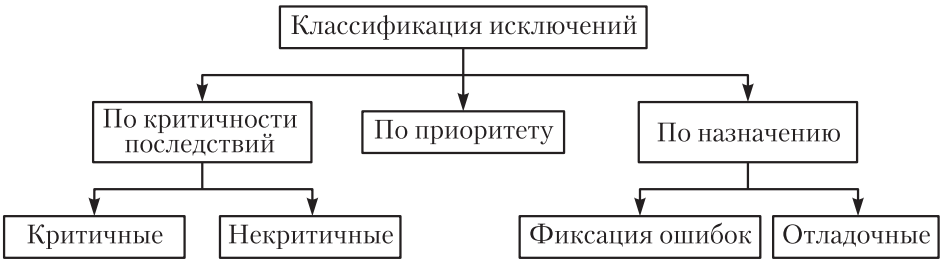


Рис. 3 Классификация ИС

Таблица 1 Перечень и классификация ИС

Тип ИС	Место возникновения	Классификация				Приоритет
		Критичность		Назначение		
		Критичная	Некритичная	Фиксация ошибок	Отладка	
Ошибка по допустимости	БП	+		+		1
	Распределитель	+		+		2
	ПС	+		+		3
Ошибка по отсутствию	БП	+		+		4
Ошибка по назначению	БП	+		+		5
Ошибка по коллизии	Распределитель	+		+		6
	Тасовщик	+		+		7
Ошибка по экспозиции	Жонглер	+		+		8
Ошибка по операции	Жонглер	+		+		9
Ошибка по пересылке	Жонглер	+		+		10
Ошибка по репликации	Распределитель		+	+	+	11
Ошибка по переполнению	Вычислитель: АЛУ		+	+	+	12
	Вычислитель: регистр B		+	+	+	13
	Вычислитель: регистр C		+	+	+	14
	БОИС		+	+	+	15
Пользовательская ИС (отладочная)	Определяется разработчиком		+	+	+	16

3.3 Отладочная информация об исключительной ситуации

Основной проблемой при фиксации ИС является сбор и сохранение информации о возникшем событии. Сохраняемая об ИС информация должна обеспечивать возможность определить:

(1) где возникла ИС (ФБ и при необходимости номер ПЯ);

(2) тип ИС;

(3) время возникновения ИС: вершина в графе вычислений.

Решение первых двух проблем достаточно очевидно: ввести регистры для сохранения кода ФБ, номера ПЯ и кода типа ИС. Сохранение же временно́й характеристики требует анализа структуры РОУ. Исходя из особенностей реализации БП, можно сделать вывод, что состояние индексных и ряда других управляющих регистров точно определяет, в каком месте графа алгоритма находится процесс вычисления. Но при этом остается неизвестен номер цикла

исполнения алгоритма, поэтому необходимо ввести соответствующий счетчик. Таким образом, путем объединения информации из регистров БП и счетчика циклов формируется временной параметр.

3.4 Функционирование блока обработки исключительных ситуаций в рекуррентном опреционном устройстве

Из вышеприведенного анализа следует, что в каждом из ФБ необходимо реализовать проверку условий возникновения ИС и разработать ФБ, отвечающий за сбор и хранение информации о возникших ИС (блок обработки исключительных ситуаций, БОИС) и обеспечивающий связь с УУ.

Из предложенной классификации ИС следует, что критичные ошибки должны приводить к немедленному завершению вычислений и передаче управления на УУ. При возникновении же некритичных ИС возможно продолжение работы с логированием информации о возникшем событии. Кроме того, любая ИС может использоваться в отладочных целях, например останавливать конвейер и переводить РОУ в режим пошагового исполнения. В связи с этим целесообразно вводить следующие режимы работы БОИС:

– при возникновении некритичных ИС:
  - игнорировать их возникновение;
  - логировать ИС без изменения хода вычислительного процесса;
  - активировать режим отладки;
– при возникновении критичных ИС:
  - остановить конвейер и генерировать прерывание;
  - активировать режим отладки.

Для минимизации промежутка времени между фиксацией ИС и переходом в режим отладки принято решение расширить функциональность БОИС способностью выполнять отладочные процедуры (например, пошаговую отладку, чтение/запись памяти и внутренних регистров), контролируемые УУ или пользователем.

Логирование информации требует введения памяти в БОИС, размер которой будет зависеть от преследуемых разработчиком целей. На ранних этапах необходимо накапливать как можно больше отладочной информации, тогда как в будущем, когда МПРВС пройдет приемо-сдаточные испытания, размер может быть уменьшен. Факт переполнения данной памяти может генерировать ИС «ошибка по переполнению».

Проблема возникает при поступлении на входы БОИС сведений о нескольких одновременно возникших ИС. Она связана с необходимостью либо увеличивать

аппаратные затраты на реализацию данного блока, либо увеличивать время его работы. В этом случае целесообразно введение двух вариантов работы БОИС:

(1) при наличии хотя бы одной критичной ИС обработке подвергается наиболее приоритетная;

(2) когда все ИС не являются критичными и режим работы предполагает логирование информации, БОИС остановит конвейер и в течение нескольких тактов запишет информацию обо всех ИС в свою память.

3.5 Взаимодействие с управляющим уровнем при возникновении исключительных ситуаций

Рекуррентное операционное устройство может обнаруживать появления ИС, но одного факта фиксации недостаточно — встает вопрос о передаче управления и информации об ИС на УУ. В первую очередь УУ должен узнать о том, что произошла ошибка, а затем, после некоторых процедур сохранения своего состояния, получить информацию об ИС и перейти к ее обработке.

Для получения УУ информации о факте ошибки существует несколько вариантов реализации:

(1) УУ с некоторой периодичностью опрашивает РОУ.

(2) интеграция с системой прерываний процессора, под управлением которого функционирует УУ.

Последний вариант является более предпочтительным, так как позволяет сократить расход вычислительных ресурсов и осуществить практически моментальную передачу информации о факте возникновения ИС на УУ.

После того как УУ оповещен об ошибке, он может либо сразу, либо после завершения текущей задачи считать подробную информацию об ИС из внутренних регистров РОУ, входящих в состав БОИС.

3.6 Обработка исключительных ситуаций на управляющем уровне

Управляющий уровень после получения информации об ИС должен корректно на нее отреагировать. Можно выделить два режима работы системы в целом: рабочий и режим отладки.

В рабочем режиме необходимо либо останавливать вычислительный процесс и выдавать пользователю сообщение об ошибке, либо, если это возможно, продолжать работу, фиксируя сведения о возникших ИС во внутренней памяти устройства. Выбор вариантов действий должен осуществляться через параметры настройки устройства.

В режиме отладки требуется выдавать сообщение об ошибке, останавливать вычислительный процесс и переходить к процедуре отладки.

По завершении обработки ИС необходимо восстановить вычислительный процесс, что невозможно без изменения внутреннего состояния РОУ. Перед тем как запустить на выполнение новый цикл вычислений, требуется провести работу по очистке конвейера: установить требующиеся значения флагов, управляющих регистров и памяти.

4 Среда разработки программного обеспечения для рекуррентного операционного устройства

При создании ПО для РОУ (капсул) разработчик должен учитывать множество нюансов, связанных с особенностями работы как всего РОУ в целом, так и его отдельных ФБ. Очевидно, что человеческий фактор (невнимательность, невозможность держать в памяти огромное количество информации из спецификации) может стать потенциальным источником ошибок в капсулах. Это означает, что для их разработки необходимо использовать средство, которое будет сводить к минимуму вероятность возникновения такого рода ошибок. Для решения этой проблемы была разработана система капсульного программирования и отладки (СКАТ) [2], которую можно отнести к такому классу ПО, как интегрированные среды разработки.

Главными задачами СКАТ являются: обеспечение разработчиков капсул инструментами для программирования и отладки; обнаружение возможных ошибок на как можно более раннем этапе разработки. К сожалению, во время создания капсулы невозможно определить все ошибки, которые могут возникнуть в процессе исполнения капсулы.

Процесс разработки ПО для РОУ итерационный, и его можно разделить на три повторяющихся этапа: программирование капсулы, исполнение ее на модели РОУ и отладка с помощью встроенных средств.

4.1 Использование системы капсульного программирования и отладки для работы с исключительными ситуациями

Основная задача СКАТ состоит в помощи разработчику при формировании капсулы, которая будет содержать минимальное количество ошибок. Ошибки синтаксиса при использовании СКАТ исключены, так как используется визуальный конструктор капсул [2].

Обнаружение большинства ИС, перечень которых приведен выше, возлагать на СКАТ не имеет смысла. Во-первых, для некоторых из них это невозможно по причине необходимости запуска вычислительно процесса. Во-вторых, любые из ИС могут возникнуть в результате ошибок или аппаратного сбоя во время перемещений капсулы между УУ и РОУ, поэтому их в любом случае нужно проверять в аппаратуре.

Определенный семантический анализ капсул в СКАТ внедрить все же необходимо по причине сложности программирования и невозможности учета всех нюансов вычислительной системы. Только с помощью СКАТ могут быть выявлены следующие ошибки:

– выход за рамки допустимой избыточности в процессе развертывания алгоритма;

– нарушение структуры капсулы:

  - отсутствие обязательных операндов;
  - взаимная противоречивость отдельных функциональных полей в операндах.

Вследствие использования в СКАТ VHDL-модели РОУ для процесса тестирования и отладки капсул, представляется возможным возложить на СКАТ некоторые функции УУ, а именно:

– получение от модели РОУ информации о возникновении ИС;

– выдачу пользователю информации, которая сможет помочь ему в устранении ошибки, вызвавшей ИС.

Информация об ИС может быть показана разработчику не только в виде состояния внутренних регистров блока фиксации ИС. На основе этих данных СКАТ может в некоторых случаях автоматически определить, какой операнд содержит в себе ошибку.

Таким образом, в скором времени планируется дополнить СКАТ следующими компонентами:

– семантическим анализатором капсул, который позволит обнаруживать некоторые логические ошибки в капсулах;

– компонентом анализа ИС, который будет по результатам анализа отчета работы модели РОУ сообщать пользователю информацию о возникших ИС и, где возможно, указывать на причину их возникновения.

Внедрение данного функционала позволит не только снизить количество возможных ошибок в капсулах, но также в некоторой степени автоматизировать процесс отладки, что является актуальной задачей для сред разработки ПО низкого уровня абстракции.

5 Заключение

В данной статье проведен анализ проблем, возникающих при разработке логики фиксации и обработки ИС в РОУ, и рассмотрены варианты их решения.

Принципы создания подобной функциональности, которые могут быть использованы практически для любой вычислительной системы, можно представить в виде последовательности шагов:

(1) составление перечня возможных ИС на основе анализа функционирования каждого из блоков вычислительной системы;

(2) классификация ИС, которая необходима для выработки правил их обработки;

(3) определение минимально необходимого набора информации, который позволит эффективно проводить процедуры отладки;

(4) определение возможных вариантов работы с ИС, и в случае если их фиксация и обработка возложены на разные аппаратные модули, то разработка протокола их взаимодействия;

(5) внедрение аппаратной поддержки обнаружения и обработки ИС в вычислительную систему.

В настоящее время ведется апробация вышеизложенных вариантов решения проблем фиксации и обработки ИС для РОУ.

Литература

1. *Степченков Ю. А., Петрухин В. С.* Особенности гибридного варианта реализации на ПЛИС рекуррентного обработчика сигналов // Системы и средства информатики. Доп. вып. — М.: ИПИ РАН, 2008. С. 118–129.
2. *Зеленов Р. А., Степченков Ю. А., Волчек В. Н., Хилько Д. В., Шнейдер А. Ю., Прокофьев А. А.* Система капсульного программирования и отладки // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1.С. 24–30.
3. *Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Механизмы обеспечения поддержки алгоритмов цифровой обработки речевых сигналов в рекуррентном обработчике сигналов // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1.С. 30–46.
4. *Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Цифровой сигнальный процессор с нетрадиционной рекуррентной потоковой архитектурой // Проблемы разработки перспективных микро- и наноэлектронных систем — 2010: Сборник трудов. — М.: ИППМ РАН, 2010. 694 с.
5. *Селлерс Ф.* Методы обнаружения ошибок в работе ЭЦВМ / Пер. с англ. Ф. Селлерс. — М.: Мир, 1972. 310 с.
6. *Клингман Э.* Проектирование микропроцессорных систем / Пер. с англ. В. А. Балыбердина, В. А. Зинченко. — М.: Мир, 1980. 576 с.

ИЕРАРХИЧЕСКИЙ МЕТОД АНАЛИЗА САМОСИНХРОННЫХ ЭЛЕКТРОННЫХ СХЕМ*

Л. П. Плеханов[1]

Аннотация: Развитию и внедрению самосинхронных схем (СС), обладающих уникальными свойствами, во многом препятствуют трудности проектирования, в частности анализ на самосинхронность «больших» схем. Предлагается иерархический метод анализа схем неограниченного размера, основанный на функциональном подходе. В литературе подобного подхода и метода не отмечено.

Ключевые слова: самосинхронные схемы; асинхронные схемы; анализ самосинхронности

1 Введение

Самосинхронные схемы обладают уникальными свойствами, недостижимыми в реализации других типов схем, синхронных или асинхронных. К ним относятся независимость поведения от задержек элементов, полное отсутствие состязаний, отказобезопасность, правильность функционирования в максимально широком диапазоне внешних условий (температуры и напряжения питания) и некоторые другие [1–3].

Однако СС-схемы пока не получили широкого распространения по ряду причин, в частности индустриальной инерции. Другой важнейшей причиной является трудность проектирования таких схем. Для обеспечения свойства самосинхронности схемы необходимо тем или иным способом вычислить и проверить все возможные состояния, в которые попадает схема в реальной работе, а также все возможные переходы между этими состояниями.

Главной и неизбежной частью проектирования СС-схем является их анализ на самосинхронность. Поэтому на практике наличие средств анализа СС-схем любого размера есть необходимое условие полноценного проектирования таких схем.

Основные существующие методы проектирования СС-схем основаны на представлении поведения схем в форме переключений сигналов — событий. Такие

*Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований ОНИТ РАН на 2012 г., проект 1.5.

[1]Институт проблем информатики Российской академии наук, LPlekhanov@ipiran.ru

методы далее будут называться *событийными*. Это метод диаграмм переходов (ДП), восходящий к Маллеру [4], и метод диаграмм изменений (ДИ), предложенный группой В. И. Варшавского [5].

Другой подход — *функциональный* — основан на анализе поведения схем в представлении логическими функциями. Анализ СС-схем при таком подходе предложен в [6], а сам подход более подробно представлен в [7].

Следует отметить, что в силу отмеченной выше объективной необходимости проверки с исчерпывающей полнотой состояний и переходов никакие методы анализа, основанные на исследовании полной системы уравнений схемы (учета всех ее элементов), не позволяют неограниченно увеличивать размер анализируемой схемы. Причина — в экспоненциальном росте вычислительной сложности с увеличением числа информационных входов схемы и переменных памяти (во всех случаях) либо еще и с увеличением числа уравнений схемы (в случае ДП).

Перспектива анализа СС-схем, следовательно, лежит в разработке иерархических методов, дающих возможность корректно сокращать исследуемые сущности (события, уравнения и т. д.).

Один из возможных подходов к такому анализу в событийных представлениях можно найти в [8]. В нем предлагается использовать ДИ, ранее построенные для блоков нижнего уровня, в довольно сложном взаимодействии с ДИ анализируемой схемы. Материал изложен как возможный путь (в форме доказательств теорем), но не разработан как практический метод.

Попытка предложить иерархический анализ сделана в [9, с. 6]. Помимо того, что обсуждение и обоснование иерархического анализа не соответствует названию статьи, приведенные в ней выкладки не обоснованы. Предлагается сокращать схему путем замены блоков на «макроэлементы», имеющие только фазовые сигналы, и анализировать на полумодулярность (независимость от задержек элементов) результирующую модель схемы. Заявляется, что свойство полумодулярности исходной схемы совпадает с полумодулярностью результирующей модели. Однако и схема, и модель описываются разными системами уравнений от разных переменных. Эти системы в статье не приводятся, и связь между их переменными, уравнениями и полумодулярностью не установлена. Далее, основные источники состязаний (нарушающие полумодулярность) — информационные бистабильные сигналы — в макроэлементы не попадают, а одни фазовые сигналы не отражают всей динамики взаимодействия блоков. Если, например, на информационных входах триггера [9, рис. 2] в составе исходной схемы есть состязания, то при переходе к фазовому макроэлементу они исчезают, т. е. не учитываются.

Функциональный подход использует другие принципы и, как заявлено в публикациях [6, 7], дает возможность иерархического анализа. Конкретным способам его достижения и посвящена настоящая статья.

2 Метод анализа

Аналогов предлагаемого (функционального) подхода и метода в литературе не найдено.

К особенностям СС-схем, определяющим как их свойства, так и способы проектирования, следует отнести специальное — самосинхронное — кодирование информации и двухфазный режим работы [1].

Любая СС-схема в конечном применении должна быть замкнутой (с помощью общей обратной связи) и самогенерирующей. Схема поэтому автоматически переходит поочередно из одной фазы (стадии) в другую, что необходимо для обеспечения самосинхронности. Фазы носят название рабочей и спейсера (промежуточной).

В событийных методах анализа схемы рассматриваются именно как замкнутые (система уравнений, соответственно, также замкнута).

Однако для целей анализа можно исследовать и разомкнутые схемы [3], что и реализуется при функциональном подходе. Соответствующий функциональный метод анализа (ФМА), использующий полную систему уравнений схемы, описан в [6]. Как одна из главных задач ФМА (помимо анализа самосинхронности) изначально ставилась задача определения параметров внешних связей схемы, требуемых для следующего верхнего уровня иерархии анализируемой схемы.

В рамках функционального подхода можно естественно подойти к иерархическому анализу: схемы в окончательном виде создаются иерархически, снизу вверх, они на этапе проектирования разомкнуты и имеют понятный для разработчика интерфейс.

Существо предлагаемого иерархического метода анализа (ИМА) состоит в том, что анализируемая схема не раскрывается до элементов (уравнений), как это требуется в существующих «полных» методах. Схема должна состоять из фрагментов (блоков), заранее прошедших ФМА. На этапе ИМА проверяются только внешние описания (интерфейсы) фрагментов схемы и соединения фрагментов.

Вычислительная сложность ИМА практически линейна по числу сигналов и фрагментов, что позволяет анализировать «поэтажно», снизу вверх, схемы любого размера.

Таким образом, общий порядок анализа СС-схем должен состоять из двух этапов: анализа схем нижнего уровня по полным уравнениям (ФМА) и иерархического анализа на всех более высоких уровнях (ИМА).

Для эффективного прохождения наиболее трудоемкого этапа — ФМА — размер анализируемых схем на нем можно сделать достаточно малым, вплоть до отдельного триггера и небольшой комбинационной ячейки. Создание небольших СС-схем в настоящий момент уже хорошо отработано.

Далее будет рассматриваться иерархический метод.

2.1 Представление разомкнутых самосинхронных схем

В событийных методах схемы представляются замкнутыми, не имеющими входов и выходов. На практике разрабатываемые схемы всегда имеют входы и выходы, но при переходе к событийному анализу требуется дополнять описание каждой схемы специальным замыканием (не всегда простым), в результате чего входы и выходы «исчезают».

Для разомкнутой СС-схемы необходимо явно обозначить ее интерфейс — внешние входы и выходы, а также снабдить их определенными атрибутами — типами и свойствами, специфичными для самосинхронности. Эти атрибуты будут отражать особенности взаимодействия внешних сигналов как с окружением, так и с внутренними элементами.

Типы внешних сигналов (*СС-типы*) можно разделить на три группы: информационные, контрольные и вспомогательные.

К информационным сигналам относятся парафазные со спейсером (*ПФС-сигналы*) и бистабильные — входы и выходы бистабильных ячеек (БСЯ) (*БС-сигналы*). Контрольные сигналы предназначены для организации переходов из одной фазы в другую. К ним относятся управляющие на входе (*У-сигналы*) и индикаторные на выходе *(И-сигналы*): У-сигналы инициируют переключения из одной фазы в другую, И-сигналы показывают завершение перехода схемы в текущую фазу.

Другие разновидности сигналов, встречающиеся иногда в схемах, — мультифазные, мультистабильные, унарные информационные — учитываются аналогично вышеприведенным и для простоты рассматриваться не будут.

Вспомогательные сигналы предназначены для вспомогательных целей (например, режимные или асинхронной предустановки) и далее также учитываться не будут.

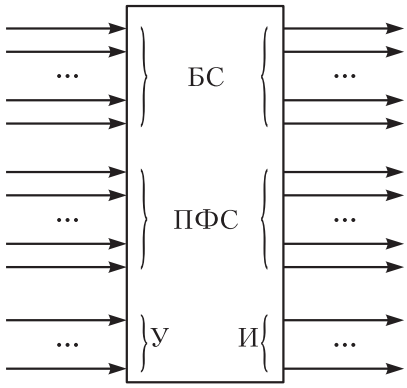


Рис. 1 Типовой интерфейс разомкнутой СС-схемы

Контрольные и ПФС-сигналы специфичны для фазы работы и вместе будут называться *фазовыми* сигналами. И на входе, и на выходе схемы должно присутствовать хотя бы по одному фазовому сигналу.

Описанный интерфейс будем называть типовым интерфейсом разомкнутой схемы (рис. 1).

Схема должна состоять из фрагментов, заранее прошедших через ФМА (СС-фрагментов). Исключения делаются для инверторов и повторителей, для которых анализ не требуется.

Схема представляется в виде списка имен фрагментов и соединений между ними, как это делается в языках описания аппаратуры (типа VHDL — VHSIC (very high-speed integrated circuits) Hardware Description Language).

Отметим некоторые особенности поведения СС-схем. В соответствии с двухфазным режимом работы любой сигнал схемы в каждой из фаз может измениться не более одного раза. А каждая пара БС-сигналов, входных или выходных, может меняться только в одной из фаз. Такая фаза называется *транзитной* для пары. При этом для входных БС-сигналов транзитная фаза — это фаза, когда их изменения разрешены по условиям самосинхронности, для выходных — когда изменения могут происходить фактически.

2.2 Исходные данные и задачи иерархического метода анализа

Предметом анализа будет разомкнутая схема (далее именуемая *главной* схемой), имеющая типовой интерфейс и состоящая из СС-фрагментов.

В интерфейсе главной схемы должны быть указаны:

(1) группы входных и выходных сигналов и их СС-типы согласно типовому интерфейсу;

(2) для входных фазовых сигналов — значения спейсеров, 0 или 1.

В описаниях каждого СС-фрагмента должны присутствовать следующие атрибуты интерфейса:

(1) группы входных и выходных сигналов и их СС-тип согласно типовому интерфейсу;

(2) для входных и выходных фазовых сигналов — значения спейсеров.

(3) для выходных фазовых сигналов — списки индицируемых ими входов и выходов фрагмента;

(4) информация, связанная с дисциплиной (порядком изменений) БС- и фазовых сигналов. Эта информация нужна для определения состязаний в главной схеме и будет объяснена в п. 2.5.

Как доказано в [1, с. 118] и более подробно описано в [3], для анализа самосинхронности разомкнутой схемы необходимо и достаточно проверить ее (а) на индицируемость и (б) на отсутствие состязаний. Кроме того, как и в ФМА, должна быть получена интерфейсная информация для следующего верхнего уровня иерархии. С учетом сказанного, задачи, которые необходимо решать в ИМА, следующие:

(1) проверка правильности соединений фрагментов;

(2) проверка индицируемости сигналов;

(3) проверка соблюдения дисциплины БС-сигналов;

(4) получение параметров интерфейса главной схемы.

2.3 Проверка правильности соединений фрагментов

В данной задаче проверяются соединения фрагментов с точки зрения самосинхронности. И-сигналы соединяются с У-сигналами, БС — с БС-сигналами, ПФС — с ПФС-сигналами.

Необходимо проверять отсутствие разбаланса фаз на входах каждого фрагмента, т. е. следить за тем, чтобы значения фазовых сигналов, подключенных к одному фрагменту, соответствовали одной и той же фазе.

Транзитные фазы соединяемых БС-сигналов должны быть согласованы со значениями спейсеров сопровождающих их фазовых сигналов.

Также необходимо следить, чтобы в цепях БС-сигналов не оказалось никаких дополнительных элементов — инверторов, повторителей или др., — так как известно, что это приводит к нарушению самосинхронности.

2.4 Проверка индицируемости сигналов

На этом шаге определяется, какие внутренние и внешние сигналы главной схемы индицируются на ее фазовых выходах. Вспомогательные сигналы не учитываются. Носителями информации об индикации выступают фазовые сигналы.

Каждому фазовому сигналу сопоставляется список индицируемых им сигналов. Изначально и сам сигнал записывается в этот список.

При проверке существенно используется свойство транзитивности индицируемости [1].

Проверка производится от входов схемы к выходам. По ходу проверки каждый фрагмент получает на свои фазовые входы списки индицирования. По полученным спискам и параметрам интерфейса (спискам индицируемых фрагментом своих входов и выходов) путем объединения формируются списки выходных фазовых сигналов фрагмента. Процесс заканчивается на фазовых выходах схемы.

Обязательным для успешной проверки является присутствие всех внутренних сигналов схемы хотя бы в одном из выходных списков индицирования. Присутствие внешних фазовых сигналов главной схемы не обязательно, так как они могут быть индицированы на верхнем уровне иерархии.

2.5 Проверка соблюдения дисциплины бистабильных сигналов

Данный шаг предназначен для обнаружения состязаний в главной схеме. При условии успешного прохождения предыдущих шагов анализа источником состязаний может стать несогласованность переключений БС-сигналов фрагментов и связанных с ними фазовых сигналов. Согласованная последовательность

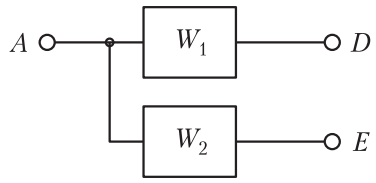


Рис. 2 К понятию очередности переключений

их переключений, обеспечивающая отсутствие состязаний, носит название *дисциплины*.

Предварительно введем понятие очередности изменений сигналов. При исследовании СС-схем из-за произвольности задержек вместо понятия времени имеет смысл использовать понятие очередности (рис. 2).

На рис. 2 W_1 и W_2 — цепочки последовательных элементов, имеющие не менее одного элемента.

По теории задержки элементов хотя и произвольны, но не равны нулю. Поэтому справедливо утверждение, что изменения сигналов D и E всегда будут происходить после изменения сигнала A.

Назовем сигнал A *инициатором* для сигналов D и E, а сигналы D и E — *континуаторами* (продолжателями) сигнала A. По смыслу СС-схем континуаторами будут фазовые сигналы: одиночные или пары ПФС-сигналов.

Таким образом, изменение инициатора всегда будет предшествовать изменению любого его континуатора. Очевидно и свойство транзитивности: континуатор континуатора будет континуатором их общего инициатора.

Будем называть также *конкурентными* сигналы, очередность изменений которых по отношению друг к другу произвольна. Конкурентными могут быть как сигналы, имеющие общий инициатор, например D и E, так и сигналы с разными инициаторами (независимые).

Отметим, что индикаторные сигналы всегда являются континуаторами тех сигналов, которые они индицируют.

Рассмотрим схематично пример соединения БС-сигналов двух фрагментов на рис. 3, где показана выходная БСЯ одного фрагмента и входная другого.

Учитывая особенности поведения СС-схем, основное правило дисциплины БС-сигналов при соединении фрагментов состоит в следующем.

В период времени, когда выходные БС-сигналы (Y_1, Y_2) могут меняться, сигнал B должен запретить изменение входной БСЯ (заблокировать ее входы).

Рассмотрим последовательность изменений сигналов в соединении.

Выходные БС-сигналы меняются в одной из фаз — транзитной. В этой фазе сигнал B блокирует подключенные БС-входы. В следующей — нетранзитной — фазе сигнал B открывает БС-входы, и фаза заканчивается, когда переходные процессы во входной БСЯ завершились, а сигнал B еще остается в разрешающем состоянии.

Как видно, в нетранзитной фазе дисциплина соблюдается, и состязаний не возникает.

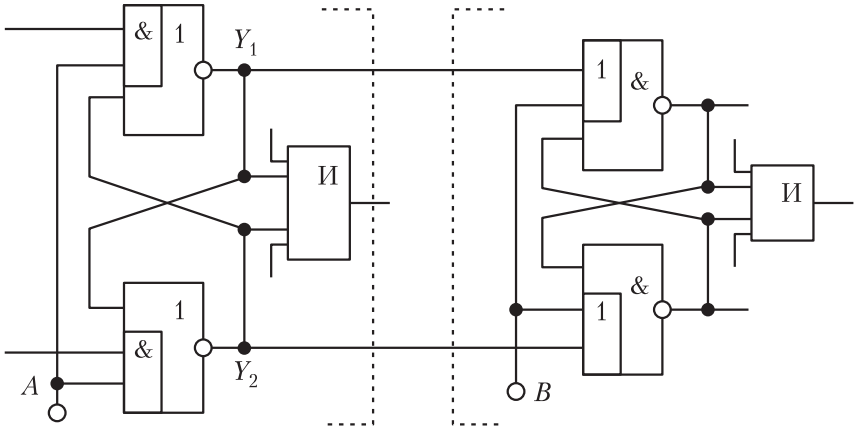


Рис. 3 Соединение БС-сигналов СС-фрагментов: Y_1 и Y_2 — БС-сигналы; A — разрешающий сигнал выхода; B — разрешающий сигнал входа; И — индикаторы; пунктир — границы фрагментов

В наступающей далее транзитной фазе изменяются и сигнал B, и выходные БС-сигналы. Чтобы не возникло состязаний, должен соблюдаться порядок этих изменений, а потому предыдущее правило можно теперь сформулировать так.

В транзитной фазе сначала должен измениться сигнал B, заблокировав входы, и лишь затем могут меняться выходные БС-сигналы.

Поскольку в нетранзитной фазе отношение очередности рассматриваемых сигналов отсутствует, правило дисциплины можно распространить на обе фазы. Данное правило уже можно выразить в схемотехнических терминах, и окончательно оно формулируется так.

В соединении БС-сигналов фрагментов выходные сигналы (Y_1, Y_2) и блокирующий сигнал B не могут быть конкурентными. Сигнал B должен быть инициатором сигналов (Y_1, Y_2), а те, в свою очередь, должны быть континуаторами сигнала B.

Это правило дисциплины будет основным при проверке возможных состязаний в иерархическом анализе.

Отметим также, что сигналы A и B должны быть фазовыми.

В реальных СС-схемах возможны четыре основных варианта соединений БС-сигналов и управляющих ими фазовых сигналов: три из них отличаются способом соединения разрешающих сигналов (рис. 4), четвертый приведен ниже.

Варианты соединений таковы.

1. Непосредственное соединение разрешающих сигналов A и B (цепочки W_1 и W_2 отсутствуют). БС-сигналы есть континуаторы сигнала A, который совпадает с B. Правило дисциплины соблюдено, и соединение корректно.

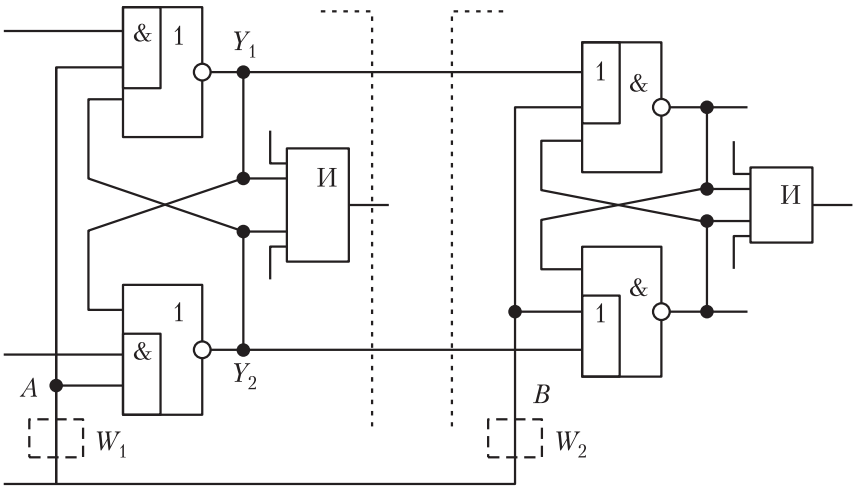


Рис. 4 Варианты соединений разрешающих сигналов: W_1 и W_2 — возможные цепочки элементов

2. Вариант с задержкой сигнала A. В наличии цепочка W_1 (обычно это усиливающие инверторы или повторители), а W_2 отсутствует. БС-сигналы (Y_1, Y_2) по-прежнему остаются континуаторами сигнала B, и соединение также корректно.
3. Соединение с задержкой сигнала B — присутствует цепочка W_2. Вне зависимости от наличия цепочки W_1 БС-сигналы и сигнал B являются конкурентными, и такое соединение *некорректно*.
4. На рис. 5 показан вариант с обратной связью. Обратная связь может быть локальной или глобальной (общей обратной связью всей схемы) с любой задержкой в цепи. Для корректности соединения сигнал B должен через обратную связь индицироваться на сигнале A, а сигнал A, уже локально, должен индицироваться на B. В противном случае будет нарушена дисциплина их изменений.

Если обратная связь глобальная, то такая индикация A на B будет реализована автоматически (при условии успешной проверки по п. 2.4), так как общая обратная связь по принципу самосинхронности должна индицировать все внутренние сигналы схемы. В остальных случаях требуется проверка индикации по полученным на предыдущем шаге спискам.

Например, сигнал B на рис. 5 может быть подсоединен к выходу индикатора левого фрагмента.

Кроме перечисленных основных вариантов возможны соединения с несколькими разрешающими сигналами в БСЯ. В этих случаях блокировка и разре-

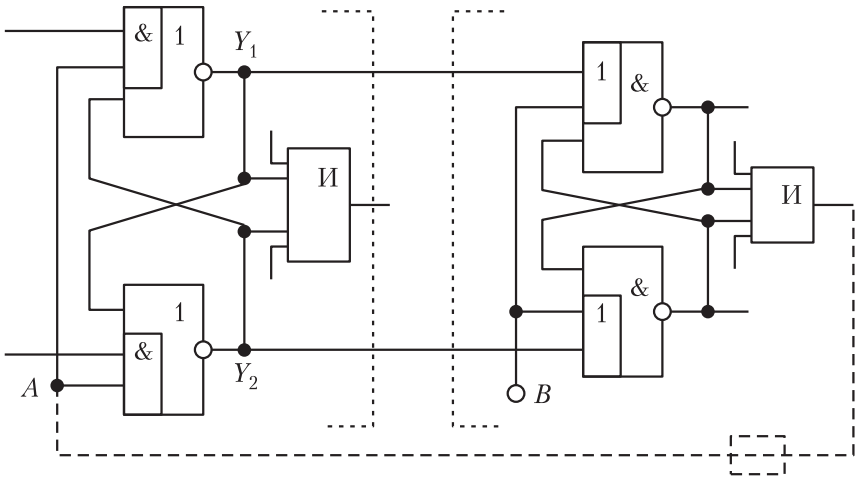


Рис. 5 Соединение БС-сигналов с использованием обратной связи

шение производятся комбинациями значений разрешающих сигналов и правило дисциплины должно учитывать как комбинации, так и очередность изменений управляющих сигналов.

Из изложенного выше видно, что в атрибутах интерфейсов СС-фрагментов должна содержаться информация о входных и выходных БСЯ: какими входами фрагментов и какими значениями они блокируются и существуют ли задержки в цепях блокировки. Эта информация готовится на более низких уровнях иерархии с помощью ИМА, а на нижнем уровне — ФМА.

Проверка дисциплины БС-сигналов заключается в установлении варианта соединения, правильности блокировки соответствующими значениями фазовых сигналов и, в случае локальной обратной связи, проверке индикации разрешающих сигналов по ранее полученным спискам индикации.

2.6 Получение параметров интерфейса главной схемы

На этом шаге готовится информация для следующего верхнего уровня иерархии об индикации всех внешних сигналов и параметрах соединений внешних БС-сигналов схемы.

Из полученных ранее списков индикации фазовых выходов выбираются внешние сигналы, которые и составляют внешние списки индикации.

Внешние БС-сигналы в силу специфики СС-схем соединяются с БС-сигналами внутренних СС-фрагментов непосредственно, без промежуточных элементов. Поэтому параметры внешних БС-сигналов и сигналов, их блокирующих, определяются по атрибутам интерфейсов подсоединенных СС-фрагментов.

Полученная информация помещается в соответствующие атрибуты интерфейса главной схемы.

Данный шаг завершает анализ самосинхронности на текущем уровне иерархии.

3 Заключение

Одной из главных трудностей проектирования СС-схем практических размеров является анализ самосинхронности. Во всех случаях требуется вычислить и проверить все рабочие состояния схемы и переходы между ними. Существующие методы требуют полного раскрытия схем, т. е. исследования уравнений всех ее элементов. При увеличении размера схемы сложность вычислений растет экспоненциально от числа входов и переменных памяти, а еще дополнительно и от числа элементов, что не позволяет анализировать схемы все увеличивающихся размеров.

Впервые представлен метод иерархического анализа СС-схем, основанный на функциональном подходе. По полным уравнениям в нем анализируются только фрагменты нижнего уровня, размер которых может быть выбран достаточно малым.

На всех уровнях иерархии выше нижнего для анализа используются не уравнения элементов, а взаимосвязи фрагментов и информация, полученная на предыдущем уровне. Сложность вычислений здесь практически линейна от числа фрагментов и сигналов.

Такой порядок позволяет наращивать снизу вверх размеры анализируемых схем и тем самым решить одну из главных проблем проектирования СС-схем, во многом тормозящую их развитие, — анализ схем неограниченных размеров.

Важность решения проблемы проектирования СС-схем любых размеров диктует необходимость ее автоматизации. Предложенный иерархический метод достаточно формализован для реализации в программных средствах. Уже разработана и функционирует программа анализа СС-схем нижнего уровня ФАЗАН [10], предоставляющая для верхнего уровня необходимую информацию. Автоматизация предложенного иерархического метода, таким образом, подготовлена и начнет осуществляться в ближайшее время.

Литература

1. Автоматное управление асинхронными процессами в ЭВМ и дискретных системах / Под ред. В. И. Варшавского. — М.: Наука, 1986. 400 с.
2. *Плеханов Л. П., Степченков Ю. А.* Экспериментальная проверка некоторых свойств строго самосинхронных схем // Системы и средства информатики. — М.: Наука, 2006. Вып. 16. С. 476–485.

3. *Плеханов Л. П.* О свойстве самосинхронности цифровых электронных схем // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2011. Вып. 21. № 1. С. 84–91.
4. *Muller D. E., Bartky W. C.* A theory of asynchronous circuits // Symposium (International) on the Theory of Switching Proceedings. — Harvard University Press, 1959. Part 1. P. 204–243.
5. *Варшавский В. И., Кишиневский М. А., Кондратьев А. Ю., Розенблюм Л. Я., Таубин А. Р.* Модели для спецификации и анализа процессов в асинхронных схемах // Техническая кибернетика, 1988. № 2. С. 171–190.
6. *Плеханов Л. П.* Анализ самосинхронности электронных схем функциональным методом // Системы и средства информатики. — М.: Наука, 2008. Вып. 18. С. 225–233.
7. *Плеханов Л. П.* Проектирование самосинхронных схем: фунциональный подход // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС-2010): Сб. трудов IV Всеросс. научно-технич. конф. — М.: ИППМ РАН, 2010. Вып. 1. С. 26–31.
8. *Kishinevsky M., Kondratyev A., Taubin A., Varshavsky V.* Concurrent hardware: The theory and practice of self-timed design. — London: John Wiley and Sons, 1993. 388 p.
9. *Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченков Д. Ю.* Разработка вычислителя, не зависящего от задержек элементов // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 5–23.
10. *Плеханов Л. П.* Программа анализа самосинхронных схем функциональным методом (ФАЗАН): Свидетельство о государственной регистрации программы для ЭВМ № 2011611102 от 01.02.2011.

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2012. Т. 22. № 1. С. 74–85

ПОВЫШЕНИЕ ОТКАЗОУСТОЙЧИВОСТИ ДАННЫХ В КЭШ-ПАМЯТИ ПУТЕМ ИХ ОБНОВЛЕНИЯ

Б. З. Шмейлин[1]

Аннотация: Предложен метод снижения уязвимости данных кэш-памяти к кратковременным отказам путем их обновления. Обновление данных планируется при анализе кода прикладной программы на этапе компиляции. Метод применим к связанным структурам данных (ССД), что показано на массивах и вложенных структурах данных. Планирование обновления данных кэш-памяти продемонстрировано на примере программы умножения матриц.

Ключевые слова: микропроцессорные системы; элементы данных кэш-памяти; уязвимость к кратковременным отказам; связанные структуры данных

1 Введение

Кэш-память является одним из основных компонентов современных микропроцессорных систем. Кэш-память данных обеспечивает процессору эффективный доступ к данным. При увеличении объема кэш-памяти первого уровня этот компонент кристалла занимает все увеличивающуюся долю среди других компонентов. Неисправности этой памяти могут исказить данные, и такие искаженные данные могут переместиться на другие компоненты в иерархии памяти, а также в процессор. В дальнейшем в статье под кэш-памятью будет пониматься кэш-память первого уровня. Одна из главных угроз надежному хранению данных в кэш-памяти — это кратковременные отказы, причиной которых являются внешние воздействия, такие как космические излучения, альфа-частицы и др. Из-за высокой степени интеграции встроенная в кристалл кэш-память наиболее подвержена кратковременным отказам.

Исследования показывают, что доля сбоев в кэш-памяти возрастает по сравнению с другими компонентами микропроцессорных систем. В последнее время такого рода неисправности возникают чаще, чем фатальные отказы из-за технологических или производственных ошибок.

Для увеличения времени наработки системы на отказ прежде всего необходимо снизить долю ошибок в кэш-памяти. Достижение этой цели представляется исключительно важной задачей.

В связи с этим проблеме снижения уязвимости элементов данных в кэш-памяти в последние годы уделяется повышенное внимание. Во всех работах,

[1]Институт проблем информатики Российской академии наук, shmeilin@mail.ru

посвященных решению этой проблемы, тем или иным способом предлагается уменьшить время пребывания немодифицированных элементов данных в кэш-памяти [1–4]. Данная работа посвящена снижению уязвимости к кратковременным отказам элементов связанных структур данных, в частности элементов массивов и элементов вложенных структур, путем обновления «чистых» (немодифицированных) строк кэш-памяти. Обновление выполняется путем записи копии строки из памяти более низкого уровня иерархии (кэш-памяти второго уровня или главной памяти). Такое обновление планируется при анализе кода прикладной программы на этапе компиляции. Компилятор встраивает в компилируемый код инструкции prefetch с адресом, определенным при планировании обновления. Инструкция prefetch предусмотрена в современных микропроцессорах, в частности в процессоре Pentium M. Отказоустойчивость элементов данных в кэш-памяти второго уровня или главной памяти обеспечивается применением корректирующих кодов.

2 Анализ уязвимости элементов данных в кэш-памяти

С целью анализа уязвимости элементов данных представим различные периоды существования элементов в кэш-памяти. К таким периодам относятся следующие промежутки времени [3]:

- FR — промежуток времени между выборкой строки в кэш (fetch — F) и чтением элемента данных (read — R);
- RR — промежуток времени между двумя чтениями элемента данных (read — R);
- WR — промежуток времени между записью (write — W) и чтением элемента данных (read — R);
- FW — промежуток времени между выборкой строки в кэш (fetch — F) и записью (write — W);
- RW — промежуток времени между чтением (read — R) и записью (write — W) элемента данных.

На рис. 1 графически показаны все эти периоды. Назовем периоды существования элементов в кэш, уязвимые к сбою, фазами риска (ФР). К ФР относятся периоды FR, RR, WR.

Следует отметить, что не все элементы данных строки используются процессором, а поэтому уязвимость к сбоям должна быть обеспечена только для элементов, ошибки в которых могут повлиять на выполнение программы. Назовем эти элементы критическими. Критическое время, ассоциирующееся с критическим элементом, — это период, в течение которого значение этого элемента данных должно быть неизменным. Критическое время для конкретного

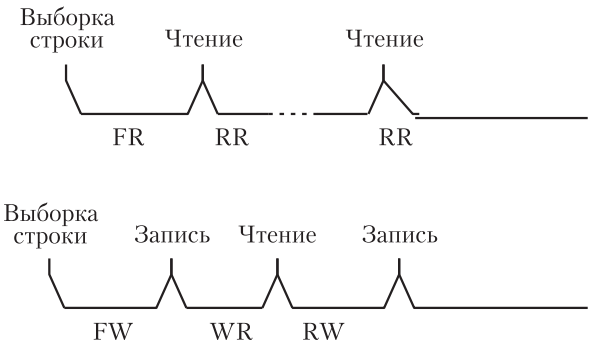


Рис. 1 Периоды существования элементов данных в кэш-памяти

элемента данных соответствует фазам риска FR, RR, WR для этого элемента. Кратковременный отказ, произошедший в интервале критического времени, может привести к ошибке выполнения приложения.

Предшествующие работы. Работа [5] посвящена анализу уязвимости элементов данных кэш-памяти на этапе компиляции. Авторами предлагается выполнить преобразование кода программы для снижения уязвимости кэш-памяти. Главная особенность рассматриваемой кэш-памяти с точки зрения ее уязвимости к кратковременным отказам заключается в том, что строка кэш-памяти охвачена контролем по четности, однако при модификации строки (записи в нее) бит четности не меняется. Это значит, что «чистая» (немодифицированная) строка неуязвима к сбоям. В работе уязвимость определяется продолжительностью интервала между повторными обращениями по чтению к одному и тому же элементу данных модифицированной строки. При этом предполагается, что в течение этого интервала времени строка не была вытеснена из кэш-памяти.

Авторами выработаны рекомендации для оптимизирующих компиляторов; целью такой оптимизации является не только повышение производительности, но и снижение уязвимости кэш-памяти. Зачастую эти требования являются противоречивыми, так как с ростом относительного числа промахов уязвимость элементов данных кэш-памяти снижается, но при этом снижается и производительность. В работе [5] показано, как достичь компромисса между ними на примере программы умножения матриц. Однако выбранный в качестве объекта исследования кэш прямого отображения с небольшим общим объемом сравнительно редко применяется, и поэтому применение результатов работы весьма ограничено.

В работе [4] предлагается выполнять обновление строк, обращение к которым наиболее вероятно в ближайшее время. Это только что использованные строки, так называемые строки MRU (most recently used). Авторы предлагают вы-

полнять статическую и динамическую предварительную выборку MRU-строк. В первом случае такая выборка планируется на этапе компиляции. Динамическая предварительная выборка выполняется по результатам моделирования работы приложения. Результаты экспериментов показывают, что для многих приложений 46% обращений выполняется к той же самой MRU-строке, а к соседним строкам — только 26%. Чем дальше удаление строк от MRU-слота, тем обращение к ним происходит реже.

При таком подходе обновление будет выполняться через слишком короткие интервалы времени, чтобы в течение них можно было ожидать ошибки в элементе данных, вызванной кратковременным отказом. В отличие от такого способа обновления в данной работе предлагается методика точного определения критического элемента данных и интервала времени, в течение которого необходимо его обновление.

В работе [4] проанализированы интервалы времени между последовательными обращениями к элементам данных кэш-памяти. Этот показатель назван inter-reference gap (IRG). Он вычислялся по 20 приложениям из набора SPEC CPU2000, причем учитывались обращения к одному и тому же элементу. В среднем по 20 приложениям IRG равен 1430 тактам. Однако такое значение IRG слишком мало, чтобы выполненное внутри такого интервала обновление заметно снизило бы уязвимость элементов данных.

Было установлено, что весьма большую долю среди всех строк кэш-памяти составляют строки с IRG $<$ 200 тактов. Можно предположить, что строки кэш-памяти с IRG $<$ 200 тактов содержат данные, обрабатываемые локальными циклами, а блоки кэш-памяти с большими значениями IRG содержат данные, обрабатываемые глобальными циклами.

Строки с IRG $<$ 200 тактов вносят очень незначительный вклад в общую уязвимость кэш-памяти (менее 3%), так как содержат элементы данных с очень частыми обращениями, в основном в слотах, близких к MRU-слотам. Поэтому в данной работе при расчете среднего значения IRG были отброшены строки с IRG $<$ 200. После перерасчета этого показателя среднее его значение по 20 приложениям составило около 4000 тактов, а по четырем приложениям оно составило $>$ 6000 тактов. В табл. 1 приведены характеристики этих приложений и средние значения IRG.

Описываемая работа ставила своей целью снижение уязвимости элементов данных кэш-памяти в приложениях с большими значениями IRG. Условия, при которых эффективно применять обновление, следующие:

- элемент данных должен быть критическим элементом;
- критическое время между повторным обращением к элементу данных должно превышать заранее установленное значение фазы риска;
- обращение процессора к элементу данных должно быть обращением по чтению.

Таблица 1 Средние значения IRG по четырем приложениям

Приложение	Характеристика приложения	Среднее значение IRG
Gcc	Оптимизирующий компилятор для процессора Motorola. Для достижения хороших результатов по оптимизации использует большой объем памяти	6414
Parser	Синтаксический анализатор предложений на английском языке. Входные данные — последовательность предложений, выход — список выявленных ошибок	8249
Swim	Программа прогноза погоды	7369
Ammp	Программа исследования молекулярной динамики на комплексе ингибитора белка	6698

В следующем разделе приведен способ описания характера обращений к элементам связанных структур данных, на основе которого планируется обновление.

3 Описание характера обращений к элементам связанных структур данных

Процесс обращения к элементам ССД будем описывать с помощью дескрипторов ССД, как это было предложено в [6]. Эти дескрипторы содержат базовый адрес структуры, ее размер, шаг обращения к ней, и они помечаются объемом работы, выполняемой программой при выполнении каждого цикла итерации. С помощью такого аппарата можно описать обращения к памяти для массивов, связанных списков, вложенных структур данных и рекурсий.

Дескриптор ССД позволяет определить факт повторного обращения к элементу данных кэш-памяти и число тактов между последовательными обращениями. Число тактов между последовательными обращениями позволяет установить, превысит ли критическое время у какого-либо элемента данных заданное значение фазы риска. Таким образом можно определить выполнение условий необходимости обновления.

В качестве ССД рассмотрим массив и вложенную структуру. Каждая из этих ССД описывается дескриптором, отражающим размещение элементов структуры в памяти и порядок обработки этих элементов приложением. Дескриптор массива — $D_m(B, L, S)$, где B — базовый адрес массива данных; L — число элементов массива; S — шаг между двумя последовательными обращениями.

Дескриптор вложенной структуры данных представлен на рис. 2. Дескриптор внешнего цикла обращений к памяти — $D_{\text{out}}(B, L_1, S_1)$, где L_1 — число элементов массива, обрабатываемых во внешнем цикле; S_1 — шаг между двумя

последовательными обращениями во внешнем цикле.

Дескриптор вложенного цикла обращений к памяти — $D_{\text{nest}}(O_2, L_2, S_2)$, где параметр O_2 — смещение относительно базового адреса массива данных при выполнении вложенного цикла; L_2 — число элементов массива, обрабатываемых во вложенном цикле; S_2 — шаг между двумя последовательными обращениями во вложенном цикле.

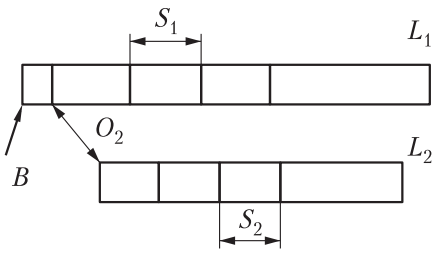


Рис. 2 Дескриптор вложенной структуры данных

Система дескрипторов ССД — это ориентированный граф, вершинами которого являются дескрипторы, а дугами — зависимости между дескрипторами. Каждый дескриптор ССД в этом графе дополняется информацией о работе приложения, выполненной при обработке структуры данных, соответствующей этому дескриптору ССД.

На примере вложенной структуры данных представим информацию о работе приложения:

```
for (i = 0; i < L1; i++)
 {
   ... = data1 [i];
  for (j = 0; j < L2; j++)
  {
    ... = data2 [j];
   }
}
```

o2

w2

w1

Здесь w_1 — объем работы приложения при выполнении одной итерации внешнего цикла; w_2 — объем работы приложения при выполнении одной итерации вложенного цикла; o_2 — объем работы приложения, выполненный им после каждой итерации внешнего цикла до первой итерации вложенного цикла.

Объем работы приложения измеряется в тактах работы процессора.

На рис. 3 представлена система дескрипторов вложенных структур данных с несколькими уровнями вложения.

Используя дескрипторы ССД, можно разработать алгоритм определения уязвимых элементов данных в кэш-памяти. Этот алгоритм состоит из двух частей:

(1) сравнение адресов обращений к памяти в различных циклах;

(2) вычисление временно́го интервала между двумя обращениями по одинаковым адресам в случае их совпадения.

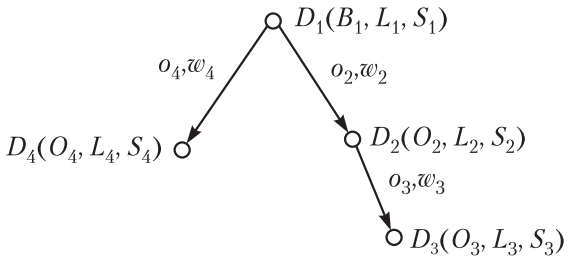


Рис. 3 Система дескрипторов вложенных структур данных

Если этот интервал превышает установленное значение фазы риска, то элемент, размещенный по данному адресу, является уязвимым и должен быть обновлен. На рис. 4 представлен алгоритм вычисления интервала времени между обращениями к элементам данных в кэш-памяти.

Момент обращения к элементу данных во внешнем цикле — $\text{time}_1 = k * w_1$.

Момент обращения к элементу данных во вложенном цикле — $\text{time}_2 = k * w_1 + o_2 + j * w_2$. Это значение получается сложением времени работы вложенного цикла — $o_2 + j * w_2$ — и времени работы внешнего цикла на итерациях от 1 до $k - k * w_1$.

```
1. for (A1 = B. i = 0; A1 < L1; A1 = +S1. i++)
   {
2.   ADR1 [i] = A1;
3.      for (ADR2 = B + O2. j = 0; ADR2 < (L1 + O2);
             ADR2 = +S2. j++)
        {
4.          for (k = 0; k < j; k++)
            {
5.                  if (ADR2 = ADR1 [k])
                        {
6. D = time2 − time1 = k * w1 + o2 + j * w2 −
                              − k * w1 = o2 + j * w2;          (1)
   }}}}
```

Рис. 4 Вычисление интервала времени между обращениями к элементам данных в кэш-памяти: 1 — цикл по адресам обращений внешнего цикла; 2 — список адресов обращений внешнего цикла; 3 — цикл по адресам обращений вложенного цикла; 4 — цикл сравнений текущего адреса обращения вложенного цикла с адресами обращений внешнего цикла; 5, 6 — при совпадении адресов вычисляется интервал времени между обращениями во внешнем и вложенном циклах

Интервал времени D между двумя последовательными обращениями к элементу данных кэш-памяти $D = \text{time}_2 - \text{time}_1 = o_2 + j * w_2$, т. е. это интервал времени между k-й итерацией внешнего цикла и работой во внутреннем цикле до момента установления равенства адресов обращений в обоих циклах.

4 Пример планирования обновления. Область применения метода обновления

В этом разделе рассматривается планирование обновления элементов данных кэш-памяти на примере умножения матриц. Умножаем матрицу A на матрицу B, результирующая матрица — C. Элемент матрицы c_{ij} равен произведению строки i матрицы A на столбец j матрицы B. Умножение возможно, если число элементов строк матрицы A равно числу элементов столбцов матрицы B.

Каждый элемент матрицы C вычисляется так:

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}\,, \quad i = 1, 2, \ldots, m\,, \;\; j = 1, 2, \ldots, q\,,$$

где n — число строк матрицы A или число столбцов матрицы B; m — число столбцов матрицы A; q — число строк матрицы B. В рассматриваемом примере матрицы заданы в виде двумерных массивов, элементы строк матриц размещены в соседних ячейках памяти. Произведение матриц выполняется следующим образом:

```
array a [M] [N]
array b [N] [Q]
array c [M] [Q]
for (k = 0; i < m; k++)    // цикл по всем строкам матрицы A
  {
    for (j = k; j < q;  j++)// цикл умножения одной строки
      c[j, 0] = 0;         // матрицы A на все столбцы матрицы B
    {
   for (i = 0; i < n; i++)// цикл умножения одной строки
        {                 // матрицы A на один столбец матрицы B
           c[j, i] = a[j, i]*b[i, j];
           c[j, i] +  = c[j, i];
         }
     }
   }
```

Программа повторно обращается к элементу $b[i, j]$ через фиксированный интервал времени. Элементы матрицы B могут стать уязвимыми в том случае,

когда интервал времени между двумя последовательными обращениями к нему превышает значение фазы риска. Задавшись этим значением, можно установить, при какой размерности матриц элементы матрицы B становятся уязвимыми.

Пусть длительность каждого чтения и каждой записи одинаковые и составляют два машинных такта (предполагается отсутствие промаха в кэш-памяти, т. е. нужная строка размещена в кэш-памяти первого уровня). Длительность каждой операции процессора составляет один машинный такт. Тогда длительность одной итерации вложенного цикла w_2 равняется сумме длительностей двух чтений ($a[i,\, j]$ и $b[j,\, i]$) и одной записи ($c[i,\, j]$) и двух операций процессора ($a[i,\, j] * b[j,\, i] + c[i,\, j]$):

$$w_2 = 3 * 2 + 2 = 8\,.$$

Объем работы приложения o_2, выполненный им после каждой итерации внешнего цикла до первой итерации вложенного цикла, равен длительности одного чтения, одной записи и одной операции процессора ($c[i,\, 0] = 0$):

$$o_2 = 2 * 2 + 1 = 5\,;$$
$$w_1 = w_2 + o_2 = 13\,.$$

По формуле (1) (см. рис. 4) интервал времени D определяется следующим образом:

$$D = o_2 + j * w_2\,.$$

В рассматриваемом случае $i = 0$, т. е. критическое время определяется работой на всех итерациях вложенного цикла. В статье [3] были рекомендованы следующие интервалы инвалидации (или обновления): либо $2K$ тактов — для приложений с относительно короткими фазами риска RR, либо 500 тактов — для приложений с относительно длительными фазами риска RR. Отсюда можно установить, при каком размере матрицы необходимо выполнять обновление с частотой $2K$ тактов или 500 тактов. Такое обновление нужно выполнять при интервале времени между повторными обращениями к элементу матрицы $D = 2K$ или $D = 500$.

Для $D = 2K$ значение индекса вложенного цикла j вычисляется следующим образом:

$$j = \frac{D}{w_2} - o_2 = \frac{2000}{8} - 5 = 250 - 5 = 145\,.$$

Для $D = 500$ это значение вычисляется следующим образом:

$$j = \frac{500}{8} - 5 = 62{,}5 - 5 = 57{,}5\,.$$

Итак, интенсивное обновление следует применять в том случае, когда число элементов строки матрицы A (столбца матрицы B) больше 58.

Здесь не рассматривались алгоритмы быстрого перемножения матриц, а данный пример приведен только в качестве иллюстрации применения метода обновления с целью снижения уязвимости элементов данных кэш-памяти.

Достоинства метода обновления элементов данных кэш-памяти:

– возможность определения адресов и времени обновления на этапе компиляции;

– благодаря универсальному показателю характера обращения к памяти в виде дескриптора ССД нет необходимости определять тип ССД: массив ли это, связанный список или другая регулярная структура;

– в отличие от предлагаемого в [4], по существу, эвристического подхода к выбору кандидатов на обновление здесь выбор определяется строго аналитически;

– реализация метода обновления данных не приводит к заметной потере производительности микропроцессорной системы.

Ограничения метода:

– на этапе компиляции реализация предлагаемого метода невозможна при косвенной адресации, а также при динамически определяемых указателях;

– выбор кандидатов на обновление предполагает, что в течение фазы риска строка кэш-памяти, содержащая элемент данных — кандидат на обновление, не была вытеснена из кэш-памяти;

– при обновлении элементов данных возникает повышенный трафик на шине памяти и связанное с этим повышенное энергопотребление.

Рассмотрим подробнее последнее ограничение — повышенное энергопотребление. В работе [1] приведены данные по потреблению энергии при работе различных приложений из набора SPEC CPU2000 при различных методах поддержания идентичности информации в кэш-памяти и главной памяти — обратной и сквозной памяти. При обратной записи модифицированная строка копируется в главную память при ее вытеснении из кэш, тогда как при сквозной записи копирование строки выполняется при каждой ее модификации. В случае обратной записи элементы более уязвимы к сбоям, но при этом снижается трафик на шине памяти и потребление энергии.

В табл. 2 и 3 приведены данные по потреблению энергии при работе приложений из набора SPEC CPU2000: в табл. 2 приведены данные для программ с высоким потреблением энергии, а в табл. 3 — с низким.

На основании приведенных здесь данных можно рекомендовать применение обновления для приложений второй группы, если энергопотребление является критическим параметром.

Таблица 2 Приложения с высоким энергопотреблением

Приложение	Характеристика приложения	Потребление энергии, Дж		
		Сквозная запись	Обратная запись	Среднее значение
Mcf	Программа планирования транспортных потоков	0,026	0,023	0,0255
Art	Программа моделирования процессов нейронной сети	0,038	0,035	0,0375
Equake	Программа моделирования сейсмического распространения волн в больших бассейнах	0,018	0,012	0,015

Таблица 3 Приложения с низким энергопотреблением

Приложение	Характеристика приложения	Потребление энергии, Дж		
		Сквозная запись	Обратная запись	Среднее значение
Gcc	Приведена в табл.1	0,007	0,001	0,004
Crafty	Программа решения шахматных задач на высокопроизводительных компьютерах	0,007	0,002	0,0045
Parser	Приведена в табл. 1	0,008	0,003	0,0055
Gap	Программа реализует язык и библиотеку, предназначенные для вычислений в группах	0,007	0,002	0,0045
Wupwise	Программа решения задач квантовой хромодинамики, Характеризуется чрезвычайно интенсивными вычислениями	0,006	0,001	0,0035
Facerec	Программа распознавания образа	0,006	0,003	0,0045

5 Заключение

Предлагаемый метод обновления элементов данных кэш-памяти обладает рядом существенных преимуществ перед другими методами повышения их отказоустойчивости, а именно:

– моменты выполнения обновления определяются на этапе компиляции и не требуют дорогостоящего моделирования работы приложения;

– реализация метода обновления данных не приводит к заметной потере производительности микропроцессорной системы.

К ограничениям применения этого метода следует отнести:

– метод применим к так называемым структурированным приложениям, а именно к приложениям, работающим со связанными структурами данных;
– реализация метода невозможна при косвенной адресации, а также при динамически определяемых указателях;
– при обновлении элементов данных кэш-памяти возникает повышенный трафик на шине памяти и связанное с этим повышенное энергопотребление.

Литература

1. *Wang S., Hu J., Ziavras S.* On the characterization and optimization of on-chip cache reliability against soft errors // IEEE Trans. Comp., 2009. Vol. 58. No. 9. P. 1171–1184.
2. *Захаров В. Н., Шмейлин Б. З.* Влияние кратковременных отказов на надежность кэш микропроцессорных систем // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 57–71.
3. *Захаров В. Н., Шмейлин Б. З.* Повышение отказоустойчивости кэш в микропроцессорных системах // Системы высокой доступности, 2011. Т. 7. Вып. 1. С. 5–12.
4. *Asadi H., Sridharan V., Tahoori M., Kaeli D.* Reducing data cache susceptibility to soft errors. www.ece.neu.edu/groups/nucar/publications/IEEETDSC.pdf.
5. *Shrivastava A., Lee J., Jeyapaul R.* Cache vulnerability equations for protecting data in embedded processor caches from soft errors // LCTES'10: Proceedings of the ACM SIGPLAN/SIGBED 2010 Conference on Languages, Compilers, and Tools for Embedded Systems. — New York: ACM, 2010. P. 143–152.
6. *Kohout N., Choi S., Kim D., Yeung D.* Multi-chain prefetching: Effective exploitation of inter-chain memory parallelism for pointer-chasing codes // 10th Annual Conference (International) Parallel Architectures and Compilation Techniques Proceedings. — Barcelona, Spain, 2001.

МОДЕЛИРОВАНИЕ ЛЕКСИЧЕСКОЙ СЕМАНТИКИ В ЗАДАЧАХ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

О. С. Кожунова[1]

Аннотация: Проведен аналитический обзор лексико-семантических методов, их примеров и возможных областей их приложения. В основу рассмотренных работ легли статьи, опубликованные в трудах *The World Congress in Computer Science, Computer Engineering, and Applied Computing* (Международного конгресса по компьютерной науке, вычислительной технике и приложениям, который ежегодно проводится в США) за 2009 и 2011 гг.

Ключевые слова: лексико-семантические методы; компьютерная лингвистика; обработка текстовых данных

1 Введение

Стремление к глубинному пониманию языка и всех языковых процессов, а также многочисленные попытки смоделировать их все чаще находят отражение в различных сферах человеческой деятельности. Поэтому сегодня методы семантического моделирования языковых процессов применяются практически в каждой информационной системе.

Тем не менее в силу различия задач в создаваемых приложениях компьютерной лингвистики и сопутствующих системах семантика языка рассматривается в определенных ракурсах. В частности, существуют подходы, направленные на моделирование структурной (синтаксической) составляющей языка вкупе со смысловыми единицами [1–3], т. е. синтактико-семантическое моделирование. Одновременно с этим развиваются лексико-семантические методы, фокус которых смещен в сторону отдельных лексических единиц и взаимодействия их смыслов [4–8]. При этом два принципиально разных подхода к моделированию языковых реалий и процессов развиваются параллельно и дополняют друг друга.

В настоящей работе предпочтение отдано лексико-семантическим методам, их примерам и возможным областям их приложения. В основу рассмотренных работ легли статьи, опубликованные в трудах *The World Congress in Computer Science, Computer Engineering, and Applied Computing* (Международного конгресса по компьютерной науке, вычислительной технике и приложениям, который ежегодно проводится в США) за 2009 и 2011 гг.

[1]Институт проблем информатики Российской академии наук, okozhunova@ipiran.ru

2 Лексико-семантические методы и их приложения

Сегодня разработчики информационных систем и приложений искусственного интеллекта все чаще обращаются к лексико-ориентированным методам компьютерной лингвистики. Часто главными акцентами создаваемых лингвистических ресурсов становятся термины предметных областей, понятия и концепты, а также их смысловые связи и отношения. Основными примерами уже существующих ресурсов такого типа являются тезаурусы, онтологии, семантические словари и т. д.

2.1 Автоматический поиск терминов и понятий

Среди статей, опубликованных в трудах конгресса *WorldComp*'2011, имеется несколько работ, в которых прослеживается тенденция к применению лексико-семантических подходов. В частности, в [9] описывается метод искусственного интеллекта, разработанный авторами для осуществления автоматического поиска терминов в Интернете в качестве ответов на сложные вопросительные предложения. Метод оперирует специально созданной для этой задачи базой понятий. В ней содержатся семантические характеристики терминов и их веса (по степени значимости в определенной предметной области и/или задачи). Веса терминов в основном используются для отображения семантических связей между терминами и узловыми понятиями.

Создание такого подхода Хирокацу Ватабе и коллегами [9] было мотивировано сложившейся ситуацией с поисковыми системами в Интернете. Поиск информации, как правило, имеет ограничения, которые накладывает сам пользователь. Но в силу интенсивного развития веб-пространства и используемых технологий результаты запросов оказываются в массиве ненужной, попутно найденной информации. Именно поэтому предоставить четкий, сфокусированный ответ на поисковый запрос специальным системам становится все сложнее. Одной из последних тенденций решения этой проблемы является вопросно-ответная форма поисковых запросов. Однако такой метод требует проведения последовательного анализа на всех стадиях, наличия базы данных и фиксации возможных смысловых (семантических связей).

С этой целью авторы вышеупомянутой работы [9] дополняют метод вопросно-ответных запросов своей базой понятий с семантическими характеристиками терминов и весами для построения ассоциаций между ними, а также механизмом обработки и отображения семантики вопросов (рис. 1).

Кратко рассмотрим основные составляющие и процедуры предложенного метода. База понятий (см. рис. 1) представляет собой базу знаний, состоящую из 120 000 концептов (понятий), которые собраны из различных источников типа языковых словарей (здесь: японских), книг и газет, и атрибутов (30 для каждого понятия), которые выражают их семантические характеристики.

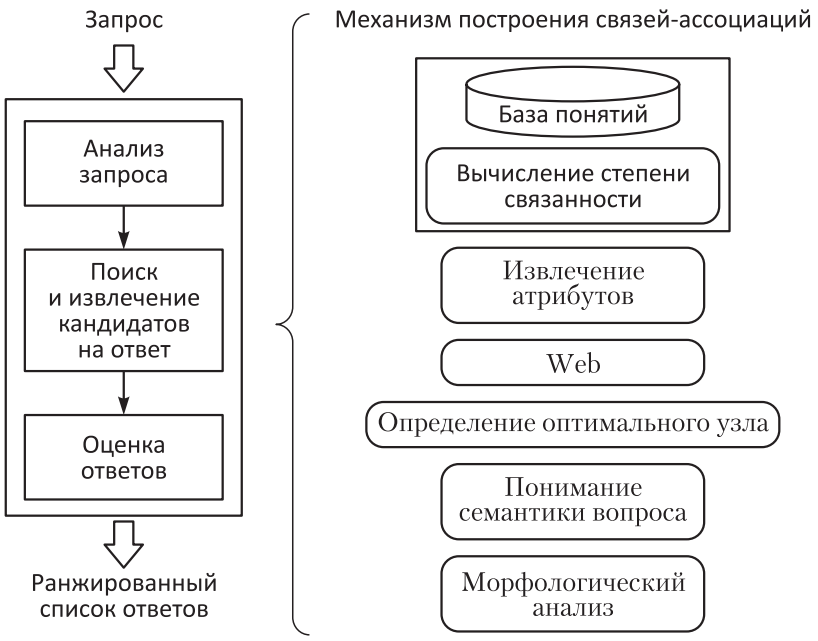


Рис. 1 Схема интеллектуальной веб-поисковой системы

Понятия наделены атрибутами наряду с весами, которые соответствуют уровню их значимости [9]:

$$A = \{(a_1, w_1), (a_2, w_2), \ldots, (a_m, w_m)\} ,$$

где A — некоторое понятие; a_i — семантический атрибут; w_i — вес.

Любой первичный атрибут a_i состоит из термов, содержащихся в наборе описаний понятия в базе понятий. Таким образом, чтобы убедиться, что первичный атрибут соответствует определенному описанию понятия, его можно извлечь из базы. Это вторичный атрибут. В базе понятий произвольное понятие определено как набор атрибутов n-го порядка [9].

Произвольный атрибут также является понятием, как описано выше. Авторы работы [9] вычисляют степень связанности между понятиями путем использования атрибутов второго порядка. Степень связанности варьируется от 0,0 до 1,0 (см. рис. 1).

Следующей частью системы [9] является блок понимания семантики вопросов. При помощи функции парсинга этот блок помогает получить целевые термины запроса из поставленного в нем вопроса. Например, если вопрос содержит вопросительное слово «кто?» или «где?», эта функция позволит получить «человек» или «место» в качестве целевых терминов запроса.

Таблица 1 Типы потенциально релевантных ответов на вопрос-запрос

Тип ответа	Типичный целевой термин
Человек	Имя человека
Организация	Компания или университет
Место	Название страны
Животное	Название животного
Растение	Название растения
. . .	. . .

Таблица 2 Извлечение поискового терма из вопроса-запроса

Вопрос	Поисковый терм
«Где-самый-северный-город-на-о.-Хонсю?»	«Самый-северный-город-на-о.-Хонсю»

На отбор ответов сильно влияет то, какого типа ответ необходим на заданный в запросе вопрос. Поэтому авторы работы [9] составили таблицу соответствия потенциально релевантных ответов на запрос и типичных целевых терминов, предварительно соотнеся их с вопросительными словами по смыслу (табл. 1).

Таким образом, типичный запрос и основные процедуры поиска ответа имеют вид:

Пример запроса: «Где находится самый северный город на о. Хонсю?»
Тип ответа: «Место».
Поисковый терм (понятие): «Хонсю — самый северный город».
Потенциально релевантные ответы: Аомори — Префектура Ома; Хоккайдо — Шиокуби-Кейп; Шимокита-Вайн. . .

Ранжирование ответов:

1. Аомори — Префектура Ома (0,0062).
2. Хоккайдо — Шиокуби-Кейп (0,0005).
3. Шимокита-Вайн (0,0000).

После соотнесения типов вопросов и ответов (см. табл. 1) происходит определение поискового терма (табл. 2).

На следующей стадии поиска ответа на вопрос-запрос происходит поиск и извлечение потенциально релевантных ответов из веб-ресурсов в соответствии со следующей процедурой [9]:

1. Поисковый терм подается на вход поисковой машины, с выхода которой выводится 100 результатов — веб-страниц с самым высоким рангом (рейтингом).

Таблица 3 Поиск и извлечение потенциально релевантных ответов

Поисковый терм «Самый-северный-город-на-о.-Хонсю»	
Потенциально релевантные ответы	Вес
Аомори — Префектура Ома	699,298
Хоккайдо — Шиокуби-Кейп	39,582
Шимокита-Вайн	13,194
. . .	. . .

2. Из полученного корпуса документов удаляется вся лишняя информация типа html-тегов, после чего он проходит обработку морфологическим анализатором для дальнейшего извлечения существительных.
3. Сложные слова и термины извлекаются в виде составных морфем (используя правило «последовательности существительных, номеров, букв алфавита — составные слова»).
4. Для извлеченных простых и составных существительных вычисляется частота встречаемости в документах и в Интернете. В результате словам присваиваются соответствующие веса.
5. Простые и составные существительные ранжируются в соответствии с их весом; 50 самых «весомых» отбираются как потенциально релевантные ответы на вопрос-запрос (табл. 3).

Поиск потенциально релевантных ответов завершается процедурой концептуализации [9]. Под концептуализацией авторы понимают процесс приписывания заданному термину набора атрибутов и веса (если термин не задан в базе понятий и эти характеристики у него отсутствуют). Такая процедура способствует вычислению степени связанности терминов (смысловой связанности и ассоциаций). Веб-документы, которые используются для концептуализации поискового терма, авторы [9] получают в результате веб-поиска, где поисковый терм, сконструированный из ключевых слов вопроса-запроса, подается на вход. Это облегчает обработку документов и извлечение необходимых семантических характеристик терминов. Это означает, что концептуализованный поисковый терм может служить для оценки состоятельности ответа на вопрос-запрос (табл. 4).

Наконец, потенциально релевантные ответы найдены, терминам и понятиям присвоены семантические характеристики и веса. С этого момента авторы подхода переходят к оценке полученных ответов.

Сначала вычисляется степень связанности потенциально релевантного ответа с поисковым термом. Полученная степень является, по сути, оценкой состоятельности потенциально релевантного ответа на вопрос-запрос, для чего

Таблица 4 Пример концептуализации (с частичным списком атрибутов)

Поисковый терм «Самый-северный-город-на-о.-Хонсю»		*Потенциально релевантный ответ «Аомори — Префектура Ома»*	
Атрибут	Вес	Атрибут	Вес
Ома	1859,522	Ома	4035,559
Хонсю	1522,526	Карта	188,496
Город	276,172	Погода	151,084

и используется в дальнейшем: $\mathrm{Do}A(\mathrm{ac}_i, \mathrm{QW})$, где $\mathrm{Do}A$ — степень связанности потенциально релевантного ответа ac_j с поисковым термом QW.

Затем происходит оценка адекватности типа вопроса:

$$\mathrm{NodeValue}\,(\mathrm{ac}_i, \mathrm{AnswerType}) = \\ = \mathrm{Do}A\,(\mathrm{ac}_i, \mathrm{AnswerType}) \cdot \log\,(\mathrm{VerbHit}\,(\mathrm{ac}_i, \mathrm{AnswerType}))\ ,$$

где $\mathrm{NodeValue}(\mathrm{ac}_i, \mathrm{AnswerType})$ — оценка узла потенциально релевантного ответа ac_i, соответствующего типу ответа AnswerType; $\mathrm{Do}A(\mathrm{ac}_i, \mathrm{AnswerType})$ — степень связанности потенциально релевантного ответа с поисковым термом (см. выше); $\mathrm{VerbHit}(\mathrm{ac}_i, \mathrm{AnswerType})$ — число «попаданий», полученных при поиске с использованием поисковой машины для связывания глагольных узлов ключевых слов с потенциально релевантными ответами [9].

Пример. Термины «Фестиваль» и «Олимпийские игры» могут одновременно принадлежать к «Событию», непосредственно после них в предложении может идти конструкция «участвовать» + «в». Если искать в Интернете «попадания», используя такую категорию связанности, можно оценить, соответствует ли определенный кандидат в ответы типу вопроса. Авторы описали узловые глаголы для всех перечисленных ими типов вопросов.

В завершение работы системы происходит оценка самих ответов [9]:

$$\mathrm{Score}(\mathrm{ac}_i) = \mathrm{Do}A\,(\mathrm{ac}_i, \mathrm{QW}) \cdot \mathrm{NodeValue}\,(\mathrm{ac}_i, \mathrm{AnswerType})\ .$$

Здесь степень связанности потенциально релевантного ответа и поискового терма является индикатором того, соответствует ли потенциально релевантный ответ вопросу-запросу, а оценка узла — индикатором соответствия потенциально релевантного ответа типу ответа.

По приблизительным оценкам точность предложенной методики составляет около 76% [9].

2.2 Семантика онтологий и лексиконов и их верификация

Следующей не менее интересной работой в данном направлении оказалась статья о верификации онтологии и лексиконов на основе онтологического семантического подхода [10]. В фокусе работы методы автоматической и экспертной верификации новых лексиконов и онтологий, позволяющие оперировать семантикой естественно-языковых единиц.

Авторы публикации [10] разработали так называемую Онтологическую семантическую технологию (*Ontological semantic technology*), которая является улучшенной версией Онтологической семантики [11] и охватывает теорию, методологию и реализацию системы, предназначенной для понимания текста на естественном языке. Эта технология основана на доступных для компьютерной обработки описаниях значений слов, перечисленных в соответствии с их синтаксическими ролями в предложении и соответствующими морфологическими характеристиками.

В основе технологии лежат репозитории лингвистических знаний и знаний из всевозможных предметных областей, построенные полуавтоматически в рамках предложенного подхода. Репозитории призваны устранять неоднозначности различных значений слов и предложений и предоставлять их репрезентации.

В числе репозиториев находятся [10]:

- онтология, содержащая независимые от выбора языка понятия и отношения между ними;
- лексикон для каждого поддерживаемого системой языка, в котором хранятся смыслы слов, привязанные к языковой онтологии (она используется для репрезентации их значений);
- словарь имен собственных (*the proper name dictionary*) с именами людей, названиями стран, организаций и др. и их описаниями, причем вся эта информация привязана к онтологическим понятиям, а единицы словаря связаны между собой;
- блок смысловых правил.

Все вышеперечисленные составляющие репозиториев используются семантическим анализатором текстов (*semantic text analyzer*), который был разработан корпорацией *RiverGlass Inc.* Этот анализатор при обработке входного текста выдает его текстовые смысловые репрезентации (*text meaning representations*). Формат репрезентаций соответствует форматам и интерпретациям онтологии. Обработанные текстовые смысловые репрезентации вносятся в *InfoStore*, динамический ресурс для хранения знаний (в рамках Онтологической семантической технологии). Информация, извлеченная из этого хранилища, используется для дальнейшей обработки и выводов [10].

Семантический анализатор текстов состоит из коллекции модулей, каждый из которых отрабатывает свои функции в соответствии с поставленными условиями. Анализатор «читает» текст по предложениям, обращаясь к лексикону и онтологии для получения возможных интерпретаций каждой части предложения, комбинирует их и выдает ранжированный массив однозначных текстовых смысловых репрезентаций. Полученные репрезентации снабжаются весами в автоматическом режиме согласно их релевантности интерпретациям каждого модуля. Таким образом, здесь существует множество возможностей некорректной обработки предложения: несовершенство самого программного обеспечения и отдельных модулей; лексикон определенного языка, на котором написан анализируемый текст и сама онтология. Поэтому верификация онтологии и лексикона становится актуальной задачей: неправильная трактовка смысла и семантики слов и предложений сказывается на результативности семантического анализатора текстов [10].

Что такое онтология с точки зрения задачи системы и технологии, которые описаны в данной работе? Неформально под онтологией, полученной в результате применения Онтологической семантической технологии, они понимают коллекцию понятий и отношений между ними, а также атрибуты, описывающие эти понятия, и механизм их категоризации.

Одним из методов онтологической верификации является так называемый метод подбора наилучшего понятия. Это означает, что некоторое программное обеспечение осуществляет поиск самого подходящего понятия для заполнения пустот в естественно-языковых структурах, как, например, в предложении типа «Я открыл дверь при помощи X», где X — неизвестное понятие, пустота [10].

Для апробации этого метода Тейлор, Хемпельман и Раскин создали словарную статью для X и поместили в нее порядка 2000 субстантивных смыслов, по одному на каждый объект и событие в онтологии. Для каждого смысла переходного глагола в своем английском лексиконе (4469 смыслов в английском лексиконе *RiverGlass*) они рассмотрели по примеру, который был введен в систему для иллюстрации случаев использования этого глагола и для замены существительным неизвестного понятия X. Затем этот пример обрабатывался ресурсами Онтологической семантической технологии. В результате выбирался наиболее релевантный смысл X для экспертной оценки носителем языка, который решал, каков найденный смысл: возможно подходящий, хорошо подходящий, подходящий наилучшим образом или неприемлемый. Как правило, лучшее решение достигалось в предложениях-примерах, где либо смысл лексемы, либо смысл ее онтологических понятий оказывался под строгими ограничениями [10]:

Пример 1. *He shucked the corn* (он почистил кукурузу).

Интерпретация семантическим анализатором текстов (с обращением к оригинальному предложению и заменой X):

(remove(theme(plant-part))
(agent(human(gender(male))))
(start-location(grain)))
(remove(theme(plant-part))
(agent(human(gender(male))))
(start-location(seednutgrain))) .

Иногда трудно выбрать корректные описания некоторого события как в отношении свойств по умолчанию, так и базовых ограничений. Кроме того, достаточно затратным оказывается и вовлечение инженера по знаниям в верификацию всевозможных изменений, происходящих с такими описаниями. Более простым решением представляется проверка того, имеют ли какой-то смысл предложения на естественном языке, связанные какими-либо базовыми ограничениями или характеристиками. Такие проверки могут выполняться и не специалистами по заданному формализму, а обычными носителями языка, на котором написаны предложения. Вышеприведенный метод подбора наилучшего понятия проводит такую верификацию для описания, определенного по умолчанию. Но необходимо также проводить проверки базовых ограничений [10].

Предположим, что существует событие E, характеризуемое свойством P с описанием F. Это описание F является индикатором того, что и само описание, и его «потомки» доступны свойству P в соответствии с моделью мира, которую представляет данная онтология. Например, событие ГУЛЯТЬ может иметь НОГУ в качестве описания свойства ИНСТРУМЕНТ, но не может иметь в качестве «предка» НОГИ понятие КОНЕЧНОСТЬ (в мире данной онтологии есть ограничение на хождение только ногами, хождение на руках неприемлемо) [10].

Для того чтобы проверить, корректно ли выбрано описание, рассматриваются «потомок» (если он существует) и «предок» интересующего описания. Простое предложение строится при помощи события E, связанного со свойством P и его описанием F:

Пример 2. [*All/most*] *animals walk on legs.*

Таким же образом строятся и предложения с «потомками» и «предками»:

Пример 3.
(a) [*All/most*] *animals walk on hoofed legs*.
(b) [*All/most*] *animals walk on limbs*.

Все подобные предложения предоставляются носителю языка, который оценивает правдоподобность знаний, выраженных в этих предложениях. В таком случае предложения не рассматриваются с точки зрения грамматической корректности [10].

Если в рассматриваемом предложении понятие-«предок» (родовое понятие) оказывается верным, предложения с «потомками» описания F генерируются аналогичным образом и просматриваются носителем языка. В рассматриваемом случае предложение «[*All/most*] *animals walk on arms*» было бы сгенерировано, если бы «[*All/most*] *animals walk on limbs*» оказалось верным [7].

После того как предложения с «предком» и «потомками» описания F проходят проверку носителем языка, его выводы отсылаются инженеру по знаниям, с тем чтобы оптимизировать степень детализации описания свойства P события E в отношении примеров, прошедших оценку носителем языка. Если носитель языка считает, что предложение с «потомком» описания F верное, как в случае с примером 3а, возникает несколько возможностей [10].

Во-первых, и описание, и его «потомки» считаются правильными. Это возможно, если все животные имеют копыта. Такой вопрос нужно задавать, если генерируются аналогичные предложения и варианты их отрицания:

Пример 4. *There are no animals whose legs are not hoofed.*

Если ответ утвердительный, степень детализации описания F должна быть пересмотрена.

Вторая возможность состоит в том, что, возможно, существуют животные без копыт. Такие примеры описаний могут быть формализованы в виде: {CHILD-OF(F)}\HOOFED-LEG. Чтобы проверить такие варианты, необходимо проверить группу предложений, аналогичных «*Some animals walk on X legs*», где $X \in$ {CHILD-OF(F)}\HOOFED-LEG. Если все эти предложения оказываются верными, степень детализации считается корректной [10].

Если же носитель языка обнаруживает, что предложение неверно, как, например, «[*All/most*] *animals walk on hoofed legs*», предполагается, что «*Some animals walk on hoofed legs*» — верно. Иначе можно считать, что либо онтология построена некорректно и признак HOOFED-LEG не должен быть легитимным «потомком» объекта LEG, либо именно это описание HOOFED-LEG должно быть исключено из инструмента события ГУЛЯТЬ [10].

Таким образом, лексикон Онтологической семантической технологии (отдельный для каждого языка) представляет собой ресурс, соединяющий слова языка со знаниями о мире, которые описаны в онтологии. Каждая лексическая единица представлена с отображением всех ее возможных смыслов с фокусом на ее семантической информации (sem-struc) и синтаксическом окружении (syn-struc). Syn-struc указывают базовую синтаксическую информацию об использовании определенного смыслового значения в предложении. Sem-struc выражает значение смысла онтологических понятий, их свойств и описаний [10].

Помимо вышеописанных методов верификации онтологии и лексикона, авторы [7] также используют ресурс, позволяющий определять лексические пропуски. Он функционирует следующим образом:

– сканирует выбранный текст при помощи семантического анализатора текстов и сообщает о словах, для которых не найдено смысла (записывает в отдельный файл по алфавиту или по числу символов), и о смысле слов, у которых он есть;
– позволяет устанавливать фильтры согласно классу слов (например, для существительных и глаголов, игнорируя при этом имена собственные);
– является полуавтоматическим ресурсом для проверки полноты новых предметных областей в онтологии.

Что показательно, этот ресурс также позволяет иллюстрировать то, как оперирует смыслом семантический анализатор текстов и сообщать об этом в различных файлах:

1. В первый файл помещаются смыслы, у которых нет примеров. Поэтому недостающие примеры предложений должны быть добавлены в онтологию, после чего смыслы оказываются в одной из трех нижеперечисленных категорий.
2. Во втором файле хранятся смыслы, у которых нет текстовых смысловых репрезентаций. Поэтому их примеры-предложения должны быть проанализированы отдельно вместе с онтологией, лексиконом и модулями семантического анализатора текстов.
3. В третьем файле расположены смыслы, которые имеют текстовые смысловые репрезентации, но такие, что не оперируют смыслом, для которого они предназначены. Исправления происходят, как и в предыдущем классе.
4. В четвертом файле — смыслы, которые не приводят к текстовым смысловым репрезентациям в корректном виде.

Каждая категория отклонений в семантике отдельных слов и связанных с ними предложений рассматривается специалистами для внесения адекватных изменений как в лексиконе, так и в онтологии [10].

Для верификации лексикона также используется визуализация виртуальной онтологии. Хотя онтология зависит от языка, любые две единицы с эквивалентными sem-struc на разных языках выражают одно и то же значение в Онтологической семантической системе. Поэтому авторы полагают, что в рамках одного языка те смыслы, что имеют эквивалентные sem-struc, являются синонимами с заданной степенью детализации [10]:

Пример 5.

(lion-n_1
(sem-struc(lion))
)
(cub-n_1
(sem-struc(lion(age(value(young)))))
) (cat-n_1

(sem-struc(cat))
)
(kitten-n_1
(sem-struc(cat(age(value(young)))))
)
(goat-n_1
(sem-struc(goat))
)
(human-n_1
(sem-struc(human))
)
(kid-n_1
(sem-struc(human(age(value(young)))))
)
(kid-n_2
(sem-struc(goat))
)

В примере 5 для каждой пары смыслов lion-n_1/cub-n_1, cat-n_1/kitten-n_1, human-n_1/kid-n_1, визуализация виртуальной онтологии позволяет увидеть узел родитель–ребенок (рис. 2). Пара goat-n_1/kid-n_2 связана с одним и тем же понятием. Даже без предварительной подготовки носитель языка заметит непоследовательность в рассмотрении взрослых млекопитающих и их детенышей. И хотя у слова kid (ребенок) есть дополнительный смысл, это не меняет сути дела.

Этот метод можно также использовать при поиске других непоследовательностей. Например, при создании различных семантических структур смыслов, которые могут быть синонимами, с пропусками некоторых лексических единиц, а также необоснованно ограниченных единиц [10].

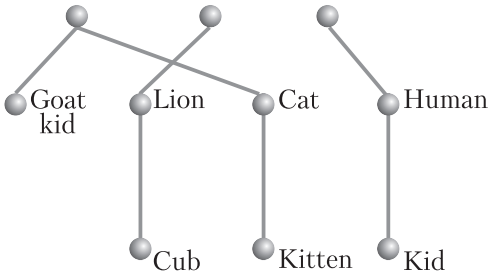


Рис. 2 Относительное расположение некоторых смыслов согласно виртуальной онтологии

2.3 Классификация веб-страниц и алгоритм извлечения атрибутов

Еще одна работа, содержащая элементы лексико-семантического подхода к анализу текстов и извлечению знаний из них, посвящена влиянию нового алгоритма извлечения атрибутов на классификацию веб-страниц [12].

Авторы мотивируют свой интерес к классификации веб-ресурсов по контенту тем, что стремительный рост числа веб-страниц в Интернете затрудняет поиск и объединение похожих страниц. Такие страницы не организованы в соответствии с каким-либо логическим порядком, поэтому результаты доступа к данным и информационный поиск часто непредсказуемы. В статье описан ряд методов и подходов, которые применялись для разрешения этой проблемы.

Широко используется в задачах текстовой классификации модель векторного пространства, в рамках которой документы представлены в виде векторов с набором идентификаторов (например, слова как термины). Эту модель часто называют «мешочной» (*bag-of-words model*). Согласно «мешочной» модели каждый документ является некоторым вместилищем слов, которые входят в него. Если пересмотреть такое представление относительно векторного пространства, то, по сути, каждый термин является измерением для векторов документа. Однако такой способ представления порождает много измерений в рамках одного документа. Авторы предлагают два эффективных способа преодоления этой проблемы: отбор атрибутов и извлечение атрибутов [12].

Алгоритмы по отбору атрибутов способствуют снижению размерности пространства следующим образом. Вместо того чтобы использовать все слова документа как атрибуты, эти алгоритмы оценивают атрибуты на основании особого классификатора, с тем чтобы найти оптимальное подмножество терминов [12]. В результате — снижение затрат на классификацию и повышение ее точности. Среди наиболее известных алгоритмов по отбору атрибутов: алгоритм частотности документов, Chi-статистика, расстояние Кульбака–Лейблера и т. п. Недостаток алгоритмов по отбору атрибутов состоит в том, что процедура отбора оценивается в соответствии с определенным классификатором. Поэтому полученное подмножество не может быть использовано другим классификатором для улучшения его же качества [13].

Алгоритмы по извлечению атрибутов не настолько ресурсоемкие, когда речь идет об охвате большого массива данных. Многомерная и неоднородная структура модели векторного пространства требует больших объемов памяти и вычислительной мощности. Цель алгоритмов по извлечению атрибутов заключается в комбинировании терминов с тем, чтобы сформировать новое описание данных с заданной степенью точности. Извлечение атрибутов позволяет спроецировать многомерные данные в пространство с более низкой размерностью. Среди наиболее известных подходов данного направлении — метод главных компонент (*principal components analysis*, PCA), Isomap и латентный семантический анализ (latent semantic analysis, LSA). Латентное семантическое индексирование

основано на Латентном семантическом анализе. Сегодня это самый широко используемый алгоритм, применяемый в задачах поиска и извлечения данных [14].

Турецкие исследователи провели обзор и анализ двух наиболее часто используемых и хорошо себя зарекомендовавших алгоритмов по извлечению атрибутов: метод главных компонент (PCA) и латентный семантический анализ (LSA).

Метод главных компонент трансформирует набор взаимосвязанных переменных, главных компонент, в набор, где их меньше. Автором метода является Пирсон (1901 г.) [13]. Чаще всего он используется для исследовательского анализа данных, а именно: для извлечения атрибутов путем сохранения характеристик набора данных, которые оказывают максимальное влияние на его разброс, т. е. для запоминания главных компонент низших порядков, которые, как правило, располагают самыми важными аспектами данных. Это выполняется путем проекции нового гиперплана с использованием внутренних весов и векторов. Метод главных компонент применяется в задачах распознавания образов.

Латентный семантический анализ был запатентован в 1988 г. [14]. Это метод, который позволяет анализировать отношения между набором текстовых документов и терминами, ключевыми словами, которые они содержат, посредством генерации множества связанных с ними понятий. Как и предполагается в названии метода, процедура сингулярного разложения раскладывает матрицу на множество меньших, латентных компонент. Латентный семантический анализ использует сингулярное разложение для поиска взаимосвязей между документами. Эта процедура разбивает матрицу терминов-документов на множество меньших компонент. Алгоритм LSA меняет одну из этих компонент (сокращает число измерений) и затем комбинирует из них новую матрицу вида, аналогичного оригинальной матрице. Латентное семантическое индексирование (latent semantic indexing, LSI), основанное на LSA, чаще всего используется в задачах информационного поиска веб-страниц и кластеризации текстовых документов, а также для классификации текстов типа «информационной фильтрации».

Алгоритм, предложенный авторами работы [12], также извлекает атрибуты из векторного пространства и проецирует их в новое гиперпространство с числом измерений, равным числу классов.

Для апробации своего алгоритма и разработки автоматического классификатора для *Dmoz Turkish directory*[1] авторы набрали тестовый набор данных. С этой целью они отобрали 28 567 не повторяющихся веб-страниц на турецком языке. Среди них было выделено 13 основных категорий (и 758 подкатегорий) таких, например, как «Наука», «Новости», «Игры» и т. п. [12].

Однако, поскольку набор данных представлен в необработанном html-формате, его необходимо было пропустить через html-парсер (авторы использовали

[1] *Dmoz Turkish directory* — разновидность информационно-поисковой машины и классификатора веб-страниц (как Yahoo!, например).

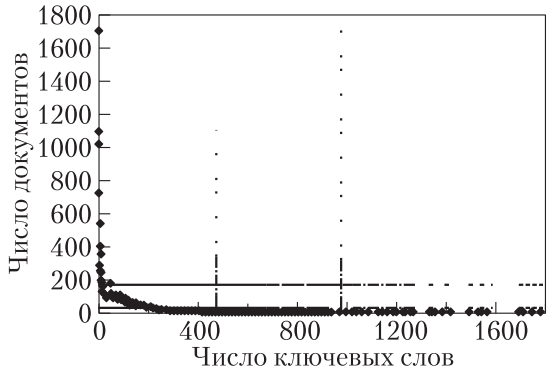


Рис. 3 Распределение терминов и документов в наборе данных

общедоступный ресурс http://htmlparser.sourceforge.net). Получив в результате содержимое веб-страниц, исследователи вычислили статистику терминов для обзора релевантных выделенным категориям документов [12].

На рис. 3 изображено распределение терминов и документов: количество тематических терминов в документах и количество документов с определенным числом терминов. Поскольку распределение неравномерное, авторы отобрали лишь те документы, где концентрация ключевых слов максимальная. Завершающим этапом подготовки набора данных стала обработка терминов для удаления их изменяемых частей (турецкий, как и русский язык, является флективным языком) и фильтрация стоп-слов (не относящихся по тематике к документам) [12].

В своем алгоритме извлечения атрибутов турецкие специалисты не использовали ни собственных векторов и значений [15], ни сингулярного разложения [16]. Они уделили особое внимание весам терминов и их вероятностным распределениям по выделенным классам. Авторы проецировали вероятности терминов на классы и суммировали эти вероятности, чтобы вычислить влияние каждого термина на каждый класс [12].

В работе турецких исследователей [12] приведен краткий пример для иллюстрации работы их алгоритма. Пусть имеется 8 терминов, 6 документов и 2 класса (рис. 4). Релевантная матрица терминов-документов приведена в табл. 5.

Ячейки табл. 5 соответствуют весам терминов, вычисленным по формуле

$$nc_{i,k} = \sum_j n_{i,j}\,, \quad D_j \in C_k\,,$$

где $n_{i,j}$ — число встречаемости термина t_i в документе D_j и вычисляется суммарное число встречаемости термина t_i в классе C_k.

Далее авторы вычисляют веса терминов по классам:

$$w_{i,k} = \log\left(nc_{i,k} + 1\right) \log\left(\frac{N}{N_i}\right),$$

где определяется вес термина t_i (в документе D_j), который влияет на класс C_k.

Это порождает матрицу $I \times K$ (I терминов на K классов), приведенную в табл. 6.

После применения формул:

$$Y_k = \sum_i w_{i,k}\,, \quad w_i \in D_j\,, \qquad (1)$$

где вычисляется суммарный эффект терминов определенного документа на класс C_k, и

$$\text{newTerm}_k = \frac{Y_k}{\sum\limits_k Y_k}\,, \qquad (2)$$

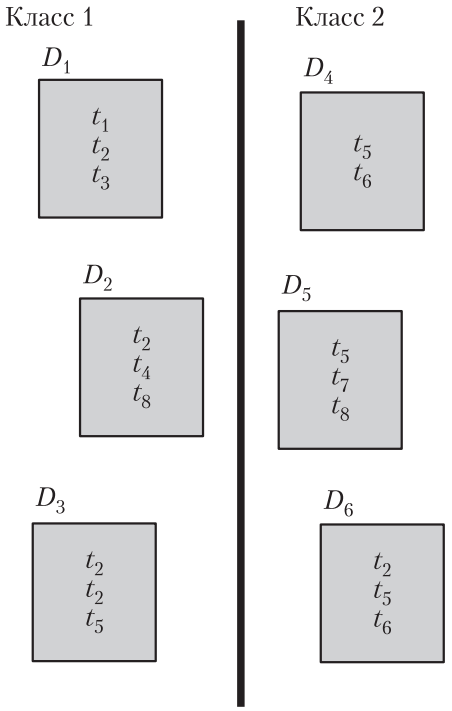


Рис. 4 Пример набора данных для демонстрации работы алгоритма

где нормализуются новые термины, была получена редуцированная матрица $J \times K$ с извлеченными характеристиками обрабатываемых документов (J документов на K классов).

Результирующая матрица приведена в табл. 7 и на рис. 5.

Таблица 5 Матрица терминов-документов для тестового набора данных

	D_1	D_2	D_3	D_4	D_5	D_6
t_1	1	0	0	0	0	0
t_2	1	1	2	0	0	1
t_3	1	0	0	0	0	0
t_4	0	1	0	0	0	0
t_5	0	0	1	1	1	1
t_6	0	0	0	1	0	1
t_7	0	0	0	0	1	0
t_8	0	1	0	0	1	0

Таблица 6 Веса терминов по классам для тестового набора данных

	C_1	C_2
t_1	0,234	0
t_2	0,123	0,053
t_3	0,234	0
t_4	0,234	0
t_5	0,053	0,106
t_6	0	0,228
t_7	0	0,234
t_8	0,144	0,144

Таблица 7 Веса документов по классам для тестового набора данных

	C_1	C_2
D_1	0,918	0,082
D_2	0,718	0,282
D_3	0,585	0,415
D_4	0,137	0,863
D_5	0,289	0,711
D_6	0,313	0,687

Извлеченные значения можно также воспринимать как вероятности принадлежности документов классам (см. рис. 5). Например, содержимое D_4 говорит о том, что этот документ на 84% принадлежит классу 2 и на 16% — классу 1. Таким образом, метод извлечения, изложенный в [12], может быть также использован в качестве самостоятельного классификатора. Когда авторы получают новый документ, их система принимает решение, к какому классу принадлежит этот документ, применяя формулы (1) и (2) к содержимому нового документа [12].

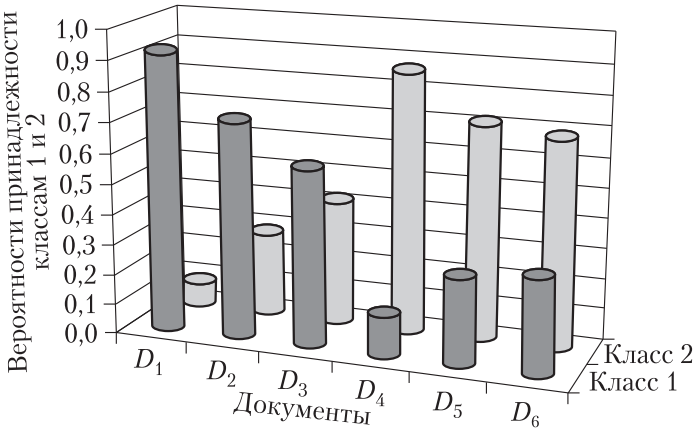


Рис. 5 Визуализация табл. 7

Результаты исследований [12] показали, что после применения редукционного алгоритма турецких исследователей время обработки классификационных процедур существенно сократилось при видимом повышении их точности (точные цифры приведены в [12]). Это было достигнуто путем сокращения числа атрибутов, соответствующих классификаторам.

2.4 Универсальный сетевой язык и семантические отношения и атрибуты

Богуславский в своей работе об особенностях гиперузлов (*hypernodes*) в UNL Interlingua уделяет достаточно много внимания именно семантическим аспектам представления знаний о языке (и на языке), причем независимо от выбора языка [17].

Универсальный сетевой язык (*the Universal Networking Language*, UNL), разработанный Хироши Учида (*Hiroshi Uchida*) в Институте передовых исследований Университета объединенных наций (*Institute for Advanced Studies of the United Nations University*), является языком представления для мультиязыковых коммуникаций в электронных сетях, при информационном поиске и в других приложениях.

Что же представляет собой выражение на языке UNL? Формально выражение на языке UNL является ориентированным гиперграфом, который соответствует предложению на естественном языке согласно объему передаваемой им информации [17]. Дуги этого графа интерпретируются как семантические отношения типа agent (агент), object (объект), time (время), reason (причина) и т. д. Узлы этого графа могут быть простыми или составными. Простые узлы являются особыми составляющими, так называемыми универсальными словами (*universal words*, UW). Универсальные слова жестко связаны с соответствующими концептами, понятиями. Они состоят из заглавного слова, которое обычно представляет собой английское слово или фразу, и множества ограничений, которые модифицируют смысловое значение заглавного слова или дополняют его какой-либо информацией. Составной узел, предмет статьи [17] (гиперузел — *hypernode*), состоит из простых или составных узлов, связанных семантическими отношениями.

Помимо отображения пропозиционального содержания (*who did what to whom* — кто что кому сделал), UNL-выражения предназначены для представления прагматической информации типа фокус, обращение/ссылка, отношение говорящего, его намерения, типы речевого акта и другие виды информации. Эта информация передается при помощи атрибутов, присоединенных к узлам, например, @imperative, @generic, @future, @obligation [17].

В качестве примера автор работы [17] приводит предложение на английском языке, его представление на UNL и его графическую репрезентацию (также средствами UNL) (пример 6, рис. 6).

Пример 6. *The actress whom we wanted to see so much did not appear on the stage in the first act*.

Предложение в виде UNL-выражения:

```
[S:1]
{org:el}
The actress whom we wanted to see so
much did not appear on the stage in the
first act.
{/org}
{unl}
mod(act(icl>dramatic_composition).@def,
first(icl>adj, ant>last))
agt(appear(icl>come>do,agt>thing,plc>
```

```
thing).@not.@past.@entry, :01)
plc(appear(icl>come>do,agt>thing,plc>
thing).@not.@past.@entry,
stage(icl>platform>thing).@def)
tim(appear(icl>come>do,agt>thing,plc>
thing).@not.@past.@entry,
act(icl>dramatic_composition).@def)
obj:01(see(icl>perceive>be, aoj>thing,
obj>thing, cob>uw):50,
actress(icl>actor>thing):39.@def.@entry)
man:01(want(icl>desire>be,cob>uw,obj>uw,
aoj>thing):Q0.@past,
so_much(icl>how):OR)
aoj:01(want(icl>desire>be,cob>uw,obj>
uw,aoj>thing):Q0.@past,we(icl>person))
obj:01(want(icl>desire>be,cob>uw,obj>
uw,aoj>thing):Q0.@past,see(icl>perceive>
be, aoj>thing, obj>thing, cob>uw):50)
{/unl}
[/S]
```

Упрощенная форма визуализации этого выражения приведена на рис. 6.

Выражение UNL какого-либо предложения, по сути, является набором отношений. Каждое отношение связывает два UW. В вышеприведенном выражении используются следующие отношения: agt (*agent* — агент, действующее лицо), obj (*object* — объект), aoj (*subject of a state* — состояние), man (*verb modifier* — глагольный модификатор), mod (*noun modifier* — модификатор существительного), tim (*time* — время) и plc (*place* — время). Большинству UW сопутствуют ограничения (выражения в скобках, следующие за заглавным словом) [17].

Эти ограничения в основном служат для указания семантического класса *UW* и сужения смыслового значения заглавного слова. Так, в примере 6 ограничение icl>come>do слова *appear* (появляться) определяет релевантное значение *appear* и одновременно показывает, что это слово связано с понятием действия как противоположного понятиям процесса или состояния [17]. Слово *first* (первый) имеет два ограничения: (1) icl>adj, которое сообщает, что *first* является понятием, обозначающим признак чего-либо (прилагательным); (2) ant>last, которое показывает, что это понятие противоположно понятию *last* (последний). Первое ограничение позволяет провести границу между этим значением *first* (первый) и другим значением (см. *first violin* — первая скрипка), для которого антонимичным понятием будет не *last* (последний), а *second* (второй) [17].

Атрибут @entry маркирует «главный» элемент структуры (у И. М. Богуславского она называется *entry node* — вводным узлом), который, как правило,

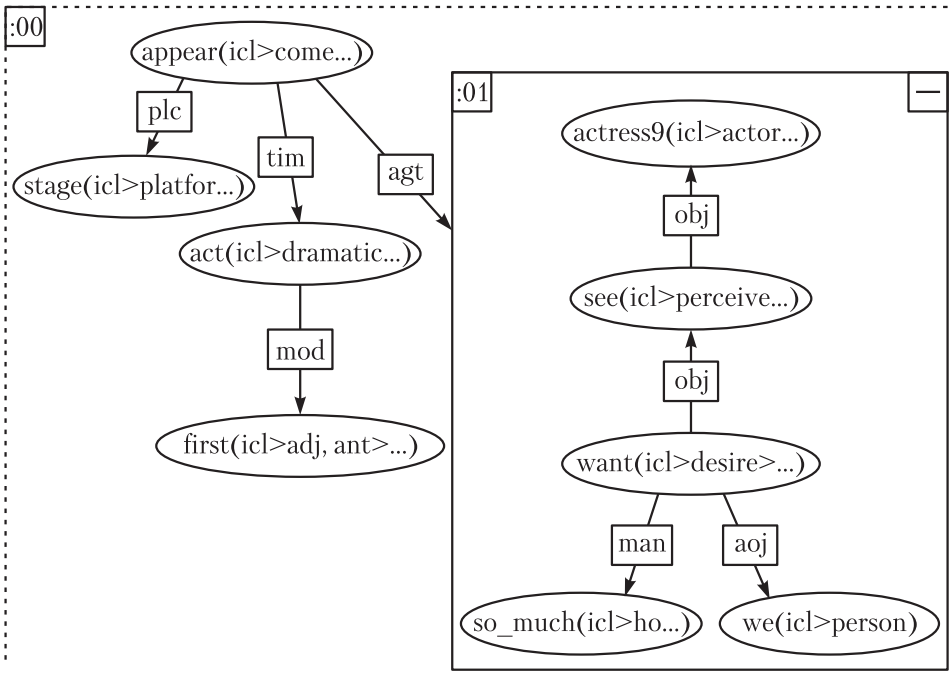


Рис. 6 Графическая форма представления UNL-выражения из примера 6

соответствует верхнему узлу предложения или определенной части предложения. Атрибут @def, приписанный в примере 6 к слову *stage* (сцена), показывает, что у слова есть определенная отсылка [17].

Выражение UNL предложения из примера 6 содержит гиперузел, который соответствует фразе «*the actress whom we wanted to see so much*» (актриса, которую мы хотели видеть так сильно), состоящей из существительного и связанного с ним придаточного предложения. Гиперузел имеет цифровой идентификатор (:01), который используется для представления гиперузла в виде отношений (как в agt(appear,:01)) и для маркирования отношений, которые он содержит (как в aoj:01(want,we)) [17].

Понятие гиперузла было введено автором [17]. Гиперузел — это множество узлов, которое в определенном смысле функционирует как одиночный узел. Таким образом, гиперузел может осуществлять все функции, свойственные одиночному узлу, т. е. может участвовать в отношениях (исходящих и входящих) и иметь свои собственные атрибуты. Богуславский описывает два главных типа гиперузлов в соответствии с задачами, которые они решают: семантически нагруженные (*semantically loaded*) и синтаксически нагруженные (*syntactic loaded*) [17].

Семантически нагруженные гиперузлы (далее — семантические) — это такие гиперузлы, которые передают информацию по разным смысловым значениям. Если такой узел не построен, то информация теряется, различий в смыслах не происходит. Поэтому семантические гиперузлы являются обязательными (см. пример 7).

Пример 7. *Culture and communication in the Internet.*

У предложения из примера 7 два смысла: (1) и культура, и коммуникации находятся в Интернете или (2) только коммуникации имеют место в Интернете (пример 7а). Для визуализации первого смысла необходимо, чтобы *culture* (культура) и *communication* (коммуникации) были включены в один гиперузел, соединенный с узлом *Internet*. Если не создано гиперузла для случая (1), теряется важная информация [17].

Пример 7а.

```
(1) and:01(communication, culture)
plc(:01, Internet)
(2) plc(communication, Internet)
and(communication, culture)
```

Семантически не нагруженные (или синтаксические) гиперузлы — это гиперузлы, которые передают информацию о синтаксической структуре предложения. Такие гиперузлы не различают смысловых нюансов, поскольку служат для удобства и понимания графов языка UNL. Синтаксические гиперузлы могут соответствовать придаточным предложениям, большим фразам, строкам и т. д. (см. пример 8). Если не строить таких узлов, структура UNL-графов будет менее прозрачной и более трудной для человеческого восприятия. При этом потери информации как таковой не происходит, т. е. синтаксические гиперузлы не являются обязательными [17].

Пример 8. *The state should take appropriate steps to return any cultural property imported after the entry into force of this convention, provided, however, that the requesting state shall pay just compensation to an innocent purchaser who has valid title to that property.*

В этом предложении может быть, по крайней мере, три синтаксических гиперузла:

(1) сложная номинативная фраза, состоящая из «*any cultural property and participial clause imported after the entry into force of this convention*»;

(2) обстоятельственное придаточное предложение «*provided, however, that the requesting state shall pay just compensation to an innocent purchaser who has valid title to that property*»;

(3) сложная номинативная фраза в составе (2), состоящая из «*an innocent purchaser and relative clause who has valid title to that property*».

Автор приводит примеры лингвистических явлений, которые можно описать при помощи гиперузлов:

1. **Выделение придаточных предложений** (см. пример 8).
2. **Расположение прилагательного в предложении:**
   so-called (*moral preoccupations*).
   Здесь гиперузел необходим для того, чтобы показать, что *so-called* (так называемый) относится ко всей фразе *moral preoccupations* (моральные предубеждения) и потому не может разделять ее на части.
3. **Общие зависимые слова** (см. пример 7).
4. **Определение, связанное с фрагментом предложения:**
   Food is cheap in England (*I believe*).
   Гиперузел использован, чтобы показать, что фраза *I believe* (я думаю, я верю) является вводной конструкцией. Этот гиперузел связан с атрибутом @parenth (*parenthetically*, т. е. для описания того, что заключено в скобки или выделено в запятых).
5. **Область действия отрицания:**
   (*They didn't sleep until midnight*.)
   Это предложение допускает две интерпретации:
   1) *they started sleeping at midnight* (они начали засыпать в полночь); нет гиперузла;
   2) *not that they slept until midnight, they woke up earlier* (не то чтобы они спали до полуночи, они проснулись раньше); здесь гиперузел содержит целое предложение и приписан к атрибуту отрицания;
6. **Обстоятельства в предложении:**
   He answered the dean very cleverly, т. е. *his answer was very clever* (его ответ был очень умным); нет гиперузла;
   Very cleverly (*he answered the dean*), т. е. *it was very clever of him to answer the dean* (с его стороны было очень умно ответить декану); здесь нужен гиперузел.

3 Заключение

В работе было подробно рассмотрено несколько публикаций, в той или иной степени описывающих актуальные, инновационные лексико-семантические

методы, примеры их реализации в различных задачах и области приложения. Анализ статей показал, что применение этих методов происходит в большом диапазоне предметных областей.

Так, в статье [9] описан алгоритм автоматического поиска терминов в Интернете в качестве ответов на сложные вопросительные предложения, для реализации которого авторами была сформирована база понятий и осуществляется трактовка семантики как вопросов и их типов, так и ответов на них (см. разд. 2.1).

Статья [10] описывает лексико-семантические методы верификации онтологии и лексиконов на основе онтологического семантического подхода, в частности методы автоматической и экспертной верификации новых лексиконов и онтологий, позволяющие оперировать семантикой естественно-языковых единиц (см. разд. 2.2).

Работа [12] содержит элементы лексико-семантического подхода к анализу текстов и извлечению знаний из них. Она посвящена влиянию нового алгоритма извлечения атрибутов на классификацию веб-страниц (см. разд. 2.3).

В завершение обзора рассмотрена статья [17], в которой приведено описание универсального сетевого языка (UNL), его гиперузлов и описанных при их помощи семантических атрибутов и отношений (см. разд. 2.4).

Таким образом, можно резюмировать, что в настоящее время особую актуальность получают глубинно-семантические подходы. Не в последнюю очередь это происходит потому, что сегодня качество и скорость обработки текстовых знаний вышли на новый уровень развития, при этом существующих подходов и средств искусственного интеллекта и прикладной лингвистики уже недостаточно. Это подтверждается и тем, что разброс инновационных лексико-семантических методов чрезвычайно велик, а в качестве областей применения выступают актуальные на сегодняшний день проблемы компьютерной лингвистики и другие сферы человеческой деятельности (проблематика интернет-приложений, классификация текстов, извлечение знаний, выстраивание связей между понятиями предметных областей на разных языках и многое другое).

Литература

1. *Carnie A.* Constituent structure // Oxford surveys in syntax and morphology. — Oxford: Oxford University Press, 2008.
2. *Kozerenko E.* Cognitive means for parallel texts alignment // WorldComp'2009 Proceedings. — Las Vegas: CREA Press, 2009. P. 473–478.
3. *Macmurray E., Leenhardt M.* Textometry information discovery: A new approach to mining textual data on the Web // WorldComp'2011 Proceedings. — Las Vegas: CREA Press, 2011. P. 605–611.
4. *Wierzbicka A.* Semantic primitives. — Frankfurt am Main: Atheneum Verlag, 1972. 235 p.

5. *Кустова Г. И., Падучева Е. В.* Словарь как лексическая база данных // Вопросы языкознания, 1994. № 4. С. 96–106.
6. *Апресян Ю. Д.* Избранные труды. Т. 1: Лексическая семантика. Синонимические средства языка. — М.: Языки русской культуры, 1995.
7. *Леонтьева Н. Н.* Категоризация единиц в русском общесемантическом словаре (РОСС) // Диалог'98: Труды Междунар. семинара по компьютерной лингвистике и ее приложениям. — Казань: Хэтер, 1998. Т. 2. С. 519–532.
8. *Kozhunova O.* Lexical and semantic methods in design of the problem-oriented linguistic resources // WorldComp'2011 Proceedings. — Las Vegas: CREA Press, 2011. P. 618–623.
9. *Watabe H., Imono M., Yoshimura E., Tsuchiya S.* An intelligent method for retrieval of verbal terms from the Web as answers in response to complex interrogative sentences // WorldComp'2011 Proceedings. — Las Vegas: CREA Press, 2011. P. 326–331.
10. *Taylor J. M., Hempelmann Ch. F., Raskin V.* Post-logical verification of ontology and lexicons: The ontological semantic technology approach // WorldComp'2011 Proceedings. — Las Vegas: CREA Press, 2011. P. 529–534.
11. *Nirenburg S., Raskin V.* Ontological semantics. — Cambridge, MA: MIT Press, 2004.
12. *Biricik G., Diri B.* Impact of a new attribute extraction algorithm on Web page classification // WorldComp'2009 Proceedings. — Las Vegas: CREA Press, 2009. P. 481–485.
13. *Pearson K.* On lines and planes of closest fit to systems of points in space // Philosophical Magazine, 1901. Vol. 2. No. 6. P. 559–572.
14. *Landauer T. K., Dumais S. T.* Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge // Psychological Review, 1997. Vol. 104. No. 2. P. 211–240.
15. *Уилкинсон Д. Х.* Алгебраическая проблема собственных значений. — М.: Наука, 1970. 564 с.
16. *Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P.* Singular value decomposition // Numerical Recipes in C. 2nd edition. — Cambridge: Cambridge University Press, 1997.
17. *Boguslavsky I.* Hypernodes in the UNL Interlingua // WorldComp'2011 Proceedings. — Las Vegas: CREA Press, 2011. P. 612–617.

PROGRAM-ORIENTED INDICATORS: PRODUCTION AND APPLICATION IN SCIENCE*

I. Zatsman[1] *and A. Durnovo*[2]

Abstract: The results of the project, whose goal is to create a tool for developing sets of indicators for monitoring and evaluating research and developments (R&D) program activities are presented. Each set of developed science and technology (S&T) indicators is oriented at the assessment of only one R&D program, i. e., at its objectives, resources, and results. The tool for developing indicators is being designed as an experts knowledge base named the proactive dictionary, which is the component of the evaluation system. This component enables experts to fix stages of indicators development, to present the results of developing different variants of indicators in graphic form, to compare and evaluate those variants, which are oriented at objectives, resources, and results of R&D program activities. The paper describes two theoretical models of the process of indicators development. It also describes a prototype of the proactive dictionary created on the base of these models. Finally, it describes the results of an experiment of development of a set of indicators by a group of the experts.

Keywords: processes of experts' knowledge creation about indicators; knowledge representation about indicators; semiotic models; program-oriented indicators; indicator denotata

Introduction

In 2004–2010, the Russian government issued a number of documents regulating the monitoring and the evaluation of R&D activities. According to those documents, the scope of indicators, which monitor and evaluate results, efficiency, and effectiveness of R&D programs, must be significantly expanded. Use of a traditional set of indicators sometimes demonstrates its irrelevance, because they do not take into account objectives and results of specific R&D programs. This has stimulated the demand for the development and application of a set of *program-oriented indicators* [1–4].

The need to develop relevant indicators has made urgent the creation of new conceptual approaches, methods, and models for processes of indicators production [5].

*This research is funded by RFH grant No. 12-02-00407. It is planned to use these research results in an international project to develop a corpus-oriented typology of the translation difficulties.

[1]Institute of Informatics Problems of the Russian Academy of Sciences, iz_ipi@a170.ipi.ac.ru

[2]Institute of Informatics Problems of the Russian Academy of Sciences, duralex49@mail.ru

The initial stages of the development of indicators are usually highly personalized. In other words, at these stages, their semantic content is often a purely personal, since it is determined only by a single expert, who is developing a new indicator in the initial stages. The semantic content of the new indicator becomes collective at later stages, at which the expert-developer begins the process of coordination of its semantic content with other experts. It is important to note that, at the very beginning in the development, an expert creates, as a rule, a new algorithm for computing the values of the indicator. This algorithm is developed by an expert according to his/her personal view of the appointment of the indicator. This view is not usually formalized. Moreover, it is often implicit, that is, it is not expressed explicitly by the expert until the writing of the algorithm. These features of the development of indicators lead to the following requirements to the modeling of the processes of their development:

(1) models have to distinguish at what stages of development the expert's knowledge is implicit and at what — explicit;

(2) models have to distinguish between the stages of personal development and the stages of coordination of the semantics of new developed indicators;

(3) models have to fix the time points of an explication in different forms (algorithm, program, text in a natural language, diagrams, etc.); and

(4) models have to fix the time points of the coordination of a personal interpretation of a developed indicator between experts and the formation of a collective interpretation.

If one considers only the first two requirements, it meets the qualitative approach to modeling, which has been suggested by Ikujiro Nonaka [6]. All four requirements are satisfied by two models, a brief description of which is presented in this article. Here, their description is focused on modeling the development process of indicators, as well as their application. However, these models are applicable not only to indicators development. It is planned to use them in an international project to develop a corpus-oriented typology of the difficulties of translation from Russian into French [7].

Thus, the paper describes two theoretical models of indicators development. It also describes a prototype of the proactive dictionary created on the base of these models and the first results of an experiment where a set of indicators is being developed by a group of experts. The development was carried out under the First Russian Academic (FRA) Program, which was adopted by the Government of the Russian Federation in February 2008 as a tool of public intervention in the area of science. The Russian Academy of Sciences (RAS) has decided to create an evaluation system prototype for providing assessment of the FRA Program as a whole and of its projects.

The incompleteness of the indicators set for the FRA Program has emerged as a result of the application of the Objectives-Resources-Results approach describing relationships between the program objectives, resources, and results [2]. According to the approach current, new, and future indicators were divided into 6 categories for (i) objectives; (ii) results; (iii) resources; ($i\nu$) efficiency (relationship between resources and results); (ν) effectiveness (relationship between objectives and results); and (νi) consistency (relationship between objectives and resources). According to the Special Report [8], the FRA Program results were divided into three classes: outputs, outcomes, and impacts; hence, result indicators were subdivided into three corresponding subcategories as well.

Besides this, a need for corresponding between 6 indicator categories and stages of the FRA Program assessment has emerged. Developers of the evaluation system prototype used the stages and substages definitions according to the Special Report [8]. These stages are ex-ante, mid-term, and ex-post. The latter, in its order, is divided into three substages: short (1–2 years after the FRA Program termination), medium (4–7 years), and long terms ($>$ 10 years).

At present time, for the FRA Program assessment, decision-makers are using two indicator categories only, namely, resources and results indicators, and among the results indicators, they are using output and outcome indicators. As for the stages, indicators on only two stages (mid-term and short term ex-post) are analyzed by decision-makers. All this demonstrates the existence of the problem of incompleteness of the indicators set [2].

In order to solve the problem and to develop new program-oriented indicators, the present authors are designing the proactive dictionary as a tool for indicator development. The dictionary is a part of the evaluation system prototype for providing assessment of the FRA Program. The design of the dictionary is based on two semiotic models for the description of the indicators development stages, including microprocesses of generation of expert knowledge about developed indicators:

- the model for the description of a frosen state of a generation microprocess, named the *frozen-state model* [9]; and
- the model for a description of dynamics of generation microprocesses, named the *time-dependent model* [10].

Two Models of Indicator Development

The idea of the frozen-state and time-dependent models has arisen in the course of study of generation microprocesses of knowledge and their creative space [6, 11–14]. These models are based on the Frege's triangle, which consists of concept, name, and denotatum vertices. For each new developed indicator, these three vertices of the triangle are the sign-meaning (the indicator concept), the sign-form (the indicator

name), and the denotatum of the sign (both an indicator computer program and source data).

According to the frozen-state model, the description of a state of new developed indicator is made by experts as a descriptor of the proactive dictionary at a discrete point in time. In the model, three vertices of the triangle are encoded by the evaluation system prototype into three following computer codes, recorded into the proactive dictionary at the discrete point in time:

(1) a semantic code for the indicator concept;

(2) an information code for the indicator name; and

(3) an object code for the indicator denotatum.

The main design of computer coding of each new emerging indicator concept, its name, and denotatum is that each developed indicator is analyzed and described in the proactive dictionary from three points of view:

(1) as the emerging indicator concept;

(2) as the variable indicator name; and

(3) as the changeable indicator denotatum.

In two models, each changeable indicator denotatum is a computer program of indicator computation together with corresponding source data, which are information resources of the evaluation system prototype. The indicator names are given and changed by the experts, developing indicators.

The coding design must take in mind that, during the development process of any indicator, its emerging concept description, name, and denotatum may be changed by the experts in great extent. Moreover, a key stage of the development process of any indicator is the coordination of a personal concept description between the experts to form a collective (group) concept description of its indicator.

The time-dependent model takes into account the variability of emerging indicator concepts, their names, and denotata. The model is based on the frozen-state model. To construct the time-dependent model, a set of points named the "Frege's space" is defined as a *trace space* for describing emerging indicator concepts and their evolution. This space is built to display development stages of indicators, including generation microprocesses of expert knowledge about developed indicators. The purpose of the Frege's space is to represent how all developed indicators change in time, using sequences of semantic, information, and object computer codes. The Frege's space has three axes of reference: semantic, information, and object code axes, respectively, as well as the fourth one — the time axis [10].

The time-dependent model describes generation microprocesses of expert knowledge about each developed indicator as *one generation trace* in the Frege's space at the successive moments of time t_i $(i = 1, 2, \ldots)$ where t_i is the ith stage of indicator development. It is supposed that at each t_i, for each developed indicator, which

changed at t_i, experts describe a frozen-state of the developed indicator and the evaluation system prototype generates three computer codes: semantic, information, and object ones. The Frege's space is a set of all points of generation traces of all developed indicators. The space is built according to the following requirement: at the discrete point in time t_i, any these three computer codes must correspond to just one frozen state and one point of the space.

Proactive Dictionary as a Part of the Evaluation System

The users of the proactive dictionary are to be only the experts, responsible for development of new indicators. Experts use the proactive dictionary as a tool for indicator development to describe a dynamics of generation microprocess of own knowledge about developed indicators. Each state of any developed indicator is described by experts as a descriptor of the proactive dictionary. At discrete points in time t_i $(i = 1, 2, \ldots)$, each expert may create or modify descriptors of the proactive dictionary. Experts can access descriptors of the proactive dictionary through the expert console software of the evaluation system prototype and receive information about any frozen state of each developed indicator. In the proactive dictionary, the following events of indicator development are fixed:

- creation of a descriptor for each uninterpreted or interpreted frozen state of any developed indicator, including its definition and a name (with indication of their authors from a group of experts);
- establishment (change) of links between descriptors; and
- establishment (change) of descriptor references to the programs that calculate the values of the developed indicators and references to their source data.

Thus, each descriptor of the projective dictionary contains an indicator frozen-state definition, its name, three computer codes and links to other descriptors, as well as references to source data, which are information resources of the evaluation system prototype (see references to information resources at t_1 and t_n in Fig. 1) and computer programs[1], which evaluate values of developed indicators (see references to program resources at t_1 and t_n).

The development of any indicator includes two main phases: personal (one expert creates an indicator) and collective (an indicator is coordinated by several experts of the group). All the experts may refuse an indicator; then it is considered to be nonactual, but still resides in the proactive dictionary. Besides the projective dictionary, which is periodically enriched by new frozen states of developed indicators, linguistic resources include a semantic dictionary, which is accessed by all users of the evaluation system prototype.

[1]These programs were developed by V. Kosarik.

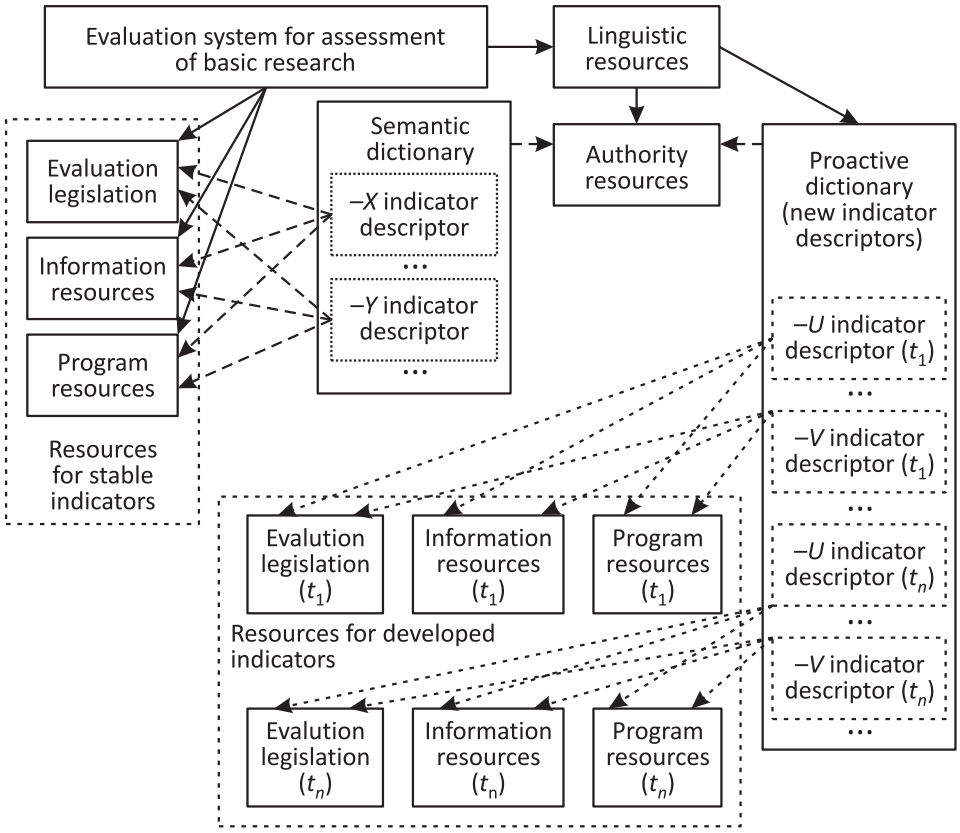


Figure 1 Semantic and proactive dictionaries of the evaluation system prototype

All descriptors of the semantic dictionary do not change in time. While the proactive dictionary deals with time-dependent indicators, the semantic dictionary describes the stable indicators and has stable links and references (see Fig. 1). Each time-dependent indicator can potentially turn into a stable one and be incorporated into the semantic dictionary if this indicator is coordinated between the experts and confirmed by decision-makers.

The proactive dictionary suggests the possibility of a return to any frozen states of developed indicators, because it keeps all their frozen states which have ever existed. For the proactive dictionary, operations of modifying or deletion of descriptors are invalid: the modification of a descriptor leads just to the generation of a new descriptor, while maintaining the old one. So, it is more convenient to consider a new descriptor as a successor of the original. As a new descriptor may be the heir

to the multiple source descriptors, such inheritance can be multiple in the projective dictionary.

Development of a Set of Indicators

Here, an experiment on the development of a set of indicators of the authors' age distribution of the articles of participants of the R&D thematic subprogram of the FRA Program named "Basic research for medicine" is performed. One of the goals of this subprogram is to attract young scientists to participate in its projects. In total, the subprogram held about 140 projects. The total number of its participants is about 1,500. One of the approved indicators of all FRA thematic subprograms represents a share of participants under the age of 39. But this indicator says nothing about the productivity of young scientists. Therefore, the present authors proposed to develop a set of indicators to calculate the authors' age distribution of the articles, dividing the participants into 14 age groups (20–24, 25–29, and up to 85–89 years).

The results of calculation of the set of developed indicators are given in the form of graphs (Fig. 2). Here, the first five stages of the experiment are described. Five experts take part in the indicator development: A, B, C, D, and E.

First stage. Expert A creates the first variant of the indicator having the following characteristics:

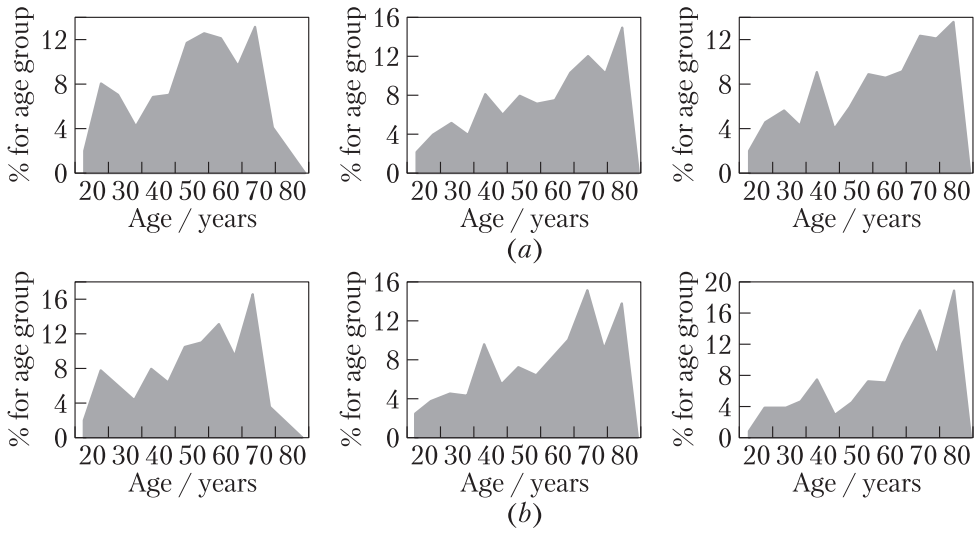


Figure 2 Results of calculation of the first (*a*) and second (*b*) sets of the developed indicators: left column — first stage; middle column — third stage; and right column — fourth stage

– to calculate the values of the indicator, the articles from all journals, issued in 2009 and entered into the database of the evaluation system prototype, are used;

– if an article has been written by co-authors, corresponding age groups receive 1 point for each author; and

– the normalization procedure using the size of age groups is not applied.

At the same time (at the first stage), experts B and C create together the second variant of the indicator with the same first and third characteristics, but the second one has a different value: corresponding age groups receive $1/N$ point for an article having N coauthors.

Second stage. Now, the expert C changes his/her mind and agrees with the expert A. In other words, the expert C refuses to coordinate the second variant of the indicator, thinking it is right to add just 1 point to each of coauthors, thus coordinating the first variant of the indicator. So, the first variant becomes a collective one, while the second variant becomes personal. The variants themselves are identical to those of the first stage (these variants are not shown in Fig. 2).

Third stage. Experts A, B, and C decide to take into account, when calculating values of their indicators, the number of people in each age group. It is reflected in changes of corresponding algorithms — the normalization is used in the calculations.

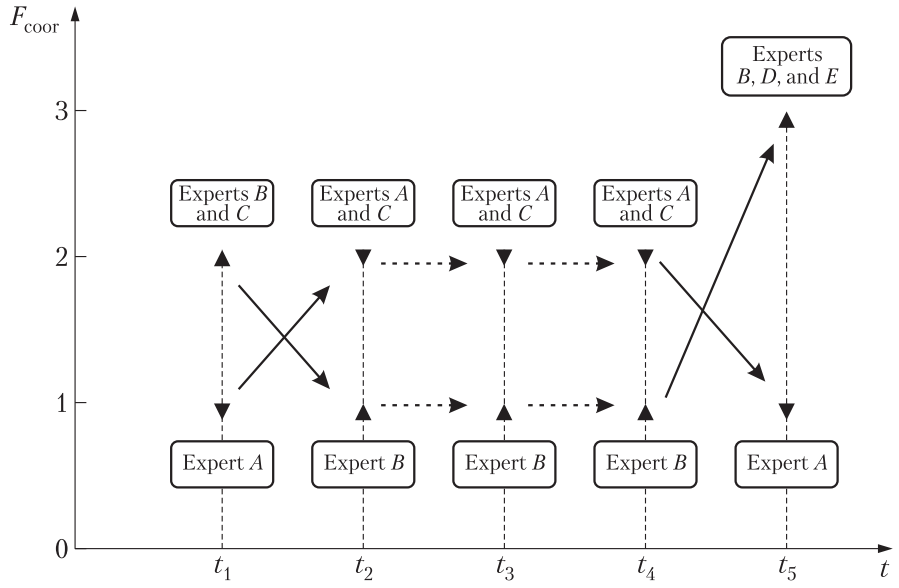


Figure 3 Coordination level function F_{coor} for the first five stages of the experiment

Fourth stage. Experts A, B, and C decide to take into account only articles printed in the journals from the rating list of the Nation Certification Commission of Russia.

Fifth stage. Now, the experts remain unchanged states of their variants of developed indicators. Yet, the points of view of some of them change. The expert C refuses the variant of the expert A; so, this concept becomes a personal concept of the expert A. Two other experts, D and E, join the second variant of the indicator. This changes its level of coordination (these variants are not shown in Fig. 2).

In the course of the experiment, the coordination level function F_{coor} was determined. The function for the first five stages of the experiment has been constructed (Fig. 3).

Concluding Remarks

Aiming to create a tool for developing the sets of indicators for monitoring and evaluating R&D activities, the proactive dictionary as a part of the evaluation system prototype has been introduced and applied.

The new notion of "Frege's space" has been introduced for quantitative representation of changeable indicator denotata, their emerging concepts and variable names during development of new indicators. The obtained results, considered in the paper, consist of:

- two theoretical models of the processes of indicators development;
- the proactive dictionary, created on the base of these models;
- results of an experiment of development of a set of indicators by a group of experts; and
- the definition of the coordination level function F_{coor}.

The proactive dictionary shows new possibilities for developing indicators for evaluating R&D thematic subprograms of the FRA Program, specifying states of developed indicators as descriptors. Development of indicators is not the only goal of the proactive dictionary design. Its implementation provides the solution of other important tasks. First, each descriptor of the proactive dictionary sets a one-to-one correspondence between the program that calculates values of the indicator and its name. In other words, each name of any indicator corresponds to a unique program calculating its values. Second, each descriptor fixes a rule of selecting source data which are used in calculating the values of the indicator. That is why the process of confirmation of any new indicator is a simultaneous approval of the following entities: its name, its rubric(s) in a classification system, its definition, the computer program of calculating its values, and the rule of selecting source data used in calculating its values.

References

1. *Zatsman I., Kozhunova O.* Evaluating for institutional academic activities: Classification scheme for R&D indicators // Abstracts book of the 10th Conference (International) on Science and Technology Indicators (STI'2008). — Vienna: ARC GmbH, 2008. P. 428–431.
2. *Zatsman I., Kozhunova O.* Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators // E-print Proceedings of the International Conference ATLC'2009 (Atlanta Conference on Science and Innovation Policy). 2009. http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf.
3. *Zatsman I., Durnovo A.* Incompleteness problem for indicators system of research program // Abstracts book of the 11th Conference (International) on Science and Technology Indicators (STI'2010). — Leiden: Universiteit Leiden, 2010. P. 309–311.
4. *Zatsman I., Durnovo A.* Modelling of processes for creation of expert knowledge for monitoring of goal-oriented program activities // Informatics and Its Applications, 2011. Vol. 5. No. 4. P. 84–98. [In Russian; abstract in English from: http://www.ipiran.ru/journal/issues/2011_04_eng/annot.asp.]
5. *Lepori B., Barre, R., Filliatreau G.* New perspectives and challenges for the design and production of S&T indicators // Research Evaluation, 2008. Vol. 17(1). P. 33–44.
6. *Nonaka I.* The knowledge-creating company // Harvard Business Review, 1991. Vol. 69(6). P. 96–104.
7. *Buntman N., Minel J.-L., Le Pesant D., Zatsman I.* Typology and computer modelling of translation difficulties // Informatics and Its Applications, 2010. Vol. 4. No. 3. P. 77–83.
8. Special report No. 9/2007 concerning "Evaluating the EU Research and Technological Development (RTD) framework" // Official J. European Union, 2008. C26(30.01.2008). P. 1–38.
9. *Zatsman I.* A semiotic model of correlations between concepts, information objects and computer codes // Informatics and Its Applications, 2009. Vol. 3. No. 2. P. 65–81. [In Russian; abstract in English from: http://www.ipiran.ru/journal/issues/2009_02_eng/annot.asp.]
10. *Zatsman I.* Time-dependent semiotic model of computer coding of concepts, information objects and denotata // Informatics and Its Applications, 2009. Vol. 3. No. 4. P. 87–101. [In Russian; abstract in English from: http://www.ipiran.ru/journal/issues/2009_04_eng/ annot.asp.]
11. *Nonaka I., Takeuchi H.* The knowledge-creating company. — New York: Oxford University Press, 1995.
12. *Wierzbicki A., Nakamori Y.* Basic dimensions of creative space // Creative space: Models of creative processes for knowledge civilization age / Eds. A. Wierzbicki and Y. Nakamori. — Heidelberg: Springer, 2006. P. 59–90.
13. *Wierzbicki A., Nakamori Y.* Knowledge sciences: Some new developments // Zeitschrift für Betriebswirtschaft, 2007. Vol. 77(3). P. 271–295.
14. *Ren H., Tian J., Nakamori Y., Wierzbicki A.* Electronic support for knowledge creation in a research institute // J. Syst. Sci. Syst. Engng., 2007. Vol. 16(2). P. 235–253.

РАЗРАБОТКА И ПРИМЕНЕНИЕ ПРОГРАММНО-ОРИЕНТИРОВАННЫХ ИНДИКАТОРОВ В СФЕРЕ НАУКИ

***И. М. Зацман*[1], *А. А. Дурново*[2]**

[1]Институт проблем информатики Российской академии наук, iz_ipi@a170.ipi.ac.ru
[2]Институт проблем информатики Российской академии наук, duralex49@mail.ru

Аннотация: Рассмотрены результаты создания моделей и средств разработки индикаторов мониторинга и оценивания программно-целевой научной деятельности. Каждый разрабатываемый набор индикаторов ориентирован на оценивание некоторой научной программы, в том числе ее целей, задач, ресурсов и результатов. Средство разработки индикаторов представляет собой базу экспертных знаний, получившую название проективного словаря, который является компонентом системы мониторинга. Этот словарь позволяет экспертам фиксировать стадии разработки индикаторов, представлять результаты поэтапной разработки в графической форме, сравнивать и оценивать варианты индикаторов целей, задач, ресурсов и результатов программы. Представлено краткое описание двух теоретических моделей процесса разработки индикаторов и опытного образца проективного словаря, созданного на основе этих моделей. Рассмотрены результаты эксперимента по разработке индикаторов группой экспертов.

Ключевые слова: процессы формирования экспертных знаний об индикаторах; представление знаний об индикаторах; семиотические модели; программно-ориентированные индикаторы; денотаты индикаторов

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2012. Т. 22. № 1. С. 121–131

ПРЕДМЕТНЫЕ СЛОВАРИ: НАЗНАЧЕНИЕ, ОСОБЕННОСТИ И ПЕРСПЕКТИВЫ

***Н. В. Сомин*[1], *И. П. Кузнецов*[2], *М. М. Шарнин*[3], *В. Г. Николаев*[4]**

Аннотация: Объектом исследования и разработки является семантико-ориентированный лингвистический процессор (ЛП) для автоматической формализации текстов на естественном языке (ЕЯ) с расширенными функциональными возможностями. Процессор обеспечивает извлечение знаний из текстов для конкретных категорий пользователей, которые интересуются определенными объектами, их признаками и связями. В данной статье рассматривается развитие таких процессоров в направлении усовершенствования лингвистических знаний, к которым относятся предметные словари. Они используются в процессе лексико-морфологического анализа (ЛМА) для придания словам и словосочетаниям семантических признаков, что снимает многие неопределенности в процессе последующего анализа текста.

Ключевые слова: извлечение знаний; лингвистический процессор; лексико-морфологический анализ; предметные словари

1 Введение

На протяжении последних 20 лет в ИПИ РАН развивается научное направление, связанное с разработкой *семантико-ориентированных ЛП*. Другое название — *объектно-ориентированные ЛП*. Такие ЛП осуществляют извлечение знаний из произвольных текстов на ЕЯ в определенной предметной области с их отображением на структуры знаний — *расширенные семантические сети* (РСС) [1–3]. Учитывается тот факт, что большинство пользователей — специалистов в своей области — интересуется лишь конкретными вещами для своих задач. Например, следователям важны фигуранты, их местожительства, телефоны, приметы, действия лиц (с указанием места, времени), связи и др. Специалистов по кадрам интересуют организации, где и кем человек работал, когда это было. Подобную информацию будем называть *информационными объектами* (другое название — *именованные сущности*), которые различаются по типам. Например, лица и фигуранты — это объекты одного типа, адреса — другого и т. д.

[1]Институт проблем информатики Российской академии наук, somin@post.ru
[2]Институт проблем информатики Российской академии наук, igor-kuz@mtu-net.ru
[3]Институт проблем информатики Российской академии наук, keywen1@mail.ru
[4]Институт проблем информатики Российской академии наук, DHLine@yandex.ru

Типология объектов, интересующих пользователя, определяется его предметной областью и классом решаемых задач.

Семантико-ориентированный ЛП осуществляет глубинный анализ текстов на ЕЯ, из которых извлекаются объекты и их связи. Такой ЛП содержит два основных блока — ЛМА и синтактико-семантического анализа (ССА).

Блок ЛМА выделяет из документа слова и предложения и выдает в виде семантической сети (ПС документа) последовательность компонентов (слов в нормальной форме, чисел, знаков) и их основные признаки — лексические, морфологические, семантические. Блок ЛМА имеет свои средства параметрической настройки и использует набор *предметных словарей* (словарь стран, регионов России, имен, профессий и др.) для придания словам и словосочетаниям дополнительных семантических признаков [4, 5].

Блок ССА путем анализа ПС документа (последовательности слов и их признаков) выделяет информационные объекты и их связи. На этой основе блок строит другую семантическую сеть, представляющую *семантическую структуру документа* (СС документа), называемую также *содержательным портретом* [1–3]. Такие портреты представляют собой структуры знаний, которые запоминаются и образуют предметную базу знаний (БЗ). Последняя является основой для реализации различных видов семантических поисков и логико-аналитических решений.

Блок ЛМА играет важнейшую роль при работе ЛП. Любые неопределенности, возникающие при работе этого блока, требуют усложнения блока ССА для их разрешения. И в то же время дополнительные признаки, формируемые блоком ЛМА (особенно семантические), во многих случаях позволяют упростить процедуру выделения объектов и связей, осуществляемую блоком ССА, обеспечить устранение неоднозначностей.

При формировании семантических признаков особую роль играют *предметные словари*. В процессе настройки ЛП на различные предметные области такие словари приобрели много новых функций, необходимых для повышения качества работы ЛП в процессе извлечения знаний. В данной статье будут рассматриваться эти функции и методики их реализации.

Отметим, что в настоящее время область «извлечение знаний из текстов» активно развивается. Рекламируется множество систем (ЛП), которые различаются глубиной анализа и качеством результатов. Многие из таких систем ориентированы на выделение ограниченного числа объектов. Как правило, это лица, организации, адреса и даты. Такие системы сравнительно легко реализуются программными средствами. Для выделения других объектов, их связей и действий требуется более глубокий анализ, где необходимо учитывать значительно большее число признаков. Кроме того, требуется специальная организация систем для их настройки на корпуса текстов. В качестве средств настройки вводятся лингвистические знания [1, 2]. В семантико-ориентированных ЛП такие зна-

ния включают в себя предметные словари, которые необходимы для глубинного анализа текстов на ЕЯ — русском и английском.

2 Функции предметных словарей

Предметные словари содержат конкретную информацию об исследуемой предметной области. В простейшем случае предметный словарь представляет собой список слов и словосочетаний (*терминов*), которые являются названиями информационных объектов или же являются ключевыми словами, необходимыми для выделения объектов. В более сложных случаях вместо терминов используются более сложные конструкции, покрывающие множество терминов (см. разд. 3).

Пример фрагмента, взятого из словаря организаций (ORG_K), представлен на рис. 1. Слова «*университет*», «*учреждение*», «*училище*» и др. являются ключевыми. Словосочетания типа «*Университет Дружбы Народов*» есть название организации.

Поиск терминов словаря в тексте какого-либо документа осуществляется блоком ЛМА. Будем называть поиск *термина* его *идентификацией*. Термин найден (или идентифицирован), если найдены соответствующие слова в тексте.

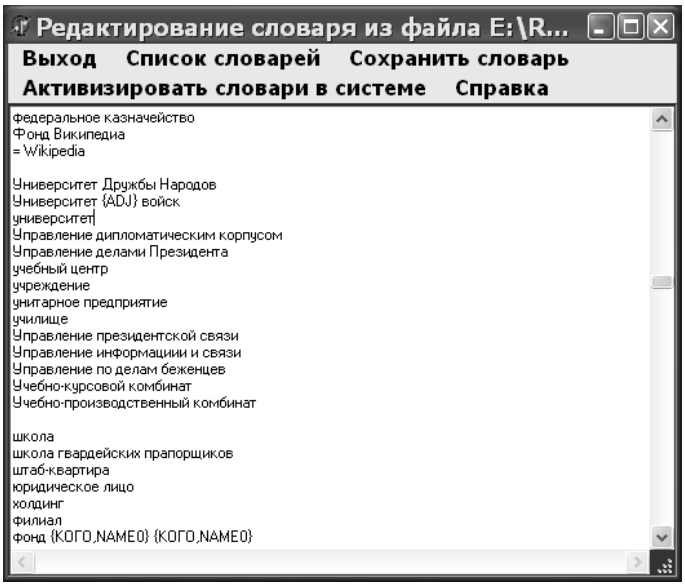


Рис. 1 Предметный словарь организаций (ORG_K)

Тогда в ПС документа этим словам присваивается семантический признак — в зависимости от того, в каком словаре этот термин находится. Например, слову «*университет*» будет присвоен признак организации — ORG_K. Система (блок ССА) опирается на такие признаки. С них начинается распознавание объектов.

При выделении объектов определенного типа используются свои словари. Например, при выделении лиц используется словарь имен собственных (NAME_K). В системе имеются словари, в которых терминами являются названия информационных объектов. Например, таким является словарь ФИО (FIO_K), содержащий список наиболее значимых или часто встречающихся лиц. При их идентификации на основе слов, найденных в тексте, сразу формируется объект, который помещается в ПС документа. Не требуется специальных правил их выделения блоком ССА (см. разд. 6).

Итак, имеет место два случая идентификации терминов. Первый, когда термин словаря — это отдельное слово, не являющееся названием объекта. Тогда при идентификации соответствующему слову текста присваивается семантический признак, который при последующем анализе инициирует распознавание объекта. Второй случай, когда термин — это словосочетание. Тогда при его идентификации в тексте ищется набор слов — такое же словосочетание. При этом найденные слова связываются между собой и образуют единое целое, которому присваивается семантический признак, определяемый словарем.

До идентификации терминов текст представляет собой последовательность слов (лексем), имеющих лишь общеязыковую семантику. Но после идентификации эти слова связываются между собой и приобретают иной статус — они или становятся названием информационного объекта, или инициируют формирование такого объекта.

В системе допускается произвольное число словарей, каждый из которых может участвовать в обработке текста. Каждый словарь в системе получает имя, определяющее семантический признак, который дается при идентификации его терминов. Набор словарей, участвующих в анализе, можно гибко менять от текста к тексту.

Каждый словарь записывается в отдельном файле. Множество словарей образует взаимосвязанную словарную систему, которая находится в отдельной директории Morf_tab (рис. 2). Словарь City_k.slv содержит список городов, словарь Drug_k.slv — названия наркотических веществ и т. д.

К настоящему времени накоплена база из 20 предметных словарей, используемых для формализации текстов в различных предметных областях. Большинство из них имеет достаточно универсальный характер и может быть использовано для решения многих задач. Опыт эксплуатации показал, что словари существенно повышают качество ССА и глубину анализа текстов.

Заметим, что система предметных словарей фактически является БЗ о предметной области, своеобразной онтологией. Однако специфика этой онтологии в

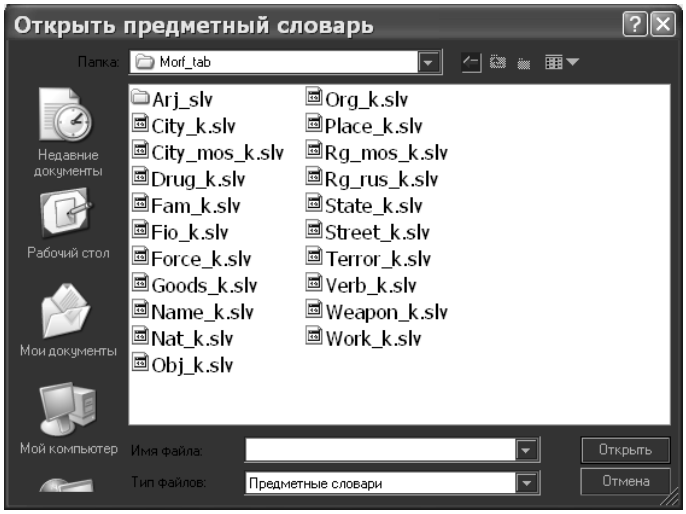


Рис. 2 Система предметных словарей

том, что она ориентирована на решение задачи выявления из текстов информационных объектов и связей между ними.

Система предметных словарей представляет собой достаточно сложную программную систему. В нее входят следующие программные блоки:

- блок ввода терминов и перевода их во внутреннюю словарную форму;
- блок идентификации словарного термина в тексте;
- блок управления словарными данными на физическом уровне (словарная система управления базами данных (СУБД)).

Рассмотрим возможности и алгоритмы работы каждого из этих блоков.

3 Возможности описания терминов

Информация, которая интересует пользователя, может быть весьма разнообразной. Это относится не только к типологии объектов, но и к объектам одного типа. Их может быть значительное количество с большим разнообразием форм записи. Имеются в виду случаи, когда у одной организации имеется множество названий. То же самое имеет место для профессий, видов оружия, автомобилей, компьютеров и др. При этом большое разнообразие затрудняет создание списка терминов словаря, необходимых для выделения объектов определенного типа. Отсюда вытекает необходимость совершенствования функциональных возможностей предметных словарей.

Во-первых, для уменьшения числа терминов оказалось необходимым ввести в словари «обобщенные слова», покрывающие группу однотипных терминов. Во-вторых, в документах часто встречается описание лиц (и других объектов), которые именуются по-разному. Такие описания нужно приводить к одному виду. И наконец, как оказалось, один и тот же термин может быть ключевым для объектов различных типов и, соответственно, помещаться в разных словарях, что делает неоднозначной его идентификацию. Эти факторы определили спектр возможностей, которыми должны обладать предметные словари.

Для реализации перечисленных возможностей в предметные словари (содержащие термины) были введены специальные конструкции. С их помощью учитываются случаи, когда один и тот же объект может быть выражен разными терминами, когда один и тот же термин принадлежит разным предметным словарям и др. Термины в словаре могут быть однословными или многословными. Они могут содержать любые символы из набора, имеющего место в современных компьютерах.

Термины в словарях задаются в каноническом виде (единственное число, именительный падеж). Впрочем, если автор словаря задаст термин как-то иначе, то блок ввода исправит термин и поставит его в каноническом виде — в блок встроена морфологическая подсистема [4, 6].

В словарях обеспечивается возможность задания группы (веера) терминов. Это делается с помощью формальных конструкций, встроенных в термин и содержащих вместо слов их признаки — лексические, морфологические (части речи и падежи) и семантические. Каждая такая конструкция играет роль «обобщенного слова», покрывающего множество слов с указанным признаком. Номенклатура признаков управляется специальным настроечным файлом и легко может пополняться. Указание «обобщенного слова» возможно для второго и последующих слов термина (только не для первого).

Рассмотрим варианты формальных конструкций.

1. В термин вместо слова вводится его лексический тип.

   Например, *аэродром* {NAMEO}, где NAMEO — признак слова с большой буквы.

   С помощью такой конструкции задаются случаи, когда за словом «*аэродром*» стоит слово с прописной буквы: *аэродром Внуково*, *аэродром Домодедово* и т. д.

2. В термине вместо слова указывается его морфологический признак.

   Например, *памятник-бюст* {КОМУ}.

   Задаются случаи, когда за словосочетанием «*памятник-бюст*» стоит слово в дательном падеже: *памятник-бюст Кутузову* и т. д.

3. Вместо слова указывается несколько лексических или морфологических признаков.

Например, *заведующий* {NOUN,ЧЕМ}, где NOUN — признак существительного.

Задаются случаи: *заведующий складом*, *заведующий отделом* и т. д.

4. Указывается несколько «обобщенных слов».

   Например, *инженер по* {NOUN,ЧЕМУ} {NOUN,ЧЕМУ}.

   Имеется в виду: *инженер по технике безопасности* и т. д.

5. В термин вместо слова вводится признак принадлежности к определенному предметному словарю. Например, *отдел* {ORG_K}, где ORG_K — имя другого словаря (словаря организаций). Таким способом задаются словосочетания *отдел ИПИ РАН*, *отдел ООО «Алмаз»* (если организации *ИПИ РАН* и *ООО «Алмаз»* введены в словарь ORG_K).

6. Синонимия терминов задается с помощью знака равенства (=).

   Пример для двух терминов-синонимов:

   литейных дел мастер

   = *мастер литейных дел*.

   Допускается вариант множества терминов-синонимов:

   Река Мойка

   = *Мойка*

   = *р. Мойка.*

   Другой пример:

   Зюганов Геннадий

   = *Зюганов Г. А.*

   = *Геннадий Зюганов*

   = *Г. Зюганов*

   = *Зюганов*.

   Последний пример взят из словаря FIO_K. При наличии в тексте отдельных фамилий типа *Зюганов* система (блок ССА) делает дополнительные проверки, чтоб в документе не было фамилий *Зюганов* с другими инициалами. Тогда формируется объект, соответствующий лицу — *Зюганов Геннадий*.

7. Дополнительное задание классов для слов, с которыми идентифицированы термины словаря. Для этого используется знак в виде двух решеток (##):

   Михалков Никита ## актер, режиссер

   = *Никита Михалков*

   = *Н. Михалков*.

   Такая запись (помимо синонимии) задает дополнительную информацию: лицо (*Михалков Никита*) относится к классу *актеров* и *режиссеров*, что представляется в ПС документа. Более того, классы трактуются (блоком ССА) как семантические признаки слов.

4 Методы повышения эффективности работы словарной системы

К блоку идентификации терминов, работающему во время обработки текста, предъявляются два важных требования:

(1) осуществлять идентификацию терминов, записанных в словаре в любом падеже и числе;

(2) обеспечивать высокую эффективность поиска терминов в тексте — так, чтобы время поиска подходящего термина не превышало 1/3 работы всего лексико-морфологического блока (ЛМА).

Первое требование было реализовано с помощью подсистемы морфологического анализа. Во время трансляции словарей каждое слово термина преобразуется в нормальную форму. Идентификация осуществляется после нормализации слов текста. Идентификация терминов со словами текста сводится к совпадению их нормальных форм, независимо от числа и падежа.

Второе требование выполняется благодаря особой логической организации словарей. Все словари (за исключением street.slv — словаря московских улиц) на физическом уровне записываются в одном файле. Это позволяет осуществлять поиск сразу по всем словарям.

Для представления термина на физическом уровне выбрана двухчастная схема вида: «первое слово – остальные слова». Иначе говоря, каждый термин представляется двумя физическими записями словаря: первая содержит первое слово термина, а вторая — остальные слова и дополнительную информацию. Отметим, что первое слово не может быть «обобщенным». Это ограничение позволяет сделать процедуру идентификации быстрой и эффективной.

Каждая физическая запись состоит из текстового поискового ключа и списка ссылок на другие записи. В текстовом ключе первой записи термина хранится нормальная форма первого слова термина. Список ссылок первой записи составляют ссылки на все возможные продолжения (с указанием номера словаря). Дело в том, что разные термины, не важно, из одного и того же словаря или из разных, могут иметь одно и то же (с точностью до числа и падежа) первое слово. Например:

общественная организация,

общественное объединение,

общественное движение «Пчелки».

Пример ссылок для таких терминов изображен на рис. 3.

В данном примере из первой физической записи (*общественный*) к каждой второй будут вести ссылки, образующие список ссылок, как показано на рис. 3. Вторая физическая запись содержит нормальные формы всех остальных слов, канонический вид термина и признаки, кодирующие обобщенные слова.

Обработка каждого слова текста на предмет вхождения в словарь осуществляется следующим образом. Нормальная форма первого слова ищется в файле предметного словаря. Эта операция осуществляется эффективно благодаря иерархической структуре индексов словаря (см. ниже). Далее по прямым ссылкам на вторые записи просматриваются все термины с данным первым словом. Поскольку число таких записей в среднем чуть больше двух, то эта операция также оказывается эффективной. Наконец, само сравнение слов производится как сравнение символьных строк, т. е. также эффективно. В результате авторам удалось приблизиться к заданным 1/3 времени обработки, несмотря на то что такой процедуре подвергается каждое слово текста.

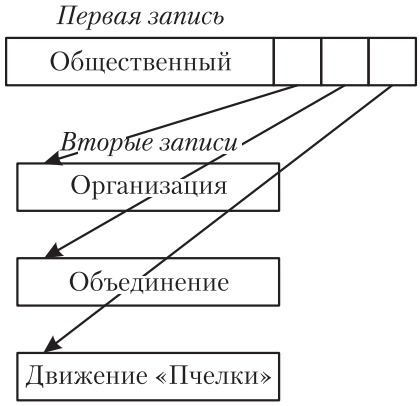


Рис. 3 Пример логической структуры предметного словаря

5 Физическая структура хранения словарей

Первой проблемой, от решения которой зависели во многом параметры словарной системы, был выбор СУБД. Было решено остановиться на разработанной ранее и модифицированной для данного случая собственной разработке, названной «*словарной СУБД*» (С-СУБД). На такое решение, главным образом, повлияло требование эффективности, а также возможности легкой подстройки под нужды ЛП.

Словарная СУБД предназначена для работы с текстовыми словарями. Каждый термин словаря получает в С-СУБД уникальный идентификатор (номер), по которому выстроен специальный индекс, благодаря чему доступ к термину по номеру происходит эффективно. Этот номер и используется в качестве ссылки на вторые физические записи. Словарная С-СУБД поддерживает операции, вызываемые командами:

(1) актуализировать (открыть) словари;

(2) вставить термин в словарь;

(3) по термину получить номер (или выяснить, что такого термина нет);

(4) по номеру получить термин;

(5) закрыть словари.

Для эффективной реализации операций *2* и *3* весь массив терминов образует иерархическую блочную структуру, в которой термины упорядочены лексикографически. Благодаря этому термин (первая физическая запись) ищется двоичным поиском. При этом наиболее интенсивно используются блоки верхних уровней иерархии. В С-СУБД реализована виртуальная память с задержкой наиболее часто использующихся блоков в оперативной памяти. В результате такой поиск (и вставка) также реализуется эффективно даже при значительных объемах словарной информации.

6 Настройка предметных словарей на особенности текста

В процессе создания ЛП его компоненты (за счет лингвистических знаний) постоянно подстраиваются с учетом особенностей предметной области. Это относится и к системе предметных словарей. В зависимости от информационных объектов, которые необходимо извлекать из текстов, производится отбор словарей из уже имеющихся, а если их недостаточно, то создание новых словарей. Однако имеется возможность более тонкой настройки словарей. Для этого разработаны два оператора, которые вызываются управляющими фрагментами:

Имя оператора	Назначение	Прагматика использования	Пример управляющего фрагмента
BAD_DIC	Задаются слова, игнорируемые при идентификации терминов	Такие слова связываются с конкретными предметными словарями	BAD_DIC(‘TERROR_K’,‘-’)
MAKE_FR	Задается особая форма генерации выходных фрагментов для указанных словарей	Для удобства обработки, осуществляемой блоком ССА	MAKE_FR(TERROR_K, WORK_K, ORG_K)

Оператор BAD_DIC задает слова, которые нужно игнорировать в тексте при идентификации терминов указанного словаря. В рассматриваемом примере при идентификации всех терминов, относящихся к словарю TERROR_K, в тексте игнорируются дефисы ‘-’.

Оператор MAKE_FR инициирует в ПС документа формирование объекта — на основе слов текста, с которыми идентифицирован термин словаря. Например, управляющий фрагмент MAKE_FR(TERROR_K, ...) при идентификации

любого термина словаря TERROR_K со словами текста (документа) использует эти слова для формирования объекта, который встраивается в ПС документа. Подробно все средства настройки ЛП изложены в [7].

7 Перспективы развития словарной подсистемы

Развитие системы предметных словарей предполагается осуществлять в двух направлениях.

1. Предполагается совершенствование механизма «обобщенных слов». Ограничение, что первое слово термина не может быть «обобщенным», в принципе можно изменить на менее жесткое: хотя бы одно слово в описании термина должно быть необобщенным. Однако это требует кардинальной перестройки всего алгоритма идентификации термина.
2. Чтобы словари стали полноценной онтологией, предполагается ввести отношения между терминами, прежде всего отношения род–вид и часть–целое. Тогда после идентификации термина блок будет выдавать еще несколько понятий, ассоциированных с найденным. Это особенно важно при постановке задачи извлечения имплицитной информации [8].

Литература

1. *Кузнецов И. П., Сомин Н. В.* Англо-русская система извлечения знаний из потоков информации в интернет-среде // Системы и средства информатики. — М.: Наука, 2007. Вып. 17. С. 236–254.
2. *Кузнецов И. П., Мацкевич А. Г.* Семантико-ориентированные системы на основе баз знаний. — М.: МТУСИ, 2007. 173 с.
3. Лаборатория компьютерной лингвистики ИПИ РАН. Официальный сайт www.IpiranLogos.com.
4. *Сомин Н. В., Соловьева Н. С., Шарнин М. М.* Система морфологического анализа: опыт эксплуатации и модификации // Системы и средства информатики. — М.: Наука, 2005. Вып. 15. С. 20–30.
5. *Кузнецов И. П., Сомин Н. В.* Особенности лексико-морфологического анализа при извлечении информационных объектов и связей из текстов естественного языка // Компьютерная лингвистика и интеллектуальные технологии: по мат-лам междунар. конф. «Диалог 2010». — М.: РГГУ, 2010. Вып. 9(16). С. 254–264.
6. *Сомин Н. В., Кузнецов И. П., Мацкевич А. Г., Николаев В. Г.* Методы и средства настройки морфо-лексического анализатора на предметную область // Системы и средства информатики. — М.: Наука, 2009. Вып. 19. С. 96–118.
7. *Кузнецов И. П., Сомин Н. В.* Средства настройки семантико-ориентированной системы на выделение и поиск объектов // Системы и средства информатики. — М.: Наука, 2008. Вып. 18. С. 119–143.
8. *Кузнецов И. П.* Семантические методы извлечения имплицитной информации // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2011. Вып. 21. № 2. С. 116–138.

ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ СУММ К НЕСИММЕТРИЧНОМУ РАСПРЕДЕЛЕНИЮ СТЬЮДЕНТА*

***В. Е. Бенинг*[1], *Л. М. Закс*[2], *В. Ю. Королев*[3]**

Аннотация: Постороены оценки точности приближения распределений специальных смешанных пуассоновских случайных сумм независимых случайных величин с ненулевыми средними несимметричным распределением Стьюдента.

Ключевые слова: случайная сумма; оценка скорости сходимости; смешанное пуассоновское распределение; несимметричное распределение Стьюдента

1 Введение

Статистические закономерности поведения информационных потоков в современных информационных и телекоммуникационных системах характеризуются наличием так называемых тяжелых хвостов. В развитие идей теоретического моделирования таких эффектов с помощью предельных теорем для обобщенных дважды стохастических пуассоновских процессов — в определенном смысле наилучших моделей хаотических процессов со случайной интенсивностью, изложенных в работах [1–3], в данной статье приводятся оценки скорости сходимости распределений специальных смешанных пуассоновских случайных сумм независимых случайных величин с ненулевыми средними к несимметричному (скошенному) распределению Стьюдента. Как показано в работе [3], такое распределение возникает как асимптотическая аппроксимация для аддитивных характеристик информационных потоков в том случае, когда случайная интенсивность соответствующего потока информативных событий имеет обратное гамма-распределение. Таким образом, приводимые в данной статье оценки могут быть полезны при определении адекватности смешанных вероятностных моделей

*Работа поддержана РФФИ, гранты 12-07-00109а, 12-07-00115а.

[1]Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, bening@cs.msu.su

[2]Альфа-банк, отдел моделирования и математической статистики, lily.zaks@gmail.com

[3]Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vkorolev@cs.msu.su

статистических закономерностей, в которых смешивающим является обратное гамма-распределение.

Известно несколько попыток описать несимметричные обобщения распределения Стьюдента. Пожалуй, наиболее успешная из них — это несимметричное (скошенное, skew) распределение Стьюдента, описанное в работе [4] как частный случай обобщенного гиперболического распределения. В статье [5] распределению, предложенному в [4], дано другое, более удобно интерпретируемое определение как специальной сдвиг-масштабной смеси нормальных законов. Согласно [5], скошенным распределением Стьюдента называется распределение с плотностью

$$p_{SS}(x;a,\sigma,\mu,\lambda)=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_0^\infty \exp\left\{-\frac{1}{2}\left(\frac{x-au}{\sigma\sqrt{u}}\right)^2\right\}\frac{h(u;\mu,\lambda)}{\sqrt{u}}\,du\,. \qquad (1)$$

Здесь $a\in\mathbb{R}$; $\sigma>0$; $\mu>0$; $\lambda>0$; $h(x;\mu,\lambda)$ — плотность обратного гамма-распределения, т. е. распределения случайной величины $V_{\mu,\lambda}^{-1}$:

$$h(x;\mu,\lambda)=\frac{\lambda^\mu}{\Gamma(\mu)}x^{-\mu-1}\exp\left\{-\frac{\lambda}{x}\right\}, \qquad (2)$$

где $V_{\mu,\lambda}$ — случайная величина с гамма-распределением с параметром формы μ и параметром масштаба λ. Напомним, что плотность распределения само́й случайной величины $V_{\mu,\lambda}$ имеет вид:

$$g(x;\mu,\lambda)=\frac{\lambda^\mu}{\Gamma(\mu)}x^{\mu-1}e^{-\lambda x}\,,\ \ x\geqslant 0\,.$$

Здесь и далее $\Gamma(\,\cdot\,)$ — эйлерова гамма-функция:

$$\Gamma(z)=\int\limits_0^\infty e^{-y}y^{z-1}\,dy,\quad z>0\,.$$

Пусть $\{X_{n,j}\}_{j\geqslant 1}$, $n=1,2,\ldots$, — последовательность серий независимых и одинаково в каждой серии распределенных случайных величин, а N_n, $n=1,2,\ldots$, — неотрицательные целочисленные случайные величины такие, что при каждом n случайная величина N_n независима от последовательности $\{X_{n,j}\}_{j\geqslant 1}$. Для натуральных k обозначим

$$S_{n,k}=X_{n,1}+\cdots+X_{n,k}\,.$$

Для определенности будем считать, что $\sum_{j=1}^{0} = 0$. Символ $\Longrightarrow$ будет обозначать сходимость по распределению. Стандартную нормальную функцию распределения будем обозначать $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz\,.$$

Скошенное распределение Стьюдента является частным случаем обобщенных гиперболических распределений, находящих широкое применение, скажем, в финансовой математике (см., например, работы [6–11], в которых установлено хорошее согласие обобщенного гиперболического распределения с данными о ценах на датских и немецких биржах, финансовыми индексами NYSE (New York Stock Exchange) и DAX (Deutscher Aktienindex), обменными курсами валют и т. д.).

Вместе с тем в прикладной теории вероятностей принято, что ту или иную модель можно считать в достаточной мере обоснованной (адекватной) только тогда, когда она является *асимптотической аппроксимацией*, т. е. когда существует довольно простая предельная схема (например, схема суммирования) и соответствующая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения [12]. «Асимптотическое» обоснование моделей типа скошенного распределения Стьюдента было дано лишь недавно в работе [3], где показано, что скошенные распределения Стьюдента могут выступать в качестве предельных в довольно простых предельных теоремах для регулярных статистик, построенных по выборкам случайного объема, в частности в схеме случайного суммирования случайных величин, и, следовательно, могут считаться *естественными* асимптотическими аппроксимациями для распределений многих процессов, например, сходных с неоднородными случайными блужданиями.

В частности, в статье [3] доказана следующая теорема.

Теорема А. *Предположим, что существуют числа* $a \in \mathbb{R}$, $\sigma^2 \in (0,\infty)$, $\mu \in (0,\infty)$, $\lambda \in (0,\infty)$ *и последовательность натуральных чисел* $\{m_n\}_{n\geqslant 1}$ *такие, что при* $n \to \infty$

$$\mathsf{P}\left(S_{n,m_n} < x\right) \longrightarrow \Phi\left(\frac{x-a}{\sigma}\right); \tag{3}$$

$$\mathsf{P}(N_n < m_n x) \Longrightarrow H(x;\mu,\lambda)\,, \tag{4}$$

где $H(x;\mu,\lambda)$ — *функция обратного гамма-распределения с параметрами* μ *и* λ, *соответствующая плотности* (2). *Тогда*

$$\mathsf{P}\left(S_{n,N_n} < x\right) \longrightarrow P_{SS}(x;a,\sigma,\mu,\lambda)\,,$$

где $P_{SS}(x;a,\sigma,\mu,\lambda)$ *— функция скошенного распределения Стьюдента, соответствующая плотности* (1).

В данной статье будет рассмотрена скорость сходимости в теореме А для довольно иллюстративного частного случая, демонстрирующего один из возможных механизмов формирования несимметричных предельных законов для случайных блужданий.

2 Основной результат

Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $\mathsf{E}\xi_1 = 0$, $0 < \mathsf{D}\xi_1 = \sigma^2 < \infty$, $\beta^3 = \mathsf{E}|\xi_1|^3 < \infty$, $a \in \mathbb{R}$, n — натуральное число.

Положим

$$X_{n,j} = \frac{\xi_j}{\sqrt{n}} + \frac{a}{n}\,.$$

В терминах случайных блужданий рассматриваемая конструкция слагаемых предполагает *одинаковый порядок малости* элементарных трендов и *дисперсий*, что характерно, например, для приращений винеровского процесса со сносом. Обозначим

$$S_n = \sum_{j=1}^{n} X_{n,j} \quad \left(= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_j + a \right).$$

В силу классической центральной предельной теоремы имеем:

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathsf{P}(S_n < x) - \Phi\left(\frac{x-a}{\sigma}\right) \right| = 0\,,$$

т. е. таким образом определенные случайные величины $X_{n,j}$ удовлетворяют условию (3) с $m_n = n$.

Пусть $U_{\mu,\lambda}$ — случайная величина с обратным гамма-распределением (2), независимая от стандартного пуассоновского процесса $M(t)$, $t \geqslant 0$. Для натурального n положим $N_n = M\left(nU_{\mu,\lambda}\right)$. Несложно видеть, что распределение случайной величины N_n является смешанным пуассоновским и имеет вид:

$$\mathsf{P}(N_n = k) = \frac{1}{k!} \int_0^\infty e^{-nz} (nz)^k h(z;\mu,\lambda)\, dz\,, \quad k = 0, 1, 2, \ldots$$

Таким образом определенная случайная величина N_n может быть интерпретирована как число событий, зарегистрированных к моменту времени n в пуассоновском процессе со случайной интенсивностью, имеющей обратное гамма-распределение. В работе [13] показано, что подобные модели демонстрируют

более высокую адекватность при описании потоков страховых требований в автостраховании, нежели классические модели с гамма-распределенной интенсивностью. Целесообразность использования обратного гамма-распределения в качестве априорного в более общих байесовских моделях отмечена в работах [14, 15].

Предположим, что случайная величина $U_{\mu,\lambda}$ и пуассоновский процесс $M(t)$ независимы от последовательности $\xi_1, \xi_2, \ldots$ Тогда, очевидно, при каждом n случайная величина N_n также будет независима от этой последовательности.

Обозначим $A_n(z) = \mathsf{P}(N_n < nz)$. Несложно видеть, что

$$A_n(z) \Longrightarrow H(x;\mu,\lambda) \quad (n \to \infty)\,.$$

Действительно, как известно, если $\Pi(x;\ell)$ — функция распределения Пуассона с параметром $\ell > 0$ и $E(x;c)$ — функция распределения с единственным единичным скачком в точке $c \in \mathbb{R}$, то

$$\Pi(\ell x;\ell) \Longrightarrow E(x;1) \quad (\ell \to \infty)\,.$$

Так как для $x \in \mathbb{R}$

$$A_n(x) = \int\limits_0^\infty \Pi(nx;nz)\, d_z H(z;\mu,\lambda)\,,$$

то по теореме Лебега о мажорируемой сходимости при $n \to \infty$

$$A_n(x) \Longrightarrow \int\limits_0^\infty E(x/z;1)\, d_z H(z;\mu,\lambda) = \int\limits_0^x d_z H(z;\mu,\lambda) = H(x;\mu,\lambda)\,,$$

т. е. так определенные случайные величины N_n удовлетворяют условию (4).

Таким образом, в силу непрерывности функции распределения $P_{SS}(x;a,\sigma,\mu,\lambda)$ из теоремы А следует, что

$$D_n \equiv \sup_{x\in\mathbb{R}} \left| \mathsf{P}\left(\sum_{j=1}^{N_n} X_{n,j} < x \right) - P_{SS}(x;a,\sigma,\mu,\lambda) \right| \longrightarrow 0 \ \ (n \to \infty)\,.$$

Скорость стремления D_n к нулю описывается следующим утверждением.

Теорема 1. *Для любого $n \geqslant 1$ справедлива оценка*

$$D_n \leqslant \frac{0{,}4532}{\sqrt{n}}\, \frac{\beta^3}{\sigma^3}\, \frac{\Gamma(\mu+1/2)}{\sqrt{\lambda}\Gamma(\mu)} + 0{,}1210 \frac{a^2}{n\sigma^2}\,.$$

Доказательство. Как уже было показано, распределение случайной величины N_n является смешанным пуассоновским. Следовательно, по теореме Фубини

$$\mathsf{P}\left(\sum_{j=1}^{N_n} X_{n,j} < x\right) = \mathsf{P}\left(\sum_{j=1}^{M(nU_{\mu,\lambda})} X_{n,j} < x\right) = \int_0^\infty \mathsf{P}\left(\sum_{j=1}^{M(nz)} X_{n,j} < x\right) h(z;\mu,\lambda)\,dz\,.$$

При этом

$$\mathsf{E}X_{n,j} = \frac{a}{n}\,; \quad \mathsf{D}X_{n,j} = \frac{\sigma^2}{n}\,; \quad \mathsf{E}|X_{n,j} - \mathsf{E}X_{n,j}|^3 = \frac{\beta^3}{n^{3/2}}\,.$$

Таким образом, при каждом $z \in (0,\infty)$

$$\mathsf{E}\sum_{j=1}^{M(nz)} X_{n,j} = az\,,$$
$$\mathsf{D}\sum_{j=1}^{M(nz)} X_{n,j} = nz\left(\frac{a^2}{n^2} + \frac{\sigma^2}{n}\right) = z\sigma^2\left(1 + \frac{a^2}{n\sigma^2}\right)\,.$$

Несложно видеть, что функция распределения $P_{SS}(x;a,\sigma,\mu,\lambda)$ имеет вид:

$$P_{SS}(x;a,\sigma,\mu,\lambda) = \int_0^\infty \Phi\left(\frac{x-az}{\sigma\sqrt{z}}\right) h(z;\mu,\lambda)\,dz\,.$$

Поэтому

$$D_n = \sup_x \left| \int_0^\infty h(z;\mu,\lambda) \left[\mathsf{P}\left(\sum_{j=1}^{M(nz)} X_{n,j} < x\right) - \Phi\left(\frac{x-az}{\sigma\sqrt{z(1+a^2/(n\sigma^2))}}\right) + \right.\right.$$
$$\left.\left. + \Phi\left(\frac{x-az}{\sigma\sqrt{z(1+a^2/(n\sigma^2))}}\right) - \Phi\left(\frac{x-az}{\sigma\sqrt{z}}\right)\right] dz \right| \leqslant I_1 + I_2\,, \quad (5)$$

где

$$I_1 = \int\limits_0^\infty h(z;\mu,\lambda) \sup_x \left| \mathsf{P}\left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) - \Phi\left(\frac{x - az}{\sigma\sqrt{z(1 + a^2/(n\sigma^2))}} \right) \right| dz\,;$$

$$I_2 = \int\limits_0^\infty h(z;\mu,\lambda) \sup_x \left| \Phi\left(\frac{x - az}{\sigma\sqrt{z(1 + a^2/(n\sigma^2))}} \right) - \Phi\left(\frac{x - az}{\sigma\sqrt{z}} \right) \right| dz\,.$$

В дальнейшем рассмотрении понадобится следующее утверждение.

Лемма 1. *Пусть случайные величины $X_1, X_2, \ldots$ одинаково распределены. Пусть N_λ — пуассоновская случайная величина с параметром $\lambda > 0$. Предположим, что случайные величины $N_\lambda, X_1, X_2, \cdots$ независимы в совокупности. Обозначим*

$$S_\lambda = X_1 + \cdots + X_{N_\lambda}\,.$$

Тогда

$$\sup_x \left| \mathsf{P}(S_\lambda < x) - \Phi\left(\frac{x - \mathsf{E}S_\lambda}{\sqrt{\mathsf{D}S_\lambda}} \right) \right| \leqslant \frac{0{,}4532}{\sqrt{\lambda}} \, \frac{\mathsf{E}|X_1 - \mathsf{E}X_1|^3}{(\mathsf{D}X_1)^{3/2}}\,.$$

Доказательство леммы 1 приведено в работе [16] (см. также [17, с. 144]).

Продолжим доказательство теоремы 1. Рассмотрим I_1. Применяя лемму 1 с учетом (4), получаем:

$$I_1 \leqslant \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \int\limits_0^\infty \frac{1}{\sqrt{z}} \, h(z;\mu,\lambda)\, dz = \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \, \mathsf{E}\sqrt{\frac{1}{U_{\mu,\lambda}}} =$$

$$= \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \, \mathsf{E}\sqrt{V_{\mu,\lambda}} = \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \, \frac{\lambda^\mu}{\Gamma(\mu)} \int\limits_0^\infty e^{-\lambda z} z^{\mu - 1/2}\, dz =$$

$$= \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \, \frac{\lambda^\mu \Gamma(\mu + 1/2)}{\lambda^{\mu+1/2}\Gamma(\mu)} = \frac{0{,}4532}{\sqrt{n}} \, \frac{\beta^3}{\sigma^3} \, \frac{\Gamma(\mu + 1/2)}{\sqrt{\lambda}\Gamma(\mu)}\,. \quad (6)$$

Рассмотрим I_2. В дальнейшем понадобится еще одно вспомогательное утверждение.

Лемма 2. *Пусть $b \in \mathbb{R}$, $0 < c < \infty$, $0 < d < \infty$. Тогда*

$$\sup_y |\Phi(y) - \Phi(cy)| \leqslant \frac{1}{\sqrt{2\pi e}} \left| \max\left\{ c,\, \frac{1}{c} \right\} - 1 \right|\,; \quad (7)$$

$$\sqrt{1 + d} - 1 \leqslant \frac{d}{2}\,. \quad (8)$$

Элементарное доказательство этого утверждения основано на применении формулы Лагранжа.

Продолжим доказательство теоремы 1. В лемме 2 положим

$$y = \frac{x - az}{\sigma\sqrt{z(1 + a^2/(n\sigma^2))}}; \quad c = \sqrt{1 + \frac{a^2}{n\sigma^2}}.$$

Тогда $c \geqslant 1$ и в силу утверждения (7) леммы 2 имеем:

$$I_2 \leqslant \frac{1}{\sqrt{2\pi e}} \left(\sqrt{1 + \frac{a^2}{n\sigma^2}} - 1 \right).$$

При этом в силу утверждения (8) леммы 2

$$\sqrt{1 + \frac{a^2}{n\sigma^2}} - 1 \leqslant \frac{a^2}{2n\sigma^2}.$$

Окончательно получаем

$$I_2 \leqslant \frac{a^2}{2\sqrt{2\pi e} n\sigma^2}. \tag{9}$$

Подставляя (6) и (9) в (5), получаем утверждение теоремы. Теорема доказана.

Если $a = 0$, $\mu = \lambda = \gamma/2$ при некотором $\gamma > 0$, то распределение $P_{SS}(x; a, \sigma, \mu, \lambda) = P_{SS}(x; 0, \sigma, \gamma/2, \gamma/2)$ с точностью до параметра масштаба σ становится классическим распределением Стьюдента с параметром («числом степеней свободы») γ (см., например, [18]). При этом аналогом леммы 1 может служить следующее утверждение.

Лемма 3. *Пусть случайные величины $X_1, X_2, \ldots$ одинаково распределены. Пусть N_ℓ — пуассоновская случайная величина с параметром $\ell > 0$. Предположим, что случайные величины $N_\ell, X_1, X_2, \ldots$ независимы в совокупности. Обозначим*

$$S_\ell = X_1 + \cdots + X_{N_\ell}.$$

Тогда

$$\sup_x \left| \mathsf{P}(S_\ell < x) - \Phi\left(\frac{x}{\sqrt{\mathsf{D}S_\ell}} \right) \right| \leqslant \frac{0{,}3041}{\sqrt{\ell}} \frac{\mathsf{E}|X_1|^3}{(\mathsf{E}X_1^2)^{3/2}}.$$

Доказательство этого утверждения приведено в [19] (см. также [17], теорема 2.4.3).

Если в случае $a = 0$ в доказательстве теоремы 1 вместо леммы 1 воспользоваться леммой 3, то в результате получится следующее утверждение, содержащее оценку скорости сходимости распределений специальных случайных сумм, описанных выше, к классическому распределению Стьюдента с γ степенями свободы.

Следствие 1. *Пусть в дополнение к условиям теоремы* 1 $a = 0$, $\mu = \lambda = \gamma/2$ *при некотором* $\gamma > 0$. *Тогда для любого* $n \geqslant 1$ *справедлива оценка:*

$$\sup_{x\in\mathbb{R}}\left|\mathsf{P}\left(\sum_{j=1}^{N_n} X_{n,j} < x\right) - P_{SS}\left(x; 0, \sigma, \frac{\gamma}{2}, \frac{\gamma}{2}\right)\right| \leqslant \frac{0{,}3041}{\sqrt{n}}\,\frac{\beta^3}{\sigma^3}\,\frac{\sqrt{2}\Gamma((\gamma+1)/2)}{\sqrt{\gamma}\Gamma(\gamma/2)}.$$

Литература

1. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008.
2. *Королев В. Ю., Шоргин С. Я.* Математические методы анализа стохастической структуры информационных потоков. — М.: ИПИ РАН, 2011.
3. *Королев В. Ю., Соколов И. А.* Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 3–11.
4. *Aas K., Haff I. H.* The generalized hyperbolic skew Student's t-distribution // J. Financial Econometrics, 2006. Vol. 4. No. 2. P. 275–309.
5. *Kim Y., McCulloch J. H.* The skew-Student distribution with application to U.S. stock market returns and the equity premium. Preprint. — Columbus: Department of Economics, Ohio State University, 2007.
6. *Eberlein E., Keller U.* Hyperbolic distributions in finance // Bernoulli, 1995. Vol. 1. No. 3. P. 281–299.
7. *Prause K.* Modeling financial data using generalized hyperbolic distributions. Preprint No. 48. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1997.
8. *Eberlein E., Keller U., Prause K.* New insights into smile, mispricing and value at risk: The hyperbolic model // J. Business, 1998. Vol. 71. P. 371–405.
9. *Barndorff-Nielsen O. E.* Processes of normal inverse Gaussian type // Finance Stochastics, 1998. Vol. 2. P. 41–68.
10. *Eberlein E., Prause K.* The generalized hyperbolic model: Financial derivatives and risk measures. Preprint No. 56. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1998.
11. *Eberlein E.* Application of generalized hyperbolic Lévy motions to finance. Preprint No. 64. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1999.
12. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. — М.–Л.: ГИТТЛ, 1949.

13. *Schiegl M.* About the justification of experience rating: Bonus Malus system and a new Poisson mixture model // Quantitative Finance Papers, 2010. 1009.4142. arXiv.org.
14. *Gelfand A. E., Smith A. F. M.* Sampling-based approaches to calculating marginal densities // J. Amer. Stat. Association, 1990. Vol. 88. P. 1219–1227.
15. *Gelman A.* Prior distributions for variance parameters in hierarchical models // Bayesian Anal., 2006. Vol. 1. No. 3. P. 515–533.
16. *Королев В. Ю., Шевцова И. Г., Шоргин С. Я.* О неравенствах типа Берри–Эссеена для пуассоновских случайных сумм // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 64–66.
17. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
18. *Бенинг В. Е., Королев В. Ю.* Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения, 2004. Т. 49, Вып. 3. С. 417–435.
19. *Korolev V., Shevtsova I.* An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // Scandinavian Actuarial J., June 04, 2010. DOI:10.1080/03461238.2010.485370.

О ТОЧНОСТИ ПРИБЛИЖЕНИЯ РАСПРЕДЕЛЕНИЯ ОЦЕНКИ РИСКА ПОРОГОВОЙ ОБРАБОТКИ ВЕЙВЛЕТ-КОЭФФИЦИЕНТОВ СИГНАЛА НОРМАЛЬНЫМ ЗАКОНОМ ПРИ НЕИЗВЕСТНОМ УРОВНЕ ШУМА*

О. В. Шестаков[1]

Аннотация: Исследуются асимптотические свойства оценки риска при пороговой обработке коэффициентов вейвлет-разложения функции сигнала. Получены некоторые оценки скорости сходимости распределения оценки риска к нормальному закону.

Ключевые слова: вейвлеты; пороговая обработка; оценка риска; нормальное распределение; оценка скорости сходимости

1 Введение

Методы обработки сигналов и изображений с помощью вейвлет-разложения применяются в самых разнообразных областях, включая геофизику, физику плазмы, вычислительную томографию, компьютерную графику и др. Основные задачи, для решения которых используется вейвлет-разложение, — это сжатие сигналов/изображений и удаление шума. При этом строится оценка сигнала или изображения, основанная на пороговой обработке вейвлет-коэффициентов, которая обнуляет коэффициенты, не превышающие заданного порога. Наличие шума неизбежно приводит к погрешностям в оцениваемом сигнале/изображении. Свойства оценки таких погрешностей (риска) исследовались в работах [1–9]. При определенных условиях оценка риска является состоятельной и асимптотически нормальной [6]. В работах [8, 9] получены некоторые оценки скорости сходимости оценки риска к нормальному закону в одномерном случае (т. е. при обработке одномерных сигналов). В данной работе эти оценки будут улучшены.

Введем необходимые понятия и обозначения. При использовании вейвлет-разложения функция $f \in L^2(\mathbf{R})$, описывающая сигнал, представляется в виде ряда из сдвигов и растяжений некоторой вейвлет-функции ψ:

*Работа выполнена при финансовой поддержке РФФИ (проекты 11-01-00515а и 11-01-12-26-офи-м), а также Министерства образования и науки РФ (государственный контракт № 14.740.11.0996).

[1]Московский государственный университет им. М. В. Ломоносова, факультет ВМК; Институт проблем информатики Российской академии наук, oshestakov@cs.msu.su

$$f = \sum_{j,k \in Z} \langle f, \psi_{j,k} \rangle \psi_{j,k} \,, \tag{1}$$

где $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ (семейство $\{\psi_{j,k}\}_{j,k \in Z}$ образует ортонормированный базис в $L^2(\mathbf{R})$). Индекс j в (1) называется масштабом (или уровнем), а индекс k — сдвигом. Функция ψ должна удовлетворять определенным требованиям [10], однако ее можно выбрать таким образом, чтобы она обладала некоторыми полезными свойствами, например была дифференцируемой нужное число раз и имела заданное число M нулевых моментов [10], т. е.

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0 \,, \quad k = 0, \ldots, M-1 \,.$$

В дальнейшем будут рассматриваться функции $f \in L^2(\mathbf{R})$ с носителем на отрезке $[0,1]$, равномерно регулярные по Липшицу с некоторым показателем $\gamma > 0$, т. е. такие функции, для которых существует константа $L > 0$ и полином P_y степени $n = \lfloor \gamma \rfloor$ такой, что для любого $y \in [0,1]$ и любого $x \in \mathbf{R}$

$$|f(x) - P_y(x)| \leqslant L\,|x - y|^{\gamma} \,.$$

Для таких функций f известно [11], что если вейвлет-функция M раз непрерывно дифференцируема ($M \geqslant \gamma$), имеет M нулевых моментов и быстро убывает на бесконечности вместе со своими производными, т. е. для всех $0 \leqslant k \leqslant M$ и любого $m \in N$ найдется константа C_m, что при всех $x \in \mathbf{R}$

$$\left|\psi^{(k)}(x)\right| \leqslant \frac{C_m}{1 + |x|^m} \,,$$

то найдется такая константа $A > 0$, что

$$\langle f, \psi_{j,k} \rangle \leqslant \frac{A}{2^{j(\gamma + 1/2)}} \,. \tag{2}$$

На практике функции сигнала всегда заданы в дискретных отсчетах. Не ограничивая общности, будем считать, что функция f задана в точках i/N ($i = 1, \ldots N$, где $N = 2^J$ для некоторого J): $f_i = f\,(i/N)$. Дискретное вейвлет-преобразование представляет собой умножение вектора значений функции f (обозначим его через $\overline{f}$) на ортогональную матрицу W, определяемую вейвлет-функцией ψ: $\overline{f}^W = W\overline{f}$ [11]. При этом если перейти к двойному индексу (j,k), то дискретные вейвлет-коэффициенты будут связаны с непрерывными следующим образом: $f^W_{j,k} \approx \sqrt{N}\langle f, \psi_{j,k} \rangle$ (см., например, [2] или [12]). В дальнейшем

для удобства будем нумеровать дискретные вейвлет-коэффициенты так же, как отсчеты функции f, одним индексом i вместо двойного индекса (j,k).

В реальных наблюдениях всегда присутствует шум. В данной работе рассматривается аддитивная модель шума

$$Y_i = f_i + z_i\,, \quad i = 1,\ldots,N\,,$$

где z_i — независимые случайные величины, имеющие нормальное распределение с нулевым средним и дисперсией σ^2. Тогда в силу ортогональности матрицы W для дискретных вейвлет-коэффициентов принимается следующая модель:

$$Y_i^W = f_i^W + z_i^W\,, \quad i = 1,\ldots,N\,,$$

где z_i^W также независимы и нормально распределены с нулевым средним и дисперсией σ^2, а f_i^W равны соответствующим непрерывным вейвлет-коэффициентам, умноженным на $\sqrt{N}$.

2 Пороговая обработка и оценка риска

Смысл пороговой обработки вейвлет-коэффициентов заключается в удалении достаточно маленьких коэффициентов, которые считаются шумом. Будем использовать так называемую мягкую пороговую обработку с порогом T. К каждому вейвлет-коэффициенту применяется функция $\rho_T(x) = \operatorname{sgn}(x)\,(|x| - T)_+$, т. е. при такой пороговой обработке коэффициенты, которые по модулю меньше порога T, обнуляются, а абсолютные величины остальных коэффициентов уменьшаются на величину порога. Погрешность (или риск) мягкой пороговой обработки определяется следующим образом:

$$R_N(f) = \sum_{i=1}^{N} \mathrm{E}\left(f_i^W - \rho_T(Y_i^W)\right)^2\,. \tag{3}$$

В выражении (3) присутствуют неизвестные величины f_i^W, поэтому вычислить значение $R_N(f)$ нельзя. Однако его можно оценить. В каждом слагаемом, если $\left|Y_i^W\right| > T$, то вклад этого слагаемого в риск составляет $\sigma^2 + T^2$, а если $\left|Y_i^W\right| \leqslant T$, то вклад составляет $(f_i^W)^2$. Поскольку $\mathrm{E}(Y_i^W)^2 = \sigma^2 + (f_i^W)^2$, величину $(f_i^W)^2$ можно оценить разностью $(Y_i^W)^2 - \sigma^2$.

Таким образом, в качестве оценки риска можно использовать следующую величину:

$$\widehat{R}_N(f,\sigma) = \sum_{i=1}^{N} F[(Y_i^W)^2, \sigma]\,,$$

где

$$F[x,\sigma] = (x-\sigma^2)\mathbb{1}(|x| \leqslant T^2) + (\sigma^2+T^2)\mathbb{1}(|x| > T^2)\,. \tag{4}$$

Для таким образом определенной оценки риска справедливо следующее утверждение [11].

Теорема 1. $\mathrm{E}\widehat{R}_N(f,\sigma) = R_N(f)$, *т. е.* $\widehat{R}_N(f,\sigma)$ *является несмещенной оценкой для* $R_N(f)$.

В работах [1, 3] было предложено использовать порог $T(\sigma) = \sigma\sqrt{2\ln N}$. Было показано, что при таком пороге риск близок к минимальному [1]. Этот порог получил название «универсальный». В дальнейшем будет использоваться именно такой вид порога.

Зачастую дисперсия σ^2 неизвестна, и ее также необходимо оценивать, при этом выражение (4) принимает вид:

$$\widehat{R}_N(f,\widehat{\sigma}) = \sum_{i=1}^{N} F\left[(Y_i^W)^2, \widehat{\sigma}\right]$$

(вместо $T(\sigma)$ используется $T(\hat{\sigma})$).

3 Оценка скорости сходимости распределения оценки риска к нормальному закону

Обычно дисперсия σ^2 (или среднеквадратичное отклонение σ) оценивается по половине всех вейвлет-коэффициентов разложения сигнала для $j = J-1$ (напомним, что $N = 2^J$), поскольку если функция f удовлетворяет требуемым условиям регулярности, то в силу (2) эти коэффициенты фактически содержат только шум.

В работах [6–9] исследуется асимптотическое поведение оценки риска $\widehat{R}_N(f)$ при использовании различных оценок $\hat{\sigma}$. Показано, что при определенных условиях оценка риска является асимптотически нормальной, и получены некоторые оценки скорости сходимости к нормальному закону. В этом параграфе эти оценки будут улучшены.

Далее для удобства будем обозначать Y_i^W через X_i, а f_i^W — через a_i. Сначала рассмотрим ситуацию, когда в качестве оценки σ^2 используется выборочная дисперсия:

$$\widehat{\sigma}^2 = \frac{1}{N/2-1}\sum_{i=N/2+1}^{N}(X_i^2 - \overline{X})^2\,,$$

где

$$\overline{X} = \frac{2}{N}\sum_{i=N/2+1}^{N} X_i\,. \tag{5}$$

Теорема 2. *Пусть f задана на отрезке $[0,1]$ и является равномерно регулярной по Липшицу с показателем $\gamma > 1/4$ и пусть оценка σ^2 задана соотношением* (5)*, тогда существуют такие константы $\tilde{C}_0$ и $\tilde{C}_1$ (зависящие от γ, A и σ), что*

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\widehat{\sigma})-R_N(f)}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|\leqslant \\ \leqslant\frac{\tilde{C}_0(\ln N)^{3/2+1/(4\gamma+2)}}{N^{1/2-1/(4\gamma+2)}}+\frac{\tilde{C}_1}{N^{2\gamma-1/2}}\,, \quad (6)$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Доказательство. Имеем

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\widehat{\sigma})-R_N(f)}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|= \\ =\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma)-R_N(f)+\widehat{R}_N(f,\widehat{\sigma})-\widehat{R}_N(f,\sigma)}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|.$$

Запишем

$$\widehat{R}_N(f,\widehat{\sigma})-\widehat{R}_N(f,\sigma)=U_N(\widehat{\sigma})+V_N\,,$$

где

$$V_N=N(\sigma^2-\widehat{\sigma}^2)\,,\ \ U_N(\widehat{\sigma})=\sum_{i=1}^{N}\{u_i(\widehat{\sigma})-u_i(\sigma)\}\,,$$

а $u_i(s)=\min\{X_i^2,T^2(s)\}+2s^2\mathbb{1}\left(|X_i|>T(s)\right)$. Тогда

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma)-R_N(f)+\widehat{R}_N(f,\widehat{\sigma})-\widehat{R}_N(f,\sigma)}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|\leqslant \\ \leqslant\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma)-R_N(f)+V_N}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|+\frac{\varepsilon_N}{\sqrt{2\pi}}+ \\ +\mathrm{P}\left(\left|\frac{U_N(\widehat{\sigma})}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)\,. \quad (7)$$

Запишем разность $\widehat{R}_N(f,\sigma)-R_N(f)$ в виде

$$\widehat{R}_N(f,\sigma)-R_N(f)=R_1(f,\sigma)+R_2(f,\sigma)\,,$$

где

$$R_j(f,\sigma)=\sum_{I_j}\left\{F[(Y_i^W)^2,\sigma]-\mathrm{E}F[(Y_i^W)^2,\sigma]\right\},\ \ j=1,2\,. \tag{8}$$

Здесь I_2 — множество тех индексов i, для которых в силу (2) выполнено $|a_i|\leqslant A/(\ln N)^{1/2}$, а I_1 — множество остальных индексов i.

Поскольку f регулярна по Липшицу с $\gamma>1/4$, число слагаемых в $R_1(f,\sigma)$ не превосходит $B_1(N\ln N)^{1/(2\gamma+1)}$, где B_1 — некоторая константа, зависящая от γ. Слагаемые в $R_1(f,\sigma)$ по модулю не превосходят $B_2\ln N$ с некоторой константой B_2. Следовательно, применяя неравенство Хеффдинга или неравенство Бернштейна [13], можно показать, что для некоторых констант B_3 и B_4

$$\mathrm{P}\left(\left|\frac{R_1(f,\sigma)}{\sigma^2\sqrt{2N}}\right|>B_3\frac{(\ln N)^{3/2+1/(4\gamma+2)}}{N^{1/2-1/(4\gamma+2)}}\right)\leqslant\frac{B_4}{N^{1/2}}\,.$$

Таким образом, найдется константа $B_5>0$ такая, что

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma)-R_N(f)+V_N}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|\leqslant$$
$$\leqslant\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{R_2(f,\sigma)+V_N}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|+\frac{B_5(\ln N)^{3/2+1/(4\gamma+2)}}{N^{1/2-1/(4\gamma+2)}}\,. \tag{9}$$

Далее, поступая как в работе [8], получаем, что для некоторых констант B_6, B_7 и B_8 справедливо

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{R_2(f,\sigma)+V_N}{\sigma^2\sqrt{2N}}<x\right)-\Phi(x)\right|\leqslant$$
$$\leqslant\frac{B_6}{N^{1/2}}+\frac{B_7(\ln N)^{1/(2\gamma+1)}}{N^{1-1/(2\gamma+1)}}+\frac{B_8}{N^{2\gamma-1/2}}\,. \tag{10}$$

Оценим теперь третье слагаемое в (7).

$$\mathrm{P}\left(\left|\frac{U_N(\widehat{\sigma})}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)\leqslant$$
$$\leqslant\mathrm{P}\left(\sup_{|s^2-\sigma^2|\leqslant\delta_N}\left|\frac{U_N(s)}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)+\mathrm{P}(\left|\widehat{\sigma}^2-\sigma^2\right|>\delta_N)\,. \tag{11}$$

Можно показать, что при выполнении условий теоремы найдется такая константа C_δ, что при выборе $\delta_N = C_\delta N^{-1/2}(\ln N)^{1/2}$ для некоторой константы $\widetilde{C}_\delta$ выполнено

$$\mathrm{P}(\left|\widehat{\sigma}^2 - \sigma^2\right| > \delta_N) \leqslant \frac{\widetilde{C}_\delta}{\sqrt{N}}\,. \tag{12}$$

Далее

$$\sup_{|s^2-\sigma^2|\leqslant\delta_N} |U_N(s)| = \sup_{|s^2-\sigma^2|\leqslant\delta_N} \sum_{i=1}^{N} \big[\min\{X_i^2, T^2(s)\} + 2s^2 \mathbb{1}\left(|X_i| > T(s)\right) -$$
$$- \min\{X_i^2, T^2(\sigma)\} + 2\sigma^2 \mathbb{1}\left(|X_i| > T(\sigma)\right)\big] \leqslant$$
$$\leqslant \sum_{i=1}^{N} \Big[\Big(T^2(\sqrt{\sigma^2+\delta_N}) - T^2(\sigma)\Big)\, \mathbb{1}\left(|X_i| > T(\sigma)\right) +$$
$$+ 2\sigma^2 \mathbb{1}\left(T(\sqrt{\sigma^2-\delta_N}) < |X_i| \leqslant T(\sigma)\right) +$$
$$+ 2\delta_N \mathbb{1}\left(|X_i| > T(\sqrt{\sigma^2-\delta_N})\right)\Big] \leqslant \sum_{i=1}^{N} \big[2\delta_N \ln N \mathbb{1}\left(|X_i| > T(\sigma)\right) +$$
$$+ 2\sigma^2 \mathbb{1}\left(T(\sqrt{\sigma^2-\delta_N}) < |X_i| \leqslant T(\sigma)\right) + 2\delta_N \mathbb{1}\left(|X_i| > T(\sqrt{\sigma^2-\delta_N})\right)\big] \equiv$$
$$\equiv \sum_{i=1}^{N} h(X_i, \delta_N) = U'_N(\delta_N) + U''_N(\delta_N)\,,$$

где суммирование в $U'_N(\delta_N)$ ведется по $i \in I_1$, а в $U''_N(\delta_N)$ — по $i \in I_2$. Имеем $|h(X_i, \delta_N)| \leqslant 2\delta_N \ln N + 2\sigma^2 + 2\delta_N$ п. в. и если $i \in I_2$, то для некоторых констант $C_E > 0$ и $C_D > 0$

$$\mathrm{E}h(X_i, \delta_N) \leqslant \frac{C_E}{N\sqrt{\ln N}} \text{ и } \mathrm{D}h(X_i, \delta_N) \leqslant \mathrm{E}h^2(X_i, \delta_N) \leqslant \frac{C_D}{N\sqrt{\ln N}}\,.$$

Следовательно, применяя неравенство Бернштейна [13], можно показать, что для некоторых констант C''_U и $\tilde{C}''_U$ справедливо

$$\mathrm{P}\left(\frac{U''_N(\delta_N)}{\sigma^2\sqrt{2N}} > C''_U N^{-1/2} \ln N\right) \leqslant \frac{\tilde{C}''_U}{\sqrt{N}}\,.$$

Если же $i \in I_1$, то, поступая так же, как при оценке R_1, можно выбрать такую константу $C'_\varepsilon > 0$ и

$$\varepsilon'_N = C'_\varepsilon \frac{(\ln N)^{1/2+1/(4\gamma+2)}}{N^{1/2-1/(4\gamma+2)}},$$

что

$$\mathrm{P}\left(\frac{U'_N(\delta_N)}{\sigma^2\sqrt{2N}} > \varepsilon'_N\right) \leqslant \frac{\tilde{C}'_U}{\sqrt{N}}$$

для некоторой константы $\tilde{C}'_U$. Таким образом, существует такая константа C_U, что

$$\mathrm{P}\left(\sup_{|s^2-\sigma^2|\leqslant\delta_N}\left|\frac{U_N(s)}{\sigma^2\sqrt{2N}}\right| > \varepsilon_N\right) \leqslant \frac{C_U}{\sqrt{N}}, \tag{13}$$

где

$$\varepsilon_N = C_\varepsilon \frac{(\ln N)^{1/2+1/(4\gamma+2)}}{N^{1/2-c1/(4\gamma+2)}}$$

и $C_\varepsilon > C'_\varepsilon$.

Объединяя (7) и (9)–(13), получаем (6). Теорема доказана.

Теперь получим оценку скорости сходимости распределения $\widehat{R}_N(f,\widehat{\sigma})$ к нормальному закону при использовании в качестве оценки σ соответствующим образом нормированного интерквартильного размаха $\hat{\sigma}_R$ и абсолютного медианного отклонения от медианы $\hat{\sigma}_M$. Преимущество использования таких оценок заключается в их робастности, т. е. нечувствительности к выбросам (см. [14, 15]), и в полной мере проявляется, когда дисперсия оценивается по выборке сигнала. Оценки $\hat{\sigma}_R$ и $\hat{\sigma}_M$ определяются следующим образом:

$$\hat{\sigma}_R = \frac{X_{N/2,3/4} - X_{N/2,1/4}}{2\xi_{3/4}}; \tag{14}$$

$$\hat{\sigma}_M = \frac{\underset{N/2+1\leqslant i\leqslant N}{\mathrm{med}}|X_i - \underset{N/2+1\leqslant j\leqslant N}{\mathrm{med}} X_j|}{\xi_{3/4}}, \tag{15}$$

где $X_{N/2,1/4}$ и $X_{N/2,3/4}$ — выборочные квантили порядка $1/4$ и $3/4$, построенные по выборке из половины всех вейвлет коэффициентов на уровне $J-1$ ($N=2^J$), $\xi_{3/4}$ – теоретическая квантиль порядка $3/4$ стандартного нормального распределения, а med обозначает выборочную медиану.

Теорема 3. *Пусть f задана на отрезке $[0,1]$ и является равномерно регулярной по Липшицу с показателем $\gamma > 1/2$ и пусть оценка σ^2 задана соотношением* (14) *или соотношением* (15)*, тогда существуют такие константы $\tilde{C}_0$ и $\tilde{C}_1$ (зависящие от γ, A и σ), что*

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\widehat{\sigma}) - R_N(f)}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| \leqslant \frac{\tilde{C}_0}{N^{\gamma-1/2}} + \frac{\tilde{C}_1(\ln N)^{3/4}}{N^{1/4}}, \tag{16}$$

где $\Phi_\Upsilon(x)$ *— функция распределения нормального закона с нулевым средним и дисперсией* $\Upsilon = [2\xi_{3/4}\phi(\xi_{3/4})]^{-2} - 1$.

Доказательство. Поступая, как в предыдущей теореме, имеем

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\widehat{\sigma}) - R_N(f)}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| =$$
$$= \sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma) - R_N(f) + \widehat{R}_N(f,\widehat{\sigma}) - \widehat{R}_N(f,\sigma)}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| \leqslant$$
$$\leqslant \sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma) - R_N(f) + V_N}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| + \frac{\varepsilon_N}{\sqrt{2\pi}} +$$
$$+ \mathrm{P}\left(\left|\frac{U_N(\widehat{\sigma})}{\sigma^2\sqrt{2N}}\right| > \varepsilon_N\right). \quad (17)$$

Здесь

$$\widehat{R}_N(f,\widehat{\sigma}) - \widehat{R}_N(f,\sigma) = U_N(\widehat{\sigma}) + V_N\,;$$
$$V_N = N(\sigma^2 - \widehat{\sigma}^2)\,; \quad U_N(\widehat{\sigma}) = \sum_{i=1}^{N}\{u_i(\widehat{\sigma}) - u_i(\sigma)\}\,,$$

а $u_i(s) = \min\{X_i^2, T^2(s)\} + 2s^2\mathbb{1}\,(|X_i| > T(s))$.

Так же, как в предыдущей теореме, запишем разность $\widehat{R}_N(f,\sigma) - R_N(f)$ в виде:

$$\widehat{R}_N(f,\sigma) - R_N(f) = R_1(f,\sigma) + R_2(f,\sigma)\,,$$

где $R_j(f,\sigma)$ определяются соотношением (8), и для первого слагаемого в (17) получим

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{\widehat{R}_N(f,\sigma) - R_N(f) + V_N}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| \leqslant$$
$$\leqslant \sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{R_2(f,\sigma) + V_N}{\sigma^2\sqrt{2N}} < x\right) - \Phi_\Upsilon(x)\right| + \frac{\tilde{B}(\ln N)^{3/2+1/(4\gamma+2)}}{N^{1/2-1/(4\gamma+2)}} \quad (18)$$

с некоторой константой $\tilde{B} > 0$.

Далее, используя разложение Бахадура для выборочных квантилей [16] и результаты работ [6, 9, 14, 17–19], а также учитывая (2), можно показать, что для некоторых констант $\tilde{B}_1$, $\tilde{B}_2$ и $\tilde{B}_3$ справедливо

$$\sup_{x\in\mathbf{R}}\left|\mathrm{P}\left(\frac{R_2(f,\sigma)+V_N}{\sigma^2\sqrt{2N}}<x\right)-\Phi_\Upsilon(x)\right|\leqslant\frac{\tilde{B}_1(\ln N)^{3/4}}{N^{1/4}}+$$
$$+\frac{\tilde{B}_2(\ln N)^{1/(2\gamma+1)}}{N^{1-1/(2\gamma+1)}}+\frac{\tilde{B}_3}{N^{\gamma-1/2}}. \quad (19)$$

Для третьего слагаемого в (17) имеем

$$\mathrm{P}\left(\left|\frac{U_N(\widehat{\sigma})}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)\leqslant\mathrm{P}\left(\sup_{|s-\sigma|\leqslant\delta_N}\left|\frac{U_N(s)}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)+\mathrm{P}(|\widehat{\sigma}-\sigma|>\delta_N). \quad (20)$$

Используя экспоненциальные неравенства для выборочных квантилей или абсолютного медианного отклонения (см. [14] и [18]), можно показать, что при выполнении условий теоремы найдется такая константа $D_\delta>0$, что при выборе $\delta_N=D_\delta N^{-1/2}(\ln N)^{1/2}$ для некоторой константы $\widetilde{D}_\delta>0$ выполнено

$$\mathrm{P}(|\widehat{\sigma}-\sigma|>\delta_N)\leqslant\frac{\widetilde{D}_\delta}{\sqrt{N}}. \quad (21)$$

Далее

$$\mathrm{P}\left(\sup_{|s-\sigma|\leqslant\delta_N}\left|\frac{U_N(s)}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right)\leqslant\mathrm{P}\left(\sup_{|s^2-\sigma^2|\leqslant\delta'_N}\left|\frac{U_N(s)}{\sigma^2\sqrt{2N}}\right|>\varepsilon_N\right), \quad (22)$$

где $\delta'_N=\delta_N^2+2\sigma\delta_N\leqslant D'_\delta N^{-1/2}(\ln N)^{1/2}$ с некоторой константой $D'_\delta>0$. Правая часть в (22) оценивается так же, как в предыдущей теореме.

Объединяя (17)–(22) с учетом того, что $\gamma-1/2<1/2-1/(4\gamma+2)$ при $\gamma<3/4$, получаем (16). Теорема доказана.

Литература

1. *Donoho D. L., Johnstone I. M.* Ideal spatial adaptation via wavelet shrinkage // Biometrika, 1994. Vol. 81. No. 3. P. 425–455.
2. *Donoho D. L., Johnstone I. M.* Adapting to unknown smoothness via wavelet shrinkage // J. Amer. Stat. Assoc., 1995. Vol. 90. No. 432. P. 1200–1224.
3. *Donoho D. L., Johnstone I. M., Kerkyacharian G., Picard D.* Wavelet shrinkage: Asymptopia? // J. R. Statist. Soc. Ser. B, 1995. Vol. 57. No. 2. P. 301–369.

4. *Marron J. S., Adak S., Johnstone I. M., Neumann M. H., Patil P.* Exact risk analysis of wavelet regression // J. Comput. Graph. Stat., 1998. Vol. 7. P. 278–309.
5. *Antoniadis A., Fan J.* Regularization of wavelet approximations // J. Amer. Stat. Assoc., 2001. Vol. 96. No. 455. P. 939–967.
6. *Маркин А. В.* Предельное распределение оценки риска при пороговой обработке вейвлет-коэффициентов // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 57–63.
7. *Маркин А. В., Шестаков О. В.* О состоятельности оценки риска при пороговой обработке вейвлет-коэффициентов // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн., 2010. № 1. С. 26–34.
8. *Шестаков О. В.* Аппроксимация распределения оценки риска пороговой обработки вейвлет-коэффициентов нормальным распределением при использовании выборочной дисперсии // Информатика и её применения, 2010. Т. 4. Вып. 4. С. 73–81.
9. *Шестаков О. В.* О скорости сходимости оценки риска пороговой обработки вейвлет-коэффициентов к нормальному закону при использовании робастных оценок дисперсии // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 122–128.
10. *Добеши И.* Десять лекций по вейвлетам. — Ижевск: Регулярная и хаотическая динамика, 2001.
11. *Mallat S.* A wavelet tour of signal processing. — NY: Academic Press, 1999.
12. *Abramovich F., Silverman B. W.* Wavelet decomposition approaches to statistical inverse problems // Biometrika, 1998. Vol. 85. No. 1. P. 115–129.
13. *Bennett G.* Probability inequalities for sums of independent random variables // J. Amer. Stat. Assoc., 1962. Vol. 57. P. 33–45.
14. *Serfling R.* Approximation theorems of mathematical statistics. —- NY: John Wiley and Sons, 1980.
15. *Hall P., Welsh A. H.* Limits theorems for median deviation // Ann. Inst. Statist. Math., 1985. Vol. 37. No. 1. P. 27–36.
16. *Bahadur R. R.* A note on quantiles in large samples // Ann. Statist., 1966. Vol. 37. No. 3. P. 577–580.
17. *DasGupta A.* Asymptotic values and expansions for the correlation between different measures of spread // J. Stat. Planning Inference, 2006. Vol. 136. No. 7. P. 2197–2212.
18. *Serfling R., Mazumder S.* Exponential probability inequality and convergence results for the median absolute deviation and its modifications // Stat. Probab. Lett., 2009. Vol. 79. No. 16. P. 1767–1773.
19. *Mazumder S., Serfling R.* Bahadur representations for the median absolute deviation and its modifications // Stat. Probab. Lett., 2009. Vol. 79. No. 16. P. 1774–1783.

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2012. Т. 22. № 1. С. 153–166

ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТ-АНАЛИЗА В КЛИМАТИЧЕСКИХ ИССЛЕДОВАНИЯХ*

М. Ш. Хазиахметов[1], *Т. В. Захарова*[2]

Аннотация: Описано применение вейвлет-анализа цифровых сигналов к одной из задач климатических исследований — поиску числовых закономерностей в поле водяного пара Земли.

Ключевые слова: вейвлет-анализ; вейвлет-разложение; вейвлет-преобразование

1 Введение

В настоящее время существует множество методов для анализа цифровых сигналов. Классическим примером такого метода является анализ Фурье, он же является одним из наиболее применимых. Преобразование Фурье раскладывает сигнал в сумму гармонических компонент с разными частотами. Однако такой метод хорошо работает лишь со стационарными сигналами, когда удается найти периодические составляющие. Кроме того, преобразование Фурье дает форму сигнала в частотной области, но оно не несет информации о том, в какой момент времени в сигнале присутствовали колебания с данной частотой. Тем самым теряется информация о локализации явлений во временно́й области.

Безусловно, есть и другие методы анализа и обработки цифровых сигналов. Значительный вклад в развитие цифровой обработки сигналов внес известный французский ученый Стефан Малла. В конце 1980-х гг. он предложил схему разложения сигнала, на которой построен современный вейвлет-анализ. Основная идея заключается в разложении сигнала на две части: аппроксимацию и детали. Данная процедура может быть последовательно применена к получаемым аппроксимациям сигнала несколько раз, тем самым осуществляется многоуровневое вейвлет-разложение исходного сигнала. Дальнейшая работа по обработке сигнала обычно проводится одним из двух способов: либо анализируется сам

*Работа выполнена при поддержке РФФИ (проекты 11-01-12026-офи-м, 11-01-00515), а также Министерства образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

[1]Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, khaziakhmetov@yandex.ru

[2]Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, lsa@cs.msu.ru

результат декомпозиции и производится его интерпретация в рамках решаемой задачи, либо выполняется необходимая обработка вейвлет-коэффициентов и восстановление сигнала с применением к модифицированным коэффициентам обратного вейвлет-преобразования.

Данная статья описывает применение вейвлет-анализа цифровых сигналов к одной из задач климатических исследований. Работа авторов заключалась в поиске числовых закономерностей в поле водяного пара Земли.

2 Базовые сведения о вейвлет-преобразовании

В данный раздел помещена информация о вейвлет-преобразовании, необходимая для понимания содержимого статьи. Более подробный материал содержится в [1–4].

2.1 Одномерное вейвлет-преобразование

2.1.1 Одномерное дискретное вейвлет-преобразование

Определение 1. Функция $\varphi(x) \in L^2(\mathbb{R})$ называется масштабирующей, если она может быть представлена в виде

$$\varphi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \varphi(2x - n)\,, \tag{1}$$

где числа $h_n, n \in \mathbb{Z}$, удовлетворяют условию:

$$\sum_{n \in \mathbb{Z}} |h_n|^2 < \infty\,.$$

Равенство (1) называют масштабирующим уравнением, а соответствующий набор коэффициентов $\{h_n\}_{n \in \mathbb{Z}}$ — масштабирующим фильтром.

Функция Хаара является масштабирующей и имеет следующий вид:

$$F(x) = \begin{cases} 1, & x \in [0, 1)\,; \\ 0 & \text{иначе}\,. \end{cases}$$

Заметим, что в то же время центрированная функция Хаара

$$F(x) = \begin{cases} 1, & x \in \left[-\dfrac{1}{2}, \dfrac{1}{2}\right) \\ 0 & \text{иначе} \end{cases};$$

масштабирующей не является.

Определение 2. Ортогональным кратномасштабным разложением пространства $L^2(\mathbb{R})$ называется последовательность замкнутых вложенных друг в друга подпространств $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots$, обладающая свойствами:

(1) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$;

(2) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$;

(3) $f(x) \in V_j \iff f(2x) \in V_{j+1}$;

(4) $\exists\varphi(x) \in V_0 :$ $\{\varphi_{0,n}(x) = \varphi(x-n)\}_{n\in\mathbb{Z}}$ образуют ортонормированный базис пространства V_0.

Из свойств 3 и 4 определения 2 непосредственно вытекает, что последовательность функций

$$\{\varphi_{j,n}(x) = \sqrt{2^j}\varphi(2^j x - n)\}_{n\in\mathbb{Z}}$$

образует ортонормированный базис в пространстве V_j.

Для каждого индекса j имеем $V_j \subset V_{j+1}$, обозначим через W_j ортогональное дополнение V_j до V_{j+1}, т. е. $V_{j+1} = V_j \oplus W_j$. Так как $V_j = V_{j-1} \oplus W_{j-1}$, то $V_{j+1} = V_{j-1} \oplus W_{j-1} \oplus W_j$, и, продолжая эту процедуру, получим ортогональное разложение V_{j+1} в виде:

$$V_{j+1} = \overline{\bigoplus_{k=-\infty}^{j} W_k}\,.$$

Тогда в силу свойства 1 определения 2 получаем ортогональное разложение пространства $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \overline{\bigoplus_{k=-\infty}^{+\infty} W_k}\,.$$

Определение 3. Если последовательность подпространств $\cdots \subset V_{-1} \subset V_0 \subset \subset V_1 \subset \cdots$ образует кратномасштабный анализ, то для каждого $j \in \mathbb{Z}$ ортогональное дополнение W_j к пространству V_j до V_{j+1} называется пространством вейвлетов, а его элементы — вейвлетами.

Вейвлеты обладают следующими свойствами:

– они образуют ортогональное разложение пространства $L^2(\mathbb{R})$;

– существует функция $\psi(x) \in W_0$, называемая материнским вейвлетом, такая что

(1) $\{\psi(x-n)\}_{n\in\mathbb{Z}}$ — ортонормированный базис пространства W_0;

(2) $\{\psi_{j,n}(x) = \sqrt{2^j}\psi(2^j x - n)\}_{n\in\mathbb{Z}}$ — ортонормированный базис пространства W_j для любого j.

Обозначим скалярное произведение функций $f(x)$ и $g(x)$ как $\langle f(x), g(x)\rangle$. В $L^2(\mathbb{R})$ имеем:

$$\langle f(x), g(x)\rangle = \int_{-\infty}^{+\infty} f(x)g(x)\,dx\,. \tag{2}$$

Пусть $f(x) \in V_{j+1}$. Тогда имеет место так называемое вейвлет-разложение:

$$f(x) = \sum_{k=-\infty}^{j} \sum_{n\in\mathbb{Z}} \langle f(x), \psi_{k,n}(x)\rangle \psi_{k,n}(x)\,.$$

На практике приходится ограничивать число уровней детализации, потому имеет место следующее представление:

$$f(x) = \sum_{k=0}^{j} \sum_{n\in\mathbb{Z}} \langle f(x), \psi_{k,n}(x)\rangle \psi_{k,n}(x) + \sum_{n\in\mathbb{Z}} \langle f(x), \varphi(x-n)\rangle \varphi(x-n)\,,$$

т. е. сигнал из пространства V_{j+1} представляется в виде элементов пространств $\{V_0, W_0, \ldots, W_j\}$.

В дискретном случае интеграл в формуле (2) превращается в сумму. Пусть $x = \{x_n\}_{n\in\mathbb{Z}}$ — оцифровка сигнала $x(t), t \in \mathbb{R}$. Тогда наиболее удобное представление для вейвлет-коэффициента $a_{j,n}$ получается в терминах дискретной свертки с цифровым фильтром $h = \{h_n\}_{n\in\mathbb{Z}}$, соответствующим вейвлету $\psi_{j,n}$:

$$a_{j,n} = \sum_{k=-\infty}^{\infty} h_k x_{n-k}\,.$$

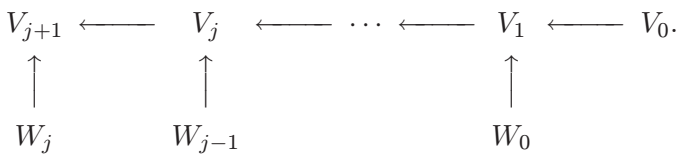
Обычно исследование сигнала начинается с представления его в терминах базиса пространства V_{j+1}. Далее производим разложение по базисам V_j и W_j, получая тем самым аппроксимирующую и детализирующую составляющие. Затем вновь разделяем полученную аппроксимацию на составляющие и т. д. Схематически этот процесс можно представить так:

$$\begin{array}{ccccccccc} V_{j+1} & \longrightarrow & V_j & \longrightarrow & \cdots & \longrightarrow & V_1 & \longrightarrow & V_0. \\ \downarrow & & \downarrow & & & & \downarrow & & \\ W_j & & W_{j-1} & & & & W_0 & & \end{array}$$

Здесь элементы разложения по пространству V_0 — грубая аппроксимация исходного сигнала, а коэффициенты, соответствующие базисным функциям W_0, — самые глобальные детали, полученные в ходе декомпозиции сигнала.

Как было отмечено во введении к статье, можно использовать коэффициенты декомпозиции, интерпретируя их соответствующим образом, и делать выводы, базируясь на них, либо выполнить над ними некоторые операции и осуществить реконструкцию сигнала.

Процесс реконструкции выполняется обратным ходом:



2.1.2 Одномерное непрерывное вейвлет-преобразование

Для одномерного дискретного вейвлет-преобразования справедлива формула:

$$\psi_{j,n}(x) = \sqrt{2^j}\psi(2^j x - n)\,, \quad j,\, n \in \mathbb{Z}\,.$$

Непрерывное преобразование строится путем разрешения произвольных вещественных значений параметров масштабирования и сдвига (в дискретном случае коэффициент масштабирования может быть только степенью двойки, а сдвиг — целочисленным). Данное обобщение дискретного преобразования позволяет выделять детали сигнала с произвольной величиной носителя.

Пусть $\psi(x)$ — некоторый вейвлет и $\psi(x) \in L^2(\mathbb{R})$. Предположим дополнительно, что справедливо

$$C_\psi = 2\pi \int\limits_{-\infty}^{+\infty} |\omega|^{-1} |\hat{\psi}(\omega)|^2 d\omega < \infty\,. \tag{3}$$

Соотношение (3) необходимо для существования обращения непрерывного преобразования.

Определим двухпараметрическое семейство вейвлетов с параметром масштаба a и параметром сдвига b как

$$\psi_{ab}(x) = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right)\,, \quad a,\, b \in \mathbb{R}\,.$$

Одномерное непрерывное вейвлет-преобразование задается формулой:

$$W_\psi\,[f]\,(a,b) = (f, \psi_{a,b}) = |a|^{-1/2} \int\limits_{-\infty}^{+\infty} f(x)\overline{\psi\left(\frac{x-b}{a}\right)}\,dx\,. \tag{4}$$

Очевидно, что коэффициенты дискретного разложения связаны с непрерывным преобразованием следующим соотношением:

$$c_{jk} = W_\psi[f]\left(\frac{1}{2^j}, \frac{k}{2^j}\right). \tag{5}$$

Следующая теорема восстановления исходной функции доказана в [2].

Теорема 1. *Если* $f(x) \in L^2(\mathbb{R})$ *и выполнено условие* (3)*, то справедлива формула обращения*

$$f(x) = C_\psi^{-1} \iint W_\psi[f](a,b)\psi_{ab}(x)\frac{dadb}{a^2}.$$

2.2 Двумерное дискретное вейвлет-преобразование

В данной части статьи рассмотрим случай функций из пространства $L^2(\mathbb{R}^2)$.

Наиболее простой путь построения многомерного обобщения для вейвлет-преобразования основан на тензорном произведении. Для $L^2(\mathbb{R}^2)$ имеем следующее представление:

$$L^2(\mathbb{R}^2) = L^2(\mathbb{R}) \otimes L^2(\mathbb{R}).$$

Таким образом, линейные комбинации функции $f(x)g(y)$ образуют плотное множество в пространстве $L^2(\mathbb{R}^2)$.

Определим V_0^2 как тензорное произведение пространств V_0 функций одной переменной:

$$V_0^2 = V_0 \otimes V_0.$$

Тогда набор функций вида

$$\{\varphi_{0,k,n}(x,y) = \varphi(x-k)\varphi(y-n)\}_{k,n\in\mathbb{Z}}$$

образует ортонормированный базис в V_0^2.

Пространства $V_j^2 = V_j \otimes V_j$ — масштабированные версии пространства V_0^2, и для них имеет место соотношение:

$$f(x,y) \in V_0^2 \iff f(2^j x, 2^j y) \in V_j^2.$$

Поэтому, как и в одномерном случае, получаем последовательность вложенных друг в друга подпространств: $\cdots \subset V_{-1}^2 \subset V_0^2 \subset V_1^2 \subset \cdots$ Далее, поскольку $V_1 = V_0 \oplus W_0$, получаем для двумерного случая:

$$\begin{aligned} V_1^2 = V_1 \otimes V_1 &= (V_0 \oplus W_0) \otimes (V_0 \oplus W_0) = \\ &= (V_0 \otimes V_0) \oplus (V_0 \otimes W_0) \oplus (W_0 \otimes V_0) \oplus (W_0 \otimes W_0) = \\ &= V_0^2 \oplus (V_0 \otimes W_0) \oplus (W_0 \otimes V_0) \oplus (W_0 \otimes W_0). \end{aligned}$$

Пространства $V_0 \otimes W_0$, $W_0 \otimes V_0$, $W_0 \otimes W_0$ в сумме образуют пространство двумерных вейвлетов W_0^2, причем

- пространство $V_0 \otimes W_0$ порождено сдвигами функции $\psi^H(x,y) = \varphi(x)\psi(y)$, будем обозначать его W_0^H — это так называемые горизонтальные детали (фиксируются горизонтальные элементы сигнала, т. е. области, однородные вдоль оси OX);
- пространство $W_0 \otimes V_0$ порождено сдвигами функции $\psi^V(x,y) = \psi(x)\varphi(y)$, будем обозначать его W_0^V — это так называемые вертикальные детали (фиксируются вертикальные элементы сигнала, т. е. области, однородные вдоль оси OY);
- пространство $W_0 \otimes W_0$ порождено сдвигами функции $\psi^D(x,y) = \psi(x)\psi(y)$, будем обозначать его W_0^D — это так называемые диагональные детали (фиксируются диагональные неоднородности сигнала).

Таким образом, справедливо разложение:

$$V_{j+1}^2 = V_j^2 \oplus W_j^H \oplus W_j^V \oplus W_j^D\,, \quad \forall j\,.$$

Нетрудно проверить, что следующие системы функций образуют ортонормированные базисы в указанных выше пространствах:

$$\begin{aligned}
&\{\varphi_{j,k,n}(x,y) = 2^j\varphi(2^jx-k)\varphi(2^jy-n)\}_{k,n\in\mathbb{Z}}\,;\\
&\{\psi^H_{j,k,n}(x,y) = 2^j\varphi(2^jx-k)\psi(2^jy-n)\}_{k,n\in\mathbb{Z}}\,;\\
&\{\psi^V_{j,k,n}(x,y) = 2^j\psi(2^jx-k)\varphi(2^jy-n)\}_{k,n\in\mathbb{Z}}\,;\\
&\{\psi^D_{j,k,n}(x,y) = 2^j\psi(2^jx-k)\psi(2^jy-n)\}_{k,n\in\mathbb{Z}}\,.
\end{aligned}$$

Соответственно, для сигнала будем иметь 4 типа коэффициентов:

$$\begin{aligned}
\langle f(x,y), \varphi_{j,k,n}(x,y)\rangle &= 2^j \int\limits_{\mathbb{R}^2} f(x,y)\varphi(2^jx-k)\varphi(2^jy-n)\,dxdy\,;\\
\langle f(x,y), \psi^H_{j,k,n}(x,y)\rangle &= 2^j \int\limits_{\mathbb{R}^2} f(x,y)\varphi(2^jx-k)\psi(2^jy-n)\,dxdy\,;\\
\langle f(x,y), \psi^V_{j,k,n}(x,y)\rangle &= 2^j \int\limits_{\mathbb{R}^2} f(x,y)\psi(2^jx-k)\varphi(2^jy-n)\,dxdy\,;\\
\langle f(x,y), \psi^D_{j,k,n}(x,y)\rangle &= 2^j \int\limits_{\mathbb{R}^2} f(x,y)\psi(2^jx-k)\psi(2^jy-n)\,dxdy\,.
\end{aligned}$$

Схемы процессов декомпозиции и реконструкции сигнала соответственно приобретают вид:

$$\begin{array}{ccccccccc} V_{j+1}^2 & \longrightarrow & V_j^2 & \longrightarrow & \cdots & \longrightarrow & V_1^2 & \longrightarrow & V_0^2, \\ \downarrow & & \downarrow & & & & \downarrow & & \\ W_j^H, W_j^V, W_j^D & & W_{j-1}^H, W_{j-1}^V, W_{j-1}^D & & & & W_0^H, W_0^V, W_0^D & & \end{array}$$

$$\begin{array}{ccccccccc} V_{j+1}^2 & \longleftarrow & V_j^2 & \longleftarrow & \cdots & \longleftarrow & V_1^2 & \longleftarrow & V_0^2. \\ \uparrow & & \uparrow & & & & \uparrow & & \\ W_j^H, W_j^V, W_j^D & & W_{j-1}^H, W_{j-1}^V, W_{j-1}^D & & & & W_0^H, W_0^V, W_0^D & & \end{array}$$

3 Задача анализа поля водяного пара Земли

3.1 Введение

Земная атмосфера очень сложна для анализа, процессы в ней крайне тяжело прогнозировать. Энергия, заключенная в ней, очень велика и находится, главным образом, в водяном паре, что связано с его большой теплоемкостью. Изучение процессов, происходящих в поле водяного пара, способно помочь объяснить и предсказать многие атмосферные явления, например тропические циклоны, обладающие большой разрушительной силой. Изучение циклонов и ураганов особенно важно вследствие их значительного воздействия на окружающую среду.

Многие атмосферные явления имеют периодическую природу, например поле водяного пара имеет годовой цикл. Однако наиболее интересные для изучения явления таким свойством не обладают. Они должны быть локализованы в пространстве и времени, а для этого нужны адекватные методы.

В проведенных исследованиях был использован вейвлет-анализ.

3.2 Постановка задачи

Для каждого дня с 01.01.1999 до 31.12.2009 включительно имеются цифровые карты плотности водяного пара на поверхности Земли размером 360×720 точек (шаг градусной сетки земного шара составлял $0{,}5^\circ$). Значение каждого пиксела — плотность (в кг$/$м^2) водяного пара в сферическом сегменте Земли с соответствующими географическими координатами в заданный день (если быть более точными, то на момент создания космического снимка), округленная до ближайшего целого. Таким образом, для анализа имелся трехмерный массив целых чисел.

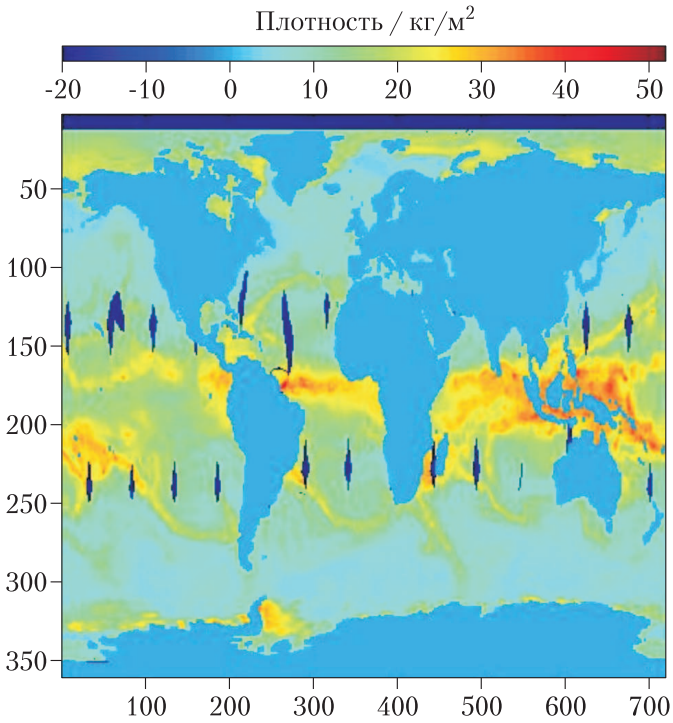


Рис. 1 Поле водяного пара Земли 01.01.1999

Следует сразу оговорить несколько особенностей данного массива. Дело в том, что алгоритм построения значений плотности водяного пара по данным съемки из космоса может быть применен только к атмосфере над поверхностью Мирового Океана, поэтому все точки, размещенные над поверхностью суши, были заполнены нулями. Кроме того, система спутников не гарантировала покрытие всей поверхности Земли снимками в течение суток, соответственно для некоторых элементов массива данные просто отсутствовали; организацией, предоставившей данные, они были заполнены значением -20 (вследствие высокой межсуточной изменчивости предложить адекватный метод интерполяции не представляется возможным).

Очевидно, наиболее удобным способом для отображения информации о плотности водяного пара в заданный день является карта, каждая точка которой окрашена в цвет, соответствующий плотности водяного пара в ней. Рисунок 1 демонстрирует пример такого изображения. На нем можно заметить все перечисленные особенности массива данных, касающиеся неполноты информации о поле водяного пара.

Таким образом, задача анализа поля водяного пара сводится к анализу данного массива на наличие закономерностей.

Предварительный анализ данных указал на довольно важный факт, который серьезно уменьшил объем исходных данных. Дисперсии значений плотности за первые 7 лет и последние 4 года отличались примерно в 2 раза (как оказалось, организация, предоставившая данные, с 2006 г. изменила алгоритм вычисления плотности), вследствие этого было решено использовать лишь первую часть данных.

3.3 Методы исследования

Данные анализировались в двух аспектах:

(1) исследовались временны́е ряды плотности водяного пара при фиксированных географических координатах;

(2) исследовались изменения плотности по времени вдоль фиксированного меридиана.

Все алгоритмы анализа массива данных были реализованы в Matlab с применением Wavelet Toolbox.

3.3.1 Исследование временны́х рядов

Итак, слои имеющегося трехмерного массива представляют собой карты плотности водяного пара. Зафиксируем значение географических координат и извлечем значения из каждой карты в этой позиции. Расположив их в порядке дат, получим временной ряд. Пример такого ряда приведен на рис. 2.

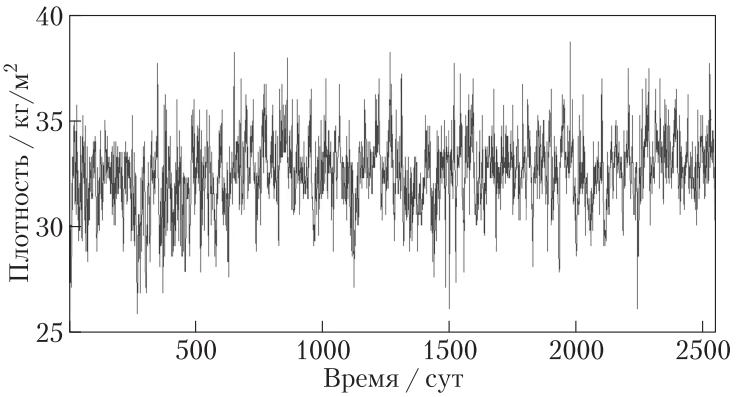


Рис. 2 Пример временно́го ряда

Wavelet Toolbox позволяет анализировать одномерные массивы с применением и дискретного, и непрерывного вейвлет-преобразования. Очевидно, второй вариант предпочтительнее, так как позволяет исследовать детали произвольной длины, а не только степени двойки (собственно, данные дискретного преобразования всегда содержатся в непрерывном преобразовании в силу (5)). В ходе экспериментов с данными были исследованы многие семейства вейвлетов, для некоторых из них были получены интересные результаты.

Наиболее значимые результаты были получены при использовании вейвлета Морле (в обозначениях Matlab — cmor1-1.5). Данный вейвлет — комплексная функция следующего вида:

$$\psi(t) = e^{2\pi i t - t^2/2}.$$

Декомпозиция каждого временно́го ряда проводилась в соответствии с формулой (4).

Коэффициенты декомпозиции образуют двумерный массив, где по столбцам находятся коэффициенты, вычисленные с одним и тем же сдвигом вейвлет-фильтра относительно начала сигнала, а по строкам — коэффициенты с одинаковой длиной фильтра. Для каждого значения географических координат был получен такой массив коэффициентов. Для каждой пары (географические координаты, коэффициент масштаба декомпозиции) с использованием преобразования Фурье была определена главная частота. Проделав данную процедуру для каждого

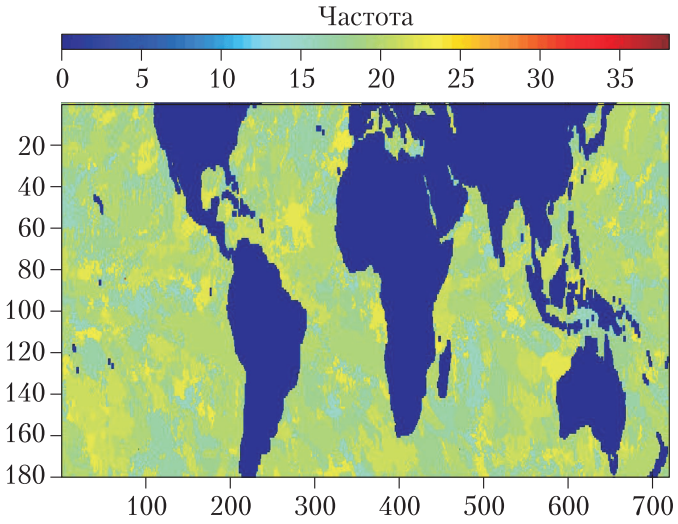


Рис. 3 Пример «карты частот» (30-дневная активность)

значения коэффициента масштаба получаем так называемые «карты частот». Впервые этот термин был введен в работе [5]. Пример такой карты приведен на рис. 3. Здесь представлена лишь область 45° с. ш. – 45° ю. ш. (она наиболее интересна с точки зрения анализа и прогнозирования атмосферных явлений).

Коэффициенты детализации вейвлет-разложения показывают изменения сигнала. Соответственно, главная частота для каждой длины фильтра в фиксированной точке карты показывает относительную изменчивость (активность в поле водяного пара) в данной области на промежутках указанной длины. Рисунок 3 демонстрирует распределение 30-дневной активности.

Алгоритм вычисления относительной активности в поле водяного пара, разработанный и используемый Институтом космических исследований РАН, был значительно дополнен. Используя полученные карты частот, удалось выявить новые, более тонкие закономерности в поле водяного пара Земли. Были определены подзоны с различной вариабельностью поля водяного пара, а положение известных зон было уточнено. Также были подтверждены все сезонные закономерности и высокая ежедневная изменчивость значений плотности водяного пара.

3.3.2 Меридиональный анализ

На практике большой интерес представляет анализ изменения по времени значений плотности водяного пара на фиксированном меридиане. Из исходного массива данных плотности были извлечены векторы значений плотности, соответствующие рассматриваемому меридиану, и из них был составлен двумерный массив. Наиболее удобно работать с такими массивами в терминах индексированного изображения. Его пример приведен на рис. 4.

Для анализа изображения были использованы биортогональные вейвлеты с параметризацией 1.1 согласно терминологии Matlab (bior1.1).

Физический смысл декомпозиции для исследуемого массива достаточно прост. Двумерное вейвлет-преобразование на каждом уровне разложения удваивает размер охватываемых деталей. В данном случае детали рассматриваются в пространстве (вдоль меридиана) и/или во времени. В соответствии со сказанным выше можно утверждать, что первый уровень декомпозиции состоит из деталей, охватывающих междневные изменения плотности и/или изменчивость вдоль меридиана между соседними областями размером $0{,}5° \times 0{,}5°$, пятый уровень можно считать соответствующим изменениям на протяжении одного месяца. Нужно иметь в виду, что детали двумерного разложения бывают трех типов: горизонтальные, вертикальные и диагональные. Горизонтальные детали отражают постоянство во времени и изменчивость вдоль меридиана, вертикальные — наоборот, отражают изменчивость во времени и постоянство вдоль меридиана, диагональные детали фиксируют пространственно-временну́ю активность.

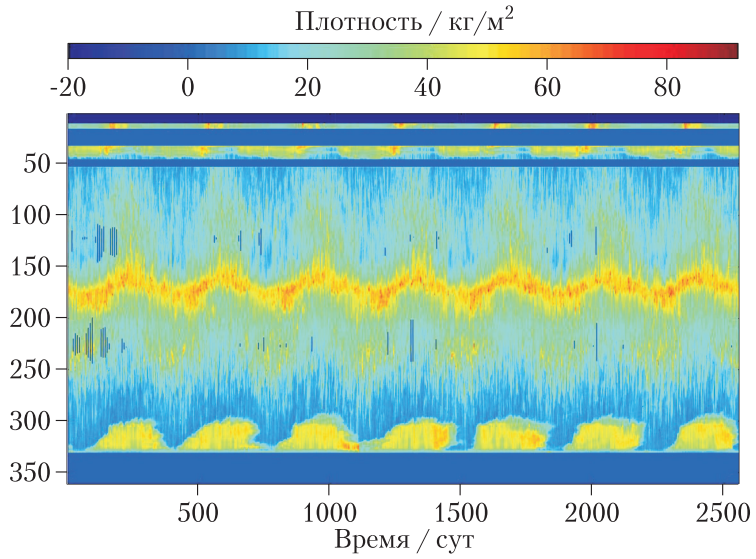


Рис. 4 Пример распределения плотности в поле водяного пара за 1999–2005 гг. на меридиане 21,5° западной долготы

Было зафиксировано множество интересных закономерностей в поле водяного пара, найденных сравнительным анализом полученных декомпозиций. Приведем наиболее значимые из них:

- на основе анализа первого уровня коэффициентов декомпозиции сделан вывод о превалировании временно́й изменчивости над межширотной, при этом наибольшая активность наблюдается в экваториальной зоне и средних широтах, в субтропиках и полярных областях она значительно слабее;
- при анализе горизонтальных деталей уровня 5 зафиксировано серьезное отличие в месячной активности соответственно в экваториальной, субтропических, приполярных зонах и в средних широтах.

3.4 Основные результаты и идеи дальнейшей работы

У авторов есть множество идей по развитию данной работы в частности и в целом по применению аппарата вейвлет-преобразования к анализу и обработке данных. Разрабатывается гипотеза о том, что активность в поле водяного пара предопределяет возникновение атмосферных явлений. Многое говорит о том, что «карты частот» могут помочь создать необходимый математический аппарат для прогнозирования таких атмосферных явлений, как тропические циклоны. В данный момент эти предположения основаны на визуальном сравнении карт частот и

траекторий циклонов, которое выявило определенную зависимость между конфигурацией высокочастотных областей (зон сильной изменчивости) и направлением движения циклонов. В перспективе предполагается более тщательное изучение этой гипотезы и ее проверка с привлечением соответствующих статистических методов.

Подводя итог вышесказанному, можно выделить следующие основные результаты работы:

- предложен новый метод для анализа данных (в частности, климатических);
- введен новый инструмент для анализа — карта частот;
- выявлены новые закономерности в поле водяного пара;
- высказано предположение о том, что на основе карт частот можно предложить методики для прогнозирования атмосферных явлений.

Литература

1. *Bogges A., Narkovich F. A.* A first course in wavelets with Fourier analysis. — Prentice Hall, 2001.
2. *Малла С.* Вэйвлеты в обработке сигналов / Пер. с англ. — М.: Мир, 2005.
3. *Смоленцев Н. К.* Основы теории вейвлетов. Вейвлеты в Matlab. — М.: ДМК Пресс, 2005.
4. *Захарова Т. В., Шестаков О. В.* Вейвлет-анализ и его приложения: Учебное пособие. — 2-е изд. — М: ИНФРА-М, 2012.
5. *Хазиахметов М. Ш., Шрамков Я. Н., Захарова Т. В.* О применении вейвлет-анализа в задачах климатических исследований // Обозрение прикладной и промышленной математики, 2011. Т. 18. Вып. 1. С. 153–155.

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2012. Т. 22. № 1. С. 167–179

УСТОЙЧИВОСТЬ МАСШТАБНЫХ СМЕСЕЙ НОРМАЛЬНЫХ ЗАКОНОВ ОТНОСИТЕЛЬНО ИЗМЕНЕНИЙ СМЕШИВАЮЩЕГО РАСПРЕДЕЛЕНИЯ*

А. К. Горшенин[1]

Аннотация: Рассмотрен вопрос изменения конечных масштабных смесей нормальных законов при малых изменениях параметров смешивающего распределения в двух практически важных моделях конечных масштабных смесей нормальных законов. Устойчивость к таким изменениям устанавливается в терминах двусторонних соотношений для расстояний Леви между смешивающими распределениями и смесями.

Ключевые слова: масштабные смеси нормальных законов; метрика Леви

1 Введение

Математические модели, основанные на конечных масштабных смесях нормальных законов, широко применяются для исследования стохастической структуры процессов при анализе информационных потоков в вычислительных или телекоммуникационных системах, при исследовании поведения биржевых цен и характеристик турбулентной плазмы. Столь разные по своей природе процессы объединяют некоторые общие черты, которые и служат обоснованием возможности применения схожих математических моделей для анализа. Так, для каждой из областей характерны непредсказуемость, неоднородность по времени (например, интенсивность взаимодействия в рамках информационной системы может изменяться в течение времени функционирования данной системы в зависимости от различных факторов; активность торгов может изменяться в течение торгового дня; в плазме наблюдается структурная турбулентность), наличие более-менее устойчивых внутренних структур, оказывающих существенное влияние на функционирование всей системы (например, стационарные объекты информационной системы, солитоны в плазме, группировки участников финансовых рынков). При проведении анализа в данных областях внутренняя структура процесса чаще всего исследователю неизвестна.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты 11-01-12026-офи-м, 10-07-00021.

[1]Институт проблем информатики Российской академии наук, agorshenin@ipiran.ru

В настоящей работе предлагаются две модели конечных масштабных смесей нормальных распределений: добавления и расщепления компоненты. Данные модели часто встречаются на практике при проведении статистического анализа с помощью различных алгоритмов разделения смесей вероятностных распределений (например, с использованием различных модификаций EM (expectation–maximization) алгоритма, сеточных методов).

Модель добавления компоненты соответствует ситуации, при которой в результате вычислений параметров смеси одним из методов декомпозиции смесей возникают компоненты с малым весом. В такой ситуации необходимо убедиться в значимости компоненты с малым весом. Модель расщепления компоненты может применяться в ситуации, в которой необходимо проверить, не являются ли полученные компоненты ошибочно «склеенными» (так как в ряде ситуаций подобные алгоритмы объединяют близкие по параметрам компоненты, что не соответствует истинному распределению исходных данных). Здесь также возникает необходимость убедиться в значимости компоненты с малым весом (так как вес какой-либо из компонент должен быть разделен между ней и некоторой дополнительной компонентой). Факт наличия устойчивости в данных моделях играет решающую роль при практическом применении в прикладных задачах, так как он позволяет считать результаты, получаемые в рамках математической модели, адекватными.

2 Постановка задачи

Предположим, что каждое из независимых наблюдений $\mathbf{X}_n = (X_1, \ldots, X_n)$ имеет распределение, представимое в виде конечной масштабной смеси нормальных законов, т. е.

$$G(x) = \sum_{i=1}^{k} p_i \Phi(x\sigma_i)\,, \tag{1}$$

где

$$\sum_{i=1}^{k} p_i = 1\,,\ \ p_i \geqslant 0\,,\ \ \sigma_i > 0\,,\ \ i = 1, \ldots, k\,,$$

через $\Phi(\cdot)$ обозначена функция распределения стандартного нормального закона:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\frac{t^2}{2}\right\} dt\,.$$

Также в дальнейшем понадобится вид функции плотности стандартного нормального закона

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

Очевидно, что функция распределения $G(x)$ из соотношения (1) может быть представлена в виде:

$$G(x) = \mathbb{E}\Phi(Ux)\,,$$

где U — дискретная случайная величина, принимающая значения σ_i с вероятностями p_i, т. е.

$$U : \begin{matrix} \sigma_1 & \sigma_2 & \dots & \sigma_k \\ p_1 & p_2 & \dots & p_k \end{matrix}. \tag{2}$$

Обозначим через $\rho(F, G)$ равномерное расстояние между функциями распределения $F(x)$ и $G(x)$:

$$\rho(F, G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)|\,. \tag{3}$$

Известно, что для решения задачи устойчивости для конечных масштабных смесей нормальных законов равномерная метрика (3) не подходит, поэтому необходимо рассматривать метрики, метризующие слабую сходимость, например метрику Леви $L(F, G)$ между функциями распределения $F(x)$ и $G(x)$, определяемую соотношением

$$L(F, G) = \inf\{h :\, G(x-h) - h \leqslant F(x) \leqslant G(x+h) + h\,,\ \ \forall x \in \mathbb{R}\}\,.$$

Модели добавления и расщепления компоненты могут быть представлены в виде:

$$G_p(x) = \mathbb{E}\Phi(U_p\, x)\,,$$

где дискретная случайная величина U_p определяется для каждой конкретной модели. Необходимо получить соотношения, связывающие расстояния Леви между смешивающими распределениями и смесями. Перейдем к рассмотрению каждой из моделей.

3 Модель добавления компоненты

Модель добавления компоненты формализуется следующим образом. Предполагается, что каждое из независимых наблюдений $\mathbf{X}_n = (X_1, \dots, X_n)$ имеет распределение, представимое в виде:

$$G_p(x) = (1-p)\sum_{i=1}^{k} p_i \Phi(x\sigma_i) + p\Phi(x\sigma)\,, \tag{4}$$

где все величины σ_i, p_i, $i = 1, \ldots, k$, считаются известными, а σ и p являются параметрами модели, при этом $\sigma > 0$, $0 \leqslant p \leqslant 1$. Без ограничения общности для определенности будем считать, что выполнены соотношения:

$$0 < \sigma \leqslant \sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_k \,. \tag{5}$$

Отметим, что условие отделенности параметров от нуля в формуле (5) также является достаточно общим и означает, что рассматриваются невырожденные нормальные законы с конечными дисперсиями.

Для данной модели дискретная случайная величина U_p имеет следующий вид:

$$U_p : \begin{matrix} \sigma & \sigma_1 & \sigma_2 & \ldots & \sigma_k \\ p & p_1(1-p) & p_2(1-p) & \ldots & p_k(1-p) \end{matrix} . \tag{6}$$

Отметим, что $L(U, U_p)$ не превосходит величины p, так как расстояние между ступеньками функций распределения составляет в точности p на сегменте $[\sigma, \sigma_1]$ и pp_i на сегментах $[\sigma_i, \sigma_{i+1}]$, $i = 1, \ldots, k-1$. Изменяться могут лишь параметры σ и p, величины σ_i, p_i, $i = 1, \ldots, k$, считаем постоянными. Однако при фиксированном параметре p и при $\sigma \to \sigma_1$ очевидно, что $L(U, U_p)$ к нулю не стремится. Таким образом, без ограничения общности считаем, что $0 \leqslant p \leqslant \sigma_1 - \sigma$. Поэтому

$$L(U, U_p) = p \,. \tag{7}$$

Тогда справедлива следующая теорема.

Теорема 1. *В рамках модели добавления компоненты* (4) *при выполнении условий* (5) *и* (7) *расстояние Леви* $L(U, U_p)$ *между смешивающими распределениями* U *из соотношения* (2) *и* U_p *из соотношения* (6) *и расстояние Леви* $L(G, G_p)$ *между истинным распределением* $G(x)$ *из соотношения* (1) *и приближающей смесью* $G_p(x)$ *из соотношения* (4) *связывают неравенства*

$$L(G, G_p) \leqslant L(U, U_p) \leqslant C_1^{[1]}(\sigma_k) L^{1/2}(G, G_p) \,,$$

где коэффициент $C_1^{[1]}(\sigma_k)$ *зависит только от известной величины* σ_k *и имеет вид:*

$$C_1^{[1]}(\sigma_k) = \frac{1}{\sqrt{\varphi(\sigma_k)}} \left(1 + \frac{\sigma_k}{\sqrt{2\pi}} \right)^{1/2} . \tag{8}$$

Доказательство. Запишем оценки снизу для равномерного расстояния между функциями распределения $G(x)$ и $G_p(x)$, воспользовавшись формулой Лагранжа, свойством монотонного убывания плотности стандартного нормального распределения $\varphi(x)$ от положительного аргумента и соотношениями (5) и (7):

$$\rho(G,G_p)=\sup_x|G(x)-G_p(x)|=\sup_x|G(x)-G(x)+p(G(x)-\Phi(x\sigma))|=$$

$$=p\sup_x|G(x)-\Phi(x\sigma)|\geqslant p|G(1)-\Phi(\sigma)|=p\left|\sum_{i=1}^k p_i(\Phi(\sigma_i)-\Phi(\sigma))\right|=$$

$$=p\left|\sum_{i=1}^k p_i(\sigma_i-\sigma)\varphi(\theta_i\sigma_i+(1-\theta_i)\sigma)\right|\geqslant p\sum_{i=1}^k p_i(\sigma_i-\sigma)\varphi(\sigma_k)\geqslant$$

$$\geqslant p(\sigma_1-\sigma)\varphi(\sigma_k)\geqslant L^2(U,U_p)\varphi(\sigma_k)\,.$$

Воспользуемся известным неравенством для метрики Леви (см., например, книгу [1]):

$$L(G,G_p)\leqslant\rho(G,G_p)\leqslant(1+\max_x G'(x))L(G,G_p)\,. \tag{9}$$

Воспользуемся правым неравенством из соотношения (9)

$$L^2(U,U_p)\varphi(\sigma_k)\leqslant\rho(G,G_p)\leqslant(1+\max_x G'(x))L(G,G_p)\,.$$

Очевидно, что

$$\max_x G'(x)=G'(0)=\sum_{i=1}^k p_i\sigma_i\varphi(0)=\frac{1}{\sqrt{2\pi}}\sum_{i=1}^k p_i\sigma_i\,.$$

Выпишем оценку сверху для $L(U,U_p)$. Имеем

$$L^2(U,U_p)\varphi(\sigma_k)\leqslant\left(1+\frac{1}{\sqrt{2\pi}}\sum_{i=1}^k p_i\sigma_i\right)L(G,G_p)\,;$$

$$L(U,U_p)\leqslant\varphi^{-1/2}(\sigma_k)\left(1+\frac{1}{\sqrt{2\pi}}\sum_{i=1}^k p_i\sigma_i\right)^{1/2}L^{1/2}(G,G_p)\leqslant$$

$$\leqslant\varphi^{-1/2}(\sigma_k)\left(1+\frac{\sigma_k}{\sqrt{2\pi}}\right)^{1/2}L^{1/2}(G,G_p)=C_1^{[1]}(\sigma_k)L^{1/2}(G,G_p)\,.$$

Оценка снизу для $L(U,U_p)$ может быть найдена из соотношений:

$$L(G,G_p)\leqslant\rho(G,G_p)=\sup_x|G(x)-G_p(x)|=$$

$$= p \sup_x \left| \sum_{i=1}^{k} p_i (\Phi(x\sigma_i) - \Phi(x\sigma)) \right| \leqslant p \sup_x \sum_{i=1}^{k} p_i |\Phi(x\sigma_i) - \Phi(x\sigma)| \leqslant$$

$$\leqslant p \sum_{i=1}^{k} p_i \sup_x |\Phi(x\sigma_i) - \Phi(x\sigma)| \leqslant p \sum_{i=1}^{k} p_i = L(U, U_p) \,. \,\square$$

Рассмотрим следующее обобщение модели (4). Пусть имеется еще одна смесь данного типа, отличающаяся от (4) только весом, т. е.

$$G_q(x) = (1-q) \sum_{i=1}^{k} p_i \Phi(x\sigma_i) + q\Phi(x\sigma) \tag{10}$$

(при этом $0 \leqslant q \leqslant 1$).

Для $G_q(x)$ дискретная случайная величина U_q имеет следующий вид:

$$U_q : \begin{matrix} \sigma & \sigma_1 & \sigma_2 & \dots & \sigma_k \\ q & p_1(1-q) & p_2(1-q) & \dots & p_k(1-q) \end{matrix} . \tag{11}$$

Предположим, рассуждая, как описано выше, что $|p-q| \leqslant \sigma_1 - \sigma$. В этом случае расстояние Леви $L(U_p, U_q)$ примет вид:

$$L(U_p, U_q) = |p-q| \,. \tag{12}$$

Тогда справедлива следующая теорема.

Теорема 2. *В рамках модели добавления компоненты* (4) *при выполнении условий* (5) *и* (12) *расстояние Леви* $L(U_p, U_q)$ *между смешивающими распределениями* U_p *из соотношения* (6) *и* U_q *из соотношения* (11) *и расстояние Леви* $L(G_p, G_q)$ *между распределениями* $G_p(x)$ *из соотношения* (4) *и* $G_q(x)$ *из соотношения* (10) *связывают неравенства:*

$$L(G_p, G_q) \leqslant L(U_p, U_q) \leqslant C_1^{[1]}(\sigma_k) L^{1/2}(G_p, G_q) \,,$$

где коэффициент $C_1^{[1]}(\sigma_k)$ *зависит только от известной величины* σ_k *и определяется формулой* (8)*.*

Доказательство. Запишем оценки снизу для равномерного расстояния между функциями распределения $G_p(x)$ и $G_q(x)$, воспользовавшись формулой Лагранжа, свойством монотонного убывания плотности стандартного нормального

распределения $\varphi(x)$ от положительного аргумента, соотношениями (5) и (12), а также выражениями, полученными в доказательстве теоремы 1. Имеем

$$\rho(G_p, G_q) = \sup_x |(q-p)\sum_{i=1}^{k} p_i \Phi(x\sigma_i) + (p-q)\Phi(x\sigma)| =$$
$$= |p-q| \sup_x \left| \sum_{i=1}^{k} p_i \Phi(x\sigma_i) - \Phi(x\sigma) \right| \geqslant |p-q| \left| \sum_{i=1}^{k} p_i (\Phi(\sigma_i) - \Phi(\sigma)) \right| \geqslant$$
$$\geqslant L^2(U_p, U_q)\varphi(\sigma_k)\,.$$

Оценим максимум производной для функций G_p и G_q. Запишем выражения, например, для функции G_p (для функции G_q оценка получается аналогично). Имеем

$$\max_x G_p'(x) = G_p'(0) = (1-p)\sum_{i=1}^{k} p_i \sigma_i \varphi(0) + p\sigma\varphi(0) =$$
$$= \frac{1-p}{\sqrt{2\pi}} \sum_{i=1}^{k} p_i \sigma_i + \frac{p}{\sqrt{2\pi}}\,\sigma \leqslant \frac{\sigma_k}{\sqrt{2\pi}}\,.$$

Пользуясь правым неравенством в формуле (9), приходим к следующему результату:

$$L(U_p, U_q) \leqslant \varphi^{-1/2}(\sigma_k) \left(1 + \frac{\sigma_k}{\sqrt{2\pi}}\right)^{1/2} L^{1/2}(G_p, G_q) = C_1^{[1]}(\sigma_k) L^{1/2}(G_p, G_q)\,.$$

Оценка снизу для $L(U_p, U_q)$ может быть найдена из следующих соотношений:

$$L(G_p, G_q) \leqslant \rho(G_p, G_q) = |p-q| \sup_x \left| \sum_{i=1}^{k} p_i \Phi(x\sigma_i) - \Phi(x\sigma) \right| \leqslant$$
$$\leqslant |p-q| \sup_x \sum_{i=1}^{k} p_i |\Phi(x\sigma_i) - \Phi(x\sigma)| \leqslant$$
$$\leqslant |p-q| \sum_{i=1}^{k} p_i \sup_x |\Phi(x\sigma_i) - \Phi(x\sigma)| \leqslant |p-q| \sum_{i=1}^{k} p_i = L(U_p, U_q)\,. \;\square$$

4 Модель расщепления компоненты

Модель расщепления компоненты формализуется следующим образом. Предполагается, что каждое из независимых наблюдений $\mathbf{X}_n = (X_1, \ldots, X_n)$ имеет распределение, представимое в виде:

$$G_p(x) = \sum_{i=1}^{k-1} p_i \Phi(x\sigma_i) + (p_k - p)\Phi(x\sigma_k) + p\Phi(x\sigma)\,, \tag{13}$$

где все величины σ_i, p_i, $i = 1, \ldots, k$, считаются известными, а σ и p являются параметрами модели, при этом $\sigma > 0$, $0 \leqslant p \leqslant p_k$. Без ограничения общности для определенности будем считать, что выполнены соотношения

$$0 < \sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_{k-1} \leqslant \sigma \leqslant \sigma_k\,. \tag{14}$$

Отметим, что условие отделенности параметров от нуля в формуле (14) также является достаточно общим и означает, что рассматриваются невырожденные нормальные законы с конечными дисперсиями.

Для данной модели дискретная случайная величина U_p имеет вид:

$$U_p : \begin{matrix} \sigma_1 & \sigma_2 & \ldots & \sigma & \sigma_k \\ p_1 & p_2 & \ldots & p & p_k - p \end{matrix}\,. \tag{15}$$

Воспользовавшись геометрической интерпретацией расстояния Леви, можно получить, что

$$L(U, U_p) = \min\{\sigma_k - \sigma, p\}\,. \tag{16}$$

В этой ситуации оба условия $\sigma \to \sigma_k$ при фиксированном параметре p и $p \to 0$ при фиксированном σ влекут справедливость соотношения $L(U, U_p) \to 0$. Тогда справедлива следующая теорема.

Теорема 3. *В рамках модели расщепления компоненты* (13) *при выполнении условий* (14) *расстояние Леви* $L(U, U_p)$ *из соотношения* (16) *между смешивающими распределениями* U *из соотношения* (2) *и* U_p *из соотношения* (15) *и расстояние Леви* $L(G, G_p)$ *между истинным распределением* $G(x)$ *из соотношения* (1) *и приближающей смесью* $G_p(x)$ *из соотношения* (13) *связывают неравенства:*

$$C_2^{[2]}(\sigma_1, \sigma_k) L(G, G_p) \leqslant L(U, U_p) \leqslant C_1^{[2]}(\sigma_k) L^{1/2}(G, G_p)\,,$$

где коэффициенты $C_j^{[2]}$, $j = 1, 2$, *не зависят от величин* p *и* σ *и имеют вид:*

$$C_1^{[2]}(\sigma_k) = \varphi^{-1/2}(\sigma_k)\left(1+\frac{\sigma_k}{\sqrt{2\pi}}\right)^{1/2}; \tag{17}$$

$$C_2^{[2]}(\sigma_1,\sigma_k) = \frac{\sigma_1\sqrt{2\pi e}}{\max\{1,\sigma_k\}}. \tag{18}$$

Доказательство. Запишем оценки снизу для равномерного расстояния между функциями распределения $G(x)$ и $G_p(x)$, воспользовавшись формулой Лагранжа, свойством монотонного убывания плотности стандартного нормального распределения $\varphi(x)$ от положительного аргумента и соотношениями (14) и (16):

$$\begin{aligned}\rho(G,G_p) &= \sup_x |G(x)-G_p(x)| = \\ &= \sup_x \left|\sum_{i=1}^k p_i\Phi(x\sigma_i) - \sum_{i=1}^k p_i\Phi(x\sigma_i) + p\Phi(x\sigma_k) - p\Phi(x\sigma)\right| = \\ &= p\sup_x |\Phi(x\sigma_k)-\Phi(x\sigma)| \geqslant p|\Phi(\sigma_k)-\Phi(\sigma)| = \\ &= p|(\sigma_k-\sigma)\varphi(\theta\sigma_k+(1-\theta)\sigma)| \geqslant p|\sigma_k-\sigma|\varphi(\sigma_k) \geqslant L^2(U,U_p)\varphi(\sigma_k).\end{aligned}$$

Для отыскания оценки сверху для $L(U,U_p)$ воспользуемся правым неравенством из соотношения (9) и найденным ранее максимумом для производной, а также неравенствами (14). Имеем

$$L^2(U,U_p)\varphi(\sigma_k) \leqslant \left(1+\frac{1}{\sqrt{2\pi}}\sum_{i=1}^k p_i\sigma_i\right) L(G,G_p),$$

откуда

$$\begin{aligned}L(U,U_p) &\leqslant \varphi^{-1/2}(\sigma_k)\left(1+\frac{1}{\sqrt{2\pi}}\sum_{i=1}^k p_i\sigma_i\right)^{1/2} L^{1/2}(G,G_p) \leqslant \\ &\leqslant \varphi^{-1/2}(\sigma_k)\left(1+\frac{\sigma_k}{\sqrt{2\pi}}\right)^{1/2} L^{1/2}(G,G_p) = C_1^{[2]}(\sigma_k)L^{1/2}(G,G_p).\end{aligned}$$

Выпишем оценку снизу для $L(U,U_p)$. С этой целью заметим, что

$$L(G,G_p) \leqslant \rho(G,G_p) = p\sup_x |\Phi(x\sigma_k)-\Phi(x\sigma)|.$$

Запишем производную функции $\Phi(x\sigma_k) - \Phi(x\sigma)$. Имеем

$$\sigma_k\varphi(x\sigma_k) - \sigma\varphi(x\sigma) = 0\,.$$

Экстремумы функционала $\Phi(x\sigma_k) - \Phi(x\sigma)$ будут в точках

$$|x| = \sqrt{\frac{2\log(\sigma_k/\sigma)}{\sigma_k^2 - \sigma^2}}\,.$$

Обозначим через x^* экстремум, доставляющий максимум функционалу $|\Phi(x\sigma_k) - \Phi(x\sigma)|$. Тогда

$$p\sup_x|\Phi(x\sigma_k) - \Phi(x\sigma)| = p|\Phi(x^*\sigma_k) - \Phi(x^*\sigma)| = \\ = p|(\sigma_k - \sigma)x^*\varphi(x^*(\theta\sigma_k + (1-\theta)\sigma))|\,.$$

Максимум функционалу $|t\varphi(at)|$ доставляют значения $\pm|a|^{-1}$, поэтому (с учетом неравенств (14))

$$p|(\sigma_k - \sigma)x^*\varphi(x^*(\theta\sigma_k + (1-\theta)\sigma))| \leqslant \\ \leqslant p(\sigma_k - \sigma)\,(2\pi e)^{-1/2}\,(\theta\sigma_k + (1-\theta)\sigma)^{-1} \leqslant p(\sigma_k - \sigma)\,(2\pi e)^{-1/2}\,\sigma_1^{-1} = \\ = L(U, U_p)\max\{p, \sigma_k - \sigma\}\,(2\pi e)^{-1/2}\,\sigma_1^{-1} \leqslant \\ \leqslant L(U, U_p)\max\{1, \sigma_k\}\,(2\pi e)^{-1/2}\,\sigma_1^{-1}\,.$$

Итак,

$$L(U, U_p) \geqslant \frac{\sigma_1\sqrt{2\pi e}}{\max\{1, \sigma_k\}}\,L(G, G_p) = C_2^{[2]}(\sigma_1, \sigma_k)L(G, G_p)\,. \quad \square$$

Рассмотрим следующее обобщение модели (13). Пусть имеется еще одна смесь данного типа, отличающаяся от (13) только весом, т. е.

$$G_q(x) = \sum_{i=1}^{k-1} p_i\Phi(x\sigma_i) + (p_k - q)\Phi(x\sigma_k) + q\Phi(x\sigma) \tag{19}$$

(при этом $0 \leqslant q \leqslant p_k$).

Для $G_q(x)$ дискретная случайная величина U_q имеет вид:

$$U_q : \begin{matrix} \sigma_1 & \sigma_2 & \dots & \sigma & \sigma_k \\ p_1 & p_2 & \dots & q & p_k - q \end{matrix}\,. \tag{20}$$

Воспользовавшись геометрической интерпретацией расстояния Леви, можно получить, что

$$L(U_p, U_q) = \min\{\sigma_k - \sigma, |p - q|\}\,. \tag{21}$$

Тогда справедлива следующая теорема.

Теорема 4. *В рамках модели расщепления компоненты* (13) *при выполнении условий* (14) *расстояние Леви* $L(U_p, U_q)$ *из соотношения* (21) *между смешивающими распределениями* U_p *из соотношения* (15) *и* U_q *из соотношения* (20) *и расстояние Леви* $L(G_p, G_q)$ *между распределениями* $G_p(x)$ *из соотношения* (13) *и* $G_q(x)$ *из соотношения* (19) *связывают неравенства*

$$C_2^{[2]}(\sigma_1, \sigma_k) L(G_p, G_q) \leqslant L(U_p, U_q) \leqslant C_1^{[2]}(\sigma_k) L^{1/2}(G_p, G_q)\,,$$

где коэффициенты $C_j^{[2]}$, $j = 1, 2$, *не зависят от величин* p *и* σ *и определяются формулами* (17) *и* (18).

Доказательство. Запишем оценки снизу для равномерного расстояния между функциями распределения $G_p(x)$ и $G_q(x)$, воспользовавшись формулой Лагранжа, свойством монотонного убывания плотности стандартного нормального распределения $\varphi(x)$ от положительного аргумента, соотношениями (14) и (21), а также выражениями, полученными в доказательстве теоремы 3. Имеем

$$\rho(G_p, G_q) = \sup_x |G_p(x) - G_q(x)| =$$

$$= |p - q| \sup_x |\Phi(x\sigma_k) - \Phi(x\sigma)| \geqslant |p - q||\Phi(\sigma_k) - \Phi(\sigma)| =$$

$$= |p - q||(\sigma_k - \sigma)\varphi(\theta\sigma_k + (1 - \theta)\sigma)| \geqslant p|\sigma_k - \sigma|\varphi(\sigma_k) \geqslant L^2(U_p, U_q)\varphi(\sigma_k)\,.$$

Оценим максимум производной для функций G_p и G_q. Имеем

$$\max_x G_p'(x) = G_p'(0) = \sum_{i=1}^{k} p_i\sigma_i\varphi(0) + p(\sigma\varphi(0) - \sigma_k\varphi(0)) =$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{k} p_i\sigma_i + \frac{p}{\sqrt{2\pi}}(\sigma - \sigma_k) \leqslant \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{k} p_i\sigma_i + \frac{p}{\sqrt{2\pi}}(\sigma_k - \sigma_k) =$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{k} p_i\sigma_i \leqslant \frac{\sigma_k}{\sqrt{2\pi}}\,.$$

Пользуясь правым неравенством в формуле (9), приходим к следующему результату:

$$L(U_p, U_q) \leqslant \varphi^{-1/2}(\sigma_k)\left(1 + \frac{\sigma_k}{\sqrt{2\pi}}\right)^{1/2} L^{1/2}(G_p, G_q) = C_1^{[2]}(\sigma_k) L^{1/2}(G_p, G_q).$$

Оценка снизу для $L(U_p, U_q)$ может быть найдена из следующих соотношений:

$$L(G_p, G_q) \leqslant \rho(G_p, G_q) = |p - q| \sup_x |\Phi(x\sigma_k) - \Phi(x\sigma)| .$$

Повторяя рассуждения из доказательства теоремы 3, получаем, что

$$L(U_p, U_q) \geqslant \frac{\sigma_1\sqrt{2\pi e}}{\max\{1, \sigma_k\}} L(G_p, G_q) = C_2^{[2]}(\sigma_1, \sigma_k) L(G_p, G_q) . \quad \square$$

5 Выводы

Теоремы 1 и 3 означают, что близость смешивающих распределений влечет близость смесей и, наоборот, близость смесей влечет близость смешивающих распределений. Теоремы 2 и 4 означают, что если рассматриваемые смеси близки по параметрам p и q, то близки и их смешивающие распределения, и наоборот: если смешивающие распределения близки по параметрам p и q, то близки и соответствующие смеси.

Данный факт позволяет использовать модели добавления и расщепления компоненты для статистической проверки гипотез о числе компонент в смеси на практике, так как обычно с помощью методов статистической декомпозиции смесей возможно получить приближающее распределение, но при этом «истинное» распределение чаще всего неизвестно. Доказанные теоремы позволяют утверждать, что при выполнении условий данных теорем приближающее и «истинное» распределения близки, а значит, математическая модель описывает данные корректно.

Итак, близость смешивающих распределений (т. е. стремление к 0 веса p дополнительной компоненты в обеих моделях) влечет близость итоговых смесей (т. е. число компонент смеси равно k, а не $k + 1$) в терминах расстояния Леви. Причем справедливо и обратное утверждение. Таким образом, устанавливается взаимно однозначное соответствие между значением параметра веса и числом компонент в смеси. Поэтому становится возможным свести задачу проверки гипотез о значении дискретного параметра, обозначающего число компонент смеси, к задаче проверки гипотез о значении непрерывного параметра, соответствующего весу компоненты. Данный результат может быть использован при построении асимптотически наиболее мощных критериев для моделей добавления и расщепления

компоненты для случая произвольных конечных сдвиг-масштабных смесей, так как полученные результаты играют значительную роль в обосновании вида гипотез в задаче статистической проверки числа компонент смеси и количественной оценки того, насколько может измениться модель при добавлении или изъятии компоненты (подробнее об этом в работах [2–4].

Полученные в статье результаты могут быть непосредственно использованы для стохастических систем, которые описываются моделями на основе конечных масштабных смесей нормальных распределений на прямой. Подходы к анализу круговых стохастических систем рассматриваются, например, в статье [5].

Литература

1. *Золотарев В. М.* Современная теория суммирования независимых случайных величин. — М.: Наука, 1986. 417 с.
2. *Бенинг В. Е., Горшенин А. К., Королев В. Ю.* Асимптотически оптимальный критерий проверки гипотез о числе компонент смеси вероятностных распределений // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 4–16.
3. *Горшенин А. К.* Проверка статистических гипотез в модели расщепления компоненты // Вестник Московского ун-та. Сер. 15. Вычислительная математика и кибернетика, 2011. № 4. С. 26–32.
4. *Gorshenin A. K.* Testing of statistical hypotheses in the splitting component model // Moscow University Comput. Math. Cybernetics, 2011. Vol. 35. No. 4. P. 176–183.
5. *Синицын И. Н.* Стохастические информационные технологии для исследования нелинейных круговых стохастических систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 78–89.

О НЕРАВНОМЕРНЫХ ОЦЕНКАХ СКОРОСТИ СХОДИМОСТИ В ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЕ*

***М. Е. Григорьева*[1], *С. В. Попов*[2]**

Аннотация: Показано, что в неравномерном аналоге неравенства Берри–Эссеена $(1+|x|^3)|F_n(xB_n)-\Phi(x)| \leqslant (C/B_n^3)\sum_{k=1}^n \beta_k$, $n \geqslant 1$, $x \in \mathbb{R}$, где $F_n(x)$ — функция распределения суммы n независимых случайных величин $X_1,\ldots,X_n$ с $\mathsf{E}X_k = 0$, $\mathsf{E}X_k^2 = \sigma_k^2$; $\beta_k = \mathsf{E}|X_k|^3 < \infty$, $k = 1,\ldots,n$; $B_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$; $\Phi(x)$ — стандартная нормальная функция распределения, абсолютная постоянная C удовлетворяет неравенству $C \leqslant 22{,}2417$.

Ключевые слова: центральная предельная теорема; неравномерная оценка скорости сходимости; неравенство Берри–Эссеена; абсолютная постоянная

1 Введение и формулировка основного результата

При анализе статистических закономерностей поведения тех или иных характеристик сложных информационных систем ключевую роль играют вероятностные модели, основанные на нормальной аппроксимации, целесообразность которой обосновывается центральной предельной теоремой теории вероятностей. Точность указанных асимптотических моделей определяется оценками скорости сходимости в центральной предельной теореме, уточнению которых и посвящена данная статья.

Пусть $X_1,\ldots,X_n$ — независимые случайные величины, заданные на некотором вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$ и удовлетворяющие условиям:

$$\mathsf{E}X_k = 0\,; \quad \mathsf{E}X_k^2 = \sigma_k^2\,; \quad \mathsf{E}|X_k|^3 = \beta_k < \infty, \quad k = 1,\ldots,n\,.$$

Обозначим

$$B_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 > 0\,; \quad L_n = \frac{1}{B_n^3}\sum_{k=1}^n \beta_k\,.$$

*Работа выполнена при поддержке РФФИ (проекты 11-01-12026-офи-м, 11-07-00112 и 11-01-00515), а также Министерства образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

[1]Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, maria-grigoryeva@yandex.ru

[2]Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, popovserg@yandex.ru

Пусть $F_n(x) = \mathsf{P}(X_1 + \cdots + X_n < x)$; $\varphi(x)$ и $\Phi(x)$ — соответственно плотность и функция распределения стандартного нормального закона. Положим

$$\Delta_n(x) = |F_n(xB_n) - \Phi(x)|\,, \quad x \in \mathbb{R}\,,\ n \geqslant 1\,.$$

Известно, что при указанных условиях существует (см., например, [1]) абсолютная положительная конечная константа C_0 такая, что

$$\sup_x \Delta_n(x) \leqslant C_0 L_n\,. \tag{1}$$

Наилучшая известная авторам оценка константы $C_0 \leqslant 0{,}5600$ приведена в [2]. В работе [3] была получена оценка, являющаяся структурным уточнением неравенства (1):

$$\sup_x \Delta_n(x) \leqslant 0{,}3197 \sum_{k=1}^{n} \frac{\beta_k + \sigma_k^3}{B_n^3}\,. \tag{2}$$

Оценка скорости сходимости $F_n(xB_n)$ к $\Phi(x)$, устанавливаемая неравенствами (1) и (2), равномерна по x. Но поскольку и $F_n(x)$, и $\Phi(x)$ — функции распределения, должно выполняться соотношение $\Delta_n(x) \to 0$ при $|x| \to \infty$. Это обстоятельство не учитывается в равномерных оценках. Вместе с тем точность нормальной аппроксимации для функции распределения сумм случайных величин именно при больших значениях аргумента представляет особый интерес, например при вычислении рисков критически больших потерь. В данной статье будут рассмотрены неравномерные оценки скорости сходимости в центральной предельной теореме.

В работах С. В. Нагаева [4] (для случая одинаково распределенных слагаемых) и А. Бикялиса [5] (для случая необязательно одинаково распределенных слагаемых) было показано, что существует такое положительное конечное число C, что

$$\sup_x \left(1 + |x|^3\right) \Delta_n(x) \leqslant C L_n\,. \tag{3}$$

Впервые верхние оценки для C были получены в работах Л. Падитца для необязательно одинаково распределенных слагаемых. В 1986 г. Падитц [6] показал, что $C \leqslant 31{,}935$. В данной работе эта оценка будет уточнена, а именно будет показано, что $C \leqslant 22{,}2417$.

Идея, лежащая в основе описанного Падитцем [7] (и уточненного Ю. С. Нефедовой и И. Г. Шевцовой [8] для одинаково распределенных слагаемых) конструктивного метода построения неравномерных оценок точности нормальной аппроксимации для распределений сумм независимых случайных величин, позволяющего получить явные оценки абсолютных констант, заключается в подходящем разбиении вещественной прямой на зоны «малых», «умеренных» и «больших» значений x. Традиционно используются разбиения вида

(i) («малые» значения x): $0 \leqslant x^2 \leqslant K^2$;

(ii) («умеренные» значения x): $K^2 \leqslant x^2 \leqslant c_n(x; a, b, c)$;

(iii) («большие» значения x): $c_n(x; a, b, c) \leqslant x^2 < \infty$,

где $K > 0$, $a > 0$, $b > c \geqslant 1$ — вспомогательные свободные параметры; $c_n(x; a, b, c)$ — некоторая функция, монотонно возрастающая по x. Положим

$$c_n(x; a, b, c) = \frac{b^2}{2(b-c)} \left[\ln |x|^3 - \ln (aL_n)\right] \equiv \frac{b^2}{2(b-c)} \ln \frac{|x|^3}{aL_n} ,$$

2 Вспомогательные утверждения

Метод доказательства основан на усечении случайных величин X_k, $k = 1, \ldots, n$:

$$\overline{X}_k = \begin{cases} X_k, & \text{если } |X_k| \leqslant y ; \\ 0 & \text{в противном случае,} \end{cases}$$

где $y > 0$ — параметр усечения. Обозначим

$$F_n^y(x) = \mathsf{P}\left(\sum_{k=1}^{n} \overline{X}_k < x\right) .$$

Также для параметра $h \geqslant 0$ введем следующее обозначение:

$$f_k(h) = \mathsf{E} e^{h\overline{X}_k} = \mathsf{E} \exp\left\{hX_k \mathbb{1}(|X_k| \leqslant y)\right\} , \quad x \in \mathbb{R} .$$

Сформулируем и докажем несколько вспомогательных утверждений.

Лемма 1. *Для произвольных параметров $h \geqslant 0$ и $y > 0$ справедливы неравенства:*

$$h\sigma_k^2 - \frac{\beta_k}{y^2}\left(1 + hy + \frac{(hy)^2}{2}\right) \leqslant m_{1,k} \equiv \mathsf{E}\overline{X}_k e^{h\overline{X}_k} \leqslant h\sigma_k^2 + \frac{\beta_k}{y^2} e^{hy} ; \qquad (4)$$

$$\sigma_k^2 - \frac{\beta_k}{y}(1 + hy) \leqslant m_{2,k} \equiv \mathsf{E}\overline{X}_k^2 e^{h\overline{X}_k} \leqslant \sigma_k^2 + \frac{\beta_k}{y} e^{hy} ; \qquad (5)$$

$$m_{3,k} \equiv \mathsf{E}|\overline{X}_k|^3 e^{h\overline{X}_k} \leqslant \beta_k e^{hy} ;$$

$$f_k(h) \equiv \mathsf{E} e^{h\overline{X}_k} \leqslant 1 + \frac{h^2\sigma_k^2}{2} + \frac{\beta_k e^{hy}}{y^3} \leqslant \exp\left\{\frac{h^2\sigma_k^2}{2} + \frac{\beta_k}{y^3} e^{hy}\right\} ; \qquad (6)$$

$$f_k(h) \geqslant 1 - \frac{h\beta_k}{y^2} . \qquad (7)$$

Доказательство основано на применении формулы Маклорена:

$$e^{h\overline{X}_k} = 1 + h\overline{X}_k + \frac{(h\overline{X}_k)^2}{2} + \sum_{m=3}^{\infty} \frac{(h\overline{X}_k)^m}{m!}\,.$$

Найдем двусторонние оценки моментов случайной величины $\overline{X}_k$. Имеем:

$$\left|\mathsf{E}\overline{X}_k\right| = |\mathsf{E}X_k - \mathsf{E}X_k\mathbb{1}(|X_k| > y)| \leqslant \mathsf{E}\,|X_k|\,\mathbb{1}(|X_k| > y) \leqslant$$
$$\leqslant y^{-2}\mathsf{E}\,|X_k|^3\,\mathbb{1}(|X_k| > y) \leqslant \frac{\beta_k}{y^2}\,;$$

$$\mathsf{E}\overline{X}_k^2 - \sigma_k^2 =$$
$$= \mathsf{E}(\overline{X}_k^2 - X_k^2) = -\mathsf{E}X_k^2\mathbb{1}(|X_k| > y) \begin{cases} \leqslant 0\,; \\ \geqslant -y^{-1}\mathsf{E}|X_k|^3\mathbb{1}(|X_k| > y) \geqslant -\dfrac{\beta_k}{y}\,; \end{cases}$$

$$\left|\mathsf{E}(\overline{X}_k)^m\right| \leqslant \mathsf{E}\left|\overline{X}_k\right|^m = \mathsf{E}\,|X_k|^m\,\mathbb{1}(|X_k| \leqslant y) \leqslant \beta_k y^{m-3}\,,\ m \geqslant 3\,.$$

Отсюда с использованием верхних оценок получаем:

$$m_{1,k} \equiv \mathsf{E}\overline{X}_k e^{h\overline{X}_k} = \mathsf{E}\left(\overline{X}_k\left(1 + h\overline{X}_k + \sum_{m=2}^{\infty}\frac{(h\overline{X}_k)^m}{m!}\right)\right) =$$
$$= \mathsf{E}\overline{X}_k + h\mathsf{E}\overline{X}_k^2 + \sum_{m=2}^{\infty}\frac{h^m\mathsf{E}\,\overline{X}_k^{m+1}}{m!} \leqslant \frac{\beta_k}{y^2} + h\sigma_k^2 + \frac{\beta_k}{y^2}\sum_{m=2}^{\infty}\frac{(hy)^m}{m!} =$$
$$= \frac{\beta_k}{y^2} + h\sigma_k^2 + \frac{\beta_k}{y^2}\left(e^{hy} - 1 - hy\right) = h\sigma_k^2 + \frac{\beta_k}{y^2}\left(e^{hy} - hy\right)\,;$$

$$m_{2,k} \equiv \mathsf{E}\overline{X}_k^2 e^{h\overline{X}_k} = \mathsf{E}\left(\overline{X}_k^2\left(1 + \sum_{m=1}^{\infty}\frac{(h\overline{X}_k)^m}{m!}\right)\right) =$$
$$= \mathsf{E}\overline{X}_k^2 + \sum_{m=1}^{\infty}\frac{h^m\mathsf{E}\,\overline{X}_k^{m+2}}{m!} \leqslant \sigma_k^2 + \frac{\beta_k}{y}\sum_{m=1}^{\infty}\frac{(hy)^m}{m!} = \sigma_k^2 + \frac{\beta_k}{y}\left(e^{hy} - 1\right)\,;$$

$$m_{3,k} \equiv \mathsf{E}|\overline{X}_k|^3 e^{h\overline{X}_k} \leqslant \sum_{m=0}^{\infty}\frac{h^m\mathsf{E}\,|\overline{X}_k|^{m+3}}{m!} \leqslant \beta_k\sum_{m=0}^{\infty}\frac{(hy)^m}{m!} = \beta_k e^{hy}\,;$$

$$f_k(h) \equiv \mathsf{E}e^{h\overline{X}_k} \leqslant 1 + h\mathsf{E}\overline{X}_k + \frac{(h\mathsf{E}\overline{X}_k)^2}{2} + \sum_{m=3}^{\infty} \frac{h^m \mathsf{E}|\overline{X}_k|^m}{m!} \leqslant$$

$$\leqslant 1 + \frac{h\beta_k}{y^2} + \frac{h^2\sigma_k^2}{2} + \frac{\beta_k}{y^3}\sum_{m=3}^{\infty}\frac{(hy)^m}{m!} = 1 + \frac{h^2\sigma_k^2}{2} + \frac{\beta_3}{y^3}\left(e^{hy} - 1 - \frac{h^2y^2}{2}\right).$$

Для доказательства (4) дополнительно заметим, что $xe^{hx} \geqslant x(1 + hx + (hx)^2/2)$, $x \in \mathbb{R}$, и тогда получим оценку:

$$m_{1,k} \equiv \mathsf{E}\overline{X}_k e^{h\overline{X}_k} \geqslant \mathsf{E}\left(\overline{X}_k(1 + h\overline{X}_k + 0{,}5(h\overline{X}_k)^2)\right) \geqslant$$

$$\geqslant -\frac{\beta_k}{y^2} + h\left(\sigma_k^2 - \frac{\beta_k}{y}\right) - \frac{h^2\beta_k}{2} = h\sigma_k^2 - \frac{\beta_k}{y^2}\left(1 + hy + \frac{(hy)^2}{2}\right).$$

Замечая, что $e^{hx} \geqslant 1 + hx$, $x \in \mathbb{R}$, получаем нижние оценки для $m_{2,k}$ и $f_k(h)$:

$$m_{2,k} \equiv \mathsf{E}\overline{X}_k^2 e^{h\overline{X}_k} \geqslant \mathsf{E}\left(\overline{X}_k^2(1 + h\overline{X}_k)\right) \geqslant \sigma_k^2 - \frac{\beta_k}{y} - \beta_k h = \sigma_k^2 - \frac{\beta_k}{y}(1 + hy);$$

$$f_k(h) \equiv \mathsf{E}e^{h\overline{X}_k} \geqslant \mathsf{E}\left(1 + h\overline{X}_k\right) \geqslant 1 - \frac{h\beta_k}{y^2}.$$

Лемма 1 доказана.

Лемма 2. *Пусть* $\alpha > 0$, $\beta > 0$ *и* $\alpha/3 + \beta \geqslant 1$. *Тогда*

$$\sum_{k=1}^{n}\left(\frac{\sigma_k}{B_n}\right)^{\alpha}\left(\frac{\beta_k}{B_n^3}\right)^{\beta} \leqslant (L_n)^{\alpha/3+\beta}.$$

Доказательство. По неравенству Ляпунова

$$\sigma_k^2 = \mathsf{E}\,|X_k - \mathsf{E}X_k|^2 \leqslant \left(\mathsf{E}\,|X_k - \mathsf{E}X_k|^3\right)^{2/3} = (\beta_k)^{2/3}.$$

Таким образом,

$$\sum_{k=1}^{n}\left(\frac{\sigma_k}{B_n}\right)^{\alpha}\left(\frac{\beta_k}{B_n^3}\right)^{\beta} \leqslant \sum_{k=1}^{n}\left(\frac{\beta_k}{B_n^3}\right)^{\alpha/3+\beta} \leqslant \left(\sum_{k=1}^{n}\frac{\beta_k}{B_n^3}\right)^{\alpha/3+\beta} = (L_n)^{\alpha/3+\beta}.$$

Лемма 2 доказана.

Лемма 3 (см. [8]). *Для параметра усечения $y > 0$ справедливо неравенство*

$$\sup_{x\in\mathbb{R}} |F_n^y(x) - F_n(x)| \leqslant \sum_{k=1}^{n} \mathsf{P}(|X_k| > y) .$$

Лемма 4 (см. [8]). 1°. *Пусть $q > 0$ и $A > 0$. Тогда*

$$\sup_{v\geqslant A} |\Phi(v) - \Phi(qv)| \leqslant \frac{1}{2} \max\left\{q^2 - 1,\ \frac{1-q^2}{q^2}\right\} (v\varphi(v))\Big|_{v=\max\{1,A\min\{1,q\}\}} .$$

2°. *Пусть $a \in \mathbb{R}$ и $A > 0$. Тогда*

$$\sup_{v\geqslant A} |\Phi(v+a) - \Phi(v)| \leqslant |a|\,\varphi(\min\{A + a,\ A\}) .$$

Лемма 5 (см. [9]). *Для любой функции распределения F с нулевым средним и единичной дисперсией*

$$\sup_{x\in\mathbb{R}} |F(x) - \Phi(x)| \leqslant \sup_{x>0} \left(\Phi(x) - \frac{x^2}{1+x^2}\right) = 0{,}54093654\ldots \equiv \varkappa .$$

Лемма 6 (см. [8]). *Для любой случайной величины X с $\mathsf{E}|X|^3 < \infty$ и $\mathsf{E}X = a$ при любом a справедливо неравенство:*

$$\mathsf{E}|X-a|^3 \leqslant \min\{L\mathsf{E}|X|^3,\ \mathsf{E}|X|^3 + 3|a|\mathsf{E}X^2 + a^2\mathsf{E}|X| - |a|^3\} ;$$
$$L = \frac{17 + 7\sqrt{7}}{27} < 1{,}3156 ,$$

причем равенство $\mathsf{E}|X-a|^3 = L\mathsf{E}|X|^3$ достигается на двухточечном распределении вида:

$$\mathsf{P}\left(X = \frac{6a}{4 - \sqrt{7} \pm \sqrt{1+2\sqrt{7}}}\right) = \frac{3 \pm \sqrt{1+2\sqrt{7}}}{6} .$$

3 Случаи (*i*) и (*iii*) — «малые» и «большие» значения x

В случае (i), т. е. для $0 \leqslant |x| \leqslant K$, в соответствии с неравенством (1) имеем:

$$|x|^3 \Delta_n(x) \leqslant C_0 K^3 L_n .$$

В случае же (iii), т. е. для

$$x^2 \geqslant \frac{b^2}{2(b-c)} \ln \frac{|x|^3}{aL_n} ,$$

справедлив следующий результат.

Теорема 1. *Предположим, что*

$$c_n(x;a,b,c) > 0 , \;\; x^2 \geqslant \max\{(2\pi)^{-1}, c_n(x;a,b,c)\} , \;\; b > c \geqslant 1 , \;\; a > 0 .$$

Тогда для любого $n \geqslant 1$

$$|x|^3 \Delta_n(x) \leqslant P(a,b,c,x) L_n , \tag{8}$$

где

$$P(a,b,c,x) = b^3 + a \exp\left\{ \frac{b^3}{a} - \frac{2(b-c)(c-1)}{b^2} x^2 \right\} .$$

Замечание 1. Очевидно, функция $P(a,b,c,x)$ монотонно не возрастает по $|x|$. Также несложно убедиться, что функция $P(a,b,c,x)$ аргумента $a > 0$ при фиксированных b, c и x достигает своего минимального значения в точке $a = b^3$:

$$\inf_{a>0} P(a,b,c,x) = b^3 \left(1 + \exp\left\{ 1 - \frac{2(b-c)(c-1)}{b^2} x^2 \right\} \right) ,$$

при этом $P(a,b,c,x) \geqslant 1$ для всех рассматриваемых значений a, b, c, x. Кроме того,

$$\inf_{a>0,\, b>c\geqslant 1} \lim_{|x|\to\infty} P(a,b,c,x) = \lim_{b\to 1+} b^3 = 1 .$$

Доказательство теоремы 1. Без ограничения общности пусть $x \geqslant 0$. Поскольку

$$\Delta_n(x) = \max\left\{ (1 - F_n(xB_n)) + (\Phi(x) - 1),\, (F_n(xB_n) - 1) + (1 - \Phi(x)) \right\}$$

и величины $F_n(xB_n) - 1$ и $\Phi(x) - 1$ неположительны для любого $x \in \mathbb{R}$, получаем:

$$\Delta_n(x) \leqslant \max\left\{ 1 - F_n(xB_n),\, 1 - \Phi(x) \right\} .$$

Для всех $x \geqslant \max\{(2\pi)^{-1/2}, \sqrt{c_n(x;a,b,c)}\}$ справедливо неравенство:

$$1 - \Phi(x) \leqslant \frac{\varphi(x)}{x} = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} \leqslant e^{-x^2/2} \leqslant \left[\frac{aL_n}{x^3} \right]^{b^2/(4(b-c))} .$$

Замечая, что $b^2/(4(b-c)) \geqslant c$ при $b \geqslant c$ и для рассматриваемого диапазона значений x

$$\frac{aL_n}{x^3} \leqslant \exp\left\{-\frac{2(b-c)x^2}{b^2}\right\} < 1\,, \tag{9}$$

окончательно заключаем:

$$1-\Phi(x) \leqslant \left[\frac{aL_n}{x^3}\right]^c\,.$$

С другой стороны, для $1-F_n(xB_n)$ имеем:

$$1-F_n(xB_n) \leqslant 1-F_n^y(xB_n) + |F_n^y(xB_n) - F_n(xB_n)|$$

с параметром усечения $y = xB_n/b$, $b>1$. Тогда по лемме 3 и по неравенству Маркова получим:

$$|F_n^y(xB_n) - F_n(xB_n)| \leqslant \sum_{k=1}^n \mathsf{P}(|X_k|>y) \leqslant \sum_{k=1}^n \frac{b^3\beta_k}{x^3B_n^3} = \frac{b^3L_n}{x^3}\,. \tag{10}$$

Обозначим

$$h = \frac{1}{y}\ln\frac{x^3}{aL_n} \equiv \frac{2(b-c)}{bxB_n}\,c_n(x;a,b,c)\,.$$

Используя неравенство Маркова и учитывая верхнюю оценку (6) для $f_k(h) = \mathsf{E}e^{h\overline{X}_k}$ из леммы 1 для указанного h, получаем:

$$\begin{aligned}1-F_n^y(xB_n) = \mathsf{P}(h\overline{X}_1+\cdots+h\overline{X}_n \geqslant hxB_n) &\leqslant \left(\prod_{k=1}^n f_k(h)\right)\exp\{-hxB_n\} \leqslant \\ &\leqslant \exp\left\{-hxB_n + \frac{1}{2}\,h^2B_n^2 + \frac{1}{y^3}\,e^{hy}\sum_{k=1}^n\beta_k\right\} = \\ &= \exp\left\{-\frac{2(b-c)}{b}c_n(x;a,b,c) + \frac{2(b-c)^2}{b^2x^2}c_n^2(x;a,b,c) + \frac{b^3}{a}\right\}\,.\end{aligned}$$

Для $x^2 \geqslant c_n(x;a,b,c)$ имеем:

$$1-F_n^y(xB_n) \leqslant \exp\left\{-\frac{2c(b-c)}{b^2}c_n(x;a,b,c) + \frac{b^3}{a}\right\} = \left(\frac{aL_n}{x^3}\right)^c\exp\left\{\frac{b^3}{a}\right\},$$

откуда видно, что мажоранта для $1 - F_n(xB_n)$ не меньше, чем мажоранта для $1 - \Phi(x)$. Далее, учитывая (9), получаем:

$$1 - F_n^y(xB_n) \leqslant a \exp\left\{\frac{b^3}{a} - \frac{2(b-c)(c-1)}{b^2}x^2\right\}\frac{L_n}{x^3}. \tag{11}$$

Суммируя оценки (10) и (11), приходим к (8). Теорема 1 доказана.

4 Случай (*ii*) — «умеренные» значения x

Переопределим величины

$$y = \gamma x B_n\,; \quad h = \frac{(1-\gamma)x}{B_n}\,, \quad \gamma \in (0,1)\,.$$

Обозначим $S_n^* = X_1^* + \cdots + X_n^*$ сумму независимых случайных величин с функциями распределения

$$\mathsf{P}(X_k^* < u) = \frac{1}{f_k(h)} \int\limits_{-\infty}^{u} e^{ht}\, d\mathsf{P}\left(\overline{X}_k < t\right)\,.$$

Легко проверить справедливость следующих равенств:

$$1 - \Phi(x) = \exp\left\{\frac{h^2 B_n^2}{2}\right\} \int\limits_{x}^{+\infty} e^{-hB_n t} d\Phi(t - hB_n)\,; \tag{12}$$

$$1 - F_n^y(xB_n) = \left(\prod_{k=1}^{n} f_k(h)\right) \int\limits_{x}^{+\infty} e^{-hB_n t} d\mathsf{P}(S_n^* < tB_n)\,. \tag{13}$$

Пусть без ограничения общности $x \geqslant 0$. Тогда

$$\begin{aligned}\Delta_n(x) = |1 - \Phi(x) + (F_n(xB_n) - F_n^y(xB_n)) + F_n^y(xB_n) - 1| \leqslant \\ \leqslant |F_n(xB_n) - F_n^y(xB_n)| + |1 - \Phi(x) + F_n^y(xB_n) - 1|\,.\end{aligned}$$

Далее воспользуемся неравенством леммы 3 и выражениями (12) и (13):

$$\Delta_n(x) \leqslant \sum_{k=1}^{n} \mathsf{P}(|X_k| > y) +$$

$$+\left|\left(\exp\left\{\frac{h^2B_n^2}{2}\right\}-\prod_{k=1}^{n}f_k(h)\right)\int\limits_x^{+\infty}e^{-hB_nt}\,d\mathsf{P}(S_n^*<tB_n)\right|+$$

$$+\left|\exp\left\{\frac{h^2B_n^2}{2}\right\}\int\limits_x^{+\infty}e^{-hB_nt}\,d\left(\Phi(t-hB_n)-\mathsf{P}(S_n^*<tB_n)\right)\right|.$$

Применяя неравенство Маркова к первому слагаемому и интегрируя по частям последнее, получаем:

$$\Delta_n(x)\leqslant\frac{L_n}{\gamma^3x^3}+\left|\prod_{k=1}^{n}f_k(h)-e^{h^2B_n^2/2}\right|\exp\{-hxB_n\}\,\mathsf{P}\,(S_n^*>xB_n)+$$

$$+2\exp\left\{\frac{h^2B_n^2}{2}-hxB_n\right\}\sup_{u\geqslant x}|\mathsf{P}(S_n^*<uB_n)-\Phi(u-hB_n)|\equiv$$

$$\equiv\frac{L_n}{\gamma^3x^3}+I_1I_2+2\exp\left\{\frac{h^2B_n^2}{2}-hxB_n\right\}I_3\,.$$

Сформулируем два утверждения, которые будут неоднократно использоваться в дальнейшем. Во-первых, если $x^2\leqslant c_n(x;a,b,c)$, то (см. (ii))

$$L_n\leqslant\frac{x^3}{a}\exp\left\{-\frac{2(b-c)}{b^2}x^2\right\}. \tag{14}$$

Во-вторых, если $x^2\geqslant K^2$, то при $x\geqslant\sqrt{r/(2s)}$, $r>0$, или при $r\leqslant 0$

$$x^r\exp\{-sx^2\}\leqslant K^r\exp\left\{-sK^2\right\},\quad s>0\,. \tag{15}$$

Всюду далее будем предполагать, что параметры $b>c\geqslant 1$, $\gamma\in(0,1)$, $K>0$ удовлетворяют следующим условиям:

$$\frac{4(b-c)}{3b^2}>\gamma(1-\gamma)\,; \tag{16}$$

$$K^2\geqslant\frac{9b^2}{4(b-c)}\,; \tag{17}$$

$$K^2\geqslant\left[\frac{10(b-c)}{3b^2}-\gamma(1-\gamma)\right]^{-1}; \tag{18}$$

$$K^2 \geqslant \frac{3}{2}\left[\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right]^{-1}; \tag{19}$$

$$K^2 \geqslant \frac{3}{(1-\gamma)^2}; \tag{20}$$

$$K^2 \geqslant \frac{4}{1-\gamma^2}. \tag{21}$$

4.1 Оценка I_1

Оценим выражение

$$I_1 = \left|\prod_{k=1}^{n} f_k(h) - \exp\left\{h^2 B_n^2/2\right\}\right| \exp\left\{-hxB_n\right\}.$$

Применяя оценку (6) из леммы 1 и неравенство $|e^x - 1| \leqslant |x|e^{|x|}$, получим:

$$\left|\prod_{k=1}^{n} f_k(h) - \exp\left\{\frac{h^2B_n^2}{2}\right\}\right| \leqslant \exp\left\{\frac{h^2B_n^2}{2}\right\}\left|\exp\left\{\frac{e^{hy}}{y^3}\sum_{k=1}^{n}\beta_k\right\} - 1\right| \leqslant$$

$$\leqslant \frac{1}{y^3}\left(\sum_{k=1}^{n}\beta_k\right)\exp\left\{\frac{e^{hy}}{y^3}\sum_{k=1}^{n}\beta_k + hy + \frac{h^2B_n^2}{2}\right\} =$$

$$= \frac{L_n}{(\gamma x)^3}\exp\left\{\frac{e^{\gamma(1-\gamma)x^2}L_n}{(\gamma x)^3} + \frac{(1-\gamma^2)x^2}{2}\right\}.$$

Воспользуемся оценкой (14):

$$\left|\prod_{k=1}^{n} f_k(h) - \exp\left\{\frac{h^2B_n^2}{2}\right\}\right| \leqslant \frac{L_n}{(\gamma x)^3}\exp\left\{\frac{(1-\gamma^2)x^2}{2}\right\}A_1(x),$$

где

$$\log A_1(x) \equiv \frac{1}{a\gamma^3}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\} \leqslant \log A_1(K)$$

при $x \geqslant K$, если выполнено условие (16). Здесь и далее символами $A(x)$, $A_1(x)$, $A_2(x), \ldots$ будут обозначаться положительные функции аргумента x, также зависящие от параметров a, b, c и γ. Таким образом,

$$I_1 \leqslant \frac{1}{\gamma^3}\exp\left\{-(1-\gamma)x^2 + \frac{(1-\gamma^2)x^2}{2}\right\}A_1(x)\frac{L_n}{x^3} \leqslant A_1(K)A_2(K)\frac{L_n}{x^3},$$

где $A_2(x) = \gamma^{-3}\exp\left\{-(1-\gamma)^2x^2/2\right\} \leqslant A_2(K)$ при $x \geqslant K$.

4.2 Оценка I_2

Оценим выражение $I_2 \equiv \mathsf{P}\left(S_n^* > xB_n\right)$. С учетом (2) имеем:

$$I_2 \leqslant \left|\mathsf{P}(S_n^* < xB_n) - \Phi\left(\frac{xB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right)\right| + \Phi\left(-\frac{xB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right) \leqslant$$
$$\leqslant 0{,}3197 \sum_{k=1}^{n} \frac{\mathsf{E}\left|X_k^* - \mathsf{E}X_k^*\right|^3 + (\mathsf{D}X_k^*)^{3/2}}{(\mathsf{D}S_n^*)^{3/2}} + \Phi\left(-\frac{xB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right).$$

Для удобства дальнейших ссылок заметим, что в силу оценок (7) и (14)

$$f_k(h) \geqslant 1 - \frac{h\beta_k}{y^2} = 1 - \frac{(1-\gamma)\beta_k}{\gamma^2 x B_n^3} \geqslant 1 - \frac{(1-\gamma)L_n}{\gamma^2 x} \geqslant$$
$$\geqslant 1 - \frac{(1-\gamma)x^2}{a\gamma^2}\exp\left\{-\frac{2(b-c)}{b^2}x^2\right\} \equiv A_3(x) \geqslant A_3(K), \quad (22)$$

если, в соответствии с (15), выполнено более сильное условие (17).

Теперь оценим $(\mathsf{E}X_k^*)^2 = (m_{1,k}/f_k(h))^2$. С учетом (4), (14) и леммы 2 имеем:

$$\sum_{k=1}^{n}(\mathsf{E}X_k^*)^2 \leqslant \frac{1}{A_3^2(K)}\sum_{k=1}^{n}\left(h\sigma_k^2 + \frac{\beta_k e^{hy}}{y^2}\right)^2 =$$
$$= \frac{B_n^2}{A_3^2(K)}\sum_{k=1}^{n}\left((1-\gamma)x\frac{\sigma_k^2}{B_n^2} + \frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}\cdot\frac{\beta_k}{B_n^3}\right)^2 =$$
$$= \frac{B_n^2}{A_3^2(K)}\left[(1-\gamma)^2x^2\sum_{k=1}^{n}\frac{\sigma_k^4}{B_n^4} + \frac{2(1-\gamma)xe^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}\cdot\sum_{k=1}^{n}\left(\frac{\sigma_k^2}{B_n^2}\cdot\frac{\beta_k}{B_n^3}\right) + \right.$$
$$\left. + \frac{e^{2\gamma(1-\gamma)x^2}}{(\gamma x)^4}\cdot\sum_{k=1}^{n}\left(\frac{\beta_k}{B_n^3}\right)^2\right] \leqslant \frac{B_n^2}{A_3^2(K)}\left[(1-\gamma)^2x^2L_n^{4/3} + \right.$$
$$\left. + \frac{2(1-\gamma)xe^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}L_n^{2/3+1} + \frac{e^{2\gamma(1-\gamma)x^2}}{(\gamma x)^4}L_n^2\right] =$$
$$= \frac{B_n^2}{A_3^2(K)}\left((1-\gamma)xL_n^{2/3} + \frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}L_n\right)^2 \leqslant B_n^2 A_4(x),$$

где

$$A_4(x) = \frac{1}{A_3^2(K)} \left[\frac{(1-\gamma)x^3}{a^{2/3}} \exp\left\{ -\frac{4(b-c)}{3b^2} x^2 \right\} + \right.$$
$$\left. + \frac{x}{a\gamma^2} \exp\left\{ -\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma) \right) x^2 \right\} \right]^2 .$$

В соответствии с (15) заключаем, что $A_4(x) \leqslant A_4(K)$, если выполнены условия:

$$K^2 \geqslant \frac{9b^2}{8(b-c)}; \qquad K^2 \geqslant \frac{1}{2} \left[\frac{2(b-c)}{b^2} - \gamma(1-\gamma) \right]^{-1},$$

вытекающие из (17) и (19). Далее с учетом (6) имеем:

$$(f_k(h))^{-1} \geqslant 2 - f_k(h) \geqslant 1 - \frac{h^2\sigma_k^2}{2} - \frac{\beta_k e^{hy}}{y^3}.$$

Таким образом, принимая во внимание также лемму 2,

$$\sum_{k=1}^{n} \frac{\sigma_k^2}{f_k(h)} \geqslant \sum_{k=1}^{n} \sigma_k^2 \left(1 - \frac{h^2\sigma_k^2}{2} - \frac{\beta_k e^{hy}}{y^3} \right) \geqslant$$
$$\geqslant B_n^2 \sum_{k=1}^{n} \left(\frac{\sigma_k^2}{B_n^2} - \frac{(1-\gamma)^2 x^2}{2} \frac{\sigma_k^4}{B_n^4} - \frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^3} \frac{\sigma_k^2}{B_n^2} \frac{\beta_k}{B_n^3} \right) \geqslant$$
$$\geqslant B_n^2 \left(1 - \frac{(1-\gamma)^2 x^2}{2} L_n^{4/3} - \frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^3} L_n^{2/3+1} \right) \geqslant B_n^2 (1 - A_5(x)),$$

где

$$A_5(x) = \frac{(1-\gamma)^2 x^6}{2a^{4/3}} \exp\left\{ -\frac{8(b-c)}{3b^2} x^2 \right\} +$$
$$+ \frac{x^2}{a^{5/3}\gamma^3} \exp\left\{ -\left(\frac{10(b-c)}{3b^2} - \gamma(1-\gamma) \right) x^2 \right\}.$$

При этом $A_5(x) \leqslant A_5(K)$, если справедливы условия (17) и (18).

Таким образом, используя нижнюю оценку (5) для $m_{2,k} = f_k(h)\mathsf{E}\,(X_k^*)^2$ и неравенства (22) и (14), получаем:

$$\mathsf{D}S_n^* = \sum_{k=1}^n \mathsf{E}\,(X_k^*)^2 - \sum_{k=1}^n (\mathsf{E}X_k^*)^2 \geqslant$$

$$\geqslant \sum_{k=1}^n \frac{\sigma_k^2}{f_k(h)} - \sum_{k=1}^n \frac{\beta_k}{f_k(h)y}(1+hy) - \sum_{k=1}^n (\mathsf{E}X_k^*)^2 \geqslant$$

$$\geqslant \left(1 - A_5(K) - \frac{1}{A_3(K)}\,\frac{L_n}{\gamma x}(1+\gamma(1-\gamma)x^2) - A_4(K)\right) B_n^2 \geqslant (1 - A_6(x))B_n^2,$$

где

$$A_6(x) = A_4(K) + A_5(K) + \frac{1}{A_3(K)}\,\frac{x^2}{a\gamma}(1+\gamma(1-\gamma)x^2)\exp\left\{-\frac{2(b-c)}{b^2}x^2\right\}.$$

При этом $A_6(x) \leqslant A_6(K)$ в силу (17). Таким образом,

$$\mathsf{D}S_n^* = \sum_{k=1}^n \mathsf{D}X_k^* \geqslant (1 - A_6(K))B_n^2 \equiv A_7(K)B_n^2\,. \tag{23}$$

Из леммы 6 получаем:

$$\sum_{k=1}^n \mathsf{E}\,|X_k^* - \mathsf{E}X_k^*|^3 \leqslant L\sum_{k=1}^n \mathsf{E}\,|X_k^*|^3\,.$$

Оценим входящие в правую часть моменты по лемме 1. Имеем:

$$\sum_{k=1}^n \mathsf{E}\,|X_k^*|^3 = \sum_{k=1}^n \frac{m_{3,k}}{f_k(h)} \leqslant \sum_{k=1}^n \frac{e^{hy}\beta_k}{A_3(K)} = \frac{1}{A_3(K)}\exp\{\gamma(1-\gamma)x^2\}B_n^3 L_n \leqslant$$

$$\leqslant \frac{B_n^3}{A_3(K)}\,\frac{x^3}{a}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\}.$$

Таким образом,

$$\sum_{k=1}^n \mathsf{E}\,|X_1^* - \mathsf{E}X_1^*|^3 \leqslant A_8(x)B_n^3\,, \tag{24}$$

где

$$A_8(x) = \frac{L}{A_3(K)}\,\frac{x^3}{a}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\}.$$

В соответствии с (15) заключаем, что $A_8(x) \leqslant A_8(K)$, если справедливо условие (19). Далее, применяя оценку (5) и неравенство Минковского, получаем:

$$\sum_{k=1}^{n}(\mathsf{D}X_k^*)^{3/2} \leqslant \sum_{k=1}^{n}\left(\mathsf{E}(X_k^*)^2\right)^{3/2} = \sum_{k=1}^{n}\left(\frac{m_{2,k}}{f_k(h)}\right)^{3/2} \leqslant$$

$$\leqslant \frac{1}{A_3^{3/2}(K)}\sum_{k=1}^{n}\left(\sigma_k^2 + \frac{e^{hy}\beta_k}{y}\right)^{3/2} \leqslant \frac{B_n^3}{A_3^{3/2}(K)}\sum_{k=1}^{n}\left(\frac{\sigma_k^2}{B_n^2} + \frac{e^{\gamma(1-\gamma)x^2}}{\gamma x}\frac{\beta_k}{B_n^3}\right)^{3/2} \leqslant$$

$$\leqslant \frac{B_n^3}{A_3^{3/2}(K)}\left[\left(\sum_{k=1}^{n}\frac{\sigma_k^3}{B_n^3}\right)^{2/3} + \frac{e^{\gamma(1-\gamma)x^2}}{\gamma x}\left(\sum_{k=1}^{n}\left(\frac{\beta_k}{B_n^3}\right)^{3/2}\right)^{2/3}\right]^{3/2} \leqslant$$

$$\leqslant \frac{B_n^3}{A_3^{3/2}(K)}\left[L_n^{2/3} + \frac{e^{\gamma(1-\gamma)x^2}}{\gamma x}L_n\right]^{3/2} \leqslant A_9(x)B_n^3\,, \quad (25)$$

где

$$A_9(x) = \frac{1}{A_3^{3/2}(K)}\left[\frac{x^2}{a^{2/3}}\exp\left\{-\frac{4(b-c)}{3b^2}x^2\right\} + \right.$$

$$\left. + \frac{x^2}{a\gamma}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\}\right]^{3/2}.$$

В соответствии с (15) заключаем, что $A_9(x) \leqslant A_9(K)$, если справедливы условия (17) и (19). Обозначим

$$A_{10}(x) = A_8(x) + A_9(x)\,.$$

Следовательно, по неравенству (2) с учетом (23)–(25) получаем:

$$\left|\mathsf{P}\left(\frac{S_n^* - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}} < u\right) - \Phi(u)\right| \leqslant 0{,}3197\frac{A_{10}(K)}{A_7^{3/2}(K)}\,. \quad (26)$$

Теперь оценим величину

$$u \equiv \frac{xB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\,.$$

Используя верхнюю оценку (4) для $m_{1,k} \equiv \mathsf{E}\overline{X}_k e^{h\overline{X}_k} = f_k(h)\mathsf{E}X_k^*$ и нижние оценки (7) и (22) для $f_k(h)$, а также лемму 2, получаем:

$$\mathsf{E}S_n^* - hB_n^2 = \sum_{k=1}^n \frac{m_{1,k}}{f_k(h)} - hB_n^2 \leqslant$$

$$\leqslant \sum_{k=1}^n \left[\frac{1}{f_k(h)}\left(h\sigma_k^2 + \frac{\beta_k e^{hy}}{y^2}\right) - h\sigma_k^2\right] = \sum_{k=1}^n \left[\frac{h\sigma_k^2(1-f_k(h))}{f_k(h)} + \frac{\beta_k e^{hy}}{f_k(h)y^2}\right] \leqslant$$

$$\leqslant \sum_{k=1}^n \left[\frac{h^2\sigma_k^2\beta_k}{A_3(K)y^2} + \frac{e^{hy}\beta_k}{A_3(K)y^2}\right] \leqslant$$

$$\leqslant \frac{B_n}{A_3(K)}\left(\frac{(1-\gamma)^2x^2L_n^{2/3+1}}{(\gamma x)^2} + \frac{e^{\gamma(1-\gamma)x^2}L_n}{(\gamma x)^2}\right) \leqslant$$

$$\leqslant \frac{B_n}{A_3(K)}\left(\frac{(1-\gamma)^2x^5}{a^{5/3}\gamma^2}\exp\left\{-\frac{10(b-c)}{3b^2}x^2\right\} + \right.$$

$$\left. + \frac{x}{a\gamma^2}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\}\right) \equiv A_{11}(x)B_n\,. \quad (27)$$

В соответствии с (15) заключаем, что $A_{11}(x) \leqslant A_{11}(K)$ при $x \geqslant K$ в силу (17) и (19).

Далее понадобится следующая верхняя оценка для $\mathsf{D}S_n^*$, которую получим с учетом верхних оценок (5) для $m_{2,k}$:

$$\mathsf{D}S_n^* = \sum_{k=1}^n \mathsf{D}X_k^* \leqslant \sum_{k=1}^n \mathsf{E}\,(X_k^*)^2 = \sum_{k=1}^n \frac{m_{2,k}}{f_k(h)} \leqslant$$

$$\leqslant \frac{1}{A_3(K)}\sum_{k=1}^n\left(\sigma_k^2 + \frac{e^{hy}\beta_k}{y}\right) \leqslant \frac{B_n^2}{A_3(K)}\left(1 + \frac{e^{\gamma(1-\gamma)x^2}}{\gamma x}L_n\right) \leqslant$$

$$\leqslant \frac{B_n^2}{A_3(K)}\left(1 + \frac{x^2}{a\gamma}\exp\left\{-\left(\frac{2(b-c)}{b^2} - \gamma(1-\gamma)\right)x^2\right\}\right) \equiv$$

$$\equiv A_{13}(x)B_n^2 \leqslant A_{13}(K)B_n^2\,,$$

если, в соответствии с (15), выполнено условие (19), которое является даже более сильным, чем необходимо.

Найдем верхнюю оценку для $\mathsf{E}S_n^* - xB_n$. Учитывая (27), получаем:

$$\mathsf{E}S_n^* - xB_n = (\mathsf{E}S_n^* - hB_n^2) + (hB_n^2 - xB_n) \leqslant (A_{11}(K) - \gamma K)B_n\,,$$

т. е. при $x \geqslant K$

$$\frac{xB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}} \geqslant \frac{\gamma K - A_{11}(K)}{\sqrt{A_{13}(K)}} \equiv A_{14}(K)\,. \tag{28}$$

Отсюда, учитывая также оценку (26), окончательно получаем:

$$I_2 \leqslant 0{,}3197\frac{A_{10}(K)}{A_7^{3/2}(K)} + \frac{1}{\sqrt{2\pi}A_{14}(K)}\exp\left\{-\frac{A_{14}^2(K)}{2}\right\} \equiv A_{15}(K)\,. \tag{29}$$

Оценка (29) справедлива при $A_{14}(K) > 0$.

4.3 Оценка I_3

Приступим к оцениванию величины $I_3 \equiv \sup\limits_{u \geqslant x} \left|\mathsf{P}\left(S_n^* < uB_n\right) - \Phi\left(u - hB_n\right)\right|$.

Имеем:

$$I_3 = \sup_{u \geqslant x}\left|\mathsf{P}\left(\frac{S_n^* - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}} < \frac{uB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right) - \Phi\left(\frac{uB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right) + \right.$$

$$\left. + \Phi\left(\frac{uB_n - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}}\right) - \Phi\left(\frac{uB_n - \mathsf{E}S_n^*}{B_n}\right) + \Phi\left(\frac{uB_n - \mathsf{E}S_n^*}{B_n}\right) - \Phi\left(u - hB_n\right)\right| \leqslant$$

$$\leqslant \sup_{v \in \mathbb{R}}\left|\mathsf{P}\left(\frac{S_n^* - \mathsf{E}S_n^*}{\sqrt{\mathsf{D}S_n^*}} < v\right) - \Phi(v)\right| +$$

$$+ \sup_{v \geqslant (xB_n - \mathsf{E}S_n^*)/\sqrt{\mathsf{D}S_n^*}}\left|\Phi(v) - \Phi\left(\frac{v\sqrt{\mathsf{D}S_n^*}}{B_n}\right)\right| +$$

$$+ \sup_{u \geqslant x}\left|\Phi\left(u - \frac{\mathsf{E}S_n^*}{B_n}\right) - \Phi\left(u - hB_n\right)\right| = I_{31} + I_{32} + I_{33}\,.$$

Для оценки I_{31} воспользуемся неравенством (2), из которого получим:

$$I_{31} \leqslant 0{,}3197\sum_{k=1}^{n}\frac{\mathsf{E}\left|X_k^* - \mathsf{E}X_k^*\right|^3 + \left(\mathsf{D}X_k^*\right)^{3/2}}{\left(\mathsf{D}S_n^*\right)^{3/2}}\,.$$

Величины $\mathsf{E}|X_k^* - \mathsf{E}X_k^*|^3$ оценим аналогично (24), только выделив множитель L_n. Имеем:

$$\sum_{k=1}^{n}\mathsf{E}\left|X_k^* - \mathsf{E}X_k^*\right|^3 \leqslant L\sum_{k=1}^{n}\mathsf{E}\left|X_k^*\right|^3 \leqslant \frac{L}{A_3(K)}\exp\{\gamma(1-\gamma)x^2\}B_n^3L_n =$$

$$= \widehat{A}_8(x)B_n^3L_n\,,$$

где

$$\widehat{A}_8(x) = \frac{L\, e^{\gamma(1-\gamma)x^2}}{A_3(K)}\,.$$

Далее, согласно (25):

$$\sum_{k=1}^{n}(\mathsf{D}X_k^*)^{3/2} \leqslant \frac{B_n^3}{A_3^{3/2}(K)}\left[L_n^{2/3} + \frac{e^{\gamma(1-\gamma)x^2}}{\gamma x}L_n\right]^{3/2} \leqslant \widehat{A}_9(x)B_n^3 L_n\,,$$

где

$$\widehat{A}_9(x) = \frac{1}{A_3^{3/2}(K)}\left[1 + \frac{1}{a^{1/3}\gamma}\exp\left\{-\left(\frac{2(b-c)}{3b^2} - \gamma(1-\gamma)\right)x^2\right\}\right]^{3/2}\,.$$

Обозначим

$$\widehat{A}_{10}(x) = \widehat{A}_8(x) + \widehat{A}_9(x)\,.$$

Воспользуемся неравенством (23), чтобы оценить величину $\mathsf{D}S_n^*$. Тогда для I_{31} окончательно получим:

$$I_{31} \leqslant 0{,}3197\frac{\widehat{A}_{10}(x)}{A_7^{3/2}(K)}\,.$$

Для I_{32} в силу (28) имеем:

$$I_{32} \equiv \sup_{v \geqslant (xB_n - \mathsf{E}S_n^*)/\sqrt{\mathsf{D}S_n^*}}\left|\Phi(v) - \Phi\left(\frac{v\sqrt{\mathsf{D}S_n^*}}{B_n}\right)\right| \leqslant$$

$$\leqslant \sup_{v \geqslant A_{14}(K)}\left|\Phi(v) - \Phi\left(\frac{v\sqrt{\mathsf{D}S_n^*}}{B_n}\right)\right|\,.$$

Используя утверждение 1° леммы 4 в предположении, что $A_{14}(K) > 0$, а также оценку $\min\{1, B_n^{-1}\sqrt{\mathsf{D}S_n^*}\} \geqslant \min\{1, \sqrt{A_7(K)}\} = \sqrt{A_7(K)}$ (см. (23)) и замечая, что функция $v\varphi(v)$ монотонно убывает при $v > 1$, получаем:

$$I_{32} \leqslant \frac{1}{2}\max\left\{\frac{\mathsf{D}S_n^*}{B_n^2} - 1,\ \frac{B_n^2 - \mathsf{D}S_n^*}{\mathsf{D}S_n^*}\right\}(v\varphi(v))\Big|_{v=\max\{1, A_{14}(K)\sqrt{A_7(K)}\}}\,.$$

Оценим $\mathsf{D}S_n^*$ (аналогично случаю I_2, только будем выделять ляпуновскую дробь L_n).

Согласно доказанному ранее,

$$\sum_{k=1}^{n}(\mathsf{E}X_k^*)^2 \leqslant \frac{B_n^2}{A_3^2(K)}\left((1-\gamma)xL_n^{2/3}+\frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}L_n\right)^2 \leqslant \widehat{A}_4(x)B_n^2L_n\,,$$

где

$$\widehat{A}_4(x)=\frac{1}{A_3^2(K)}\left[\frac{(1-\gamma)x^{3/2}}{a^{1/6}}\exp\left\{-\frac{(b-c)}{3b^2}x^2\right\}+\right.$$
$$\left.+\frac{1}{a^{1/2}\gamma^2x^{1/2}}\exp\left\{-\left(\frac{(b-c)}{b^2}-\gamma(1-\gamma)\right)x^2\right\}\right]^2.$$

На основе предыдущих вычислений можно также получить:

$$\sum_{k=1}^{n}\frac{\sigma_k^2}{f_k(h)}\geqslant B_n^2\left(1-\frac{(1-\gamma)^2x^2}{2}L_n^{4/3}-\frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^3}L_n^{2/3+1}\right)=B_n^2(1-\widehat{A}_5(x)L_n)\,,$$

где

$$\widehat{A}_5(x)=\frac{(1-\gamma)^2x^3}{2a^{1/3}}\exp\left\{-\frac{2(b-c)}{3b^2}x^2\right\}+$$
$$+\frac{1}{a^{2/3}\gamma^3x}\exp\left\{-\left(\frac{4(b-c)}{3b^2}-\gamma(1-\gamma)\right)x^2\right\}.$$

Таким образом,

$$\mathsf{D}S_n^*=\sum_{k=1}^{n}\mathsf{E}\,(X_1^*)^2-\sum_{k=1}^{n}(\mathsf{E}X_1^*)^2\geqslant$$
$$\geqslant\sum_{k=1}^{n}\frac{\sigma_k^2}{f_k(h)}-\sum_{k=1}^{n}\frac{\beta_k}{f_k(h)y}(1+hy)-\sum_{k=1}^{n}(\mathsf{E}X_k^*)^2\geqslant B_n^2(1-\widehat{A}_6(x)L_n)\,,$$

где

$$\widehat{A}_6(x)=\widehat{A}_4(x)+\widehat{A}_5(x)+\frac{1+\gamma(1-\gamma)x^2}{A_3(K)\gamma x}\,.$$

Замечая, что $\widehat{A}_6(x)>0$, и учитывая оценку $\mathsf{D}S_n^*\geqslant A_7(K)B_n^2$ (см. (23)), получаем:

$$\frac{B_n^2-\mathsf{D}S_n^*}{\mathsf{D}S_n^*}\leqslant\frac{\widehat{A}_6(x)}{A_7(K)}L_n\,.$$

С другой стороны, в силу леммы 1

$$\frac{\mathsf{D}S_n^*}{B_n^2} - 1 \leqslant \sum_{k=1}^n \frac{m_{2,k}}{B_n^2 f_k(h)} - 1 = \frac{1}{B_n^2}\sum_{k=1}^n \left(\frac{m_{2,k}}{f_k(h)} - \sigma_k^2\right) \leqslant$$

$$\leqslant \frac{1}{B_n^2}\sum_{k=1}^n \frac{m_{2,k} - \sigma_k^2 f_k(h)}{f_k(h)} \leqslant \frac{1}{B_n^2}\sum_{k=1}^n \frac{e^{hy}\beta_k}{f_k(h)y} + \frac{1}{B_n^2}\sum_{k=1}^n \frac{\sigma_k^2(1-f_k(h))}{f_k(h)} \leqslant$$

$$\leqslant \frac{1}{A_3(K)\gamma x}\left(e^{\gamma(1-\gamma)x^2}\sum_{k=1}^n \frac{\beta_k}{B_n^3} + \frac{1-\gamma}{\gamma}\sum_{k=1}^n \frac{\sigma_k^2}{B_n^2}\frac{\beta_k}{B_n^3}\right) \leqslant$$

$$\leqslant \frac{L_n}{A_3(K)\gamma x}\left(e^{\gamma(1-\gamma)x^2} + \frac{1-\gamma}{\gamma}L_n^{2/3}\right) \leqslant A_{12}(x)L_n\,,$$

где

$$A_{12}(x) = \frac{1}{A_3(K)\gamma x}\left(e^{\gamma(1-\gamma)x^2} + \frac{(1-\gamma)x^2}{a^{2/3}\gamma}\exp\left\{-\frac{4(b-c)}{3b^2}x^2\right\}\right).$$

Таким образом, для I_{32} окончательно получаем:

$$I_{32} \leqslant \frac{1}{2}\max\left\{A_{12}(x),\, \frac{\widehat{A}_6(x)}{A_7(K)}\right\} L_n\left(v\varphi(v)\right)\Big|_{v=\max\{1,A_{14}(K)\sqrt{A_7(K)}\}}\,.$$

Наконец, рассмотрим величину I_{33}:

$$I_{33} = \sup_{u\geqslant x}\left|\Phi\left(u - \frac{\mathsf{E}S_n^*}{B_n}\right) - \Phi\left(u - hB_n\right)\right| =$$

$$= \sup_{v\geqslant x-hB_n}\left|\Phi\left(v + \frac{1}{B_n}(hB_n^2 - \mathsf{E}S_n^*)\right) - \Phi\left(v\right)\right|.$$

Учитывая, что $x - hB_n = x\gamma \geqslant \gamma K > 0$, и пользуясь утверждением 2° леммы 4, получим:

$$I_{33} \leqslant \frac{1}{B_n}\left|\mathsf{E}S_n^* - hB_n^2\right|\varphi\left(\min\left\{\gamma K + \frac{1}{B_n}(hB_n^2 - \mathsf{E}S_n^*),\, \gamma K\right\}\right).$$

Воспользуемся нижней оценкой (4) для $m_{1,k} \equiv \mathsf{E}\overline{X}_k e^{h\overline{X}_k} = f_k(h)\mathsf{E}X_k^*$:

$$hB_n^2 - \mathsf{E}S_n^* = \sum_{k=1}^n\left(h\sigma_k^2 - \frac{m_{1,k}}{f_k(h)}\right) = \sum_{k=1}^n \frac{h\sigma_k^2 f_k(h) - m_{1,k}}{f_k(h)} \leqslant$$

$$\leqslant \sum_{k=1}^n \frac{h\sigma_k^2(f_k(h)-1)}{A_3(K)} + \sum_{k=1}^n \frac{\beta_k}{A_3(K)y^2}\left(1 + hy + \frac{(hy)^2}{2}\right).$$

Оценим выражение $h\sigma_k^2(f_k(h)-1)$ с помощью неравенства (6):

$$\sum_{k=1}^{n} h\sigma_k^2(f_k(h)-1) \leqslant \sum_{k=1}^{n}\left(\frac{h^2\sigma_k^2}{2}+\frac{\beta_k e^{hy}}{y^3}\right)h\sigma_k^2 =$$

$$= \frac{(hy)^3 B_n^3}{y^2}\sum_{k=1}^{n}\left(\frac{1}{2\gamma x}\,\frac{\sigma_k^4}{B_n^4}+\frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^4(1-\gamma)^2x^2}\,\frac{\sigma_k^2}{B_n^2}\,\frac{\beta_k}{B_n^3}\right)\leqslant$$

$$\leqslant \frac{(hy)^3 B_n^3 L_n}{y^2}\left(\frac{1}{2\gamma x}\,(L_n)^{1/3}+\frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^4(1-\gamma)^2x^2}\,(L_n)^{2/3}\right)\leqslant$$

$$\leqslant \sum_{k=1}^{n}\frac{(hy)^3\beta_k}{y^2}\,A(x)\,;$$

$$A(x) \equiv \frac{1}{2\gamma a^{1/3}}\exp\left\{-\frac{2(b-c)}{3b^2}x^2\right\}+$$

$$+\frac{1}{\gamma^4(1-\gamma)^2a^{2/3}x^4}\exp\left\{-\left(\frac{4(b-c)}{3b^2}-\gamma(1-\gamma)\right)x^2\right\}.$$

При этом $A(x)\leqslant A(K)$ для $x^2\geqslant K^2$, если выполнено условие (16). Далее, в предположении, что $A(K)\leqslant 1/6$, имеем:

$$hB_n^2-\mathsf{E}S_n^* \leqslant \frac{1}{A_3(K)}\sum_{k=1}^{n}\left(\frac{(hy)^3}{6}\,\frac{\beta_k}{y^2}+\frac{\beta_k}{y^2}\left(1+hy+\frac{(hy)^2}{2}\right)\right)=$$

$$=\frac{1}{A_3(K)}\left(1+hy+\frac{(hy)^2}{2}+\frac{(hy)^3}{6}\right)\sum_{k=1}^{n}\frac{\beta_k}{y^2}\leqslant\frac{1}{A_3(K)}\sum_{k=1}^{n}\frac{\beta_k e^{hy}}{y^2}\,.$$

С другой стороны, при оценке величины I_2 было показано, что

$$\mathsf{E}S_n^*-hB_n^2\leqslant\frac{1}{A_3(K)}\sum_{k=1}^{n}\left(\frac{h^2\sigma_k^2\beta_k}{y^2}+\frac{e^{hy}\beta_k}{y^2}\right).$$

Сравнивая эту оценку с оценкой для $hB_n^2-\mathsf{E}S_n^*$, получаем в силу (17):

$$\left|\mathsf{E}S_n^*-hB_n^2\right|\leqslant\frac{1}{A_3(K)}\sum_{k=1}^{n}\left(\frac{h^2\sigma_k^2\beta_k}{y^2}+\frac{e^{hy}\beta_k}{y^2}\right)\leqslant$$

$$\leqslant\frac{B_nL_n}{A_3(K)}\left(\frac{(1-\gamma)^2x^2}{(\gamma x)^2}L_n^{2/3}+\frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^2}\right)\leqslant\widehat{A}_{11}(x)B_nL_n\,,$$

где

$$\widehat{A}_{11}(x) = \frac{1}{A_3(K)} \left(\frac{(1-\gamma)^2 x^2}{a^{2/3}\gamma^2} \exp\left\{ -\frac{4(b-c)}{3b^2} x^2 \right\} + \frac{e^{\gamma(1-\gamma)x^2}}{(\gamma x)^2} \right).$$

С помощью полученной при рассмотрении I_2 оценки $(hB_n^2 - \mathsf{E}S_n^*) \geqslant$ $\geqslant -A_{11}(K)B_n$, где $A_{11}(K) > 0$, величина I_{33} оценивается следующим образом:

$$I_{33} \leqslant \frac{\widehat{A}_{11}(x)}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\gamma K - A_{11}(K) \right]^2 \right\} L_n \,.$$

Таким образом, получаем оценку $I_3 \leqslant A_{16}(x) L_n$, где

$$\begin{aligned} A_{16}(x) = 0{,}3197 \frac{\widehat{A}_{10}(x)}{A_7^{3/2}(K)} + \\ + \frac{\max\{1,\ A_{14}(K)\sqrt{A_7(K)}\}}{\sqrt{8\pi}} \max\left\{ A_{12}(x),\ \frac{\widehat{A}_6(x)}{A_7(K)} \right\} \times \\ \times \exp\left\{ -\frac{1}{2} \left[\max\left\{ 1,\ A_{14}(K)\sqrt{A_7(K)} \right\} \right]^2 \right\} + \\ + \frac{\widehat{A}_{11}(x)}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\gamma K - A_{11}(K) \right]^2 \right\}. \end{aligned}$$

В итоге для каждого x из рассматриваемого диапазона (ii) справедливо неравенство:

$$\Delta_n(x) \leqslant \frac{L_n}{\gamma^3 |x|^3} + I_1 I_2 + 2 I_3 e^{-(1-\gamma^2)x^2/2} \leqslant A_{17}(x) \frac{L_n}{|x|^3} \,,$$

где

$$\begin{aligned} A_{17}(x) = A_{17}(x, a, b, c, \gamma) = \\ = \frac{1}{\gamma^3} + A_1(K) A_2(K) A_{15}(K) + 2 A_{16}(|x|) |x|^3 e^{-(1-\gamma^2)x^2/2} \,. \end{aligned}$$

Покажем, что для $x \geqslant K$ при всех рассматриваемых значениях a, b, c и γ выполняется неравенство $A_{17}(x, a, b, c, \gamma) \leqslant A_{17}(K, a, b, c, \gamma)$. Для этого, пользуясь соотношением (15), покажем, что для всех $x \geqslant K$

$$A_{16}(x) x^3 \exp\left\{ -\frac{(1-\gamma^2)x^2}{2} \right\} \leqslant A_{16}(K) K^3 \exp\left\{ -\frac{(1-\gamma^2)K^2}{2} \right\}.$$

Заметим прежде всего, что в силу (21) функция $x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$ монотонно убывает при $x \geqslant K$.

Далее рассмотрим выражение $\widehat{A}_{10}(x)x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$. В него входит одно слагаемое вида $x^r \exp\{-sx^2\}$: $x^3 \exp\left\{-(1-\gamma)^2x^2/2\right\}$. Оно убывает в силу (20).

В выражениях $A_{12}(x)x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$ и $\widehat{A}_{11}(x)x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$ представляющие интерес слагаемые имеют вид: $x^2 \exp\left\{-(1-\gamma)^2x^2/2\right\}$; $x \exp\left\{-4(b-c)x^2/(3b^2)\right\} x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$; $x^2 \exp\left\{-4\times\right.$ $\left.\times(b-c)x^2/(3b^2)\right\} x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$; $x \exp\left\{-(1-\gamma)^2x^2/2\right\}$, — и они убывают в силу (17) и (20).

Рассмотрим «наихудшие» слагаемые вида $x^r \exp\{-sx^2\}$, входящие в $\widehat{A}_6(x)x^3 \exp\{-(1-\gamma^2)x^2/2\}$: $(x + \widehat{A}_4(x) + \widehat{A}_5(x))x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$. Первое слагаемое убывает в силу (21). Слагаемые вида $x^r \exp\{-sx^2\}$, присутствующие в выражении $\widehat{A}_4(x)x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$, имеют вид: $x \exp\left\{-(b-c)x^2/(3b^2)\right\} x^2 \exp\left\{-(1-\gamma^2)x^2/4\right\}$; $\exp\left\{-0{,}5\left(2(b-c)/b^2-\right.\right.$ $\left.\left.-\gamma(1-\gamma)\right)x^2\right\} x \exp\left\{-(1-\gamma)^2x^2/4\right\}$, — и они убывают, если выполнены укусловия (16), (17), (20) и (21). Аналогичные рассуждения применимы к обоим слагаемым вида $x^r \exp\{-sx^2\}$ в выражении $\widehat{A}_5(x)x^3 \exp\left\{-(1-\gamma^2)x^2/2\right\}$.

Резюмируя сказанное, получаем следующее утверждение.

Теорема 2. *Предположим, что* $K > 0$, $K^2 \leqslant x^2 \leqslant c_n(x; a, b, c)$, $0 < \gamma < 1$, $a > 0$, $b > c \geqslant 1$ *таковы, что* $A(K) \leqslant 1/6$, $A_3(K) > 0$, $A_7(K) > 0$, $A_{14}(K) > 0$ *и выполнены условия* (16)–(21). *Тогда для всех* $n \geqslant 1$

$$|x|^3 \Delta_n(x) \leqslant A_{17}(K, a, b, c, \gamma) L_n \,.$$

Замечание 2. Несложно убедиться, что при $K \to \infty$ величины $A(K) \to 0$, $A_3(K) \to 1$, $A_7(K) \to 1$, $A_{14}(K) \to \infty$, так что все условия теоремы 2, кроме (16), выполняются автоматически для любых фиксированных a, b, c и γ. Кроме того,

$$\inf_{a,\,b,\,c,\,\gamma} \lim_{K \to \infty} A_{17}(K, a, b, c, \gamma) = \inf_{a,\,b,\,c,\,\gamma} \gamma^{-3} = \lim_{\gamma \to 1-} \gamma^{-3} = 1 \,.$$

При $\gamma \to 1$ справедливость условия (16) также становится очевидной для любых фиксированных $b > c \geqslant 1$.

5 Основные результаты

Теорема 3. *Для константы* U *в неравенстве*

$$\sup_{x \in \mathbb{R}} |x|^3 \Delta_n(x) \leqslant U L_n \,, \quad n \geqslant 1 \,,$$

справедлива оценка

$$U = \inf_{K,\,a,\,b,\,c,\,\gamma} \max\left\{C_0K^3,\ A_{17}(K,a,b,c,\gamma),\ P(a,b,c,K)\right\}, \qquad (30)$$

где инфимум берется по множеству всех значений вспомогательных параметров a, b, c, γ *и* K, *удовлетворяющих условиям теоремы* 2, *а также условиям* $K \geqslant 1/\sqrt{2\pi}$ *и* $a \leqslant P(a,b,c,K)/\varkappa$ (*величина* $\varkappa = 0{,}54\ldots$ *определена в лемме* 5).

Следствие 1. Неравенство (3) при всех $n \geqslant 1$ справедливо с $C = U + C_0$.

Результаты вычислений величин U и C:

K_0	a_0	b_0	c_0	γ_0	U	C
3,3830	12,9655	1,6737	1,0170	0,4880	21,6817	22,2417

Алгоритм оптимизации в (30) реализован с использованием пакета Matlab 7.10.0, при этом применялась процедура fminsearch(. . .) от аргументов K, a, b, c и γ.

Доказательство теоремы 3. Покажем, что для любых значений параметров K, a, b, c и γ, удовлетворяющих условиям теоремы, при всех $x \in \mathbb{R}$ и $n \geqslant 1$ выполняется неравенство $|x|^3\Delta_n(x) \leqslant \leqslant \max\{C_0K^3,\ A_{17}(K,a,b,c,\gamma),\ P(a,b,c,K)\}L_n$. Зафиксируем K и x. Возможны два варианта:

(1) $|x| \leqslant K$, тогда $|x|^3\Delta_n(x) \leqslant C_0K^3L_n$;

(2) $|x| > K$, тогда для заданных a, b, c и $n \geqslant 1$ имеется три возможности:

(2.1) $c_n(x;a,b,c) \leqslant 0$, но это условие означает, что $aL_n \geqslant |x|^3$, откуда по лемме 5 получаем, что $\Delta_n(x) \leqslant \varkappa \leqslant P(a,b,c,K)/a \leqslant \leqslant P(a,b,c,K)L_n/|x|^3$;

(2.2) $0 < c_n(x;a,b,c) < x^2$, т. е. $x^2 \geqslant \max\{(2\pi)^{-1},\ c_n(x;a,b,c)\}$, что отвечает условиям теоремы 1, согласно которой $|x|^3\Delta_n(x) \leqslant \leqslant P(a,b,c,x)L_n \leqslant P(a,b,c,K)L_n$, так как $|x| \geqslant K$, а функция $P(a,b,c,x)$ монотонно не возрастает по x;

(2.3) $x^2 \leqslant c_n(x;a,b,c)$. Учитывая также условие $x^2 \geqslant K^2$, по теореме 2 получаем $|x|^3\Delta_n(x) \leqslant A_{17}(K,a,b,c,\gamma)L_n$.

Теорема 3 доказана.

Литература

1. *Петров В. В.* Суммы независимых случайных величин. — М.: Наука, 1972. 416 с.
2. *Шевцова И. Г.* Уточнение оценок скорости сходимости в теореме Ляпунова // Докл. РАН, 2010. Т. 435. Вып. 1. С. 26–28.

3. *Королев В. Ю., Шевцова И. Г.* Новая моментная оценка скорости сходимости в теореме Ляпунова // Теория вероятностей и ее применения, 2010. Т. 55. Вып. 3. С. 577–582.
4. *Нагаев С. В.* Некоторые предельные теоремы для больших уклонений // Теория вероятностей и ее применения, 1965. Т. 10. Вып. 2. С. 231–254.
5. *Бикялис А.* Оценки остаточного члена в центральной предельной теореме // Литов. матем. сб., 1966. Т. 6. Вып. 3. С. 323–346.
6. *Paditz L.* Über eine Fehlerabschätzung im zentralen Grenzwertsatz // Wiss. Z. Hochschule für Verkehrswesen "Friedrich List." — Dresden, 1986. Vol. 33. No. 2. P. 399–404.
7. *Paditz L.* On the analytical structure of the constant in the nonuniform version of the Esseen inequality // Statistics, 1989. Vol. 20. No. 3. P. 453–464.
8. *Нефедова Ю. С., Шевцова И. Г.* О неравномерных оценках скорости сходимости в центральной предельной теореме // Теория вероятностей и ее применения, 2012. Т. 57. Вып. 1.
9. *Бхаттачария Р. Н., Ранга Рао Р.* Аппроксимация нормальным распределением. — М.: Наука, 1982. 286 с.

ABSTRACTS

DEVELOPMENT OF SOFTWARE TOOLS FOR ANALYSIS OF NONLINEAR MULTICHANNEL STOCHASTIC SYSTEMS

I. N. Sinitsyn[1], *A. R. Korepanov*[2], *V. V. Belousov*[3], *and T. D. Konashenkova*[4]

[1]IPI RAN, sinitsin@dol.ru
[2]IPI RAN, ekorepanov@ipiran.ru
[3]IPI RAN, vbelousov@ipiran.ru
[4]IPI RAN, tkonashenkova@ipiran.ru

The methodological aspects of the advanced software tools "CStS-ANALYSIS" for the analysis of nonlinear multichannel circular stochastic systems are discussed. The testing examples are given.

Keywords: analytical modeling; circular random variable; circular stochastic process; coefficients of circular orthogonal expansions; "CStS-ANALYSIS"; MATLAB, nonlinear multichannel stochastic system; one- and multidimensional densities; circular orthogonal expansion; standard density; wrapped normal density

PERSONIFIED TRANSFORMATION OF COLOR IMAGES PRESENTATIONS ON A PC MONITOR

O. P. Arkhipov and Z. P. Zykova

Oryol Branch of IPI RAN, ofran@orel.ru

The method used to increase the accuracy of hue reproduced in polygraphy is based on application of equicontrast graded transformation of hue scales in color space of "standard" observer. A similar problem is considered for the color space of arbitrary personal computer user in case of color representation to computer monitor. The proposed method is based on application of personified graded equicontrast transformation of hue scales. The examples to illustrate the efficiency of the method are provided.

Keywords: hue reproducing; contrast; gradation; hue scales; equicontrast graded transformation

TOOLS FOR SELF-TIMED CELLS CHARACTERIZATION

Yu. Diachenko[1], *N. Morozov*[2], *D. Stepchenkov*[3], *and Yu. Stepchenkov*[4]

[1]IPI RAN, diaura@mail.ru
[2]IPI RAN, NMorozov@ipiran.ru
[3]IPI RAN, Stepchenkov@mail.ru
[4]IPI RAN, YStepchenkov@ipiran.ru

A procedure for characterization of the standard cells (extraction of time and electrical parameters for creating functional model) intended for self-timed circuit design is presented. It is shown that specific behavior of self-timed circuits additionally constrains characterization process as well as leads to changing the values of some parameters of a cell model. A software tool for standard cell characterization and the results of its beta testing are described.

Keywords: self-timed cells; characterization; functional model; software tools

EXCEPTIONS FIXATION IN RECURRENT DATAFLOW PROCESSOR

R. Zelenov[1], *A. Prokofyev*[2], *Yu. Stepchenkov*[3], *and V. Volchek*[4]

[1]IPI RAN, graf.developer@gmail.co
[2]IPI RAN, a.a.prokofyev@mail.ru
[3]IPI RAN, YStepchenkov@ipiran.ru
[4]IPI RAN, v_volchek@inbox.ru

The problems related to fixation and handling of exceptions in the multicore dataflow computing system are identified. Possible variants of exceptions detection and debugging information generation for its transfer on processing are analyzed. The principles of fixation logic creation and exceptions handling, applicable for all kinds of computing systems, are formulated.

Keywords: exceptions; interrupt requests; dataflow architecture; recurrence

HIERARCHICAL METHOD OF THE ANALYSIS OF SPEED-INDEPENDENT ELECTRONIC CIRCUITS

L. Plekhanov

IPI RAN, LPlekhanov@ipiran.ru

Development and implementation of speed-independent circuits possessing unique properties are largely discouraged by design difficulties, in particular, the analysis

for speed-independence of "large" circuits. A method of the analysis of circuits of unlimited size based on the functional approach is suggested. In the literature, such approach and method are not observed, and the method is proposed here for the first time.

Keywords: speed-independent circuits; self-timed circuits; analysis for speed-independence

THE FAULT TOLERANCE INCREASING OF CACHE-MEMORY DATA BY THEIR REFRESHING

B. Shmeilin

IPI RAN, shmeilin@mail.ru

For reducing vulnerability to transient failures of cache-memory data, a refresh method is suggested. The data refreshing planning is based on the applications code analysis at compilation stage. The method is applicable to linked data structures that is shown on arrays and nested structures. The data refresh planning is demonstrated by an example of the matrix multiplication program.

Keywords: microprocessor systems; cache-memory data; vulnerability to transient failures; linked data structures

LEXICAL SEMANTICS MODELING IN THE TASKS OF COMPUTATIONAL LINGUISTICS

O. S. Kozhunova

IPI RAN, okozhunova@ipiran.ru

An analytical review of lexical and semantic methods, their examples, and spheres of application is presented. Among the considered articles, those published in the 2009 and 2011 Proceedings of The World Congress in Computer Science, Computer Engineering, and Applied Computing were chosen.

Keywords: lexical and semantic methods; computational linguistics; text data processing

PROGRAM-ORIENTED INDICATORS: PRODUCTION AND APPLICATION IN SCIENCE

I. Zatsman[1] *and A. Durnovo*[2]

[1]IPI RAN, iz_ipi@a170.ipi.ac.ru
[2]IPI RAN, duralex49@mail.ru

The results of the project whose goal is to create a tool for developing sets of indicators for monitoring and evaluating research and developments (R&D) program activities are presented. Each set of developed science and technology indicators is oriented at the assessment of only one R&D program, i. e., at its objectives, resources, and results. The tool for developing indicators is being designed as an experts knowledge base named the proactive dictionary, which is the component of the evaluation system. This component enables experts to fix stages of indicators development, to present the results of developing different variants of indicators in graphic form, to compare and evaluate variants oriented at objectives, resources, and results of R&D program activities. The paper describes two theoretical models of the process of indicators development. It also describes a prototype of the proactive dictionary created on the base of these models. Finally, it describes the results of an experiment of development of a set of indicators by a group of the experts.

Keywords: processes of experts' knowledge creation about indicators; knowledge representation about indicators; semiotic models; program-oriented indicators; indicator denotata

SUBJECT VOCABULARIES: PECULIARITIES, APPLICATIONS, AND PROSPECTS

N. V. Somin[1]*, I. P. Kuznetsov*[2]*, M. M. Sharnin*[3]*, and V. G. Nikolaev*[4]

[1]IPI RAN, somin@post.ru
[2]IPI RAN, igor-kuz@mtu-net.ru
[3]IPI RAN, keywen1@mail.ru
[4]IPI RAN, DHLine@yandex.ru

The semantic-oriented linguistic processor which extracts the knowledge from natural language texts and forms the structures in the Knowledge Base is considered. Significant part of this processor is the block of lexical morphological analysis including the subject vocabularies. The paper discusses the direction of development the vocabularies which are used to analyze the context of words and assigns new

semantic signs for them. It gives the possibilities to eliminate various ambiguities in the process of syntactical semantic analysis.

Keywords: knowledge extraction; linguistic processor; lexical morphological analysis; subject vocabulary

ESTIMATES OF THE RATE OF CONVERGENCE OF THE DISTRIBUTIONS OF RANDOM SUMS TO THE SKEW STUDENT DISTRIBUTION

***V. E. Bening*[1], *L. M. Zaks*[2], *and V. Yu. Korolev*[3]**

[1]Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, bening@cs.msu.su

[2]Department of Modeling and Mathematical Statistics, Alpha-Bank, lily.zaks@gmail.com

[3]Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

The estimates are constructed for the accuracy of the approximation of the distributions of special random sums of independent random variables with nonzero means by the skew Student distribution.

Keywords: random sum; convergence rate estimate; mixed Poisson distribution; skew Student distribution

ON THE ACCURACY OF NORMAL APPROXIMATION FOR RISK ESTIMATE DISTRIBUTION WHEN THRESHOLDING SIGNAL WAVELET COEFFICIENTS IN CASE OF UNKNOWN NOISE LEVEL

O. V. Shestakov

M. V. Lomonosov Moscow State University; IPI RAN, oshestakov@cs.msu.su

Asymptotic properties of risk estimate when thresholding signal wavelet coefficients are studied. Some rates of convergence of risk estimate distribution to the normal law are obtained.

Keywords: wavelets; thresholding; risk estimate; normal distribution; rate of convergence

APPLICATION OF WAVELET-ANALYSIS TO CLIMATIC RESEARCH

***M. Sh. Khaziakhmetov*[1] *and T. V. Zakharova*[2]**

[1]M. V. Lomonosov Moscow State University, khaziakhmetov@yandex.ru
[2]M. V. Lomonosov Moscow State University, lsa@cs.msu.ru

The application of wavelet analysis of digital signals to a problem of climatic research — the search of numerical regularities in the field of the Earth water vapor is described.

Keywords: wavelet-analysis; wavelet-expansion; wavelet-transform

STABILITY OF NORMAL SCALE MIXTURES WITH RESPECT TO VARIATIONS IN MIXING DISTRIBUTION

A. K. Gorshenin

IPI RAN, agorshenin@ipiran.ru

The problem of stability of finite-scale mixtures of normal distributions with respect to small parameter variations for two practically important cases is examined. The stability is set in terms of two-sided inequalities for the Lévy metrics between mixing distributions and mixtures.

Keywords: scale mixtures of normal distributions; Lévy metric

ON NONUNIFORM ESTIMATES OF THE RATE OF CONVERGENCE IN THE CENTRAL LIMIT THEOREM

***M. E. Grigorieva*[1] *and S. V. Popov*[2]**

[1]M. V. Lomonosov Moscow State University, maria-grigoryeva@yandex.ru
[2]M. V. Lomonosov Moscow State University, popovserg@yandex.ru

It is shown that in the nonuniform analog of the Berry–Esseen inequality $(1+|x|^3)|F_n(xB_n) - \Phi(x)| \le \left(C/B_n^3\right) \sum_{k=1}^{n} \beta_k$, $n \ge 1$, $x \in \mathbb{R}$, where $F_n(x)$ is the distribution function of the sum of n independent random variables $X_1, \ldots, X_n$ with $\mathsf{E}X_k = 0$, $\mathsf{E}X_k^2 = \sigma_k^2$; $\beta_k = \mathsf{E}|X_k|^3 < \infty$, $k = 1, \ldots, n$; $B_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$; $\Phi(x)$ is the standard normal distribution function, the absolute constant C satisfies the inequality $C \le 22.2417$.

Keywords: central limit theorem; nonuniform estimate of convergence rate; Berry–Esseen inequality; absolute constant

ОБ АВТОРАХ

Архипов Олег Петрович (р. 1948) — кандидат технических наук, директор Орловского филиала ИПИ РАН

Белоусов Василий Владимирович (р. 1977) — кандидат технических наук, заведующий сектором ИПИ РАН

Бенинг Владимир Евгеньевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Волчек Виктор Николаевич (р. 1985) — инженер-исследователь, ИПИ РАН

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН

Григорьева Мария Евгеньевна (р. 1986) — аспирантка кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Дурново Александр Адрианович (р. 1949) — ведущий программист, ИПИ РАН

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Закс Лилия Михайловна (р. 1989) — главный специалист отдела моделирования и математической статистики Альфа-банка

Захарова Татьяна Валерьевна (р. 1967) — кандидат физико-математических наук, научный сотрудник кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Зацман Игорь Моисеевич (р. 1952) — кандидат технических наук, заведующий отделом ИПИ РАН

Зеленов Роман Альбертович (р. 1986) — инженер-исследователь, ИПИ РАН

Зыкова Зоя Петровна (р. 1953) — кандидат физико-математических наук, заведующая лабораторией Орловского филиала ИПИ РАН

Кожунова Ольга Сергеевна (р. 1982) — кандидат технических наук, заведующая сектором ИПИ РАН

Конашенкова Татьяна Дмитриевна (р. 1964) — ведущий программист, ИПИ РАН

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, заведующий сектором ИПИ РАН

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ведущий научный сотрудник ИПИ РАН

Кузнецов Игорь Петрович (р. 1938) — доктор технических наук, профессор, главный научный сотрудник ИПИ РАН

Морозов Николай Викторович (р. 1956) — старший научный сотрудник, ИПИ РАН

Николаев Владимир Григорьевич (р. 1948) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Плеханов Леонид Петрович (р. 1943) — кандидат технических наук, старший научный сотрудник, ИПИ РАН

Попов Сергей Владимирович (р. 1986) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Прокофьев Александр Александрович (р. 1986) — инженер-исследователь, ИПИ РАН

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заведующий отделом ИПИ РАН

Сомин Николай Владимирович (р. 1947) — кандидат физико-математических наук, ведущий научный сотрудник ИПИ РАН

Степченков Дмитрий Юрьевич (р. 1973) — научный сотрудник, ИПИ РАН

Степченков Юрий Афанасьевич (р. 1951) — кандидат технических наук, заведующий отделом ИПИ РАН

Хазиахметов Максим Шамилевич (р. 1989) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Шарнин Михаил Михайлович (р. 1959) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Шестаков Олег Владимирович (р. 1976) — кандидат физико-математических наук, ассистент кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Шмейлин Борис Захарьевич (р. 1939) — кандидат технических наук, доцент, старший научный сотрудник ИПИ РАН

ABOUT AUTHORS

Arkhipov Oleg P. (b. 1948) — Candidate of Science (PhD) in technology, Director, Oryol Branch of the Institute of Informatics Problems, Russian Academy of Sciences

Belousov Vasiliy V. (b. 1977) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

Bening Vladimir E. (b. 1954) — 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

Diachenko Yuri G. (b. 1958) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Durnovo Alexander A. (b. 1949) — leading programmer, Institute of Informatics Problems, Russian Academy of Sciences

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Grigorieva Maria E. (b. 1986) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Khaziakhmetov Maxim Sh. (b. 1989) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov State University

Konashenkova Tatyana D. (b. 1964) — managing programmer, Institute of Informatics Problems, Russian Academy of Sciences

Korepanov Edward R. (b. 1966) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Kozhunova Olga S. (b. 1982) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

Kuznetsov Igor P. (b. 1938) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences

Morozov Nikolai V. (b. 1956) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Nikolaev Vladimir G. (b. 1948) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Plekhanov Leonid P. (b. 1943) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences
Popov Sergey V. (b. 1986) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov State University
Prokofyev Alexander A. (b. 1986) — engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences
Sharnin Mikhail M. (b. 1959) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences
Shestakov Oleg V. (b. 1976) — Candidate of Science (PhD) in physics and mathematics, assistant 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
Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Institute of Informatics Problems, Russian Academy of Sciences
Shmeilin Boris Z. (b. 1939) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences
Somin Nikolay V. (b. 1947) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences
Stepchenkov Yuri A. (b. 1951) — Candidate of Science (PhD) in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences
Stepchenkov Dmitri Yu. (b. 1973) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences
Volchek Victor N. (b. 1985) — engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences
Zakharova Tatyana V. (b. 1967) — Candidate of Science (PhD) in physics and mathematics, scientist, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University
Zaks Lily M. (b. 1989) — principal officer, Department of Modeling and Mathematical Statistics, Alpha-Bank
Zatsman Igor M. (b. 1952) — Candidate of Science (PhD) in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences
Zelenov Roman A. (b. 1986) — engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences
Zykova Zoya P. (b. 1953) — Candidate of Science (PhD) in physics and mathematics, Head of Laboratory, Oryol Branch of the Institute of Informatics Problems, Russian Academy of Sciences

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий. Журнал издается на русском языке. Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

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 «Системы и средства информатики» 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 авторских прав 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, полные 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. Корректура 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 — A4; 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 LaTex (.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 и http://www.ipiran.ru/journal/issues/2011_01_eng/authors.asp, аннотация (не более 100 слов), ключевые слова, литература. 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 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