

Информатика и её применения

Том 7 Выпуск 4 Год 2013

СОДЕРЖАНИЕ

Исследование динамики многомерных стохастических систем на основе энтропийного моделирования А. Н. Тырсин, О. В. Ворфоломеева	3
Предельная теорема для геометрических сумм независимых неодинаково распределенных случайных величин и ее применение к прогнозированию вероятности катастроф в неоднородных потоках экстремальных событий М. Е. Григорьева, В. Ю. Королев, И. А. Соколов	11
Распределение времени выхода из множества состояний перегрузки в системе $M M 1 \langle L, H \rangle \langle H, R \rangle$ с гистерезисным управлением нагрузкой Ю. В. Гайдамака, А. В. Печинкин, Р. В. Разумчик, А. К. Самуйлов, К. Е. Самуйлов, И. А. Соколов, Э. С. Сопин, С. Я. Шоргин	20
Об одной задаче оптимального управления нагрузкой на сервер М. Г. Коновалов	34
Оптимизация функций Lab-контрастного градационного преобразования О. П. Архипов, З. П. Зыкова	44
Метод извлечения библиографической информации из полнотекстовых описаний изобретений И. М. Зацман, В. А. Хавансков, С. К. Шубников	52
О сходимости распределений случайных сумм к скошенным экспоненциально-степенным законам М. Е. Григорьева, В. Ю. Королев	66
Обращение сферического преобразования Радона в классе дискретных случайных функций О. В. Шестаков, М. Г. Кузнецова, И. А. Садовой	75
Информационно-аналитическая автоматизированная система «Мегалит» в оптимизации диагностики и лечения мочекаменной болезни М. П. Кривенко, С. А. Голованов, П. А. Савченко, А. В. Сивков, А. П. Сучков	82
Анализ однородности данных о химическом составе камней при уролитолизе М. П. Кривенко, С. А. Голованов, А. В. Сивков	94
О методе прогнозирования и классификации для цензурированных данных Т. В. Захарова, Е. М. Абрамова	105

Информатика и её применения

Том 7 Выпуск 4 Год 2013

СОДЕРЖАНИЕ

Conceptual declarative problem specification and solving in data intensive domains <i>L. Kalinichenko, S. Stupnikov, A. Vovchenko, and D. Kovalev</i>	112
Probabilistic methods for self-correcting hardware design <i>S. Dolev, S. Frenkel, and D. E. Tamir</i>	140
Об авторах	148
Авторский указатель за 2013 г.	150
2013 Author Index	154

Технический редактор *Л. Кокушкина*

Художественный редактор *Д. Комиссарова*

Сдано в набор 01.10.13. Подписано в печать 15.12.13. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 19,3. Уч.-изд. л. 18,0. Тираж 100 экз.
Заказ № 4699

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43

Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов

Москва 121099, Шубинский пер., д. 6.

ИССЛЕДОВАНИЕ ДИНАМИКИ МНОГОМЕРНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ НА ОСНОВЕ ЭНТРОПИЙНОГО МОДЕЛИРОВАНИЯ*

А. Н. Тырсин¹, О. В. Ворфоломеева²

Аннотация: Описан энтропийный подход к моделированию динамики стохастических систем. В его основе лежит представление системы в виде многомерного случайного вектора. Показано, что изменение энтропии многомерной стохастической системы может быть выражено через дисперсии и условные корреляции компонент случайного вектора. Это позволяет обнаружить причину изменения энтропии системы и оценить этот случайный вектор количественно. Получено, что энтропия стохастической системы складывается из двух компонент, которые характеризуют ее свойства. Первая компонента определяет предельную энтропию, соответствующую полной независимости элементов системы, и характеризует рассмотрение целостного объекта как состоящего из частей (аддитивность). Вторая компонента отражает степень взаимосвязей между элементами системы, характеризуя свойства системы как целого (целостность). Описанный подход делает возможным использование энтропийной модели в задачах диагностики и контроля состояния стохастических систем, а также эффективного управления ими. К достоинствам предложенного подхода следует отнести простоту реализации и интерпретации математической модели, универсальность и применимость к стохастическим системам различной природы, возможность ее использования на малых выборках данных. Приведен пример практического применения математической модели.

Ключевые слова: многомерная случайная величина; энтропия; динамика; стохастическая система; дисперсия; корреляция

DOI: 10.14357/19922264130401

1 Введение

Энтропия является фундаментальным свойством любых систем с неоднозначным, или вероятностным, поведением [1]. В настоящее время достаточно распространено использование энтропии для описания поведения открытых стохастических систем в различных областях [2–6]. Однако, несмотря на частое использование этого термина, использование энтропии для моделирования открытых систем, в отличие от термодинамики, недостаточно формализовано и носит в основном качественный характер. Отсутствуют достаточно простые и адекватные математические модели, позволяющие связать энтропию с фактическими характеристиками состояний стохастических систем.

Известно [7], что энтропия непрерывной случайной величины X (дифференциальная энтропия) определяется по формуле:

$$H(X) = - \int_{-\infty}^{+\infty} f(x) \ln f(x) dx, \quad (1)$$

где $f(x)$ — плотность распределения случайной величины X . Полученная по формуле (1) энтропия называется энтропией закона распределения или дифференциальной энтропией.

Представим стохастическую систему S в виде многомерной случайной величины $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$. Будем считать, что данное представление является адекватной математической моделью системы S . Каждый элемент Y_i вектора \mathbf{Y} является одномерной случайной величиной, которая характеризует функционирование соответствующего элемента исследуемой системы. Элементы могут быть как взаимозависимыми, так и не зависеть друг от друга. Совместную дифференциальную энтропию многомерной случайной величины \mathbf{Y} будем определять по формуле [7]:

$$H(\mathbf{Y}) = - \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{Y}}(x_1, x_2, \dots, x_m) \times \ln f_{\mathbf{Y}}(x_1, x_2, \dots, x_m) dx_1 dx_2 \dots dx_m. \quad (2)$$

* Работа выполнена при поддержке проекта 12-М-127-2049 фундаментальных исследований УрО РАН.

¹ Научно-инженерный центр «Надежность и ресурс больших систем и машин» УрО РАН, г. Екатеринбург, at2001@yandex.ru

² Челябинский государственный университет, ya.olga.work@yandex.ru

где $f_Y(x_1, x_2, \dots, x_m)$ — совместная плотность распределения случайных величин Y_1, Y_2, \dots, Y_m .

Аналитическое нахождение энтропии $H(Y)$ в настоящее время получено лишь для совместного нормального распределения [8]. Рассмотрение других распределений затруднено отсутствием меры нелинейной корреляционной взаимосвязи случайных величин с иными распределениями, аналогичной определителю корреляционной матрицы для совместного нормального распределения. В [9] предпринята попытка оценить взаимосвязь случайных величин через совместную энтропию. Но при этом требуется вычислить саму энтропию многомерной случайной величины по ограниченной выборке, что весьма затруднительно, особенно если законы распределения не известны. А необходимо, наоборот, выразить совместную энтропию многомерной случайной величины через характеристики ее компонент.

Таким образом, актуальна задача разработки и формального обоснования энтропийного подхода к моделированию открытых стохастических систем. Поэтому рассмотрим более общий случай, когда случайный вектор Y не имеет многомерного нормального распределения.

2 Энтропия многомерной непрерывной случайной величины

Теорема 1. Пусть X_1, X_2 — две непрерывные случайные величины, определенные на всей числовой оси и описываемые однотипами законами распределения с плотностями $f_1(x) = f(x; \mu_1, \lambda_1)$, $f_2(x) = f(x; \mu_2, \lambda_2)$ соответственно, где μ_1, μ_2 и λ_1, λ_2 — параметры положения и масштаба случайных величин X_1 и X_2 . Тогда разность дифференциальных энтропий случайных величин X_1 и X_2 равна:

$$H(X_2) - H(X_1) = \ln \frac{\lambda_2}{\lambda_1}. \quad (3)$$

Доказательство. Выразим плотность вероятности случайной величины X_2 через плотность вероятности случайной величины X_1 :

$$f(x; \mu_2, \lambda_2) = \frac{\lambda_1}{\lambda_2} f\left(\frac{\lambda_1}{\lambda_2}(x + \mu_2 - \mu_1); \mu_1, \lambda_1\right).$$

С учетом последнего соотношения разность дифференциальных энтропий случайных величин X_1 и X_2 равна:

$$\begin{aligned} H(X_2) - H(X_1) &= \\ &= - \int_{-\infty}^{+\infty} f(x; \mu_2, \lambda_2) \ln f(x; \mu_2, \lambda_2) dx + \\ &+ \int_{-\infty}^{+\infty} f(x; \mu_1, \lambda_1) \ln f(x; \mu_1, \lambda_1) dx = \\ &= - \frac{\lambda_1}{\lambda_2} \int_{-\infty}^{+\infty} f\left(\frac{\lambda_1}{\lambda_2}(x + \mu_2 - \mu_1); \mu_1, \lambda_1\right) \times \\ &\times \ln \left[\frac{\lambda_1}{\lambda_2} f\left(\frac{\lambda_1}{\lambda_2}(x + \mu_2 - \mu_1); \mu_1, \lambda_1\right) \right] dx + \\ &+ \int_{-\infty}^{+\infty} f(x; \mu_1, \lambda_1) \ln f(x; \mu_1, \lambda_1) dx = \\ &= - \ln \frac{\lambda_1}{\lambda_2} \int_{-\infty}^{+\infty} f(x; \mu_2, \lambda_2) dx - \\ &- \int_{-\infty}^{+\infty} f(t; \mu_1, \lambda_1) \ln f(t; \mu_1, \lambda_1) dt + \\ &+ \int_{-\infty}^{+\infty} f(x; \mu_1, \lambda_1) \ln f(x; \mu_1, \lambda_1) dx, \end{aligned}$$

где

$$t = \frac{\lambda_1}{\lambda_2}(x + \mu_2 - \mu_1).$$

Отсюда имеем:

$$H(X_2) - H(X_1) = - \ln \frac{\lambda_1}{\lambda_2} + H(X_1) - H(X_1) = \ln \frac{\lambda_2}{\lambda_1}.$$

Следствие 1. Пусть в условиях теоремы 1 X_1 и X_2 — две непрерывные случайные величины, имеющие конечные дисперсии. Поскольку среднее квадратическое отклонение непрерывной случайной величины, если оно существует, пропорционально параметру масштаба, то формулу (3) можно записать в виде:

$$H(X_2) - H(X_1) = \ln \frac{\sigma_2}{\sigma_1} = \ln \sigma_2 - \ln \sigma_1, \quad (4)$$

где σ_1 и σ_2 — средние квадратические отклонения случайных величин X_1 и X_2 .

Следствие 2. Дифференциальная энтропия непрерывной случайной величины X , имеющей конечную дисперсию σ_X^2 , равна:

$$H(X) = \ln \sigma_X + C = \ln e^C \sigma_X,$$

где $C = H(\overset{\circ}{X}) = H(X/\sigma_X)$ — энтропия случайной величины $\overset{\circ}{X}$ с единичной дисперсией и тем же распределением, что и у случайной величины X .

Действительно, из (4) получим

$$H(X) - H(\overset{\circ}{X}) = \ln \sigma_X,$$

откуда

$$H(X) = \ln \sigma_X + H(\overset{\circ}{X}) = \ln \sigma_X + C = \ln e^C \sigma_X.$$

Отметим, что константа C характеризует энтропию закона распределения. Она может быть выражена через введенный в [10] энтропийный коэффициент k закона распределения как $C = \ln 2k$.

Теорема 2. Пусть имеется две системы непрерывных случайных величин $\mathbf{Y}^{(1)} = (Y_1^{(1)}, Y_2^{(1)}, \dots, Y_m^{(1)})$ и $\mathbf{Y}^{(2)} = (Y_1^{(2)}, Y_2^{(2)}, \dots, Y_m^{(2)})$, каждые соответствующие компоненты которых $Y_i^{(1)}, Y_i^{(2)}$ ($i = 1, 2, \dots, m$) определены на всей числовой оси, имеют конечные дисперсии и описываются однотипными законами распределения с некоторыми параметрами положения и масштаба. Тогда разность совместных энтропий систем случайных величин $\mathbf{Y}^{(2)} = (Y_1^{(2)}, Y_2^{(2)}, \dots, Y_m^{(2)})$ и $\mathbf{Y}^{(1)} = (Y_1^{(1)}, Y_2^{(1)}, \dots, Y_m^{(1)})$ равна:

$$\begin{aligned} \Delta H(\mathbf{Y}) &= H(\mathbf{Y}^{(2)}) - H(\mathbf{Y}^{(1)}) = \\ &= \sum_{k=1}^m \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}} + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}, \end{aligned} \quad (5)$$

где

$$\sigma_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}} = \sigma_{Y_k^{(j)}} \sqrt{1 - R_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}}^2};$$

$R_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}}^2$ — коэффициенты детерминации соответствующих регрессионных зависимостей, $k = 2, 3, \dots, m, j = 1, 2$.

Доказательство. Совместная энтропия $H(\mathbf{Y})$ системы случайных величин \mathbf{Y} согласно свойству иерархической аддитивности [11] равна

$$\begin{aligned} H(\mathbf{Y}) &= H(Y_1) + H(Y_2/Y_1) + H(Y_3/Y_1Y_2) + \dots \\ &\quad \dots + H(Y_m/Y_1 \dots Y_{m-1}). \end{aligned} \quad (6)$$

Рассмотрим две системы непрерывных случайных величин $\mathbf{Y}^{(1)}$ и $\mathbf{Y}^{(2)}$, каждые соответствующие компоненты $Y_i^{(1)}, Y_i^{(2)}$ ($i = 1, 2, \dots, m$) которых определены на всей числовой оси, имеют конечные дисперсии и описываются однотипными законами

распределения с некоторыми параметрами положения и масштаба. Тогда изменение энтропии с учетом (6) равно:

$$\begin{aligned} \Delta H(\mathbf{Y}) &= H(\mathbf{Y}^{(2)}) - H(\mathbf{Y}^{(1)}) = \\ &= H(Y_1^{(2)}) - H(Y_1^{(1)}) + H(Y_2^{(2)}/Y_1^{(2)}) - \\ &\quad - H(Y_2^{(1)}/Y_1^{(1)}) + \dots + H(Y_m^{(2)}/Y_1^{(2)} \dots Y_{m-1}^{(2)}) - \\ &\quad - H(Y_m^{(1)}/Y_1^{(1)} \dots Y_{m-1}^{(1)}). \end{aligned}$$

Условное математическое ожидание $E[Y_2/Y_1 = x]$ является регрессией Y_2 на Y_1 с коэффициентом детерминации R_{Y_2/Y_1}^2 . Поэтому дисперсия случайной величины Y_2/Y_1 равна дисперсии $\sigma_{Y_2/Y_1}^2 = \sigma_{Y_2}^2(1 - R_{Y_2/Y_1}^2)$ остаточной случайной компоненты регрессии $E[Y_2/Y_1 = x]$ [12]. Аналогично величина $E[Y_k/Y_1Y_2 \dots Y_{k-1} = \mathbf{x}]$ является регрессией Y_k на случайные величины Y_1, Y_2, \dots, Y_{k-1} . Дисперсия случайной величины $Y_k/Y_1Y_2 \dots Y_{k-1}$ равна дисперсии остаточной случайной компоненты регрессии $E[Y_k/Y_1Y_2 \dots Y_{k-1} = \mathbf{x}]$ [12]. Поэтому

$$\sigma_{Y_k/Y_1Y_2 \dots Y_{k-1}}^2 = \sigma_{Y_k}^2(1 - R_{Y_k/Y_1Y_2 \dots Y_{k-1}}^2), \quad (7)$$

где $R_{Y_k/Y_1Y_2 \dots Y_{k-1}}^2$ — коэффициенты детерминации соответствующих регрессионных зависимостей, $k = 2, 3, \dots, m$.

Отсюда с учетом (4) и (7) получим

$$\begin{aligned} \Delta H(\mathbf{Y}) &= \ln \frac{\sigma_{Y_1^{(2)}}}{\sigma_{Y_1^{(1)}}} + \sum_{k=2}^m \ln \frac{\sigma_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}}{\sigma_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}} = \\ &= \ln \frac{\sigma_{Y_1^{(2)}}}{\sigma_{Y_1^{(1)}}} + \sum_{k=2}^m \ln \frac{\sigma_{Y_k^{(2)}} \sqrt{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}}{\sigma_{Y_k^{(1)}} \sqrt{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}} = \\ &= \sum_{k=1}^m \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}} + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}. \end{aligned}$$

Обозначив

$$\begin{aligned} \Delta H(\mathbf{Y})_{\Sigma} &= \sum_{k=1}^m \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}}; \\ \Delta H(\mathbf{Y})_{\mathbf{R}} &= \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}, \end{aligned}$$

представим формулу (5) как

$$\Delta H(\mathbf{Y}) = \Delta H(\mathbf{Y})_{\Sigma} + \Delta H(\mathbf{Y})_{\mathbf{R}}, \quad (8)$$

где $\Delta H(\mathbf{Y})_{\Sigma}$ и $\Delta H(\mathbf{Y})_{\mathbf{R}}$ — приращения энтропии за счет изменения дисперсий и корреляций случайных величин Y_1, Y_2, \dots, Y_m .

Следствие 1. Если случайный вектор \mathbf{Y} является гауссовским, то получим рассмотренный в [13] частный случай

$$\Delta H(\mathbf{Y}) = \sum_{k=1}^m \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}} + \frac{1}{2} \ln \frac{|\mathbf{R}_{\mathbf{Y}^{(2)}}|}{|\mathbf{R}_{\mathbf{Y}^{(1)}}|}, \quad (9)$$

где $|\mathbf{R}_{\mathbf{Y}^{(j)}}|$ — определитель корреляционной матрицы $\mathbf{R}_{\mathbf{Y}^{(j)}}$ случайного вектора $\mathbf{Y}^{(j)}$, $j = 1, 2$.

Следствие 2. Совместная дифференциальная энтропия $H(\mathbf{Y})$ системы непрерывных случайных величин $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ равна

$$H(\mathbf{Y}) = \sum_{k=1}^m H(Y_k) + \frac{1}{2} \sum_{k=2}^m \ln \left(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2 \right). \quad (10)$$

Действительно, подставив в (5) вместо $\mathbf{Y}^{(1)}$ и $\mathbf{Y}^{(2)}$ соответственно две системы непрерывных случайных величин $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_m)$ и $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$, каждые соответствующие компоненты \tilde{Y}_i, Y_i ($i = 1, 2, \dots, m$) которых определены на всей числовой оси, имеют конечные дисперсии и описываются одинаковыми законами распределения, причем \tilde{Y}_i ($i = 1, 2, \dots, m$) являются взаимно независимыми, получим:

$$H(\mathbf{Y}) - H(\tilde{\mathbf{Y}}) = \sum_{k=1}^m \ln \frac{\sigma_{Y_k}}{\sigma_{\tilde{Y}_k}} + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k/Y_1 \dots Y_{k-1}}^2}{1 - R_{\tilde{Y}_k/\tilde{Y}_1 \dots \tilde{Y}_{k-1}}^2}.$$

Поскольку $\forall k H(\tilde{Y}_k) = H(Y_k)$, $\sigma_{Y_k} = \sigma_{\tilde{Y}_k}$, $R_{\tilde{Y}_k/\tilde{Y}_1 \tilde{Y}_2 \dots \tilde{Y}_{k-1}}^2 = 0$, то

$$H(\mathbf{Y}) = H(\tilde{\mathbf{Y}}) + \frac{1}{2} \sum_{k=2}^m \ln \left(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2 \right) = \sum_{k=1}^m H(Y_k) + \frac{1}{2} \sum_{k=2}^m \ln \left(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2 \right).$$

Формула (10) является обобщением приведенного в [13] соотношения для энтропии многомерного нормального распределения

$$H(\mathbf{Y}) = \sum_{k=1}^m H(Y_k) + \frac{1}{2} \ln |\mathbf{R}_{\mathbf{Y}}|, \quad (11)$$

где $|\mathbf{R}_{\mathbf{Y}}|$ — определитель корреляционной матрицы $\mathbf{R}_{\mathbf{Y}}$. Поэтому, так же как и для (11), и в общем случае согласно (10) энтропия многомерной

случайной величины складывается из двух составляющих:

$$H(\mathbf{Y}) = H(\mathbf{Y})_{\Sigma} + H(\mathbf{Y})_{\mathbf{R}}, \quad (12)$$

где

$$H(\mathbf{Y})_{\Sigma} = \sum_{k=1}^m H(Y_k);$$

$$H(\mathbf{Y})_{\mathbf{R}} = (1/2) \sum_{k=1}^m \ln \left(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2 \right).$$

3 Исследование изменения состояния стохастической системы на основе энтропийной модели

Аддитивные представления (12) и (8) в виде двух компонент как самой энтропии, так и ее изменения показывают ее дуализм. Компонента $H(\mathbf{Y})_{\Sigma}$ определяет предельную энтропию, соответствующую полной независимости элементов системы. Поэтому условно назовем ее энтропией хаотичности. Величина $H(\mathbf{Y})_{\mathbf{R}}$ равна энтропии за счет совместной корреляционной взаимосвязи между элементами системы, ее условно можно назвать энтропией самоорганизации. Следует заметить, что дуализм энтропии в том или ином виде отмечался в ряде публикаций [1, 2, 7, 9, 13, 14].

Таким образом, изменение энтропии происходит аддитивным образом: с одной стороны — за счет изменения дисперсий, а с другой стороны — из-за изменения коррелированности случайных величин Y_1, Y_2, \dots, Y_m . Следовательно, причины роста и уменьшения энтропии системы могут быть различными. Например, энтропию системы можно увеличить (уменьшить) посредством увеличения (уменьшения) дисперсий $\sigma_{Y_k}^2$ или уменьшения (увеличения) коэффициентов детерминации $R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2$ компонент вектора \mathbf{Y} .

Выражение (5) позволяет обнаружить причину изменения энтропии системы и оценить его количественно. Это делает возможным использование энтропийной модели в задачах контроля и диагностики состояния стохастических систем. Пусть стохастическая система представима в виде случайного вектора \mathbf{Y} . Тогда на основе модели (5) можно осуществлять мониторинг состояния стохастической системы путем анализа изменения ее энтропии. Это можно сделать следующим образом. Будем считать, что две системы непрерывных случайных величин $\mathbf{Y}^{(1)}$ и $\mathbf{Y}^{(2)}$ соответствуют предыдущему и текущему периодам функционирования системы.

Тогда, отслеживая изменение $\Delta H(\mathbf{Y})$ энтропии в целом и ее компонент $\Delta H(\mathbf{Y})_{\Sigma}$, $\Delta H(\mathbf{Y})_{\mathbf{R}}$, можно сделать выводы о состоянии системы. Анализ изменения каждой из случайных величин Y_k

$$\Delta H(\mathbf{Y})_{\Sigma,k} = \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}};$$

$$\Delta H(\mathbf{Y})_{\mathbf{R},k} = \frac{1}{2} \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}$$

позволит выявить те элементы системы (компоненты системы \mathbf{Y}), которые оказали наибольшее влияние на изменение энтропии всей системы.

Поскольку $R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2 \geq R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2 \geq R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-2}^{(2)}}^2 \geq R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-2}^{(1)}}^2 \geq \dots \geq R_{Y_k^{(2)}/Y_1^{(2)}}^2 \geq R_{Y_k^{(1)}/Y_1^{(1)}}^2$ [12], то оценивать вклад произвольного l -го элемента в изменение энтропии самоорганизации целесообразно через их предельные значения:

$$\Delta H(\mathbf{Y})_{\mathbf{R},l}^* = \frac{1}{2} \ln \frac{1 - R_{Y_l^{(2)}/Y_1^{(2)} \dots Y_{l-1}^{(2)} Y_{l+1}^{(2)} \dots Y_m^{(2)}}^2}{1 - R_{Y_l^{(1)}/Y_1^{(1)} \dots Y_{l-1}^{(1)} Y_{l+1}^{(1)} \dots Y_m^{(1)}}^2},$$

$$l = 1, 2, \dots, m.$$

Выше был рассмотрен случай, когда пары случайных величин $Y_i^{(1)}$, $Y_i^{(2)}$ ($i = 1, 2, \dots, m$) имели однотипные распределения. Если данное предположение не выполняется, то тогда изменение энтропии хаотичности согласно (10) равно:

$$\Delta H(\mathbf{Y})_{\Sigma} = H(\mathbf{Y}^{(2)})_{\Sigma} - H(\mathbf{Y}^{(1)})_{\Sigma} =$$

$$= \sum_{k=1}^m \Delta H(\mathbf{Y})_{\Sigma,k} = \sum_{k=1}^m \left[H(Y_k^{(2)}) - H(Y_k^{(1)}) \right],$$

т. е. потребуются определять энтропии одномерных случайных величин по выборочным данным. Отметим, что в настоящее время предложен ряд алгоритмов для решения данной задачи [15–18].

4 Пример мониторинга состояния многомерных стохастических систем

Рассмотрим задачу мониторинга состояния стохастической системы на модельном примере. Пусть некоторая стохастическая система S моделируется в виде случайного вектора $\mathbf{Y} = (Y_1, Y_2, Y_3)$. Для упрощения будем считать его нормальным. Рассмотрим динамику функционирования системы на основе модели (5). Пусть в предыдущем и текущем периодах функционирования системы имеем

случайные векторы $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ с ковариационными матрицами, равными

$$\Sigma^{(1)} = \begin{pmatrix} 1,114 & 1,131 & 0,494 \\ 1,131 & 3,310 & 3,205 \\ 0,494 & 3,205 & 5,348 \end{pmatrix};$$

$$\Sigma^{(2)} = \begin{pmatrix} 1,796 & 1,713 & 0,381 \\ 1,713 & 3,589 & 3,199 \\ 0,381 & 3,199 & 4,841 \end{pmatrix}.$$

Согласно (5), (8) имеем:

$$\frac{\sigma_{Y_1^{(2)}}}{\sigma_{Y_1^{(1)}}} = 1,270; \quad \ln \frac{\sigma_{Y_1^{(2)}}}{\sigma_{Y_1^{(1)}}} = 0,239;$$

$$\frac{\sigma_{Y_2^{(2)}}}{\sigma_{Y_2^{(1)}}} = 1,041; \quad \ln \frac{\sigma_{Y_2^{(2)}}}{\sigma_{Y_2^{(1)}}} = 0,040;$$

$$\frac{\sigma_{Y_3^{(2)}}}{\sigma_{Y_3^{(1)}}} = 0,951; \quad \ln \frac{\sigma_{Y_3^{(2)}}}{\sigma_{Y_3^{(1)}}} = -0,050;$$

$$\Delta H(\mathbf{Y})_{\Sigma} = \sum_{k=1}^3 \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}} = 0,229;$$

$$\frac{1 - R_{Y_2^{(2)}/Y_1^{(2)}}^2}{R_{Y_2^{(1)}/Y_1^{(1)}}^2} = \frac{1 - 0,455}{1 - 0,347} = 0,835;$$

$$\frac{1 - R_{Y_3^{(2)}/Y_1^{(2)} Y_2^{(2)}}^2}{1 - R_{Y_3^{(1)}/Y_1^{(1)} Y_2^{(1)}}^2} = \frac{1 - 0,866}{1 - 0,673} = 0,409;$$

$$\frac{1}{2} \ln \frac{1 - R_{Y_2^{(2)}/Y_1^{(2)}}^2}{1 - R_{Y_2^{(1)}/Y_1^{(1)}}^2} = -0,090;$$

$$\frac{1}{2} \ln \frac{1 - R_{Y_3^{(2)}/Y_1^{(2)} Y_2^{(2)}}^2}{1 - R_{Y_3^{(1)}/Y_1^{(1)} Y_2^{(1)}}^2} = -0,447;$$

$$\Delta H(\mathbf{Y})_{\mathbf{R}} = \frac{1}{2} \sum_{k=2}^3 \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2} = -0,537;$$

$$\Delta H(\mathbf{Y}) = \Delta H(\mathbf{Y})_{\Sigma} + \Delta H(\mathbf{Y})_{\mathbf{R}} =$$

$$= 0,229 - 0,537 = -0,308.$$

Таким образом, энтропия системы в текущем периоде уменьшилась на 0,308, причем энтропия хаотичности выросла на 0,229 ($\Delta H(\mathbf{Y})_{\Sigma} = 0,229$), а энтропия самоорганизации сократилась на 0,537 ($\Delta H(\mathbf{Y})_{\mathbf{R}} = -0,537$). Это означает, что в текущем периоде в системе преобладала тенденция снижения энтропии самоорганизации.

Анализ изменения каждой из компонент $\Delta H(\mathbf{Y})_{\Sigma,k}$ и $\Delta H(\mathbf{Y})_{\mathbf{R},k}$ показывает, что на рост энтропии хаотичности повлиял первый элемент системы, а на снижение энтропии самоорганизации — третий элемент.

5 Заключение

1. Предложено энтропийное моделирование динамики многомерных стохастических систем. В его основе лежит представление системы в виде случайного вектора, каждая из компонент которого представляет собой непрерывную случайную величину.
2. Получены аналитические выражения для энтропии многомерной случайной величины и ее динамики.

Основные достоинства предложенного подхода:

- простота реализации и интерпретации математической модели;
- энтропийная модель применима при решении задач диагностики и контроля состояния стохастических систем, а также эффективного управления ими;
- универсальность и применимость к стохастическим системам различной природы;
- возможность использования на малых выборках данных.

Литература

1. *Климонтович Ю. Л.* Введение в физику открытых систем. — М.: Янус-К, 2002. 284 с.
2. *Вильсон А. Дж.* Энтропийные методы моделирования сложных систем / Пер. с англ. — М.: Наука, 1978. 248 с. (*Wilson A. G.* Entropy in urban and regional modeling. — London: Pion, 1970. 166 p.)
3. *Трубецков Д. И., Мчедлова Е. С., Красичков Л. В.* Введение в теорию самоорганизации открытых систем. — М.: Физматлит, 2002. 200 с.
4. *Романовский Ю. М., Степанова Н. В., Чернавский Д. С.* Математическое моделирование в биофизике. — Москва—Ижевск: Институт компьютерных исследований, 2003. 402 с.
5. *Прангишвили И. В.* Энтропийные и другие системные закономерности: Вопросы управления сложными системами. — М.: Наука, 2003. 428 с.

6. *Скоробогатов С. М.* Катастрофы и живучесть железобетонных сооружений (классификация и элементы теории). — Екатеринбург: УрГУПС, 2009. 512 с.
7. *Шеннон К.* Работы по теории информации и кибернетике / Пер с англ. С. Карпова. — М.: ИИЛ, 1963. 830 с.
8. *Cover T. M., Thomas J. A.* Elements of information theory. — N.Y.: Wiley, 1991. 563 p.
9. *Pena D., Van der Linde A.* Dimensionless measures of variability and dependence for multivariate continuous distributions // Commun. Stat.: Theor. M., 2007. Vol. 36. Issue 10. P. 1845–1854.
10. *Новицкий П. В.* Основы информационной теории измерительных устройств. — Ленинград: Энергия, 1968. 248 с.
11. *Стратонович Р. Л.* Теория информации. — М.: Советское радио, 1975. 424 с.
12. *Greene W. H.* Econometric analysis. — 7th ed. — Prentice Hall, 2011. 1230 p.
13. *Тырсин А. Н., Соколова И. С.* Энтропийно-вероятностное моделирование гауссовских стохастических систем // Математическое моделирование, 2012. Т. 24. № 1. С. 88–102.
14. *Николис Г., Пригожин И.* Самоорганизация в неравновесных системах: от диссипативных структур к упорядоченности через флуктуации / Пер с англ. — М.: Мир, 1979. 512 с. (*Nikolis G., Prigogine I.* Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations. — N.Y.: John Wiley & Sons, 1977. 512 p.)
15. *Beirlant J., Dudewicz E. J., Gyorfi L., van der Meulen E. C.* Nonparametric entropy estimation: an overview // Int. J. Math. Stat. Sci., 1997. Vol. 6. Issue 1. P. 17–39.
16. *Stowell D., Plumbley M. D.* Fast multidimensional entropy estimation by k - d partitioning // IEEE Signal Proc. Lett., 2009. Vol. 16. Issue 6. P. 537–540.
17. *Тырсин А. Н., Клявин И. А.* Повышение точности оценки энтропии случайных экспериментальных данных // Системы управления и информационные технологии, 2010. № 1(39). С. 87–90.
18. *Noughabi H. A., Arghami N. R.* A new estimator of entropy // J. Iran. Stat. Soc., 2010. Vol. 9. Issue 1. P. 53–64.

Поступила в редакцию 13.05.13

STUDY OF THE DYNAMICS OF MULTIDIMENSIONAL STOCHASTIC SYSTEMS BASED ON ENTROPY MODELING

A. N. Tyrsin¹ and O. V. Vorfolomeeva²

¹Science and Engineering Center “Reliability and Resource of Large Systems and Machines,” Ural Branch, Russian Academy of Sciences, Yekaterinburg 620049, Russian Federation

²Chelyabinsk State University, Chelyabinsk 454001, Russian Federation

Abstract: A new entropy approach of modeling of dynamics of stochastic systems is described. It is based on the representation of the system in the form of a multidimensional stochastic vector. It is shown that the change in entropy of a multivariate stochastic system can be expressed in terms of dispersions and conditional correlations of a component of a random vector. This allows to reveal the cause of the change in the entropy of the system and to evaluate it quantitatively. It was found that the entropy of a stochastic system consists of two components that characterize its properties. The first component determines the limit entropy corresponding to the full independence of the elements of the system and defines the consideration of the integral object as consisting of components (additivity). The second component reflects the extent of interrelation between the elements of the system, defining the properties of the system as a whole (integrity). This approach makes it possible to use entropy models in the diagnostics and control of stochastic systems as well as efficient management. The advantages of the proposed approach include the simplicity of implementation and interpretation of the mathematical model, the universality and adaptability for stochastic systems of different nature, the possibility of its use on small samples of data. The article contains an example of the practical application of a mathematical model.

Keywords: multidimensional random variable; entropy; dynamics; stochastic system; dispersion; correlation

DOI: 10.14357/19922264130401

Acknowledgments

The work was supported by the project 12-M-127-2049 of Basic Researches of the Ural Branch of the Russian Academy of Sciences.

References

1. Klimontovich, Ju. L. 2002. *Vvedenie v fiziku otkrytykh sistem [Introduction to the physics of open systems]*. Moscow: Janus-K Publ. 284 p.
2. Wilson A. G. 1970. *Entropy in urban and regional modeling*. L.: Pion. 166 p.
3. Trubeckov, D. I., E. S. Mchedlova, and L. V. Krasichkov. 2002. *Vvedenie v teoriyu samoorganizatsii otkrytykh sistem [Introduction to the theory of self-organization of open systems]*. Moscow: Publishing House of Physical-Mathematical Literature. 200 p.
4. Romanovskij, Ju. M., N. V. Stepanova, and D. S. Chernavskij. 2003. *Matematicheskoe modelirovanie v biofizike [Mathematical modeling in biophysics]*. Moscow–Izhevsk: Computer Research Institute. 402 p.
5. Prangishvili, I. V. 2003. *Entropiynye i drugie sistemnye zakonomernosti: Voprosy upravleniya slozhnymi sistemami [Entropic and other system regularities: Questions of management of complex systems]*. Moscow: Nauka. 428 p.
6. Skorobogatov, S. M. 2009. *Katastrofy i zhivuchest' zhelezobetonnykh sooruzheniy (klassifikatsiya i elementy teorii) [Catastrophes and serviceability of reinforced concrete buildings (classification and elements of theory)]*. Ekaterinburg: Ural State University of Railway Transport. 512 p.
7. Shannon C. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27(3):379–423; 4:623–656.
8. Cover, T. M., and J. A. Thomas. 1991. *Elements of information theory*. N.Y.: Wiley. 563 p.
9. Pena, D., A. Van der Linde. 2007. Dimensionless measures of variability and dependence for multivariate continuous distributions. *Commun. Stat. Theor. M.* 36(10):1845–1854.
10. Novickij, P. V. 1968. *Osnovy informatsionnoy teorii izmeritel'nykh ustroystv [Fundamentals of information theory measuring devices]*. Leningrad: Energy Publ. 248 p.
11. Stratonovich, R. L. 1975. *Teoriya informatsii [Information theory]*. Moscow: Soviet Radio Publ. 424 p.
12. Greene, W. H. 2011. *Econometric analysis*. 7th ed. Prentice Hall. 1230 p.
13. Tyrsin, A. N., and I. S. Sokolova. 2012. Entropiynno-veroyatnostnoe modelirovanie gaussovskikh stokhastiches-

- skikh sistem [Entropy-propabilistic modeling of Gaussian stochastic systems]. *Matematicheskoe Modelirovanie* [Mathematical Modeling] 24(1):88–102.
14. Nikolis, G., and I. Prigogine. 1977. *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. N.Y.: John Wiley & Sons. 512 p.
 15. Beirlant, J., E.J. Dudewicz, L. Györfi, and E.C. van der Meulen. 1997. Nonparametric entropy estimation: an overview. *Int. J. Math. Stat. Sci.* 6(1):17–39.
 16. Stowell, D., and M.D. Plumbley. 2009. Fast multidimensional entropy estimation by k - d partitioning. *IEEE Signal Proc. Lett.* 16(6):537–540.
 17. Tyrsin, A. N., and I. A. Kljavin. 2010. Povyshenie tochnosti otsenki entropii sluchaynykh eksperimental'nykh dannyykh [Increase in the accuracy of the estimate of the entropy of a random experimental data]. *Sistemy Upravleniya i Informatsionnye Tekhnologii* [Control Systems and Information Technologies] 1(39):87–90.
 18. Noughabi, H. A., and N. R. Arghami. 2010. A new estimator of entropy. *J. Iran. Stat. Soc.* 9(1):53–64.

Received May 13, 2013

Contributors

Tyrsin Alexander N. (b. 1961) — Doctor of Science in technology, leading researcher, Science and Engineering Center “Reliability and Resource of Large Systems and Machines,” Ural Branch of the Russian Academy of Sciences, Yekaterinburg 620049, Russian Federation; at2001@yandex.ru

Vorfolomeeva Olga V. (b. 1987) — PhD student, Mathematical Faculty, Chelyabinsk State University, Chelyabinsk 454001, Russian Federation; ya.olga.work@yandex.ru

ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ ГЕОМЕТРИЧЕСКИХ СУММ НЕЗАВИСИМЫХ НЕОДИНАКОВО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ВЕЛИЧИН И ЕЕ ПРИМЕНЕНИЕ К ПРОГНОЗИРОВАНИЮ ВЕРОЯТНОСТИ КАТАСТРОФ В НЕОДНОРОДНЫХ ПОТОКАХ ЭКСТРЕМАЛЬНЫХ СОБЫТИЙ*

М. Е. Григорьева¹, В. Ю. Королев², И. А. Соколов³

Аннотация: Рассматривается задача прогнозирования вероятностей катастроф в неоднородных потоках экстремальных событий. Статья развивает и обобщает некоторые методы, предложенные авторами в предыдущих работах. Поток экстремальных событий рассматривается как маркированный точечный случайный процесс с необязательно одинаково распределенными интервалами между точками (событиями). Основой предлагаемых обобщений служат предельные теоремы для геометрических случайных сумм независимых неодинаково распределенных случайных величин и теория Балкемы – Пикандса – Де Хаана. Рассмотрена конструкция, в рамках которой в качестве предельного распределения для геометрических случайных сумм независимых неодинаково распределенных случайных величин возникает распределение Вейбулла–Гнеденко. Эффективность методов иллюстрируется на примере их применения к прогнозированию момента столкновения Земли с потенциально опасным астероидом на основе данных Центра по малым планетам Гарвардского университета.

Ключевые слова: катастрофа; экстремальное событие; случайная сумма; геометрическая сумма; закон больших чисел; распределение Вейбулла–Гнеденко; теорема Балкемы–Пикандса–Де Хаана; обобщенное распределение Парето

DOI: 10.14357/19922264130402

1 Введение. Постановка задачи. Определение экстремального процесса

В данной статье рассматривается задача прогнозирования *вероятностных* характеристик катастроф в неоднородных потоках экстремальных событий. Рассмотрим некоторую систему, подвергающуюся влиянию некоторого фактора. Предположим, что сила воздействия этого фактора на систему в каждый момент времени характеризуется некоторым числом, причем это число изменяется во времени. Это может быть:

- финансовая система, которая характеризуется финансовым индексом, таким как DAX, NIKKEI, NASDAQ и т. п.; при этом резкие колебания индекса неблагоприятны и свидетельствуют о тех или иных кризисных явлениях;
- экологическая система, например среда обитания человека, в частности состояние жилых по-

мещений в местностях, подверженных наводнениям или землетрясениям в сейсмоопасных зонах, которое зависит соответственно от силы подземных толчков и уровня подъема воды;

- социальная система, которая подвержена политической неустойчивости или воздействию террористических организаций; при этом в качестве числовой характеристики активности неблагоприятных воздействий может выступить, к примеру, число упоминаний некоторых соответствующих ключевых слов или фраз в социальных информационных сетях или сетях связи;
- наконец, общее состояние планеты Земля, зависящее от расстояния, на которое подлетают к ней потенциально опасные космические объекты — астероиды или кометы.

При этом естественно возникает задача прогнозирования катастроф. Однако без применения специализированных методов, специально ориен-

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515-а, 11-07-00112-а, 12-07-00115-а).

¹ Parexel International, maria-grigoryeva@yandex.ru

² Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, victoryukorolev@yandex.ru

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

тированных на противодействие конкретным рискам, практически никогда нельзя абсолютно точно предсказать силу воздействия фактора на систему в каждый момент времени в будущем. Другими словами, будущее развитие фактора непредсказуемо, вследствие чего значение числа, характеризующего силу воздействия фактора на систему, рассматриваемое как функция времени, целесообразно рассматривать как *случайный процесс*. Поэтому задача прогнозирования самого момента катастрофы сводится к прогнозированию значения случайного процесса (т. е. его значения на вполне определенном элементе множества элементарных исходов) специальными методами, что чрезвычайно трудно и при рассмотрении современных сложных стохастических систем практически не реализуемо с приемлемой точностью.

В то же время возникает вполне реальная и важная задача прогнозирования *распределения* указанного случайного процесса в те или иные моменты времени, т. е. задача прогнозирования его статистических свойств. В результате решения этой задачи появляется возможность правильно оценить уровни угрозы в каждой конкретной ситуации.

Некоторым методам решения последней задачи и посвящена данная статья.

Предположим, что очень большие изменения случайного процесса, характеризующего воздействие фактора на систему, неблагоприятно влияют на систему и могут вызвать ее необратимые изменения. Вместе с тем малые флуктуации случайного процесса, характеризующего воздействие фактора на систему, вполне допустимы (в таких случаях говорят о «фоновом значении» фактора). Поэтому с целью предсказания катастроф разумно рассматривать не все изменения случайного процесса, а лишь такие, величина которых превышает некоторый *потенциально опасный порог*.

Будем говорить, что моменты превышений изменениями случайного процесса потенциально опасного порога в совокупности с самими значениями этих превышений образуют *экстремальный случайный процесс*. Другими словами, экстремальным процессом будем называть маркированный точечный процесс $\{(\tau_i, X_i)\}_{i \geq 1}$, где $\{\tau_i\}_{i \geq 1}$ — точечный случайный процесс, а $\{X_i\}_{i \geq 1}$ — случайные величины. Далее по смыслу задачи будет предполагаться, что $X_i > 0, i = 1, 2, \dots$

Среди всех превышений случайным процессом потенциально опасного порога лишь некоторые очень большие влекут катастрофические последствия. Поэтому наряду с *потенциально опасным порогом* рассмотрим *критический порог*, превышение которого экстремальным процессом и будем считать *катастрофой*.

Для удобства точку отсчета (нуль временной шкалы) поместим в то время, которое будем считать «настоящим». Тем самым «настоящее» характеризуется значением $t = 0$.

Поскольку по условию экстремальный процесс считается случайным, то *нельзя* точно предсказать момент наступления очередной катастрофы. Однако можно вычислить или оценить *вероятности наступления катастрофы* в течение некоторого интервала времени $[0, \tau)$, где $\tau > 0$. Если T — момент наступления катастрофы, то событие «катастрофа наступила в течение интервала времени $[0, \tau)$ » эквивалентно тому, что $T < \tau$. В качестве *исходных данных* будем использовать информацию о развитии экстремального процесса на некотором интервале времени $[t_0, t_1]$, где $t_0 < t_1 < 0$.

Простейшее (примитивное) решение задачи об отыскании вероятности наступления катастрофы в течение интервала времени $[0, \tau)$ при условии $\tau < t_1 - t_0$ выглядит так.

Разобьем интервал времени $[t_0, t_1]$ на непересекающиеся подынтервалы длиной τ . Пусть внутри интервала $[t_0, t_1]$ поместилось N_τ подынтервалов длиной τ . Подсчитаем количество подынтервалов, внутри каждого из которых наступила хотя бы одна катастрофа. Пусть таких подынтервалов оказалось ровно n_τ . Тогда для вероятности наступления катастрофы в течение интервала времени $[0, \tau)$ справедлива оценка:

$$P(T < \tau) \approx \frac{n_\tau}{N_\tau}, \quad (1)$$

основанная на классическом определении вероятности как (предела) частоты.

Недостатки такой оценки очевидны. Например, n_τ просто может оказаться равным нулю, что дает тривиально оптимистичную оценку. Далее, и N_τ , и n_τ могут быть (и, как правило, являются) слишком маленькими, чтобы обеспечить приемлемую точность оценки. Более того, одной-единственной катастрофы может оказаться достаточно для полного уничтожения системы, так что дальнейший сбор информации просто может оказаться невозможным.

2 Метод прогнозирования вероятностей катастроф в неоднородных потоках экстремальных событий

2.1 Особенности метода

К сожалению, именно оценками типа (1) зачастую пользуются на практике для расчетов, свя-

занных с так называемыми большими рисками в страховании, например при страховании промышленных рисков, связанных с крупными авариями и экологическими катастрофами. В данной статье описан метод оценивания указанных вероятностей наступления катастроф, основанный на довольно сложных математических моделях, но свободный от указанных недостатков. Особенность этого метода заключается в том, что для того, чтобы прогнозировать возможности наступления катастроф, необязательно иметь статистику самих *катастроф*.

Простейший вариант этого метода описан в работах [1–3] и книгах [4, 5], где предполагалось, что экстремальный процесс является маркированным процессом восстановления. В указанных работах предполагалось, что моменты τ_1, τ_2, \dots превышений исходным процессом потенциально опасного порога образуют процесс восстановления. Это означает, что случайные величины

$$\zeta_i = \tau_i - \tau_{i-1}, \quad i = 1, 2, \dots, \quad \tau_0 = 0, \quad (2)$$

независимы и имеют одинаковое распределение, т. е. подчиняются одним и тем же статистическим закономерностям. Другими словами, интенсивность потока экстремальных событий считалась постоянной. В то же время в реальных сложных системах, которые в подавляющем большинстве случаев не являются информационно и/или энергетически замкнутыми и подвержены влиянию внешней среды, интенсивности потоков информативных событий не являются постоянными. Например, при адекватном прогнозировании поведения финансовых индексов ключевую роль играет представление о том, что интенсивности потоков информативных событий на финансовых рынках являются случайными [6]. Отказ от предположения о постоянстве интенсивности потока экстремальных событий естественно приводит к необходимости предположить, что случайные величины (2) имеют неодинаковое распределение. Именно такое обобщение методов, предложенных в работах [2–5], и рассматривается в данной работе.

Обозначим величину превышения исходным процессом потенциально опасного порога в момент τ_i символом X_i , $i = 1, 2, \dots$. Будем считать что X_1, X_2, \dots — независимые и одинаково распределенные случайные величины. Это означает, что значения этих случайных величин подчиняются одним и тем же статистическим закономерностям, характеризуемым *функцией распределения*

$$F(x) = P(X_i < x), \quad -\infty < x < \infty, \quad i = 1, 2, \dots$$

Будем считать, что последовательность X_1, X_2, \dots статистически независима от последовательности τ_1, τ_2, \dots .

Пусть x_0 — критический порог, превышение которого значением X_i и есть катастрофа (т. е. катастрофическое событие формально записывается в виде неравенства $X_i \geq x_0$).

Очевидно, что время T наступления катастрофы (т. е. время первого превышения уровня x_0 какой-либо из величин X_i) можно представить в виде геометрической случайной суммы

$$T = \sum_{j=1}^N \zeta_j, \quad (3)$$

где случайные величины ζ_j определены соотношением (2), а N — это случайная величина, имеющая геометрическое распределение с параметром

$$P(X_i < x_0) = F(x_0).$$

Это означает, что

$$P(N = k) = (F(x_0))^{k-1} (1 - F(x_0)), \quad k = 1, 2, \dots$$

При этом в силу независимости последовательностей X_1, X_2, \dots и τ_1, τ_2, \dots число N слагаемых в сумме (3) независимо от самих слагаемых ζ_1, ζ_2, \dots . При этом принципиальным отличием геометрических случайных сумм, рассматриваемых здесь, от геометрических сумм в традиционном понимании (см., например, [7, 8]) является то, что в данном случае слагаемые имеют *неодинаковое* распределение, тогда как в указанных классических книгах изучались геометрические суммы *одинаково* распределенных слагаемых и, соответственно, использовались методы, ориентированные именно на такую ситуацию.

В рамках подхода, рассматриваемого в данной статье, краеугольными камнями являются два теоретических результата. Первый из них — версия закона больших чисел для случайных сумм неодинаково распределенных случайных величин (см. теорему 1 ниже), обосновывающая использование распределения Вейбулла–Гнеденко в качестве модели распределения интервалов времени между катастрофами. Второй — теорема Балкема–Пикандса–Де Хаана (см. теорему 3 ниже), обосновывающая использование обобщенного распределения Парето в качестве модели распределения критических значений неблагоприятного фактора. Эти два общих результата являются основой предлагаемого метода.

2.2 Вспомогательные результаты

Пусть ξ_1, ξ_2, \dots — необязательно одинаково распределенные случайные величины. Для каждого натурального $n \geq 1$ положим

$$S_n = \xi_1 + \xi_2 + \dots + \xi_n.$$

Рассмотрим последовательность целочисленных неотрицательных случайных величин $\{N_n\}_{n \geq 1}$ и будем считать, что при каждом n случайные величины N_n, ξ_1, ξ_2, \dots независимы в совокупности. Более того, предположим, что

$$N_n \rightarrow \infty \text{ по вероятности при } n \rightarrow \infty. \quad (4)$$

Условие (4) означает, что $P(N_n \leq m) \rightarrow 0$ при $n \rightarrow \infty$ для любого $m > 0$. Везде далее символ \implies будет обозначать сходимость по распределению.

Лемма 1. Пусть для некоторой последовательности положительных чисел $\{b_n\}_{n \geq 1}$ выполнены условия $b_n \rightarrow \infty$ при $n \rightarrow \infty$ и

$$\frac{S_n}{b_n} \implies 1, \quad n \rightarrow \infty. \quad (5)$$

Предположим, что выполнено условие (4). Для того чтобы при $n \rightarrow \infty$ имела место сходимость случайных сумм S_{N_n} , нормированных некоторой последовательностью положительных чисел $\{d_n\}_{n \geq 1}$ такой, что $d_n \rightarrow \infty$ при $n \rightarrow \infty$, к некоторой случайной величине Z :

$$\frac{S_{N_n}}{d_n} \implies Z, \quad (6)$$

необходимо и достаточно, чтобы

$$\frac{b_{N_n}}{d_n} \implies Z, \quad n \rightarrow \infty. \quad (7)$$

Замечание 1. В силу вырожденности распределения предельной случайной величины в (5) сходимость по распределению (5) оказывается эквивалентной сходимости по вероятности: для любого $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{b_n} - 1 \right| > \epsilon \right) = 0,$$

которую иногда легче проверять.

Доказательство леммы 1 приведено в [9].

Замечание 2. Лемма 1 является версией закона больших чисел для случайных сумм. Согласно классическим законам больших чисел при увеличении числа слагаемых в рассматриваемых «средних арифметических» информация о конкретном виде распределений слагаемых затухает, стягиваясь в информацию об одном лишь числе. Точно такой же эффект наблюдается в лемме 1: при рассмотрении «случайных средних арифметических» информация о распределениях слагаемых затухает, так что предельное распределение «случайного среднего арифметического» определяется видом предельного распределения для случайного индекса (числа слагаемых в сумме) при надлежащей нормировке.

Для общности пусть $x_n = x_{0,n}$ — (возрастающая) последовательность критических порогов такая, что

$$p_n \equiv 1 - F(x_n) \rightarrow 0 \quad (n \rightarrow \infty). \quad (8)$$

Тогда в данном случае случайная величина $N = N_n$ имеет геометрическое распределение с параметром $q_n = 1 - p_n$. При этом условие (8) гарантирует выполнение условия (4). Более того, $EN_n = p_n^{-1}$ и, как хорошо известно,

$$\lim_{n \rightarrow \infty} \sup_{y \geq 0} |P(p_n N_n \geq y) - e^{-y}| = 0.$$

Предположим, что постоянные b_n , обеспечивающие выполнение условия (5), имеют вид $b_n = bn^\gamma$ при некоторых $b > 0$ и $\gamma > 0$. При этом значения $\gamma > 1$ соответствуют той ситуации, когда случайные величины ζ_i «в среднем» возрастают, т. е. экстремальные события происходят все реже и реже, значения $\gamma < 1$ соответствуют той ситуации, когда случайные величины ζ_i «в среднем» убывают, т. е. экстремальные события происходят все чаще и чаще, а значение $\gamma = 1$ соответствует той ситуации, когда интенсивность потока экстремальных событий «в среднем» постоянна, например в поведении интенсивности наблюдаются проявления цикличности, причем периоды изменения интенсивности заметно меньше периода фиксации наблюдений.

Теперь выберем нормирующие постоянные d_n так, чтобы геометрическая случайная сумма S_{N_n} имела нетривиальное предельное распределение. Из леммы 1 вытекает, что если с учетом выбранной формы постоянных b_n и соотношения (7) постоянные d_n выбрать в виде $d_n = bp_n^{-\gamma}$, то для любого $y > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(\frac{b_{N_n}}{d_n} < y \right) &= \lim_{n \rightarrow \infty} P \left((p_n N_n)^\gamma < y \right) = \\ &= \lim_{n \rightarrow \infty} P \left(p_n N_n < y^{1/\gamma} \right) = 1 - \exp \left\{ -y^{1/\gamma} \right\}. \end{aligned}$$

При этом согласно лемме 1 такое же распределение Вейбулла–Гнеденко с показателем $1/\gamma$ является предельным и для геометрической случайной суммы независимых неодинаково распределенных случайных величин S_{N_n} , причем в силу непрерывности предельного распределения Вейбулла–Гнеденко сходимость (6) равномерна по $x \in \mathbb{R}$. Оформим сказанное в виде следующего утверждения.

Теорема 1. Предположим, что случайная величина N_n имеет геометрическое распределение с параметром p_n , причем $p_n \rightarrow 0$ при $n \rightarrow \infty$. Предположим, что существуют конечные $\gamma > 0$ и $b > 0$ такие, что

$$\frac{S_n}{bn^\gamma} \implies 1 \quad (n \rightarrow \infty).$$

Тогда

$$\lim_{n \rightarrow \infty} \sup_{x \geq 0} \left| P(p_n^\gamma S_{N_n} \geq bx) - \exp\{-x^{1/\gamma}\} \right| = 0.$$

2.3 Описание метода

Итак, учитывая сделанные предположения о нормирующих постоянных, можно заключить, что при достаточно больших значениях x_0

$$P(T < t) \approx 1 - \exp\left\{-[1 - F(x_0)] \left(\frac{t}{b}\right)^{1/\gamma}\right\}, \quad t > 0. \quad (9)$$

Применение описываемого метода вычисления временных характеристик катастроф в неоднородных потоках экстремальных событий заключается в следующем. Пусть $\epsilon \in (0, 1)$ — произвольное число. Решение уравнения

$$P(T < t) = \epsilon$$

относительно t обозначим $t(\epsilon)$. Если распределение случайной величины T имеет вид (9), то, очевидно,

$$t(\epsilon) = b \left[\frac{\ln(1 - \epsilon)}{F(x_0) - 1} \right]^\gamma.$$

Смысл значения $t(\epsilon)$ — это то время, вероятность наступления катастрофы до которого равна ϵ . Из соображений здравого смысла особый интерес представляют значения ϵ , близкие к нулю (соответствующее значение $t(\epsilon)$ — это то время, до которого катастрофа, скорее всего, не наступит), близкие к единице (соответствующее значение $t(\epsilon)$ — это то время, до которого катастрофа, скорее всего, наступит), а также $\epsilon = 1/2$ (соответствующее значение $t(1/2)$ — это «среднее» время до наступления катастрофы).

Особо следует сказать, что при прогнозировании «среднего» или «ожидаемого» времени до катастрофы можно использовать как медиану $t(1/2)$ случайной величины T , которая определяется как решение уравнения

$$1 - \exp\left\{-[1 - F(x_0)] \left(\frac{t}{b}\right)^{1/\gamma}\right\} = \frac{1}{2}$$

относительно t и, очевидно, равна

$$t\left(\frac{1}{2}\right) = b \left[\frac{\ln 2}{1 - F(x_0)} \right]^\gamma,$$

так и математическое ожидание

$$ET = \frac{b\Gamma(1 + \gamma)}{[1 - F(x_0)]^\gamma}.$$

При этом необходимо отметить, что, например, в случае $\gamma = 1$ медиана $t(1/2)$ случайной величины T почти в полтора раза (точнее, в $(\ln 2)^{-1}$ раз) меньше математического ожидания ET .

При этом параметры b и γ легко оценить методом наименьших квадратов. Предположим, что в нашем распоряжении имеется выборка Z_1, Z_2, \dots, Z_n предыдущих значений случайных величин ζ_j . Нормирующая функция $b_k = bk^\gamma$ параметра k имеет смысл тренда, или основной тенденции поведения реализации $R_k = Z_1 + \dots + Z_k$ случайной функции S_k . С целью линеаризации регрессионной задачи прологарифмируем b_k и R_k , обозначим $\beta = \log b$ и получим приближенные равенства

$$\log R_k \approx \beta + \gamma \log k, \quad k = 1, \dots, n, \quad (10)$$

в правой части которых стоят линейные функции параметров β и γ . Используя стандартный метод наименьших квадратов оценивания параметров линейной регрессии (10), получим оценки

$$\gamma \approx \hat{\gamma} = \frac{n \sum_{k=1}^n (\log k \cdot \log R_k) - \log n! \sum_{k=1}^n \log R_k}{n \sum_{k=1}^n (\log k)^2 - (\log n!)^2};$$

$$b = \exp\{\beta\} \approx \exp\left\{\frac{1}{n} \left(\sum_{k=1}^n \log R_k - \hat{\gamma} \log n! \right)\right\}.$$

Чтобы получить оценку величины $1 - F(x_0)$, необходимо построить разумную и адекватную параметрическую математическую модель (приближение) для функции $F(x)$. С этой целью используем метод построения асимптотических аппроксимаций для $F(x)$ при больших x_0 , основанный на теореме Балкема — Пикандса — Де Хаана и называемый методом превышений порога (POT-метод, POT = Peaks Over Threshold).

Пусть случайная величина ζ имеет функцию распределения $F(x)$. В рамках рассматриваемого метода прогнозирования катастроф как превышений экстремальным процессом критических уровней большой интерес представляет описание условного распределения превышения случайной величиной ζ некоторого (большого) порога u :

$$F_u(y) = P(\zeta - u < y | \zeta > u), \quad 0 \leq y \leq x_F - u,$$

где $y = (x - u)$ — превышение порога и $x_F = \sup\{x \in \mathbb{R} | F(x) < 1\} \leq \infty$. Функция этого условного распределения F_u может быть выражена через F :

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}.$$

Если порог u достаточно велик, то большинство реализаций случайной величины ζ лежит между 0 и u , так что оценить F в этом промежутке несложно. Но оценить F_u проблематично, так как соответствующих наблюдений мало. На помощь приходит следующая теорема.

Теорема 2 [10, 11]. *Функция распределения F принадлежит области макс-притяжения распределения, предельного для экстремальных значений тогда и только тогда, когда существует измеримая функция $\sigma(u) > 0$ такая, что*

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y < x_F - u} |F_u(y) - G_{\delta, \sigma(u)}(y)| = 0,$$

где $G_{\delta, \sigma}(y)$ — функция обобщенного распределения Парето:

$$G_{\delta, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\delta}{\sigma} y\right)^{-1/\delta}, & \delta \neq 0; \\ 1 - e^{-y/\sigma}, & \delta = 0. \end{cases}$$

Условием теоремы удовлетворяет большинство используемых на практике распределений. Параметр δ показывает, насколько тяжел хвост: чем больше δ , тем тяжелее хвост. Например, при моделировании финансовых данных обычно используется $\delta \geq 0$.

3 Пример применения метода прогнозирования вероятностных характеристик глобальных катастроф

Для иллюстрации рассмотрим интересную задачу, связанную с определением риска глобальных катаклизмов, вызванных столкновением Земли с довольно большими небесными телами (астероидами, кометами).

Известно, что такие объекты приближаются к Земле относительно часто. Исследования проводились на основании данных Центра по малым планетам Гарвардского университета, представленных в книге Остина Аткинсона [12]. Это таблица, в которой содержатся предсказания дат приближения к Земле на расстояние не более 0,2 астрономической единицы (а.е.) комет и малых планет на ближайшие 33 года начиная с июня 1999 г. Всего таких предсказаний 191, для каждого из них известно минимальное расстояние, на которое малая планета подойдет к Земле, год и месяц предполагаемого

сближения. Приводимые ниже вычисления проведены на основе метода, предложенного в данной статье, и уточняют результаты, приведенные в [2–5]. Уточнение достигнуто за счет того, что здесь используется не модель экспоненциального распределения, а более гибкая модель распределения Вейбулла–Гнеденко.

Расстояние, на которое очередной (i -й) космический объект приблизится к Земле, будем считать реализацией случайной величины Q_i , распределение которой, вообще говоря, неизвестно и подлжит определению (оцениванию). При этом, в отличие от рассматривавшейся ранее формальной модели экстремального процесса, интерес представляет не *максимальное*, а *минимальное* значение величин Q_i . Снимем формальные противоречия, полагая $X_i = Q_i^{-1}$.

Формализуем сказанное. Имеем выборку $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, $n = 191$, независимых одинаково распределенных случайных величин. Эти величины обратны расстояниям между Землей и потенциально опасными астероидами. Предполагается, что известны расстояния от центра Земли до центров астероидов. Все подсчеты ведутся в а.е., 1 а.е. = 149,6 млн км. Радиус R Земли равен $R = 6400 \text{ км} \approx 4,278075 \cdot 10^{-5} \text{ а.е.} = 0,00004278075 \text{ а.е.}$

Как уже говорилось, в указанной книге [12] приведены данные лишь о тех потенциально опасных астероидах, которые приближаются к Земле менее чем на 0,2 а.е. Таким образом, наблюдается не «полный» набор величин X_1, X_2, \dots , а лишь те из них, которые превосходят $(0,2)^{-1} = 5 \text{ (а.е.)}^{-1}$. Будем считать, что порог $u = 5 \text{ а.е.}^{-1}$ достаточно велик, чтобы аппроксимация, устанавливаемая теоремой Балкемы – Пикандса – Де Хаана для распределения превышений такого порога, была достаточно адекватна, так что в качестве модели распределения случайных величин X_i можно взять обобщенное распределение Парето

$$F(x; \alpha, \sigma) = \begin{cases} 0, & \text{если } x < 5; \\ 1 - \frac{C}{(x - \alpha)^\sigma}, & \text{если } x \geq 5. \end{cases}$$

Как показали вычисления, проведенные в работах [2, 3], эта модель действительно демонстрирует высочайшее согласие с указанными данными. При этом критический порог x_0 , превышение которого случайной величиной X_i означает катастрофу (столкновение астероида с Землей), равен

$$x_0 = \frac{1}{R} = \frac{1}{6400} \text{ км}^{-1} = 23374,9993 \text{ а.е.}^{-1}.$$

Для статистического оценивания параметров C , α и σ в работах [2, 3] использовалось несколько

методов, но наилучшее согласие данных наблюдалось с моделью, построенной на основе оценок максимального правдоподобия:

$$\hat{\alpha} = -3,165; \hat{\sigma} = 2,37; C = 96,757.$$

Оценки наименьших квадратов для параметров b и γ , построенные по выборке Z_1, \dots, Z_{191} временных промежутков между экстремальными сближениями астероидов с Землей оказались равными

$$b \approx 0,1728 \text{ года} \approx 2,0733 \text{ мес.}; \quad \gamma \approx 1,0012.$$

В результате применения описанного метода к вычислению оценок временных характеристик катастрофы, связанной со столкновением Земли с астероидом, получены следующие значения.

- время \underline{t} , до которого с вероятностью 0,9999 столкновение Земли с астероидом не произойдет, примерно равно 1235 годам;
- время \bar{t} , до которого с вероятностью 0,9999 столкновение Земли с астероидом заведомо произойдет, примерно равно 111 154 073 годам;
- «среднее» время t^* до столкновения Земли с астероидом примерно равно 12 071 039 годам (при этом в качестве ожидаемого времени катастрофы использовалось математическое ожидание).

Литература

1. *Королев В. Ю., Соколов И. А.* Некоторые вопросы анализа катастрофических рисков, связанных с неоднородными потоками экстремальных событий // Системы и средства информатики. Спец. вып. Математические методы и модели информатики. Стохастические технологии и системы. — М.: ИПИ РАН, 2005. С. 109–125.
2. *Королев В. Ю., Соколов И. А., Гордеев А. С., Григорьева М. Е., Попов С. В., Чебоненко Н. А.* Некоторые методы анализа временных характеристик катастроф в неоднородных потоках экстремальных событий // Системы и средства информатики. Спец. вып. Математические методы в информационных технологиях. — М.: ИПИ РАН, 2006. С. 5–23.
3. *Королев В. Ю., Соколов И. А., Гордеев А. С., Григорьева М. Е., Попов С. В., Чебоненко Н. А.* Некоторые методы прогнозирования временных характеристик рисков, связанных с катастрофическими событиями // Актуарий, 2007. № 1. С. 34–40.
4. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008. 200 с.
5. *Королев В. Ю., Шоргин С. Я.* Математические методы анализа стохастической структуры информационных потоков. — М.: ИПИ РАН, 2011. 130 с.
6. *Королев В. Ю., Черток А. В., Корчагин А. Ю., Горшеннин А. К.* Вероятностно-статистическое моделирование информационных потоков в сложных финансовых системах на основе высокочастотных данных // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 12–21.
7. *Kalashnikov V.* Geometric sums: Bounds for rare events with applications. — Dordrecht–Boston–London: Kluwer Academic Publs., 1997. 288 p.
8. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и дополн. — М.: Физматлит, 2011. 620 с.
9. *Королев В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. I // Теория вероятностей и ее применения, 1994. Т. 39. Вып. 2. С. 313–333.
10. *Balkema A., de Haan L.* Residual life time at great age // Ann. Probab., 1974. Vol. 2. P. 792–804.
11. *Pickands J.* Statistical inference using extreme order statistics // Ann. Stat., 1975. Vol. 3. P. 119–131.
12. *Аткинсон О.* Столкновение с Землей / Пер с англ. — СПб.: Амфора/Эврика, 2001. 400 с. (*Atkinson O.* Impact Earth: Asteroids, comets and meteors — the growing threat. — Virgin Publ., 1999. 256 p.)

Поступила в редакцию 20.10.13

A LIMIT THEOREM FOR GEOMETRIC SUMS OF INDEPENDENT NONIDENTICALLY DISTRIBUTED RANDOM VARIABLES AND ITS APPLICATION TO THE PREDICTION OF THE PROBABILITIES OF CATASTROPHES IN NONHOMOGENEOUS FLOWS OF EXTREMAL EVENTS

M. E. Grigor'eva¹, V. Yu. Korolev^{2,3}, and I. A. Sokolov³

¹Parexel International, Moscow 121609, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation

³Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: The problem of prediction of the probabilities of catastrophes in nonhomogeneous flows of extremal events is considered. The paper develops and generalizes some methods proposed by the authors in their previous works. The flow of extremal events is considered as a marked point stochastic process with not necessarily identically distributed intervals between points (events). The proposed generalizations are based on limit theorems for geometric sums of independent not necessarily identically distributed random variables and the Balkema – Pickands – De Haan theory. Within the framework of the construction under consideration, the Weibull–Gnedenko distribution appears as a limit law for geometric sums of independent not necessarily identically distributed random variables. The efficiency of the proposed methods is illustrated by the example of their application to the problem of prediction the time of the impact of the Earth with a potentially dangerous asteroid based on the data of the IAU (International Astronomical Union) Minor Planet Center.

Keywords: catastrophe; extremal event; random sum; geometric sum; law of large numbers; Weibull–Gnedenko distribution; Balkema – Pickands – De Haan theorem; generalized Pareto distribution

DOI: 10.14357/19922264130402

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (Projects Nos. 11-01-00515-a, 11-07-00112-a, and 12-07-00115-a).

References

1. Korolev, V. Yu., and I. A. Sokolov. 2005. Nekotorye voprosy analiza katastroficheskikh riskov, svyazannykh s neodnorodnymi potokami ekstremal'nykh sobytii [Some problems of the analysis of catastrophic risks related to nonhomogeneous flows of extremal events]. *Sistemy i sredstva informatiki. Spetsial'nyy vypusk "Matematicheskie metody v informatsionnykh tekhnologiyakh"* [Systems and means of informatics. Special issue "Mathematical methods and models of informatics"]. Moscow: IPI RAN. 109–125.
2. Korolev, V. Yu., I. A. Sokolov, A. S. Gordeev, M. E. Grigor'eva, S. V. Popov, and N. A. Chebonenko. 2006. Nekotorye metody analiza vremennykh kharakteristik katastrof v neodnorodnykh potokakh ekstremal'nykh sobytii [Some methods for the analysis of temporal characteristics of catastrophes in nonhomogeneous flows of extremal events]. *Sistemy i sredstva informatiki. Spetsial'nyy vypusk "Matematicheskie metody v informatsionnykh tekhnologiyakh"* [Systems and means of informatics. Special issue "Mathematical methods in information technologies"]. Moscow: IPI RAN. 5–23.
3. Korolev, V. Yu., I. A. Sokolov, A. S. Gordeev, M. E. Grigor'eva, S. V. Popov, and N. A. Chebonenko. 2007. Nekotorye metody prognozirovaniya vremennykh kharakteristik riskov, svyazannykh s katastroficheskimi sobyitiyami [Some methods for the prediction of the temporal characteristics of risks related to catastrophic events]. *Aktuariy [Actuary]* 1:34–40.
4. Korolev, V. Yu., and I. A. Sokolov. 2008. *Matematicheskie modeli neodnorodnykh potokov ekstremal'nykh sobytii* [Mathematical models of nonhomogeneous flows of extremal events]. Moscow: TORUS PRESS. 200 p.
5. Korolev, V. Yu., and S. Ya. Shorgin. 2011. *Matematicheskie metody analiza stokhasticheskoy struktury informatsionnykh potokov* [Mathematical methods for the analysis of the

- stochastic structure of information flows*]. Moscow: IPI RAN. 130 p.
6. Korolev, V. Yu., A. V. Chertok, A. Yu. Korchagin, and A. K. Gorshenin. 2013. Veroyatnostno-statisticheskoe modelirovanie informatsionnykh potokov v slozhnykh finansovykh sistemakh na osnove vysokochastotnykh dannyykh [Probability and statistical modeling of information flows in complex financial systems based on high-frequency data]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):12–21.
 7. Kalashnikov, V. 1997. *Geometric sums: Bounds for rare events with applications*. Dordrecht–Boston–London: Kluwer Academic Publs. 288 p.
 8. Korolev, V. Yu., V. E. Bening, and S. Ya. Shorgin. 2011. *Matematicheskie osnovy teorii riska [Mathematical foundations of risk theory]*. 2nd ed. Moscow: Fizmatlit. 620 p.
 9. Korolev, V. Yu. 1994. Convergence of random sequences with the independent random indices. I. *Theory Probab. Appl.* 39(2):282–297.
 10. Balkema, A., and L. de Haan. 1974. Residual life time at great age. *Ann. Probab.* 2:792–804.
 11. Pickands, J. 1975. Statistical inference using extreme order statistics. *Ann. Stat.* 3:119–131.
 12. Atkinson, A. 1999. *Impact Earth: Asteroids, comets and meteors — the growing threat*. Virgin Publ. 256 p.

Received October 20, 2013

Contributors

Grigorieva Maria E. (b. 1986) — biostatistician II, Parexel International, Moscow 121609, Russian Federation; maria-grigoryeva@yandex.ru

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; Moscow 119991, Russian Federation; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; victoryukorolev@yandex.ru

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; isokolov@ipiran.ru

РАСПРЕДЕЛЕНИЕ ВРЕМЕНИ ВЫХОДА ИЗ МНОЖЕСТВА СОСТОЯНИЙ ПЕРЕГРУЗКИ В СИСТЕМЕ $M|M|1|\langle L, H \rangle|\langle H, R \rangle$ С ГИСТЕРЕЗИСНЫМ УПРАВЛЕНИЕМ НАГРУЗКОЙ*

Ю. В. Гайдамака¹, А. В. Печинкин², Р. В. Разумчик³, А. К. Самуйлов⁴, К. Е. Самуйлов⁵, И. А. Соколов⁶, Э. С. Сопин⁷, С. Я. Шоргин⁸

Аннотация: Одним из наиболее простых в реализации и эффективных решений проблемы перегрузок, обеспечивающим наименьшее число переключений режимов функционирования системы, является гистерезисное управление нагрузкой. В статье предложен аналитический метод исследования параметров гистерезисного управления. В качестве математической модели рассмотрена система массового обслуживания (СМО) $M|M|1|\langle L, H \rangle|\langle H, R \rangle$ с двумя петлями гистерезисного управления, где H — порог обнаружения перегрузки; L — порог снижения перегрузки; R — порог сброса нагрузки. Получены два метода вычисления преобразования Лапласа–Стилтьеса (ПЛС) времени возврата системы из множества состояний перегрузки в множество состояний нормальной нагрузки: первый — путем решения системы уравнений с ПЛС неизвестных времен возврата для каждого состояния перегрузки; второй — с помощью рекуррентного представления ПЛС времен возврата в виде дробно-рациональных функций. Оба метода позволяют при вычислениях эффективно применять инструментальные программные средства общего назначения, что показано на численном примере.

Ключевые слова: перегрузка сервера; система массового обслуживания (СМО); гистерезисное управление нагрузкой; время возврата в множество состояний нормальной нагрузки; преобразование Лапласа–Стилтьеса (ПЛС); функция распределения

DOI: 10.14357/19922264130403

1 Введение

Проблема защиты жизненно важных узлов телекоммуникационной сети от перегрузок вновь стала критичной в сетях связи последующих поколений (NGN, Next Generation Network). В сетях второго поколения (2G) с коммутацией каналов перегрузки были вызваны, главным образом, поведением пользователей, порождающим взрывной рост трафика в часы наивысшей нагрузки. В современных сетях 3G и 4G основной причиной непредсказуемого по объему трафика, с обработкой которого не справляется самое современное оборудование, стал в первую очередь стремительный рост числа телекоммуникационных услуг, характеризующихся высокими требованиями к производительности сетевых узлов и серверов различного назначения. Примером проявления проблемы являются пере-

грузки серверов протокола SIP (Session Initiation Protocol), порождаемые лавинообразным потоком запросов пользователей на предоставление широкополосных услуг [1]. Так, при установлении соединения по протоколу SIP даже в простейшем случае установления речевого соединения один запрос пользователя требует передачи и обработки несколькими серверами до семи сообщений протокола SIP.

Различные сценарии перегрузок SIP-серверов описаны в документах Рабочей группы по инженерным проблемам сети Интернет (IETF, Internet Engineering Task Force), играющих роль международных стандартов [1–5] для сетей на базе протокола IP. Типичным проявлением перегрузки служит лавинный перезапуск, который происходит, когда слишком много пользователей одновременно пытаются зарегистрироваться на SIP-серверах. Примером служит

* Работа выполнена при поддержке РФФИ (проекты №№ 11-07-00112 и 12-07-00108).

¹ Российский университет дружбы народов, ygaidamaka@sci.pfu.edu.ru

² Институт проблем информатики Российской академии наук, apchinkin@ipiran.ru

³ Институт проблем информатики Российской академии наук, rrazumchik@ieee.org

⁴ Российский университет дружбы народов, asam1988@gmail.com

⁵ Российский университет дружбы народов, ksam@sci.pfu.edu.ru

⁶ Институт проблем информатики Российской академии наук, isokolov@ipiran.ru

⁷ Российский университет дружбы народов, sopin-eduard@yandex.ru

⁸ Институт проблем информатики Российской академии наук, sshorgin@ipiran.ru

сценарий так называемого «манхэттенского перезапуска» (англ. Manhattan reboots scenario), когда в результате аварии произошло отключение электричества в этом крупнейшем районе города и после восстановления электроснабжения все SIP-терминалы одновременно пытались зарегистрироваться на серверах, создав тем самым большой поток сообщений REGISTER.

Для успешного управления перегрузками, по сути, требуется ответить на два вопроса: как определить начало перегрузки и как ее устранить. Наиболее естественным решением является введение порогового управления очередью сообщений (заявок) на обработку сервером подобно тому, как это было сделано в рекомендациях Международного союза электросвязи (ITU, International Telecommunications Union) для протоколов сетевого и канального уровней общеканальной системы сигнализации № 7 (ОКС7) [6].

В [7–9] был сделан краткий обзор и анализ механизма гистерезисного управления нагрузкой, применяемого в ОКС7, а также разработана математическая модель локального управления перегрузками в сети SIP-серверов. Управление перегрузками осуществляется путем введения трех порогов в очереди на обработку сигнальных сообщений сервером — порога H обнаружения перегрузки, порога L снижения перегрузки и порога R сброса нагрузки. Пока общее число сообщений в очереди не превышает $(H - 1)$, сервер функционирует в режиме нормальной нагрузки. Если длина очереди стала равной H , система переходит в режим снижения нагрузки и остается в этом режиме до тех пор, пока длина очереди не достигнет значения $(L - 1)$ или $(R - 1)$. При уменьшении длины очереди до значения $(L - 1)$ система возвращается в режим нормальной нагрузки. Если очередь увеличилась до значения R , включается режим сброса нагрузки, в котором система находится до тех пор, пока длина очереди не станет равной H , после чего система возвращается в режим снижения нагрузки. Значения порогов выбираются так, что $0 < L < H < R$, и поэтому между парами порогов $\langle L, H \rangle$ и $\langle H, R \rangle$ возникает так называемый эффект гистерезиса [10, 11] в виде двух петель — по одной на каждую пару порогов. В ОКС7 гистерезисное управление было введено для сокращения числа переключений системы управления из режима перегрузки в режим нормальной нагрузки [12, 13], при этом задача решалась путем выбора значений порогов с целью минимизации среднего времени возврата системы из режима перегрузки в режим нормальной нагрузки.

Исследованию СМО с гистерезисным управлением посвящено достаточно много работ, при-

чем чаще всего встречаются работы по системам с гистерезисным обслуживанием [14–16] и реже встречаются работы по системам с гистерезисным управлением входящим потоком заявок [17, 18]. Объемный обзор результатов по гистерезисному управлению содержится в работах [19, 20], а наиболее близки к исследованиям настоящей статьи модель и методы исследования, разработанные в [21], где также можно найти обширный список источников по проблеме анализа СМО с гистерезисным управлением интенсивностью входящего потока заявок (далее для краткости — с гистерезисным управлением нагрузкой). В работах [7–9] с участием части авторов данной статьи проведен обзор работ, посвященных математическому и имитационному моделированию систем с гистерезисным управлением нагрузкой. В этих же статьях, а также в [22–32] были построены и исследованы марковские модели в виде СМО с пуассоновским входящим потоком и экспоненциально распределенной длительностью обслуживания заявок.

С точки зрения показателей качества обслуживания SIP-сервера интерес представляет время перехода случайного процесса, описывающего функционирование системы, из множества состояний перегрузки и сброса нагрузки в множество состояний нормальной нагрузки. Эту случайную величину принято называть временем возврата в режим нормальной нагрузки, или для краткости — временем возврата, а ее характеристики, такие как математическое ожидание или 95%-ная квантиль, подлежат минимизации при заданных ограничениях на нагрузочные и структурные параметры системы. В [32] для марковского случая получен алгоритм расчета среднего времени возврата и численно решена задача его минимизации для близких к реальным исходных данных. В данной статье, в отличие от известных результатов, предложен метод расчета ПЛС времени возврата, что позволяет вычислять функцию распределения (ФР) и находить квантили этой величины, не прибегая к средствам имитационного моделирования. Кроме того, результаты промежуточных вычислений позволяют эффективно проводить анализ и других важных вероятностно-временных характеристик исследуемой СМО.

Для решения задачи были разработаны два метода. Первый метод основан на решении системы уравнений, где неизвестными являются ПЛС времени возврата из каждого состояния множеств перегрузки и сброса нагрузки. Второй метод основан на рекуррентном представлении ПЛС времени возврата через дробно-рациональные функции. В обоих случаях получены алгоритмы вычисления ПЛС времени возврата в таком виде, что они могут быть достаточно просто запрограммированы, например, с

помощью систем символьных вычислений, встроенных в известные инструментальные программные средства, такие как MATLAB [33], Mathematica [34], Maple [35] и т. п.

Статья организована следующим образом. В разд. 2 детально описана исследуемая модель СМО, введены все необходимые понятия и обозначения. В разд. 3 предложены два метода вычисления ПЛС времени возврата. В разд. 4 представлен пример численного анализа, а в заключении подведены итоги статьи и обсуждаются полученные результаты.

2 Постановка задачи

Приведем детальное описание СМО, которая будет рассматриваться далее. Система представляет собой однолинейную СМО с двумя типами заявок. Входящие потоки заявок являются независимыми пуассоновскими с интенсивностью $\lambda_k, k = 1, 2$, поступления заявок k -го типа. Через $\lambda = \lambda_1 + \lambda_2$ обозначим суммарную интенсивность входящего потока. Время обслуживания заявки любого типа распределено по экспоненциальному закону с параметром μ .

Процесс обслуживания происходит следующим образом. Пусть в начальный момент в системе отсутствуют заявки. Тогда до того момента, когда в системе впервые окажется H заявок, к обслуживанию принимаются заявки обоих типов. Но как только число заявок становится равным H , прекращается прием заявок второго типа и принимаются только заявки первого типа. Так продолжается до того момента, когда число заявок в системе не становится равным $L - 1$ или R . В первом случае снова начинается прием заявок второго типа, а во втором — прекращается прием всех заявок (в том числе и первого типа), причем прием заявок первого типа

возобновляется в тот момент, когда число заявок в системе становится равным H . Схематическое изображение функционирования рассматриваемой СМО приведено на рис. 1.

Введем обозначения:

$\tilde{V}_n(s), n = \overline{0, H-1}$, — ПЛС времени до того момента, когда в системе впервые окажется H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки любого типа;

$V_n^*(s), n = \overline{H+1, R}$, — ПЛС времени до того момента, когда в системе впервые останется $(L-1)$ заявок, при условии, что в начальный момент в системе было n заявок и не принимались заявки (любого типа);

$V_n(s), n = \overline{L, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется $(L-1)$ заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа.

Заметим, что $V_H(s)$ представляет собой не что иное, как выраженное в терминах ПЛС распределение времени возврата (в режим нормальной нагрузки) — одного из основных используемых на практике показателей качества обслуживания SIP-сервера, о чем уже говорилось во введении. Далее, $\tilde{V}_{L-1}(s)$ — ПЛС времени от момента попадания системы в состояние с $(L-1)$ заявками до момента первого после этого попадания в состояние с H заявками, или времени пребывания функционирующего в стационарном режиме SIP-сервера в множестве состояний нормальной нагрузки. И, наконец, $T(s) = \tilde{V}_{L-1}(s)V_H(s)$ — ПЛС времени между соседними попаданиями системы в состояние с $(L-1)$ заявками с обязательным попаданием в состояние с H заявками, или цикла функционирования SIP-сервера, т. е. суммарное время пребывания

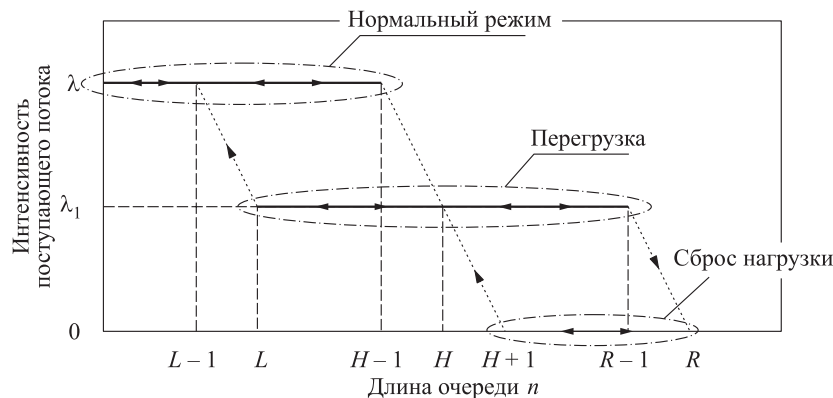


Рис. 1 Схематическое изображение функционирования системы

SIP-сервера в множествах состояний нормальной нагрузки и перегрузки.

Основным результатом настоящей статьи является решение следующих двух задач:

- (1) найти распределение времени до того момента, когда в системе впервые останется H заявок, при условии, что в начальный момент в системе было n , $n = \overline{0, H-1}$, заявок и принимались заявки любого типа, т. е. в терминах ПЛС найти $\tilde{V}_n(s)$, $n = \overline{0, H-1}$;
- (2) найти распределение времени до того момента, когда в системе впервые останется $(L-1)$ заявок, при условии, что в начальный момент в системе было n , $n = \overline{H+1, R}$, заявок и не принимались заявки (любого типа), или при условии, что было n заявок и принимались заявки только первого типа, т. е. в терминах ПЛС найти $V_n^*(s)$, $n = \overline{H+1, R}$, и $V_n(s)$, $n = \overline{L, R-1}$.

Очевидно, дифференцируя $\tilde{V}_n(s)$, $V_n^*(s)$ и $V_n(s)$ нужное число раз в точке $s = 0$, можно получить моменты любого порядка.

Найденное в терминах ПЛС распределение времени возврата используется далее для численных расчетов, причем предлагается два варианта формул. Первый вариант предпочтительно употреблять в том случае, когда обращение ПЛС производится по точкам. Второй вариант, дающий ответ в виде дробно-рациональной функции, удобен тогда, когда имеются хорошие программы разложения дробно-рациональной функции на сумму простейших дробей.

Отметим, что используемый в работе метод, более подробное изложение которого можно найти в [36], элементарно переносится на решение задачи нахождения ПЛС ФР времени перехода из любого состояния в любое другое. Кроме того, заметим, что $\tilde{V}_n(s)$ представляет собой не что иное, как ПЛС времени до первой потери заявки в системе $M|M|1|(H-2)$, исходя из состояния n , и хорошо известно (см., например, результат для более общего случая в [37]). Здесь выводы этого ПЛС приводятся для лучшего понимания дальнейших выкладок.

3 Вычисление преобразования Лапласа–Стилтьеса времени выхода

3.1 Вариант 1

Вычислим сначала ПЛС $\tilde{V}_n(s)$, $n = \overline{1, H-1}$. Эти ПЛС удовлетворяют уравнению:

$$\begin{aligned} \tilde{V}_n(s) &= \\ &= \frac{\mu + \lambda}{s + \mu + \lambda} \left(\frac{\mu}{\mu + \lambda} \tilde{V}_{n-1}(s) + \frac{\lambda}{\mu + \lambda} \tilde{V}_{n+1}(s) \right) = \\ &= \frac{1}{s + \mu + \lambda} \left[\mu \tilde{V}_{n-1}(s) + \lambda \tilde{V}_{n+1}(s) \right], \\ & \quad n = \overline{1, H-1}. \end{aligned} \quad (1)$$

Граничные условия для системы уравнений (1) определяются формулами:

$$\tilde{V}_0(s) = \frac{\lambda}{s + \lambda} \tilde{V}_1(s); \quad (2)$$

$$\tilde{V}_H(s) = 1. \quad (3)$$

Решение системы уравнений (1) имеет вид:

$$\tilde{V}_n(s) = \tilde{c}_1 \tilde{z}_1^{H-n}(s) + \tilde{c}_2 \tilde{z}_2^{H-n}(s), \quad n = \overline{0, H},$$

где $\tilde{z}_1 = \tilde{z}_1(s)$ и $\tilde{z}_2 = \tilde{z}_2(s)$ — решения уравнения

$$\lambda - (s + \mu + \lambda)z + \mu z^2 = 0,$$

т. е.

$$z_{1,2} = \frac{s + \mu + \lambda \pm \sqrt{(s + \mu + \lambda)^2 - 4\lambda\mu}}{2\mu},$$

а $\tilde{c}_1 = \tilde{c}_1(s)$ и $\tilde{c}_2 = \tilde{c}_2(s)$ определяются из граничных условий (2) и (3). Из (3) находим

$$\tilde{c}_2 = 1 - \tilde{c}_1$$

и, значит,

$$\begin{aligned} \tilde{V}_n(s) &= \tilde{c}_1 \tilde{z}_1^{H-n}(s) - (1 - \tilde{c}_1) \tilde{z}_2^{H-n}(s), \\ & \quad n = \overline{0, H-1}. \end{aligned} \quad (4)$$

Подставляя (4) в (2), получаем после элементарных преобразований

$$\begin{aligned} \tilde{c}_1 &= \left\{ [(s + \lambda)\tilde{z}_2(s) - \lambda]\tilde{z}_2^{H-1}(s) \right\} / \left\{ [(s + \lambda)\tilde{z}_1(s) - \right. \\ & \quad \left. - \lambda]\tilde{z}_1^{H-1}(s) + [(s + \lambda)\tilde{z}_2(s) - \lambda]\tilde{z}_2^{H-1}(s) \right\}. \end{aligned}$$

Для нахождения $V_n^*(s)$, $n = \overline{H+1, R}$, и $V_n(s)$, $n = \overline{L, R-1}$, введем сначала следующие вспомогательные функции:

$W_n^*(s)$, $n = \overline{H+1, R}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, при условии, что в начальный момент в системе было n заявок и не принимались заявки (любого типа);

$w_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, и вероятность того, что до этого момента в системе не было R заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа;

$\bar{w}_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые окажется R заявок, и вероятность того, что до этого момента в системе не было H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа.

Очевидно,

$$W_n^*(s) = \left(\frac{\mu}{s + \mu} \right)^{n-H}, \quad n = \overline{H+1, R}.$$

Функции $w_n(s)$, $n = \overline{H+1, R-1}$, удовлетворяют уравнению

$$w_n(s) = \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\mu}{\mu + \lambda_1} w_{n-1}(s) + \frac{\lambda_1}{\mu + \lambda_1} w_{n+1}(s) \right) = \frac{1}{s + \mu + \lambda_1} [\mu w_{n-1}(s) + \lambda_1 w_{n+1}(s)], \quad n = \overline{H+1, R-1}, \quad (5)$$

с граничными условиями

$$w_H(s) = 1; \quad (6)$$

$$w_R(s) = 0. \quad (7)$$

Решение системы уравнений (5) имеет вид:

$$w_n(s) = c_1 z_1^{R-n}(s) + c_2 z_2^{R-n}(s), \quad n = \overline{H, R},$$

где z_1, z_2 — решения уравнения

$$\lambda_1 - (s + \mu + \lambda_1)z + \mu z^2 = 0,$$

т. е.

$$z_{1,2} = \frac{s + \mu + \lambda_1 \pm \sqrt{(s + \mu + \lambda_1)^2 - 4\lambda_1\mu}}{2\mu}.$$

Из граничного условия (7) имеем

$$c_1 = -c_2 = c$$

и, следовательно,

$$w_n(s) = c [z_1^{R-n}(s) - z_2^{R-n}(s)], \quad n = \overline{H, R}.$$

Наконец, из граничного условия (6) получаем

$$c [z_1^{R-H}(s) - z_2^{R-H}(s)] = 1,$$

т. е.

$$c = \frac{1}{z_1^{R-H}(s) - z_2^{R-H}(s)}.$$

Далее, для ПЛС $\bar{w}_n(s)$, $n = \overline{H+1, R-1}$, справедливо уравнение

$$\begin{aligned} \bar{w}_n(s) &= \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\mu}{\mu + \lambda_1} \bar{w}_{n-1}(s) + \frac{\lambda_1}{\mu + \lambda_1} \bar{w}_{n+1}(s) \right) = \\ &= \frac{1}{s + \mu + \lambda_1} [\mu \bar{w}_{n-1}(s) + \lambda_1 \bar{w}_{n+1}(s)], \quad n = \overline{H+1, R-1}, \quad (8) \end{aligned}$$

с граничными условиями

$$\bar{w}_H(s) = 0; \quad (9)$$

$$\bar{w}_R(s) = 1. \quad (10)$$

Решение системы уравнений (8) имеет вид:

$$\bar{w}_n(s) = \bar{c}_1 \bar{z}_1^{n-H}(s) + \bar{c}_2 \bar{z}_2^{n-H}(s), \quad n = \overline{H, R},$$

где \bar{z}_1, \bar{z}_2 — решения уравнения

$$\mu - (s + \mu + \lambda_1)\bar{z} + \lambda_1\bar{z}^2 = 0,$$

т. е.

$$\begin{aligned} \bar{z}_{1,2} &= \frac{s + \mu + \lambda_1 \pm \sqrt{(s + \mu + \lambda_1)^2 - 4\lambda_1\mu}}{2\lambda_1} = \\ &= \frac{1}{z_{2,1}}. \quad (11) \end{aligned}$$

Из граничных условий (9) и (10) имеем

$$\bar{c}_1 = -\bar{c}_2 = \bar{c}$$

и

$$\bar{c} = \frac{1}{z_1^{R-H}(s) - z_2^{R-H}(s)}.$$

Введем теперь $W_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа. Тогда

$$W_n(s) = w_n(s) + \bar{w}_n(s)W_R^*(s), \quad n = \overline{H+1, R-1}.$$

Вычислим $V_n(s)$ при $n = \overline{L, H}$. Напомним, что $V_n(s)$, $n = \overline{L, H}$, — ПЛС времени до того момента, когда в системе впервые останется $(L-1)$ заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа. Преобразование Лапласа–Стилтьеса $V_n(s)$ при $n = \overline{L, H-1}$ удовлетворяет уравнению

$$\begin{aligned} V_n(s) &= \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\mu}{\mu + \lambda_1} V_{n-1}(s) + \frac{\lambda_1}{\mu + \lambda_1} V_{n+1}(s) \right) = \frac{1}{s + \mu + \lambda_1} [\mu V_{n-1}(s) + \\ &+ \lambda_1 V_{n+1}(s)], \quad n = \overline{L, H-1}, \quad (12) \end{aligned}$$

с граничными условиями

$$V_{L-1}(s) = 1; \quad (13)$$

$$V_H(s) = \frac{\mu + \lambda_1}{+\mu + \lambda_1} \cdot \left(\frac{\mu}{\mu + \lambda_1} V_{H-1}(s) + \frac{\lambda_1}{\mu + \lambda_1} W_{H+1}(s) V_H(s) \right),$$

второе из которых перепишем в виде

$$[s + \mu + \lambda_1 - \lambda_1 W_{H+1}(s)] V_H(s) = \mu V_{H-1}(s). \quad (14)$$

Решение системы уравнений (12) имеет вид:

$$V_n(s) = \bar{c}_1 \bar{z}_1^{n-L+1}(s) + \bar{c}_2 \bar{z}_2^{n-L+1}(s), \quad n = \overline{L-1, H},$$

где \bar{z}_1, \bar{z}_2 определяются формулой (11). Из граничных условий (13) и (14) имеем

$$\bar{c}_2 = 1 - \bar{c}_1,$$

$$\bar{c}_1 = \left(\mu \bar{z}_2^{H-L}(s) - [s + \mu + \lambda_1 - \lambda_1 W_{H+1}(s)] \bar{z}_2^{H-L+1}(s) \right) / \left([s + \mu + \lambda_1 - \lambda_1 W_{H+1}(s)] [\bar{z}_1^{H-L+1}(s) - \bar{z}_2^{H-L+1}(s)] - \mu [\bar{z}_1^{H-L}(s) - \bar{z}_2^{H-L}(s)] \right).$$

Теперь можно привести окончательные выражения для ПЛС $V_n^*(s)$ при $n = \overline{H+1, R}$ и ПЛС $V_n(s)$ при $n = \overline{H+1, R-1}$:

$$V_n^*(s) = W_n^*(s) V_H(s), \quad n = \overline{H+1, R};$$

$$V_n(s) = W_n(s) V_H(s), \quad n = \overline{H+1, R-1}.$$

3.2 Вариант 2

В этом подразделе будут найдены выражения для тех же самых ПЛС $\tilde{V}_n(s)$, $n = \overline{0, H-1}$, ПЛС $V_n^*(s)$, $n = \overline{H+1, R}$, и ПЛС $V_n(s)$, $n = \overline{L, R-1}$, что и в подразд. 3.1, но только в виде дробно-рациональных функций.

Как и прежде, начнем с $\tilde{V}_n(s)$.

Обозначим через $\tilde{w}_n(s)$, $n = \overline{0, H-1}$, ПЛС времени до того момента, когда в системе впервые окажется $(n+1)$ заявок, при условии, что в начальный момент в системе было n заявок и принимались все заявки. Тогда для $\tilde{w}_n(s)$, $n = \overline{0, H-1}$, справедливы соотношения:

$$\tilde{w}_0(s) = \frac{\lambda}{s + \lambda};$$

$$\tilde{w}_n(s) = \frac{\mu + \lambda}{s + \mu + \lambda} \left[\frac{\lambda}{\mu + \lambda} + \frac{\mu}{\mu + \lambda} \tilde{w}_{n-1}(s) \tilde{w}_n(s) \right] = \frac{1}{s + \mu + \lambda} [\lambda + \mu \tilde{w}_{n-1}(s) \tilde{w}_n(s)],$$

$$n = \overline{1, H-1}. \quad (15)$$

Из уравнения (15) получаем

$$\tilde{w}_n(s) = \frac{\lambda}{s + \mu + \lambda - \mu \tilde{w}_{n-1}(s)}, \quad n = \overline{1, H-1}.$$

Запишем ПЛС $\tilde{w}_n(s)$, $n = \overline{0, H-1}$, в виде дробно-рациональной функции

$$\tilde{w}_n(s) = \frac{\tilde{q}_n(s)}{\tilde{r}_n(s)}, \quad n = \overline{0, H-1}, \quad (16)$$

где $\tilde{q}_n(s)$, $n = \overline{0, H-1}$, — полином n -й степени, а $\tilde{r}_n(s)$, $n = \overline{0, H-1}$, — полином $(n+1)$ -й степени. Полиномы $\tilde{q}_n(s)$ и $\tilde{r}_n(s)$ можно последовательно вычислить из следующих соотношений:

$$\tilde{q}_0(s) = \lambda; \quad \tilde{r}_0(s) = s + \lambda;$$

$$\tilde{q}_n(s) = \lambda \tilde{r}_{n-1}(s), \quad n = \overline{1, H-1}; \quad (17)$$

$$\tilde{r}_n(s) = (s + \mu + \lambda) \tilde{r}_{n-1}(s) - \mu \tilde{q}_{n-1}(s), \quad n = \overline{1, H-1}.$$

Теперь можно привести формулу для $\tilde{V}_n(s)$, $n = \overline{0, H-1}$:

$$\tilde{V}_n(s) = \prod_{i=n}^{H-1} \tilde{w}_i(s), \quad n = \overline{0, H-1}. \quad (18)$$

Подставляя в (18) вместо ПЛС $\tilde{w}_i(s)$ его значение по формуле (16), а вместо ПЛС $\tilde{q}_i(s)$ — его значение по формуле (17), окончательно получаем

$$\tilde{V}_n(s) = \frac{\lambda^{H-n-1} \tilde{q}_n(s)}{\tilde{r}_{H-1}(s)}, \quad n = \overline{0, H-1}.$$

Перейдем к нахождению $V_n^*(s)$, $n = \overline{H+1, R}$, и $V_n(s)$, $n = \overline{L, R-1}$. Введем следующие вспомогательные функции:

$w_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется $(n-1)$ заявок, и вероятность того, что до этого момента в системе не было R заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа;

$\bar{w}_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые окажется $(n+1)$ заявок, и вероятность того, что до этого момента в системе не было H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа;

$W_n^*(s)$, $n = \overline{H+1, R}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, при условии, что в начальный момент в системе было n заявок и не принимались заявки (любого типа);

$\tilde{W}_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, и вероятность того, что до этого момента в системе не было R заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа;

$\overline{W}_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые окажется R заявок, и вероятность того, что до этого момента в системе не было H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа;

$W_n(s)$, $n = \overline{H+1, R-1}$, — ПЛС времени до того момента, когда в системе впервые останется H заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа.

Преобразование Лапласа–Стилтьеса $w_n(s)$, $n = \overline{H+1, R-1}$, удовлетворяет соотношению

$$w_n(s) = \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\mu}{\mu + \lambda_1} + \frac{\lambda_1}{\mu + \lambda_1} w_{n+1}(s) w_n(s) \right), \quad n = \overline{H+1, R-1},$$

из которого имеем

$$w_n(s) = \frac{\mu}{s + \mu + \lambda_1 - \lambda_1 w_{n+1}(s)}, \quad n = \overline{H+1, R-1},$$

где положено

$$w_R(s) = 0.$$

Отсюда получаем следующие соотношения для последовательного вычисления полиномов $q_n(s)$ и $r_n(s)$ в представлении $w_n(s) = q_n(s)/r_n(s)$:

$$\begin{aligned} q_{R-1}(s) &= \mu; \\ r_{R-1}(s) &= s + \mu + \lambda_1; \\ q_n(s) &= \mu r_{n+1}(s), \quad n = \overline{H+1, R-2}, \\ r_n(s) &= (s + \mu + \lambda_1) r_{n+1}(s) - \lambda_1 q_{n+1}(s), \\ &\quad n = \overline{H+1, R-2}. \end{aligned}$$

Преобразование Лапласа–Стилтьеса $\overline{w}_n(s)$, $n = \overline{H+1, R-1}$, удовлетворяет соотношению

$$\overline{w}_n(s) = \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\lambda_1}{\mu + \lambda_1} + \frac{\mu}{\mu + \lambda_1} \overline{w}_{n-1}(s) \overline{w}_n(s) \right), \quad n = \overline{H+1, R-1},$$

что дает равенство

$$\overline{w}_n(s) = \frac{\lambda_1}{s + \mu + \lambda_1 - \mu \overline{w}_{n-1}(s)}, \quad n = \overline{H+1, R-1},$$

где положено

$$\overline{w}_H(s) = 0.$$

Это приводит к следующим выражениям для полиномов $\overline{q}_n(s)$ и $\overline{r}_n(s)$ в представлении $\overline{w}_n(s) = \overline{q}_n(s)/\overline{r}_n(s)$:

$$\begin{aligned} \overline{q}_{H+1}(s) &= \lambda_1; \\ \overline{r}_{H+1}(s) &= s + \mu + \lambda_1; \\ \overline{q}_n(s) &= \lambda_1 \overline{r}_{n-1}(s), \quad n = \overline{H+2, R-1}; \\ \overline{r}_n(s) &= (s + \mu + \lambda_1) \overline{r}_{n-1}(s) - \mu \overline{q}_{n-1}(s), \\ &\quad n = \overline{H+2, R-1}. \end{aligned}$$

Далее,

$$W_n^*(s) = \left(\frac{\mu}{s + \mu} \right)^{n-H}, \quad n = \overline{H+1, R};$$

ПЛС $\tilde{W}_n(s)$, $n = \overline{H+1, R-1}$, и $\overline{W}_n(s)$, $n = \overline{H+1, R-1}$, определяются выражениями:

$$\tilde{W}_n(s) = \prod_{i=H+1}^n w_i(s) = \frac{\mu^{n-H-1} q_{H+1}(s)}{r_n(s)}, \quad n = \overline{H+1, R-1};$$

$$\overline{W}_n(s) = \prod_{i=n}^{R-1} \overline{w}_i(s) = \frac{\lambda_1^{R-n-1} \overline{q}_{R-1}(s)}{\overline{r}_n(s)}, \quad n = \overline{H+1, R-1},$$

а ПЛС $W_n(s)$, $n = \overline{H+1, R-1}$, имеет вид:

$$\begin{aligned} W_n(s) &= \tilde{W}_n(s) + \overline{W}_n(s) W_R^*(s) = \\ &= \frac{\mu^{n-H-1} q_{H+1}(s)}{r_n(s)} + \frac{\lambda_1^{R-n-1} \mu^{R-H} \overline{q}_{R-1}(s)}{(s + \mu)^{R-H} \overline{r}_n(s)} = \\ &= \frac{Q_n(s)}{R_n(s)}, \quad n = \overline{H+1, R-1}, \end{aligned}$$

где $Q_n(s)$ и $R_n(s)$ находятся по формулам

$$Q_n(s) = (s + \mu)^{R-H} \mu^{n-H-1} q_{H+1}(s) \overline{r}_n(s) + \lambda_1^{R-n-1} \mu^{R-H} \overline{q}_{R-1}(s) r_n(s), \quad n = \overline{H+1, R-1};$$

$$R_n(s) = (s + \mu)^{R-H} r_n(s) \overline{r}_n(s), \quad n = \overline{H+1, R-1}.$$

Обозначим теперь через $W_n(s)$, $n = \overline{L, H}$, ПЛС времени до того момента, когда в системе впервые останется $(n - 1)$ заявок, при условии, что в начальный момент в системе было n заявок и принимались заявки только первого типа. Тогда справедливо уравнение:

$$W_n(s) = \frac{\mu + \lambda_1}{s + \mu + \lambda_1} \left(\frac{\lambda_1}{\mu + \lambda_1} W_{n+1}(s)W_n(s) + \frac{\mu}{\mu + \lambda_1} \right), \quad n = \overline{L, H},$$

из которого находим

$$W_n(s) = \frac{\mu}{s + \mu + \lambda_1 - \lambda_1 W_{n+1}(s)}, \quad n = \overline{L, H}.$$

Таким образом, имеем следующие выражения для полиномов $Q_n(s)$ и $R_n(s)$ в представлении $W_n(s) = Q_n(s)/R_n(s)$, $n = \overline{L, H}$:

$$\begin{aligned} Q_n(s) &= \mu R_{n+1}(s), \quad n = \overline{L, H}; \\ R_n(s) &= (s + \mu + \lambda_1)R_{n+1}(s) - \lambda_1 Q_{n+1}(s), \\ & \quad n = \overline{L, H}. \end{aligned}$$

Окончательно получаем:

$$\begin{aligned} V_n(s) &= \prod_{i=L}^n W_i(s) = \frac{\mu^{n-L} Q_n(s)}{R_L(s)}, \quad n = \overline{L, H}; \\ V_n^*(s) &= W_n^*(s) V_H(s), \quad n = \overline{H+1, R}; \\ V_n(s) &= W_n(s) V_H(s), \quad n = \overline{H+1, R-1}. \end{aligned}$$

4 Вычислительный алгоритм

Полученные в предыдущем разделе математические соотношения были использованы для численных расчетов, причем в качестве рассчитываемой характеристики была взята наиболее интересная с точки зрения показателей качества обслуживания SIP-сервера ФР времени возврата, в том числе математическое ожидание и 95%-ная квантиль времени возврата.

В этом разделе приведем алгоритм вычисления ПЛС $V_H(s)$ времени возврата только для варианта 2 (см. подразд. 3.2), поскольку проведенные расчеты показали, что и точность, и трудоемкость обоих вариантов для всех предложенных исходных данных практически совпали. Кроме того, так как расчеты проводились в среде MATLAB с использованием символьных вычислений, то действия над многочленами производились автоматически. Поэтому приводимый ниже алгоритм не содержит формул для определения многочленов дробно-рациональных представлений ПЛС $V_H(s)$ и промежуточных

рассчитываемых ПЛС. Далее, поскольку алгоритм предназначен для вычисления только характеристик, связанных с временем возврата, здесь не представлены формулы для нахождения других параметров, хотя эти формулы практически не отличаются от приведенных ниже. Наконец, в рамках шагов предложенного алгоритма удобно показать также, как вычисляется математическое ожидание $T = -V_H'(0)$ времени возврата.

Алгоритм состоит из следующих шагов.

Шаг 1. Последовательно по n от $n = R - 1$ до $n = H + 1$ вычисляются ПЛС $w_n(s)$ и производные $w_n'(0)$ по формулам:

$$\begin{aligned} w_{R-1}(s) &= \frac{\mu}{s + \mu + \lambda_1}; \\ w_n(s) &= \frac{\mu}{s + \mu + \lambda_1 - \lambda_1 w_{n+1}(s)}, \quad n = \overline{H+1, R-2}; \\ w_{R-1}'(0) &= -\frac{\mu}{(\mu + \lambda_1)^2}; \\ w_n'(0) &= -\frac{\mu[1 - \lambda_1 w_{n+1}'(0)]}{[\mu + \lambda_1 - \lambda_1 w_{n+1}(0)]^2}, \quad n = \overline{H+1, R-2}. \end{aligned}$$

Далее понадобятся только $w_{H+1}(s)$ и $w_{H+1}'(0)$.

Шаг 2. Последовательно по n от $n = H + 1$ до $n = R - 1$ вычисляются ПЛС $\bar{w}_n(s)$ и производные $\bar{w}_n'(0)$ по формулам:

$$\begin{aligned} \bar{w}_{H+1}(s) &= \frac{\lambda_1}{s + \mu + \lambda_1}; \\ \bar{w}_n(s) &= \frac{\lambda_1}{s + \mu + \lambda_1 - \mu \bar{w}_{n-1}(s)}, \quad n = \overline{H+2, R-1}; \\ \bar{w}_{H+1}'(0) &= -\frac{\lambda_1}{(\mu + \lambda_1)^2}; \\ \bar{w}_n'(0) &= -\frac{\lambda_1[1 - \mu \bar{w}_{n-1}'(0)]}{[\mu + \lambda_1 - \mu \bar{w}_{n-1}(0)]^2}, \quad n = \overline{H+2, R-1}. \end{aligned}$$

Шаг 3. Вычисляются ПЛС $W_R^*(s)$ и производная $W_R^{*'}(0)$ по формулам:

$$\begin{aligned} W_R^*(s) &= \left(\frac{\mu}{s + \mu} \right)^{R-H}; \\ W_R^{*'}(0) &= -\frac{R-H}{\mu}. \end{aligned}$$

Шаг 4. Вычисляются ПЛС $\bar{W}_{H+1}(s)$ и производная $\bar{W}_{H+1}'(0)$ по формулам:

$$\begin{aligned} \bar{W}_{H+1}(s) &= \prod_{i=H+1}^{R-1} \bar{w}_i(s); \\ \bar{W}_{H+1}'(0) &= \bar{W}_{H+1}(0) \sum_{i=H+1}^{R-1} \frac{\bar{w}_i'(0)}{\bar{w}_i(0)}. \end{aligned}$$

Шаг 5. Вычисляются ПЛС $W_{H+1}(s)$ и производная $W'_{H+1}(0)$ по формулам:

$$W_{H+1}(s) = w_{H+1}(s) + \overline{W}_{H+1}(s)W_R^*(s);$$

$$W'_{H+1}(0) = w'_{H+1}(0) + \overline{W}'_{H+1}(0)W_R^*(0) + \overline{W}_{H+1}(0)W_R^{*'}(0).$$

Шаг 6. Последовательно по n от $n = H$ до $n = L$ вычисляются ПЛС $W_n(s)$ и производные $W'_n(0)$ по формулам:

$$W_n(s) = \frac{\mu}{s + \mu + \lambda_1 - \lambda_1 W_{n+1}(s)}, \quad n = \overline{L, H};$$

$$W'_n(0) = -\frac{\mu[1 - \lambda_1 W'_{n+1}(0)]}{[\mu + \lambda_1 - \lambda_1 W_{n+1}(0)]^2}, \quad n = \overline{L, H}.$$

Шаг 7. Вычисляются ПЛС $V_H(s)$ и производная $V'_H(0)$ по формулам:

$$V_H(s) = \prod_{i=L}^H W_i(s),$$

$$T = -V'_H(0) = -\sum_{i=L}^H W'_i(0).$$

Дальнейшие шаги, связанные с вычислением плотности распределения времени возврата с помощью обратного преобразования Лапласа, а затем ФР и 95%-ной квантили времени возврата зависят от используемого программного обеспечения. Приведенный выше алгоритм удобно реализовать в среде MATLAB, используя встроенные возможности символьных вычислений, интегрирования и др.

5 Пример расчетов

Приведем пример численного анализа характеристик времени возврата для одного из возможных

наборов исходных данных, разработанного авторами данной статьи в рамках решения прикладной задачи по анализу показателей качества обслуживания протокола SIP [8]. Рассматривается случай, когда $L = 74$, $H = 85$, $R = 100$, суммарная интенсивность входящего потока $\lambda = 240$ заявок/с, причем интенсивность λ_2 потока заявок второго типа связана с λ следующим образом: $\lambda_2 = q\lambda$. Здесь q — вероятность сброса заявки, т. е. вероятность поступления в систему заявки второго типа. Среднее время обслуживания заявки любого типа $\mu^{-1} = 5$ мс.

На рис. 2, а приведен график плотности распределения $p(x)$ времени возврата, построенный для значений $q = 0,3; 0,5; 0,7; 0,9$. График ФР $F(x)$ времени возврата построен на рис. 2, б для тех же самых значений q . Наконец, на рис. 3 показан график 95%-ной квантили $t_{0,95}$ и среднего значения T времени возврата в зависимости от величины вероятности сброса $q \in [0,3, 0,9]$.

Из графиков видно, что с ростом вероятности q сброса заявки время возврата уменьшается, причем его среднее значение T всегда меньше значения $t_{0,95}$ 95%-ной квантили. В [8] численно решалась задача нахождения значений порогов L и H таких, что при фиксированных значениях всех остальных параметров среднее время возврата является минимальным. Полученные в данной статье алгоритмы позволяют решать более сложные задачи оптимизации параметров гистерезисного управления нагрузкой SIP-сервера, в том числе задачу минимизации 95%-ной квантили времени возврата.

При расчетах было отмечено, что наиболее трудоемко вычисление значений 95%-ной квантили, оно может занять до 30 мин машинного времени на компьютере с процессором Intel Core i5 @ 3,30 ГГц и 8 ГБ оперативной памяти. Численный анализ и результаты имитационного моделирования пока-

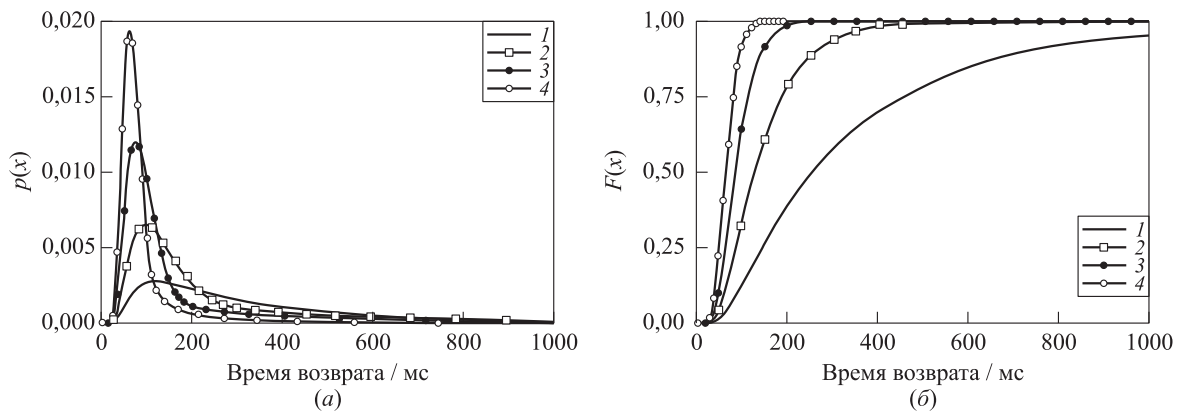


Рис. 2 Плотность распределения $p(x)$ (а) и функция распределения $F(x)$ (б) времени возврата: 1 — $q = 0,3$; 2 — $0,5$; 3 — $0,7$; 4 — $q = 0,9$

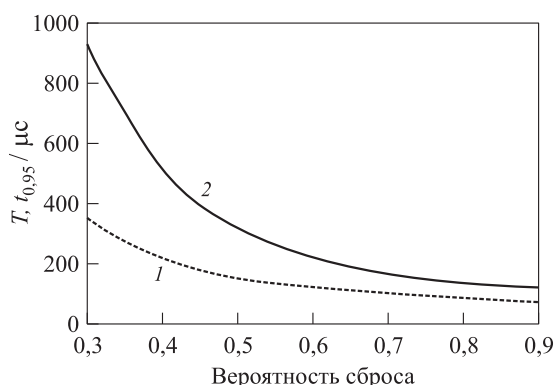


Рис. 3 Среднее значение T времени возврата (1) и 95%-ная квантиль $t_{0,95}$ (2)

зали достоверность обоих вариантов метода вычислений ПЛС ФР времени возврата для исследуемой СМО.

6 Заключение

В статье предложены новые аналитические методы исследования СМО с гистерезисным управлением входящей нагрузкой. По сравнению с известными ранее результатами получен эффективный метод, предназначенный для анализа и расчета не только математического ожидания времени возврата системы из состояний перегрузки в состояние нормальной нагрузки, но и других его характеристик, в том числе ФР и квантилей. Отметим, что 95%-ную квантиль рекомендовано использовать как показатель качества телекоммуникационных систем международными стандартизирующими организациями, такими как ИТУ и IETF. Разработка программных средств в среде MATLAB и численный анализ на близких к реальным исходных данных показали эффективность разработанных в статье методов, предназначенных для вычисления ПЛС ФР времени возврата, причем время вычисления обоими способами оказалось практически одинаковым, а сложность вычислений такова, что для проведения вычислительного эксперимента достаточно использования персонального компьютера с процессором Intel Core i5 @ 3,30 ГГц и 8 ГБ оперативной памяти. Дальнейшие исследования будут направлены на решение задач нахождения оптимальных пороговых значений для управления перегрузками SIP-серверов.

Литература

- Hilt V., Noel E., Shen C., Abdelal A. Design considerations for Session Initiation Protocol (SIP) overload control // Internet Engineering Task Force RFC-6357, 2011. [Электронный ресурс] Режим доступа: <http://tools.ietf.org/html/rfc6357>, свободный (дата обращения 01.10.2013).
- Hilt V., Rosenberg J., Schulzrinne H., et al. SIP: Session Initiation Protocol // Internet Engineering Task Force RFC-3261, 2002. [Электронный ресурс] Режим доступа: <http://tools.ietf.org/html/rfc3261>, свободный (дата обращения 01.10.2013).
- Rosenberg J. Requirements for management of overload in the Session Initiation Protocol // Internet Engineering Task Force RFC-5390, 2008. [Электронный ресурс] Режим доступа: <http://tools.ietf.org/html/rfc5390>, свободный (дата обращения 01.10.2013).
- Gurbani V., Hilt V., Schulzrinne H. Session Initiation Protocol overload control // Internet Engineering Task Force Draft, 2013. [Электронный ресурс] Режим доступа: <http://tools.ietf.org/pdf/draft-ietf-soc-overload-control-13.pdf>, свободный (дата обращения 01.10.2013).
- Noel E., Williams P.M. Session Initiation Protocol rate control // Internet Engineering Task Force Draft, 2013. [Электронный ресурс] Режим доступа: <http://tools.ietf.org/pdf/draft-ietf-soc-overload-rate-control-05.pdf>, свободный (дата обращения 01.10.2013).
- ITU-T Recommendation Q.704. 1996. Signalling System No.7 — Message Transfer Part, Signalling network functions and messages. [Электронный ресурс] Режим доступа: <http://www.itu.int/rec/T-REC-Q.704-199607-1/en>, свободный (дата обращения 01.10.2013).
- Абаев П. О., Гайдамака Ю. В., Самуйлов К. Е. Гистерезисное управление сигнальной нагрузкой в сети SIP-серверов // Вестник РУДН. Серия Математика. Информатика. Физика, 2011. № 4. С. 55–73.
- Абаев П. О., Гайдамака Ю. В., Печинкин А. В., Разумчик Р. В., Шоргин С. Я. Simulation of overload control in SIP server networks // 26th Conference (European) on Modelling and Simulation ECMS Proceedings. — Koblenz, 2012. P. 533–539.
- Абаев П., Гайдамака Ю., Самуйлов К. Queuing model for loss-based overload control in a SIP server using a hysteretic technique // Internet of things, smart spaces, and next generation networking / Eds. S. Andreev, S. Balandin, Ye. Koucheryavy. — Lecture notes in computer science ser. — Heidelberg: Springer-Verlag, 2012. Vol. 7469. P. 371–378.
- Gebhart R. F. A queuing process with bilevel hysteretic service-rate control // Nav. Res. Logist. Q., 1967. Vol. 14. P. 55–68.
- Красносельский М. А., Покровский А. В. Системы с гистерезисом. — М.: Наука, 1983. 272 с.
- Yum T., Yen H. Design algorithm for a hysteresis buffer congestion control strategy // IEEE Conference (International) on Communications Proceedings, 1983. P. 499–503.
- Brown P., Chemouil P., Delosme B. A congestion control policy for signalling networks // 7th IeCC Proceedings, 1984. P. 717–724.

14. Golubchik L., Lui J. C. S. Bounding of performance measures for a threshold-based queueing system with hysteresis // *Newsl. ACM SIGMETRICS Performance Evaluation Rev.*, 1997. Vol. 25. No. 1. P. 147–157.
15. Sindal R., Tokekar S. Modeling and analysis of voice/data call admission control scheme in CDMA cellular network for variation in soft handoff threshold parameters // 16th IEEE Conference (International) on Networks (ICON 2008) Proceedings. P. 1–6.
16. Жерновский К. Ю., Жерновский Ю. В. Система $M^0/G/1$ с гистерезисным переключением интенсивности обслуживания // *Информационные процессы*, 2012. Т. 12. № 3. С. 176–190.
17. Takagi H. Analysis of a finite-capacity $M/G/1$ queue with a resume level // *Perform. Evaluation*, 1985. Vol. 5. P. 197–203.
18. Benaboud H., Mikou N. Analysis by queueing model of multi-threshold mechanism in ATM switches // 5th IEEE Conference (International) on High Speed Networks and Multimedia Communications (HSNMC 2002) Proceedings. P. 147–151.
19. Dshalalow J. H. Queues with state dependent parameters // *Frontiers in queueing: Models and applications in science and engineering* / Ed. J. H. Dshalalow. — Probability and stochastic ser. — CRC Press, 1997. P. 61–116.
20. Bekker R. Queues with Levy input and hysteretic control // *Queueing Syst.*, 2009. Vol. 63. No. 1. P. 281–299.
21. Roughan M., Pearce C. A martingale analysis of hysteretic overload control // *Adv. Perform. Anal. J. Teletraffic Theory Perform. Anal. Communication Syst. Networks*, 2000. Vol. 3. No. 1. P. 1–30.
22. Abaev P., Gaidamaka Yu., Samouylov K. Modeling of hysteretic signalling load control in next generation networks // *Internet of things, smart spaces, and next generation networking* / Eds. S. Andreev, S. Balandin, Ye. Koucheryavy. — Lecture notes in computer science ser. — Heidelberg: Springer-Verlag, 2012. Vol. 7469. P. 440–452.
23. Абаев П. О., Разумчик Р. В. Моделирование работы SIP-сервера с помощью системы массового обслуживания с гистерезисом и прогулками в дискретном времени // *T-Comm — Телекоммуникации и транспорт*, 2012. № 7. С. 5–8.
24. Гайдамака Ю. В., Самуйлов К. Е., Сопин Э. С. Модель одной системы массового обслуживания типа $M/G/1$ с гистерезисным управлением входящим потоком // *T-Comm — Телекоммуникации и Транспорт*, 2012. № 7. С. 60–62.
25. Abaev P., Pechinkin A., Razumchik R. On analytical model for optimal SIP server hop-by-hop overload control // 4th Congress (International) on Ultra Modern Telecommunications and Control Systems ICUMT-2012 Proceedings. — IEEE, 2012. P. 299–304. doi:10.1109/ICUMT.2012.6459680.
26. Abaev P., Pechinkin A., Razumchik R. Analysis of queueing system with constant service time for SIP server hop-by-hop overload control // *Modern Probab. Meth. Anal. Telecommunication Networks Communications Computer Information Sci.*, 2013. Vol. 356. P. 1–10. doi:10.1007/978-3-642-35980-4_1.
27. Abaev P., Pechinkin A., Razumchik R. On mean return time in queueing system with constant service time and bi-level hysteric policy // *Modern Probab. Meth. Anal. Telecommunication Networks Communications Computer Information Sci.*, 2013. Vol. 356. P. 11–19. doi:10.1007/978-3-642-35980-4_2.
28. Abaev P., Gaidamaka Yu., Samouylov K., Shorgin S. Design and software architecture of SIP server for overload control simulation // 27th Conference (European) on Modelling and Simulation (ECMS 2013) Proceedings. Aalesund, Norway, 2013. P. 533–539. doi:10.7148/2013-0580.
29. Pechinkin A. V., Razumchik R. V. Approach for analysis of finite $M_2/M_2/1/R$ with hysteric policy for SIP server hop-by-hop overload control // 27th Conference (European) on Modelling and Simulation (ECMS 2013) Proceedings. Aalesund, Norway, 2013. P. 573–579. doi:10.7148/2013.
30. Shorgin S., Samouylov K., Gaidamaka Yu., Etezov Sh. Polling system with threshold control for modeling of SIP server under overload // 18th Conference (International) on Systems Science (ICSS 2013) Proceedings. Advances in intelligent systems and computing ser., 2014. Vol. 240. P. 97–107. doi:10.1007/978-3-319-01857-7_10.
31. Гайдамака Ю. В. Модель с пороговым управлением нагрузкой для анализа серверов протокола SIP в режиме перегрузок // *Автоматика и вычислительная техника*, 2013. № 4. С. 65–75.
32. Abaev P., Gaidamaka Yu., Samouylov K., Pechinkin A., Razumchik R., Shorgin S. Hysteretic control technique for overload problem solution in network of SIP servers // *Comput. Inform.*, 2014 (in press). Vol. 33. No. 1.
33. MathWorks — MATLAB and Simulink for Technical Computing. [Электронный ресурс] Режим доступа <http://www.mathworks.com>, свободный (дата обращения 01.10.2013).
34. Wolfram Mathematica: Программное обеспечение для технических вычислений. [Электронный ресурс] Режим доступа: <http://www.wolfram.com/mathematica>, свободный (дата обращения 01.10.2013).
35. Maple — Technical Computing Software for Engineers, Mathematicians, Scientists, Instructors, and Students. [Электронный ресурс] Режим доступа: <http://www.maplesoft.com/products/maple>, свободный (дата обращения 01.10.2013).
36. Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S. Queueing theory. — Utrecht, Boston: VSP, 2004. 446 p.
37. Kempa W. On time to buffer saturation in a $GI/M/1/N$ -type queue // *Int. J. Adv. Telecommunications Electrotechnics Signals Syst.*, 2012. Vol. 1. No. 2–3. С. 60–66.

Поступила в редакцию 26.09.13

THE DISTRIBUTION OF THE RETURN TIME FROM THE SET OF OVERLOAD STATES TO THE SET OF NORMAL LOAD STATES IN A SYSTEM $M|M|1|\langle L, H \rangle|\langle H, R \rangle$ WITH HYSTERETIC LOAD CONTROL

Yu. V. Gaidamaka¹, A. V. Pechinkin², R. V. Razumchik², A. K. Samuylov¹, K. E. Samouylov¹, I. A. Sokolov², E. S. Sopin¹, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia, Moscow 117198, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: An analytical method for studying the parameters of the hysteretic control, which is implemented as one of the effective solutions to the overload problem in the network of SIP-servers, is suggested. As a mathematical model, the queuing system $M|M|1|\langle L, H \rangle|\langle H, R \rangle$ with two loops hysteretic control was developed, where H is the overload onset threshold, L is the overload abatement threshold, and R is the discard threshold. Two methods of calculating the Laplace–Stieltjes transform of the distribution function of the return time from the set of overload system states to the set of normal load system states were obtained. The first method consists in solving a system of equations with return times for each state of the set of overload system states as unknowns, the second deals with the recurrence for the Laplace–Stieltjes transform of the distribution function of the return time for each state of the set of overload system states as rational fractional expressions. Both methods allow the effective calculations with standard software tools, as shown in the numerical example.

Keywords: SIP-server overload; queueing system; hysteretic load control; return time to normal load states; Laplace–Stieltjes transform; distribution function

DOI: 10.14357/19922264130403

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (projects Nos. 11-07-00112 and 12-07-00108).

References

1. Hilt, V., E. Noel, C. Shen, and A. Abdelal. Design considerations for Session Initiation Protocol (SIP) overload control. 2011. Internet Engineering Task Force RFC-6357. Available at: <http://tools.ietf.org/html/rfc6357> (accessed October 1, 2013).
2. Hilt, V., J. Rosenberg, H. Schulzrinne, *et al.* SIP: Session Initiation Protocol. 2002. Internet Engineering Task Force RFC-3261. Available at: <http://tools.ietf.org/html/rfc3261> (accessed October 1, 2013).
3. Rosenberg, J. Requirements for management of overload in the Session Initiation Protocol. 2008. Internet Engineering Task Force RFC-5390. Available at: <http://tools.ietf.org/html/rfc5390> (accessed October 1, 2013).
4. Gurbani, V., V. Hilt, and H. Schulzrinne. Session Initiation Protocol overload control. 2013. Internet Engineering Task Force Draft. Available at: <http://tools.ietf.org/pdf/draft-ietf-soc-overload-control-13.pdf> (accessed October 1, 2013).
5. Noel, E., and P.M. Williams. Session Initiation Protocol rate control. 2013. Internet Engineering Task Force Draft. Available at: <http://tools.ietf.org/pdf/draft-ietf-soc-overload-rate-control-05.pdf> (accessed October 1, 2013).
6. ITU-T Recommendation Q.704: Signalling system No. 7 — Message Transfer Part, Signalling network functions and messages. 1996. Available at: <http://www.itu.int/rec/T-REC-Q.704-199607-1/en> (accessed October 1, 2013).
7. Abaev, P., Yu. Gaidamaka, and K. Samouylov. 2011. Gisterezisnoe upravlenie signal'noy nagruzkoy v seti SIP-serverov [Hysteretic overboard control in a SIP signaling network]. *Vestnik RUDN. Seriya Matematika, Informatika, Fizika [Bulletin of Peoples' Friendship University of Russia. Mathematics. Informatics. Physics ser.]* 4:55–73.
8. Abaev, P., Yu. Gaidamaka, A. Pechinkin, R. Razumchik, and S. Shorgin. 2012. Simulation of overload control in SIP server networks. *26th Conference (European) on Modelling and Simulation ECMS Proceedings*. Koblenz.

- 533–539.
9. Abaev, P., Yu. Gaidamaka, and K. Samouylov. 2012. Queuing model for loss-based overload control in a SIP server using a hysteretic technique. *Internet of things, smart spaces, and next generation networking*. Eds. S. Andreev, S. Balandin, and Ye. Koucheryavy. Lecture notes in computer science ser. Heidelberg: Springer-Verlag. 7469:371–378.
 10. Gebhart, R. F. 1967. A queuing process with bilevel hysteretic service-rate control. *Nav. Res. Logist. Q.* 14:55–68.
 11. Krasnoselskii, M., and A. Pokrovskii. 1989. *Systems with hysteresis*. Berlin – Heidelberg – New York – London – Paris – Tokio: Springer-Verlag. 410 p.
 12. Yum, T., and H. Yen. 1983. Design algorithm for a hysteresis buffer congestion control strategy. *IEEE Conference (International) on Communications Proceedings*. 499–503.
 13. Brown, P., P. Chemouil, and B. Delosme. 1984. A congestion control policy for signalling networks. *7th IeCC Proceedings*. 717–724.
 14. Golubchik, L., and J. Lui. 1997. Bounding of performance measures for a threshold-based queuing system with hysteresis. *Newsl. ACM SIGMETRICS Performance Evaluation Rev.* 25(1):147–157.
 15. Sindal, R., and S. Tokekar. 2008. Modeling and analysis of voice/data call admission control scheme in CDMA cellular network for variation in soft handoff threshold parameters. *16th IEEE Conference (International) on Networks (ICON 2008) Proceedings*. 1–6.
 16. Zhernovyi, K., and Yu. Zhernovyi. 2012. Sistema $M^0/G/1$ s gisterizatsionnym pereklyucheniem intensivnosti obsluzhivaniya [The system $M^0/G/1$ with a hysteretic switching intensity of service]. *Informatsionnye Protssesy – Information Processes* 12(3):176–190.
 17. Takagi, H. 1985. Analysis of a finite-capacity $M/G/1$ queue with a resume level. *Perform. Evaluation* 5:197–203.
 18. Benaboud, H., and N. Mikou. 2002. Analysis by queuing model of multi-threshold mechanism in ATM switches. *5th IEEE Conference (International) on High Speed Networks and Multimedia Communications (HSNMC 2002) Proceedings*. 147–151.
 19. Dshalalow, J. H. 1997. Queuing systems with state dependent parameters. *Frontiers in queuing: Models and applications in science and engineering*. Ed. J. H. Dshalalow. Probability and stochastic ser. CRC Press. 61–116.
 20. Bekker, R. 2009. Queues with Levy input and hysteretic control. *Queueing Syst.* 63(1):281–299.
 21. Roughan, M., and C. Pearce. 2000. A martingale analysis of hysteretic overload control. *Adv. Perform. Anal. J. Teletraffic Theory Performance Anal. Communication Syst. Networks.* 3(1):1–30.
 22. Abaev, P., Yu. Gaidamaka, and K. Samouylov. 2012. Modeling of hysteretic signalling load control in next generation networks. *Internet of things, smart spaces, and next generation networking*. Eds. S. Andreev, S. Balandin, and Ye. Koucheryavy. Lecture notes in computer science ser. Heidelberg: Springer-Verlag. 7469:440–452.
 23. Abaev, P., and R. Razumchik. 2012. Modelirovanie raboty SIP servera s pomoshch'yu sistemy massovogo obsluzhivaniya s gisterizatsionnoy i progulkami v diskretnom vremeni [Modeling of SIP-server with hysteric overload control as discrete time queueing system]. *T-Comm – Telekommunikacii i transport [T-Comm – Telecommunications and Transport]* 7:5–8.
 24. Gaidamaka, Yu., K. Samouylov, and E. Sopin. 2012. Model' odnoy sistemy massovogo obsluzhivaniya tipa $M/G/1$ s gisterizatsionnym upravleniem vkhodyashchim potokom [On queuing system of $M|G|1$ type with hysteretic input flow control]. *T-Comm – Telekommunikatsii i transport [T-Comm – Telecommunications and Transport]* 7:60–62.
 25. Abaev, P., A. Pechinkin, and R. Razumchik. 2012. On analytical model for optimal SIP server hop-by-hop overload control. *4th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings*. IEEE. 299–304. doi:10.1109/ICUMT.2012.6459680.
 26. Abaev, P., A. Pechinkin, and R. Razumchik. 2013. Analysis of queuing system with constant service time for SIP server hop-by-hop overload control. *Modern Probab. Meth. Anal. Telecommunication Networks Communications Computer Information Sci.* 356: 1–10. doi:10.1007/978-3-642-35980-4_1.
 27. Abaev, P., A. Pechinkin, and R. Razumchik. 2013. On mean return time in queuing system with constant service time and bi-level hysteric policy. *Modern Probab. Meth. Anal. Telecommunication Networks Communications Computer Information Sci.* 356:11–19. doi:10.1007/978-3-642-35980-4_2.
 28. Abaev, P., Yu. Gaidamaka, K. Samouylov, and S. Shorgin. 2013. Design and software architecture of SIP server for overload control simulation. *27th Conference (European) on Modelling and Simulation Proceedings*. Aalesund, Norway. 533–539. doi:10.7148/2013-0580.
 29. Pechinkin, A., and R. Razumchik. 2013. Approach for analysis of finite $M_2/M_2/1/R$ with hysteric policy for SIP server hop-by-hop overload control. *27th Conference (European) on Modelling and Simulation Proceedings*. Aalesund, Norway. 573–579. doi:10.7148/2013.
 30. Shorgin, S., K. Samouylov, Yu. Gaidamaka, and Sh. Ete-zov. 2013. Polling system with threshold control for modeling of SIP server under overload. *18th Conference (International) on Systems Science Proceedings*. Wroclaw: Springer International Pubs. 240:97–107. doi:10.1007/978-3-319-01857-7_10.
 31. Gaidamaka, Yu. 2013. Model' s porogovym upravleniem nagruzkoy dlya analiza serverov protokola SIP v rezhime peregruzok [On model with threshold load control for SIP server overload analysis]. *Avtomatika i Vychislitel'naya Tekhnika [Automatic Control and Computer Sciences]* 4:65–75.
 32. Abaev, P., Yu. Gaidamaka, K. Samouylov, A. Pechinkin, R. Razumchik, and S. Shorgin. 2014 (in press). Hysteretic control technique for overload problem solution in network of SIP servers. *Comput. Inform.* 33(1).
 33. MathWorks — MATLAB and Simulink for Technical Computing. Available at: <http://www.mathworks.com> (accessed October 1, 2013).

34. Wolfram Mathematica: Programmnoe obespechenie dlya tekhnicheskikh vychisleniy [Software for numerical computations]. Available at: <http://www.wolfram.com/mathematica> (accessed October 1, 2013).
35. Maple — Technical Computing Software for Engineers, Mathematicians, Scientists, Instructors and Students. Available at: <http://www.maplesoft.com/products/maple> (accessed October 1, 2013).
36. Bocharov, P. P., C. D'Apice, A. V. Pechinkin, and S. Salerno. 2004. *Queueing theory*. Utrecht, Boston: VSP. 446 p.
37. Kempa, W. 2012. On time to buffer saturation in a $GI/M/1/N$ -type queue. *Int. J. Adv. Telecommunications Electrotechnics Signals Syst.* 1(2-3):60–66.

Received September 26, 2013

Contributors

Gaidamaka Yuliya V. (b. 1971) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia, Moscow 117198, Russian Federation; ygaidamaka@sci.pfu.edu.ru

Pechinkin Alexander V. (b. 1946) — Doctor of Science in physics and mathematics; principal scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; professor, Peoples' Friendship University of Russia, Moscow 117198, Russian Federation; apechinkin@ipiran.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; rrazumchik@ieee.org

Samuylov Andrey K. (b. 1988) — PhD student, Peoples' Friendship University of Russia, Moscow 117198, Russian Federation; asam1988@gmail.com

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia, Moscow 117198, Russian Federation; ksam@sci.pfu.edu.ru

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; isokolov@ipiran.ru

Sopin Eduard S. (b. 1987) — PhD student, Peoples' Friendship University of Russia, Moscow 117198, Russian Federation; sopin-eduard@yandex.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; sshorgin@ipiran.ru

ОБ ОДНОЙ ЗАДАЧЕ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ НАГРУЗКОЙ НА СЕРВЕР*

М. Г. Коновалов¹

Аннотация: Рассматривается относительно простая постановка задачи об управлении нагрузкой на сервер с фиксированным количеством мест обслуживания и потенциально бесконечной очередью. Управление заключается в принятии решения о приеме либо об отклонении каждого вновь поступающего задания. Накопление очереди сопряжено с возможной потерей качества обслуживания, поскольку срок выполнения заданий ограничен. В то же время отклонение заявок влечет потерю дохода. Доказано, что в случае экспоненциально распределенного времени выполнения заданий и для входных потоков, описываемых процессом восстановления с произвольным распределением времени между поступлением заданий, оптимальной является простая пороговая стратегия. Зависимость предельного среднего дохода от значения порога является унимодальной. Это обстоятельство значительно облегчает поиск оптимального целочисленного значения порога. Экспериментальный анализ показывает, что указанная зависимость имеет место также для произвольного распределения времени выполнения заданий и для входных потоков, описываемых как марковски модулируемые процессы (ММП) общего вида.

Ключевые слова: ограничение нагрузки; пороговое управление; потоки заданий

DOI: 10.14357/19922264130404

1 Введение

Эта статья продолжает тематику повышения производительности вычислительных комплексов за счет применения более эффективных алгоритмов распределения потоков заданий [1–4]. В отличие от упомянутых работ, в которых рассматривались сравнительно сложные модели систем, допускающие оптимизацию по многим параметрам, в данном случае на простой модели анализируется только аспект, связанный с ограничением входной нагрузки.

Суть рассматриваемой задачи в следующем. Имеется система (сервер), содержащая конечное число обслуживающих мест (процессоров), на которых происходит выполнение заданий, поступающих извне случайным образом. В каждый момент появления очередной заявки должно быть принято одно из двух возможных решений: задание может быть либо оставлено в системе для последующей обработки, либо отвергнуто. Принятое на обслуживание задание сразу помещается на свободный процессор, если таковой имеется, либо становится в очередь, которая предполагается потенциально неограниченной. Продвижение в очереди осуществляется по принципу «первый пришел — первый обслуживается». Время выполнения задания на процессоре является случайным. Существенное значение имеет общее время пребывания задания

в системе, складывающееся из времени ожидания в очереди и времени непосредственного выполнения, поскольку для каждого задания существует предельный срок, до истечения которого оно должно быть выполнено (так называемый дедлайн).

Для оценки качества работы системы вводятся числовые оценки. Так, каждое принятое на обслуживание задание сопровождается «доходом», выраженным некоторым числом. Точно так же любое принятое, но не выполненное в срок задание влечет «штраф», определенный некоторой числовой величиной. В итоге можно ставить задачу о выборе такой стратегии управления доступом заданий в систему, при которой предельный средний доход будет максимальным.

При выборе стратегии следует учитывать две разнонаправленные тенденции. С одной стороны, желательно принимать как можно больше заданий, поскольку это приносит доход, но одновременно нельзя слишком наращивать очередь, поскольку это может привести к массовому превышению дедлайна и получению большого штрафа.

Для максимального упрощения модели условимся еще, что в момент принятия решения об очередной заявке неизвестно ни время, требующееся для его выполнения на процессоре, ни заданный срок выполнения. Обе эти характеристики становятся известными только в момент оконча-

* Работа выполнена при поддержке РФФИ (гранты № 11-07-00112 и № 13-07-00665).

¹ Институт проблем информатики Российской академии наук, mkonov@ipiran.ru

ния обслуживания, и в этот же момент начисляется штраф, если дедлайн превышен. Сделанное предположение говорит о том, что наблюдаемая часть траектории процесса, на которой может основываться стратегия управления нагрузкой, сводится исключительно к последовательности моментов поступления заявок и соответствующей последовательности принимаемых решений.

Описанная ситуация представляет собой едва ли не самую простую постановку задачи, мыслимую в связи с проблемой регулирования доступа в систему обслуживания, которая имеет множество оттенков и, соответственно, много названий (управление перегрузкой, контроль доступа, управление потоками и т.д.). Источник проблемы, а также основные приложения разрабатываемых алгоритмов — разнообразные информационные и телекоммуникационные системы и сети. Не пытайтесь дать характеристику этого обширного направления, укажем в качестве примера на содержащие обзоры работы [5, 6]. Отметим, что многочисленные публикации по этой теме насыщены описанием содержательной, технической стороны. Рассматриваемые в них модели объектов сложны для аналитического изучения, а получаемые решения имеют в основном эвристический характер. Решение упрощенной задачи, возможно, позволит яснее представить механизмы, лежащие в основе управления перегрузками.

Материал оставшейся части статьи распределен следующим образом. В разд. 2 дается точная постановка задачи, которая была описана наглядно во введении. Раздел 3 содержит некоторые результаты, полученные математическим путем. В разд. 4 приведены результаты вычислительных экспериментов, иллюстрирующие и дополняющие разд. 3. В заключении статьи подводится итоговое обсуждение и предлагается направление дальнейших исследований.

2 Формальная постановка задачи

Будем считать, что процесс поступления заявок (входной поток) принадлежит классу так называемых ММП. Произвольный ММП зададим с помощью набора символов $M = (P, D, F, G, H)$, означающих следующее: P — матрица $S \times S$ переходных вероятностей марковской цепи, имеющей состояния $1, \dots, S$; $D = \{D_s, s = 1, \dots, S\}$ — набор функций распределения времен пребывания в различных состояниях; $F = \{F_s, s = 1, \dots, S\}$ — набор функций распределения времен между последовательными поступлениями заявок в различных состояниях; G — функция распределения времени

выполнения задания; H — функция распределения дедлайна (функции G и H одинаковы для всех заявок, поступающих из данного потока).

Процесс протекает в непрерывном времени, $t \geq 0$. Состояние марковской цепи P в начальный момент полагаем равным $s(0) = s_0$ (для целей статьи выбор начального значения не существен). Спустя случайное время θ , значение которого определяется распределением $D_{s_0} \in \mathcal{D}$, цепь с вероятностью $P_{s_0 s'}$ переходит в состояние $s(\theta) = s'$ и остается в нем случайное время θ' , имеющее распределение $D_{s'} \in \mathcal{D}$, после чего состояние цепи с вероятностью $P_{s' s''}$ становится равным $s(\theta + \theta') = s''$ и т.д.

На любом промежутке пребывания марковской цепи в любом состоянии s заявки поступают через случайные независимые промежутки времени, имеющие одинаковое распределение $F_s \in \mathbf{F}$. Если поступившая заявка принята, то связанное с ней задание либо занимает свободное место обслуживания (при его наличии), либо становится в конец очереди (если свободных мест нет). Предполагается, что имеется N мест обслуживания, которые по-другому будем называть процессорами. Непосредственное выполнение задания на процессоре занимает случайное время, имеющее всегда одно и то же распределение G . Каждая заявка, принятая на обслуживание, приносит доход, который полагаем равным 1. По окончании обслуживания общее время, проведенное заданием в системе и складывающееся из времени ожидания в очереди и времени непосредственного выполнения на процессоре, сравнивается со значением дедлайна, которое определяется распределением H . Если дедлайн превышен, то обслуживающая система получает штраф, равный $C > 1$.

Обозначим через τ_1, τ_2, \dots последовательные моменты поступления заявок. Каждому индексу n в последовательности τ_n сопоставим переменную y_n таким образом, что

$$y_n = \begin{cases} 1, & \text{если заявка, поступившая} \\ & \text{в момент } \tau_n, \text{ принята;} \\ 0, & \text{если эта заявка отклонена.} \end{cases}$$

Последовательность $y = \{y_n\}$ называется стратегией (управления нагрузкой на сервер). Она является, вообще говоря, случайной, и по смыслу задачи каждое y_n может зависеть произвольным образом от наблюдаемой части траектории процесса до момента τ_n включительно.

Пусть имеется некоторый ММП M и задана некоторая стратегия y . Пара (M, y) порождает вероятность $\mathbf{P} = \mathbf{P}_{M,y}$, определенную на подходящем

измеримом пространстве. Интеграл по мере \mathbf{P} будем обозначать через \mathbf{E} .

Пусть θ_n означает время, проведенное в системе заданием, принятым на обслуживание в момент τ_n , и пусть $\delta_n \in H$ — дедлайн этого задания. Обозначим через

$$g_n = (1 - CI_{\{\theta_n > \delta_n\}}) I_{\{y_n = 1\}} \quad (1)$$

доход, связанный с обслуживанием заявки n .

Основная величина, связанная с рассматриваемой системой, которая представляет интерес, — это предельный (среднеарифметический) доход

$$\varphi = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n g_k.$$

Все числовые характеристики случайной величины φ определяются параметрами M и σ (а также изложенными выше неформальными правилами, касающимися постановки в очередь и пр.). Общая задача заключается в максимизации φ в том или ином вероятностном смысле. Более конкретно, положим

$$w(M, y) = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \mathbf{E}g_k. \quad (2)$$

Пусть заданы: множество ММП $\mathcal{M} = \{M\}$; множество (допустимых) стратегий $\mathcal{Y} = \{y\}$; число $\varepsilon \geq 0$. Требуется указать стратегию $y(M, \varepsilon) \in \Sigma$ такую, что для всех $M \in \mathcal{M}$ выполняется неравенство

$$w(M, y(M, \varepsilon)) \geq \sup_{y \in \mathcal{Y}} w(M, y) - \varepsilon.$$

В случае, когда \mathcal{M} содержит единственный процесс M , имеем классическую задачу отыскания ε -оптимального алгоритма (в данном случае оптимизирующего нагрузку на сервер при конкретном входном потоке M). Если же \mathcal{M} содержит более одного элемента, то говорят о поиске стратегии, адаптивной по отношению к множеству \mathcal{M} . В этом случае требование «равномерной по \mathcal{M} » оптимальности стратегии $y(M, \varepsilon)$ вызвано, как правило, тем, что априорная информация об объекте управления неопределенна и ограничивается описанием всего класса \mathcal{M} в целом.

3 Теоретический анализ

В дальнейшем будем пользоваться следующими обозначениями:

$\Gamma_{a,b}$ — гамма-распределение с параметрами a, b и плотностью $\gamma_{a,b}(t) = (a^b/\Gamma(b))t^{b-1}e^{-at}$, $t \geq 0$, где $\Gamma(b) = \int_0^\infty x^{b-1}e^{-x} dx$, $b > 0$;

$\Gamma_a = \Gamma_{a,1}$ — экспоненциальное распределение с параметром a и плотностью $\gamma_a(t) = ae^{-at}$, $t \geq 0$;

$N_{a,b}$ — нормальное распределение со средним a и дисперсией b ;

$\Pi_{a,b}$ — распределение Парето с параметрами a, b (его плотность имеет вид ab^a/t^{a+1} , $t \geq b$);

$\xi \stackrel{d}{=} \eta$ — случайные величины ξ и η имеют одинаковые распределения;

$\xi \in F$ (или $\xi \in f$) — случайная величина ξ имеет функцию распределения F (соответственно плотность распределения f);

I_A — индикатор события A .

Для аналитического рассмотрения сделаем два предположения.

П1. Времена выполнения заданий имеют одинаковое экспоненциальное распределение, параметр которого, не ограничивая общности, можно положить равным 1. Таким образом, далее в этом разделе полагаем $G = \Gamma_1$.

П2. Марковская цепь P имеет всего одно состояние, так что входной поток представляет собой процесс восстановления $R = (F, G, H)$ с независимыми промежутками между поступлениями заявок, имеющими одинаковое распределение F , и одинаковыми для всех заданий распределениями времени выполнения G и дедлайна H .

(Предположение П2, в отличие от П1, не является принципиально важным для получения основных выводов этого раздела и сделано для сокращения выкладок.)

Пусть $v(t)$ означает общее количество заданий в системе в момент t , включая задания, выполняющиеся на процессорах, и задания, ожидающие в очереди, и пусть $v_n = v(\tau_n)$, $\tilde{v}_n = v(\tau_n + 0)$. Тогда имеем соотношение:

$$\tilde{v}_n = v_n + y_n,$$

где y_n определено формулой (1).

Введем обозначение κ_n для количества заданий, выполненных на промежутке времени $(\tau_n, \tau_{n+1}]$, $n = 1, 2, \dots$. Ясно, что $\kappa_n = \tilde{v}_n - v_{n+1}$ и что $0 \leq \kappa_n \leq \tilde{v}_n$.

Обратимся к описанию (дискретного) распределения величины κ_n , обозначая через p_m его компоненты, которые полностью определяются значением величины \tilde{v}_n . Рассмотрим процесс выполнения заданий на промежутке времени $(\tau_n, \tau_{n+1}]$ при условии $\tilde{v} = i$, при котором индекс m пробегает значения $0, 1, \dots, i$.

Пусть вначале $0 \leq i \leq N$. В этом случае до наступления момента τ_{n+1} процессоры, на которых завершилось выполнение задания, вновь не загружаются. Вероятность выполнения задания на одном занятом процессоре в течение указанного промежутка времени составляет, в соответствии с предположением П1, величину

$$p = \int_0^{\infty} (1 - e^{-t}) dF(t).$$

Следовательно, в рассматриваемом случае κ_n имеет биномиальное распределение и

$$p_m = q_1(i, m) = \binom{i}{m} p^m (1-p)^{i-m}, \quad m = 0, 1, \dots, i \leq N. \quad (3)$$

Перейдем к случаю $i > N$. В этой ситуации все освобождающиеся процессоры мгновенно загружаются вновь — так происходит до тех пор, пока число выполненных заданий не достигнет значения $r = i - N > 0$. При этом если загружены все N процессоров, то появление свободных мест происходит согласно пуассоновскому закону с параметром N . После того как выполнено r заданий, картина аналогична предыдущему случаю, причем в варианте $i = N$.

Итак, для $0 \leq m \leq r$ имеем:

$$p_m = q_2(m) = \int_0^{\infty} \frac{(Nt)^m e^{-Nt}}{m!} dF(t). \quad (4)$$

Пусть $m > r$. Чтобы выполнилось более чем r заданий, необходимо, чтобы момент i -го окончания обслуживания ζ_r не превосходил момента окончания рассматриваемого промежутка между двумя поступлениями заданий. (Согласно П2, длина промежутка между поступлением заданий $\tau^{(n)} = \tau_{n+1} - \tau_n$ имеет для всех n одинаковое распределение, $\tau^{(n)} \stackrel{d}{=} \tau \in F$.) В свою очередь, время ζ_r складывается из времен между последовательными освобождениями N одновременно занятых процессоров, $\zeta_r = \xi_1^{(N)} + \dots + \xi_r^{(N)}$, где $\xi_j^{(N)} \in \Gamma_N$, $j = 1, \dots, r$, — независимые случайные величины. Таким образом, $\zeta_r \in \Gamma_{N,r}$ [7, гл. 7, § 6.2]. Приходим к выражению:

$$p_m = q_3(i, m) = \int_0^{\infty} dF(t) \int_0^t \gamma_{N,r}(s) \binom{N}{m-r} \times [p(t-s)]^{m-r} [1-p(t-s)]^{N-m+r} ds, \quad (5)$$

где $\gamma_{N,r}(s)$ — плотность гамма-распределения $\Gamma_{N,r}$; $r = i - N$, $m = r + 1, \dots, r + N$.

Рассмотрим далее время пребывания в системе заявки, принятой на обслуживание. Его можно представить как

$$\theta_n = T_n + \xi_n,$$

где T_n — время ожидания до появления свободного процессора, а ξ_n — время непосредственного выполнения задания. Вторая из этих величин, как уже говорилось, для каждого n определяется независимо: $\xi_n \stackrel{d}{=} \xi \in \Gamma_1$. Что касается первого слагаемого, то его распределение полностью определяется значением v_n : при фиксированном значении $v_n = i$

$$T_n \stackrel{d}{=} \begin{cases} 0, & \text{если } i < N; \\ \xi_1^{(N)} + \dots + \xi_{i-N+1}^{(N)}, & \text{если } i \geq N, \end{cases}$$

где $\xi_j^{(N)}$ — независимые случайные величины, определенные при выводе формулы (5). Они имеют экспоненциальное распределение с параметром N и характеризуют времена последовательного выполнения заданий, поступивших на обслуживание раньше, чем рассматриваемое задание. Отсюда следует, что условное распределение величины T_n при условии $v_n = i \geq N$ является гамма-распределением $\Gamma_{N,i-N+1}$. Следовательно, плотность распределения времени, проведенного в системе заданием, принятым на обслуживание в тот момент, когда в системе уже находилось $i \geq N$ заданий, представляет собой свертку

$$\gamma_i(t) = \gamma_{1,1}(t) * \gamma_{N,i-N+1}(t) = e^{-t} * \gamma_{N,i-N+1}(t).$$

Для $0 \leq i \leq N$ время, проведенное в системе, совпадает со временем выполнения задания, поэтому $\gamma_i(t) = e^{-t}$.

Ключевое для этого раздела замечание заключается в том, что определенная в начале раздела последовательность v_n образует управляемую марковскую цепь с множеством состояний $\{0, 1, 2, \dots\}$, множеством управлений $\{0; 1\}$ и (управляемыми) вероятностями перехода, не зависящими от n :

$$Q_{ij}^{(k)} = \mathbf{P}(v_{n+1} = v_n + \sigma_n - \kappa_n = j | v_n = i, \sigma_n = \kappa) = \mathbf{P}(\kappa_n = i - j + k), \quad k = 0 \vee 1,$$

Используя функции $q_1(i, m)$, $q_2(m)$, $q_3(i, m)$, полученные в формулах (3)–(5) для условных распределений κ_n , получим следующие выражения для элементов матриц $Q^{(k)}$:

$$Q_{ij}^{(k)} = \begin{cases} q_1(i, i-j+k) & \text{для } 0 \leq i-j+k \leq i \leq N; \\ q_2(i-j+k) & \text{для } i > N, \\ & N+k \leq j \leq i+k; \\ q_3(i, i-j+k) & \text{для } i > N, \\ & 0 \leq j \leq N+k; \\ 0 & \text{для остальных случаев.} \end{cases}$$

Сопоставим каждой паре «состояние i – управление k » одношаговый доход:

$$g_{ik} = \begin{cases} 1 - CI_{\{\vartheta_i > \delta\}}, & \text{если } k = 1; \\ 0, & \text{если } k = 0, \end{cases}$$

где $\vartheta_i \in \gamma_i(x)$, $\delta \in H$. (Словами: доход в состоянии i равен 0, если заявка не принята; равен 1, если заявка принята и обслужена до наступления дедлайна; равен $1 - C < 0$, если заявка принята, но обслужена позже требуемого срока.)

Совокупности управляемых матриц перехода и одношаговых доходов задают однородную управляемую марковскую цепь с доходами и потенциально счетным множеством состояний. Поэтому оптимальная в смысле максимизации предельного среднего дохода стратегия найдется среди однородных марковских стратегий, которые в данном случае имеют вид:

$$\sigma = (\sigma_0, \sigma_1, \sigma_2, \dots),$$

где σ_i — вероятность принять задание на обслуживание при условии, что в момент поступления заявки в системе уже находится i заданий.

Для того чтобы исключить из рассмотрения тривиальный случай, когда системе вообще не выгодно ничего делать, сделаем еще одно предположение.

П3. Одношаговый доход за управление 1 в состоянии 0 положителен. Иными словами, вероятность q того, что время выполнения одного задания на процессоре не превысит дедлайн, удовлетворяет неравенству

$$q \leq \frac{1}{C}.$$

В терминах функций распределения предположение П3 имеет вид:

$$\int_0^\infty (1 - e^{-t}) dH(t) \leq \frac{1}{C}.$$

Исходя из предположения П3, заключаем, что всякая заявка, поступившая в момент, когда есть свободные процессоры, должна быть принята, т. е. $\sigma_0 = \sigma_1 = \dots = \sigma_{N-1} = 1$, однако значения остальных членов последовательности σ менее очевидны.

Обозначим через $h = \inf \{i : \sigma_i = 0\}$ номер первой нулевой компоненты в векторе σ .

Если $h < \infty$, то значения $\sigma_{h+1}, \sigma_{h+2}, \dots$ не существенны, поскольку в этом случае число заданий, одновременно находящихся в системе, никогда не превышает h . В этом случае число h будем называть порогом, а стратегию $\sigma = \sigma(h)$ — пороговой стратегией.

Легко вычислить средний одношаговый доход в состоянии i за применение управления 1. Он равен $r_i = 1 - C\beta_i$, где $\beta_i = \int_0^\infty [1 - H(t)]\gamma_i(t) dt$ — вероятность превышения дедлайна для задания, поступившего в систему в момент, когда в ней уже находилось i заданий. Отсюда получаем выражение для среднего одношагового дохода в состоянии i :

$$R_i = \sigma_i r_i. \quad (6)$$

Поскольку ожидаемое время пребывания задания в системе монотонно не убывает с ростом значения i , то функция R_i является монотонно невозрастающей по i (а при всех $i \geq N$, очевидно, строго монотонно убывающей), причем, очевидно, $\lim_{i \rightarrow \infty} R_i = 1 - C$.

Обсудим целесообразность применения управления 1 («заявка принята») в некотором состоянии i марковской цепи v_n . Пусть в некоторый момент поступления заявки в системе находилось i заданий. Сравним поведение системы после этого момента для случаев: заявка была принята (I) и заявка была отвергнута (II). Выполнение заданий, которые уже находились в системе к рассматриваемому моменту, не зависит от выбранного в этот момент управления и протекает одинаково (в вероятностном смысле) в обоих случаях. В то же время заявки, поступающие после этого момента, находятся в худшем положении в случае I по сравнению со случаем II, поскольку перед ними в очереди было поставлено на одно задание больше. Постановка задания на обслуживание сопровождается получением дохода R_i , который может компенсировать отрицательный эффект от дополнительного элемента в очереди. Чем меньше значение R_i , тем, очевидно, меньше компенсация. Если доход R_i отрицателен, то принятие заявки заведомо невыгодно. Учитывая соображения, высказанные в предыдущем абзаце относительно функции R_i , приходим к следующему выводу: оптимальная стратегия найдется в классе Σ пороговых стратегий, причем порог h^* оптимальной стратегии удовлетворяет неравенству

$$h^* \leq \min \{i : R_i < 0\}.$$

Заметим, что верхняя оценка оптимального значения порога не зависит от входного потока.

Легко понять, что любая пороговая стратегия σ с порогом h порождает однородную марковскую

цепь с единственным эргодическим классом состояний $\{0, \dots, h\}$ и предельным распределением на этом множестве, $\pi = \pi(\sigma) = \{\pi_i\}$. Распределение π определяется из условия

$$\pi = Q\pi, \quad (7)$$

где $Q = Q(\sigma)$ — отвечающая стратегии σ переходная матрица, элементы которой имеют вид:

$$Q_{ij} = \sigma_i Q_{ij}^{(1)} + (1 - \sigma_i) Q_{ij}^{(0)}.$$

Предельный средний доход (2) за такую стратегию принимает вид:

$$w(\sigma) = \sum_{i=0}^h \pi_i R_i. \quad (8)$$

Задача оптимизации функции предельного среднего дохода на множестве однородных марковских стратегий сводится, как хорошо известно, с помощью замены переменных к задаче линейного программирования (в данном случае эта замена имеет вид $\rho_i = \pi_i \sigma_i$). Отсюда приходят к выводу, что оптимальная стратегия существует среди вырожденных стратегий. В рассматриваемом случае достаточно рассматривать пороговые стратегии, для которых $\sigma_i = 1$ при $i < h$, а $\sigma_h = 0$. Именно такие стратегии будем называть далее простыми пороговыми стратегиями, взаимно однозначно сопоставляя их с соответствующими числами h .

Содержание этого раздела позволяет предложить способ нахождения оптимальной пороговой стратегии. Он заключается в нахождении с помощью перебора такого значения $0 \leq h \leq h^*$, при котором максимален предельный средний доход $w(h)$ за простую пороговую стратегию с порогом h , определяемый по формуле (8). При этом вычисление среднего дохода осуществляется с помощью выражений (6) и (7).

В этой работе реализация указанного способа оптимизации нагрузки не обсуждается. Заметим только, что она может оказаться трудоемкой, поскольку формулы для вычисления элементов матрицы Q достаточно сложны. В то же время есть, по-видимому, возможность сократить полный перебор, если воспользоваться вогнутостью функции $w(h)$, которая была обнаружена экспериментально.

4 Результаты вычислительных экспериментов

Все упоминаемые ниже численные результаты получены на персональном компьютере с помощью имитационной модели системы, описанной в предыдущих разделах.

В предыдущем разделе был определен средний одношаговый доход r_i в состоянии i за обслуживание принятой заявки 1. Проиллюстрируем зависимость этой величины (а также связанной с ней вероятности превышения дедлайна β_i) от размера очереди i в момент поступления заявки.

На рис. 1 изображены зависимости r_i (сплошные линии), а также β_i (штриховые линии) от i для случаев, когда число процессоров $N = 1, 2, 5, 10$. Значение $i = -1$ соответствует ситуации, когда в момент прихода заявки в системе есть свободные процессоры. Распределение времени выполнения задания во всех случаях имеет вид $F \in \Gamma_1$, распределение дедлайна — $H \in \Gamma_{1/30}$, а постоянная $C = 10$. Как видно из графиков, характер зависимости одинаков для различного числа процессоров. Масштаб горизонтальной оси изменяется по мере увеличения количества процессоров, что соответствует сопутствующему увеличению пропускной способности системы.

На рис. 2 показано сравнительное поведение функции r_i для четырех систем: 1 — $N = 2$, $G \in \Gamma_{0,5}$; 2 — $N = 4$, $G \in \Gamma_{0,25}$; 3 — $N = 5$, $G \in \Gamma_{0,2}$; 4 — $N = 10$, $G \in \Gamma_{0,1}$ (во всех случаях $H \in \Gamma_{1/30}$). С помощью этого рисунка обнаруживается следующее обстоятельство. Параметр a распределения G можно интерпретировать как производительность процессора. Для всех систем выполняется условие $Na = 1$. Тем не менее производительность систем разная, причем она уменьшается с уменьшением N . Например, система 4, в которой в пять раз больше процессоров, чем в системе a , но они работают в 5 раз медленнее, является экономически нецелесообразной, поскольку при любом размере очереди средний одношаговый доход отрицательный.

Обратимся теперь к результатам, связанным с использованием простых пороговых стратегий, каждая из которых однозначно характеризуется неотрицательным целым числом h — значением порога. Как было показано в предыдущем разделе, для систем с экспоненциальным распределением времени выполнения заданий такие стратегии дают максимальный предельный средний доход.

Проведена серия экспериментов на имитационной модели с целью получения зависимости $w(h)$ предельного среднего дохода от значения порога. Рассматривались системы с числом процессоров $N = 1, 2, 5, 10$. В качестве входного потока рассматривались описанные в предыдущем разделе процессы восстановления $R = (F, G, H)$ (см. предположение П2). Функция распределения интервалов между поступлениями заявок F выбиралась из двух типов распределений (экспоненциального и Парето), причем каждый раз так, чтобы средняя нагрузка на один процессор составляла одну заявку в едини-

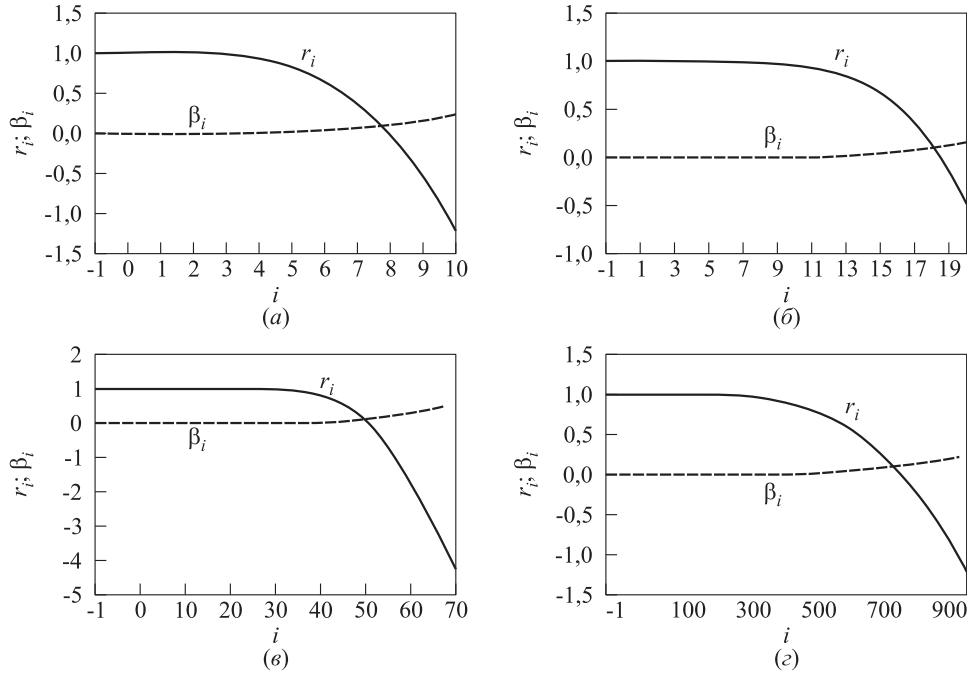


Рис. 1 Одношаговый доход в зависимости от размера очереди при числе процессоров $N = 1$ (а); 2 (б); 5 (в); 10 (г)

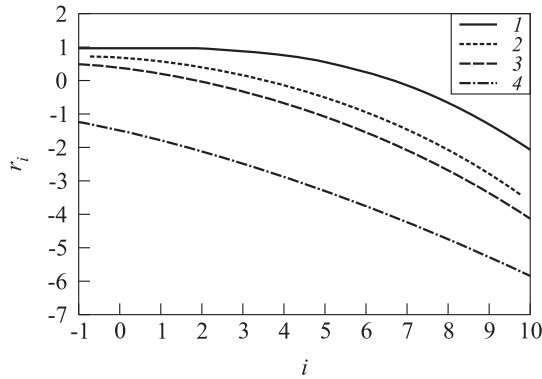


Рис. 2 Сравнительное поведение $r_i(i)$ для систем: 1 — $N = 2, G \in \Gamma_{0,5}$; 2 — $N = 4, G \in \Gamma_{0,25}$; 3 — $N = 5, G \in \Gamma_{0,2}$; 4 — $N = 10, G \in \Gamma_{0,1}$. Вычислительную мощность выгоднее концентрировать в меньшем числе процессоров

цу времени. Из тех же типов распределений выбиралась функция распределения времени выполнения задания на процессоре G , причем так, чтобы среднее время выполнения задания составляло единицу времени. Для каждого значения N рассматривались четыре возможные комбинации типов распределений функций F и G :

(1) $N = 1$:

$$(F \in \Gamma_1, G \in \Gamma_1);$$

$$(F \in \Gamma_1, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

$$(F \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})}, G \in \Gamma_1);$$

$$(F \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})}, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

(2) $N = 2$:

$$(F \in \Gamma_2, G \in \Gamma_1);$$

$$(F \in \Gamma_2, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

$$(F \in \Pi_{1+\sqrt{2}, 1/(2+\sqrt{2})}, G \in \Gamma_1);$$

$$(F \in \Pi_{1+\sqrt{2}, 1/(2+\sqrt{2})}, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

(3) $N = 5$:

$$(F \in \Gamma_5, G \in \Gamma_1);$$

$$(F \in \Gamma_5, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

$$(F \in \Pi_{1+\sqrt{2}, 0, 2\sqrt{2}/(2+\sqrt{2})}, G \in \Gamma_1);$$

$$(F \in \Pi_{1+\sqrt{2}, 0, 2\sqrt{2}/(2+\sqrt{2})}, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

(4) $N = 10$:

$$(F \in \Gamma_{10}, G \in \Gamma_1);$$

$$(F \in \Gamma_{10}, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})});$$

$$(F \in \Pi_{1+\sqrt{2}, 0, 1\sqrt{2}/(2+\sqrt{2})}, G \in \Gamma_1);$$

$$(F \in \Pi_{1+\sqrt{2}, 0, 1\sqrt{2}/(2+\sqrt{2})}, G \in \Pi_{1+\sqrt{2}, \sqrt{2}/(1+\sqrt{2})}).$$

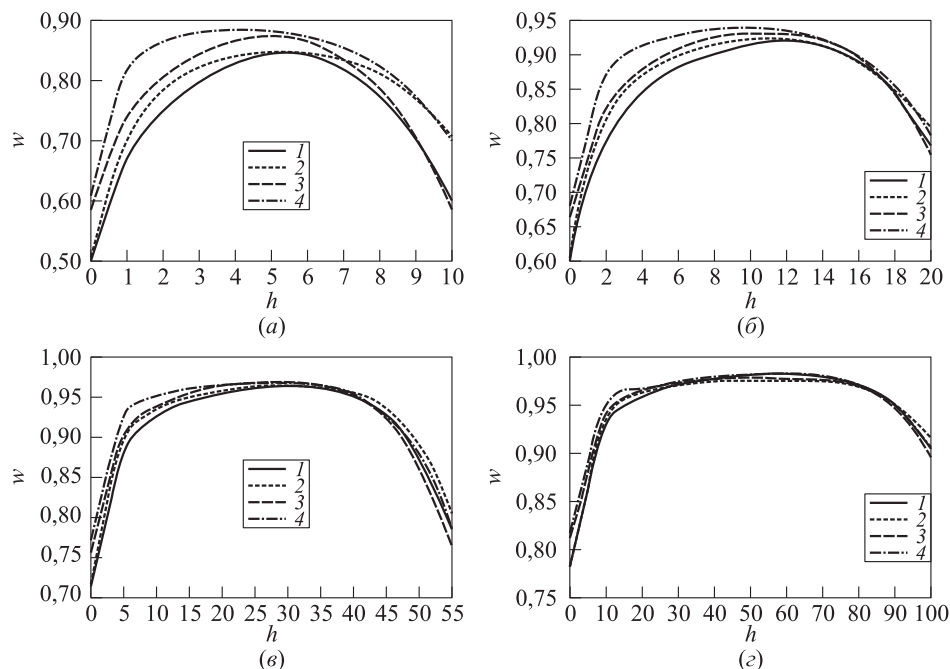


Рис. 3 Зависимость целевой функции от значения порога при числе процессоров $N = 1$ (а) 2 (б); 5 (в); 10 (г); 1 — Э–Э; 2 — Э–П; 3 — П–Э; 4 — П–П

В этих формулах параметры распределения Парето подобраны так, чтобы для каждого значения N во всех четырех скобках дисперсии соответствующих компонент совпадали.

Для всех вариантов $H \in N_{15,5}$, $C = 10$.

На рис. 3 показаны графики зависимости $w(h)$ для всех указанных 16 вариантов системы. Для обозначения принадлежности функций распределения F и G определенному типу (экспоненциальному — Э или Парето — П) используются обозначения Э–Э, Э–П, П–Э и П–П.

Во всех случаях качественный характер зависимости оказался одинаковым — это унимодальные вогнутые функции. С ростом числа процессоров обнаруживаются две тенденции: сдвиг максимума в область больших значений порога и увеличение плато на графиках в области максимума. Отсюда следует, что среди систем с единичной средней нагрузкой на один сервер допустимая очередь больше для тех из них, у которых больше мест обслуживания. При этом для систем с большим числом процессоров не так сказывается погрешность в определении оптимального значения порога.

Важным обстоятельством представляется то, что различие графиков для разных типов функций распределения F и G незначительно. К тому же это различие уменьшается с увеличением N . В связи с этим возникает вопрос: существуют ли системы с иной формой зависимости предельного среднего

дохода от значения порога, определяющего простую стратегию? Многочисленные эксперименты дали отрицательный ответ. В этих экспериментах варьировалось число процессоров, тип и параметры функций распределения F, G, H , а также параметры входного ММП. При этом в качестве типов распределений использовались экспоненциальное, равномерное, нормальное распределение, а также распределение Парето и Вейбулла. Во всех случаях характер упомянутой зависимости оставался неизменным и представлял собой вогнутую унимодальную функцию. Приведем только один заключительный пример.

Система содержит два процессора ($N = 2$). Входной поток представляет собой ММП со следующими параметрами: $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (переходная матрица состояний входного потока); $D_1 \in \Gamma_{0,001}$, $D_2 \in \Pi_{4,75}$ (распределения времен пребывания в состояниях); $F_1 \in \Gamma_1$, $F_2 \in \Gamma_{10}$ (распределения времен между поступлениями заявок); $G \in \Pi_{2,005,201/401}$ (распределение времени выполнения задания); $H \in N_{15,5}$ (распределение дедлайна); $C = 10$ (штраф за превышение дедлайна).

Рассматриваемая система выполняет задания за время, имеющее распределение Парето со средним значением 1 и большой дисперсией, равной 100. Особенность входного потока заключается в том, что периоды сравнительно небольшой нагрузки

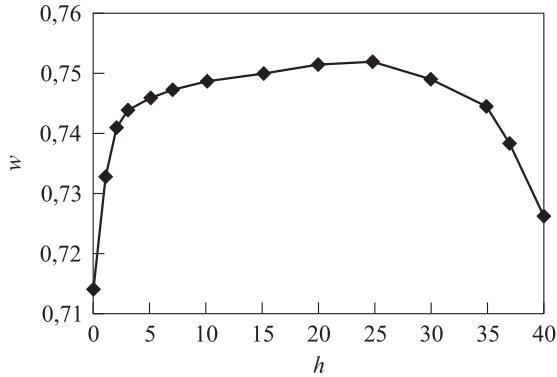


Рис. 4 Зависимость целевой функции от значения порога для более сложного входного потока

(состояние 1 цепи P) сменяются периодами, когда нагрузка в 10 раз больше (состояние 2 цепи P). При этом время пребывания в состоянии 2 хотя и меньше в среднем в 10 раз, чем время пребывания в состоянии 1, но имеет распределение Парето с большой дисперсией, равной 1250. Таким образом, поведение системы отличается большой «нерегулярностью» по сравнению с чисто «экспоненциальной» моделью. Тем не менее приведенный на рис. 4 график функции $w(h)$, полученный с помощью компьютерной имитации работы системы, качественно повторяет графики на рис. 3.

5 Заключение

В статье изучалась работа сервера, который на конечном числе процессоров выполняет задания, поступающие из случайного потока. Прием задания на обслуживание приносит доход, но сопряжен с возможным штрафом в случае невыполнения задания в срок. Сформулирована задача оптимизации доступа заданий в систему с точки зрения увеличения предельного среднего дохода.

В предположении экспоненциального распределения времени выполнения заданий для широкого класса входных потоков, описываемых как процесс восстановления с произвольным распределением времени между поступлением заданий, доказано, что оптимальная стратегия находится в конечном множестве простых пороговых стратегий. При этом вычислительные эксперименты на имитационной модели показали, что зависимость предельного среднего дохода от значения порога представляет собой унимодальную функцию, что значительно облегчает нахождение оптимального порога.

На основе экспериментальных данных можно высказать гипотезу, что характер упомянутой зависимости сохраняется для произвольного распреде-

ления времени выполнения заданий и для входных потоков, описываемых как ММП общего вида. Это существенно облегчает нахождение оптимального значения порога, позволяя избежать полного перебора.

В связи с изучаемой задачей остается открытым важный вопрос: можно ли увеличить предельный средний доход за счет расширения множества простых пороговых стратегий, если распределение времени выполнения заданий отличается от экспоненциального? В этом случае процесс v_n , обозначающий количество заданий в системе, перестает быть марковским. Естественно было бы ожидать, что оптимальная стратегия тогда будет более сложная — например, рандомизированная или зависящая от более глубокой предыстории, чем текущее значение v_n . Однако эксперименты показывают, что этого не происходит, по крайней мере, при полной наблюдаемости процесса v_n . Эти результаты требуют отдельного изложения, хотя бы из необходимости описать, как осуществляется оптимизация на множестве стратегий, более изощренных, чем простые пороговые стратегии.

Указанное направление для продолжения исследований представляется интересным, и не только с теоретической точки зрения. На этом пути можно было бы подойти к обоснованию тех алгоритмов, которые реально применяются на практике, например гистерезисных стратегий ограничения нагрузки [8].

Литература

1. Коновалов М. Г. О планировании потоков в системах вычислительных ресурсов // Информатика и её применения, 2010. Т. 4. Вып. 2. С. 3–12.
2. Kononov M. Multiagent model for jobs flows planning and pricing in distributed computing systems // 2010 Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT): Proceedings of ICUMT-T, ICUMT-CS and associated workshops. IEEE, Catalog Number CFP1063G-CDR, 2010. CD-ROM. ISBN 978-1-4244-7286-4. Report 1569339397.
3. Коновалов М. Г., Малашенко Ю. Е., Назарова И. А. Управление заданиями в гетерогенных вычислительных системах // Известия РАН. Теория и системы управления, 2011. Т. 50. № 2. С. 43–61.
4. Коновалов М. Г. Оптимизация работы вычислительно-го комплекса с помощью имитационной модели и адаптивных алгоритмов // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 37–48.
5. Welzl M. Network congestion control. — N.Y.: Wiley, 2005.
6. Hong Y., Huang C., Yan J. A comparative study of SIP overload control algorithms // Network and traffic engineering in emerging distributed computing applications /

- Eds. J. Abawajy, M. Pathan, M. Rahman, A. K. Pathan, and M. M. Deris. — IGI Global, 2012. P. 1–20. <http://arxiv.org/ftp/arxiv/papers/1210/1210.1505.pdf>.
7. *Боровков А. А.* Теория вероятностей. — 5-е изд. — М.: Либроком, 2009. 656 с.
8. *Abaev P. O., Gaidamaka Y. V., Pechinkin A. V., Razumchik R. V., Shorgin S. Ya.* Simulation of overload control in SIP server networks // ECMS 2012: 26th European Conference on Modelling and Simulation Proceedings. — Koblenz, Germany: Digitaldruck Pirrot GmbH, 2012. P. 533–539.

Поступила в редакцию 21.10.13

ABOUT ONE TASK OF OVERLOAD CONTROL

M. G. Konovalov

Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: The article considers the relatively simple task of congestion control. On the server with a finite number of places of service and potentially infinite queue, jobs are running, coming from the random flow. Control means the adoption of the decision on admission or rejection of each newly incoming job. Accumulation of the queue may result in loss of quality of service, because the period of execution of jobs is limited. At the same time, the rejection of application causes the loss of income. It is proved that in the case of exponentially distributed service time and for input flows, described as the renewal process with an arbitrary interarrival time distribution, optimum is a simple threshold strategy. The dependence of the limiting average income on the threshold value is unimodal. This circumstance greatly facilitates the search for the optimal integer value of the threshold. Experimental analysis shows that this dependence has a place for arbitrary distribution of service time and for general type of Markov modulated input flows.

Keywords: congestion control; overload control; threshold strategy; job flow

DOI: 10.14357/19922264130404

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (grants Nos. 11-07-00112 and 13-07-00665).

References

- Konovalov, M. G. 2010. O planirovanii potokov v sistemakh vychislitel'nykh resursov [On task flow planning in computational resource systems]. *Informatika i ee Primeneniya — Inform. Appl.* 4(2):3–12.
- Konovalov, M. 2010. Multiagent model for jobs flows planning and pricing in distributed computing systems. *2010 Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). ICUMT-T, ICUMT-CS and Associated Workshops Proceedings*. CD-ROM. IEEE, Catalog Number CFP1063G-CDR. Report 1569339397.
- Konovalov, M. G., Yu. E. Malashenko, and I. A. Nazarova. 2011. Job control in heterogeneous computing systems. *J. Comput. Syst. Sc. Int.* 50(2):220–237.
- Konovalov, M. G. 2012. Optimizatsiya raboty vychislitel'nogo kompleksa s pomoshch'yu imitatsionnoy modeli i adaptivnykh algoritmov [Computer system optimization using simulation model and adaptive algorithms]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):37–48.
- Welzl, M. 2005. *Network congestion control*. NY: Wiley.
- Hong, Y., C. Huang, and J. Yan. 2012. A comparative study of SIP overload control algorithms. *Network and traffic engineering in emerging distributed computing applications*. Eds. J. Abawajy, M. Pathan, M. Rahman, A. K. Pathan, and M. M. Deris. IGI Global. 1–20. Available at: <http://arxiv.org/ftp/arxiv/papers/1210/1210.1505.pdf> (accessed November 5, 2013).
- Borovkov, A. A. 2009. *Teoriya veroyatnostey [Probability theory]*. 5th ed. Moscow: Librokom. 656 p.
- Abaev, P. O., Y. V. Gaidamaka, A. V. Pechinkin, R. V. Razumchik, and S. Ya. Shorgin. 2012. Simulation of overload control in SIP server networks. *ECMS 2012: 26th European Conference on Modelling and Simulation Proceedings*. Koblenz, Germany. 533–539.

Received October 21, 2013

Contributors

Konovalov Mikhail G. (b. 1950) — Doctor of Science in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; mkonovalov@ipiran.ru

ОПТИМИЗАЦИЯ ФУНКЦИЙ LAB-КОНТРАСТНОГО ГРАДАЦИОННОГО ПРЕОБРАЗОВАНИЯ

О. П. Архипов¹, З. П. Зыкова²

Аннотация: Рассмотрена задача персонифицированного преобразования распределения контрастов на ступенчатых тоновых шкалах. Решение задачи необходимо для управления отображениями RGB-изображений на цветных периферийных устройствах персональных электронных вычислительных машин (ПЭВМ) с целью улучшения их восприятия в части детализации. Поскольку изменение соотношения Lab-контрастов цифрового описания отображений пикселей влечет за собой подобное, хотя и, возможно, менее ярко выраженное изменение соотношения реальных контрастов отображений, то для решения задачи достаточно подобрать подходящее распределение Lab-контрастов. Для приближенного вычисления функции Lab-контрастного градационного преобразования ступенчатых тоновых шкал по образцу рассматривались два семейства параметрических алгоритмов. Параметры одного из них — подмножества пикселей шкалы, а другого — множества градаций. Задача выбора оптимальных параметров решена путем сравнения погрешности вычисления функции распределения Lab-контрастов на ступенчатых тоновых шкалах на типичных примерах. Приведен пример, демонстрирующий эффективность применения соответствующей функции Lab-контрастного градационного преобразования ступенчатых тоновых шкал. Показано, что при выборе подходящего образца можно не только избежать искажения детализации, но и добиться ее улучшения.

Ключевые слова: цветовоспроизведение; цветовосприятие; Lab-координаты; контраст; градации

DOI: 10.14357/19922264130405

1 Введение

Lab-контрастное градационное преобразование ступенчатых тоновых шкал является необходимым компонентом информационной технологии равноконтрастной градационной скелетизации цветового пространства восприятия произвольным пользователем вывода цветных изображений на периферийные устройства ПЭВМ [1–9].

Информационная технология применяется на практике в обычных условиях функционирования офисных компьютерных систем с цветной периферией для персонифицированного управления детализацией RGB-изображений при их печати и выводе на монитор.

Целью такого управления является улучшение восприятия цветных изображений в части детализации не только для пользователей, цветовосприятие которых близко к стандарту, но и для тех пользователей, которые имеют такие аномалии цветового зрения, как частичная цветовая слепота.

Lab-контрастное градационное преобразование ступенчатых тоновых шкал по образцу является инструментом влияния на персонифицированное распределение контрастов в отображениях цвет-

ных изображений (представлениях изображений на периферийных устройствах ПЭВМ), позволяющим улучшить детализацию отображений в восприятии произвольных пользователей [6–9].

В основе влияния лежит тот факт, что при корректном цифровом описании цветовоспроизведения и цветовосприятия изменение соотношения Lab-контрастов цифрового описания отображений пикселей влечет за собой подобное, хотя и, возможно, менее ярко выраженное изменение соотношения реальных контрастов отображений.

Пусть x_1, x_2 — RGB-пиксели, Ψ_n — функция цветовосприятия n -го наблюдателя, а RGB-функция φ_n — цифровое описание Ψ_n .

Обозначим через $\{\Psi_n(x_i)\}$ отображения пикселей в цветовом пространстве n -го наблюдателя, $\{\varphi_n(x_i)\}$ — цифровое описание этих отображений, $C_n(\Psi_n(x_1), \Psi_n(x_2))$ — контраст отображений RGB-пикселей x_1 и x_2 в цветовом пространстве n -го наблюдателя. Пусть $E(\varphi_n(x_1), \varphi_n(x_2))$ — Lab-контраст этих пикселей:

$$E(\varphi_n(x_1), \varphi_n(x_2)) = \sqrt{(L_{n,1} - L_{n,2})^2 + (a_{n,1} - a_{n,2})^2 + (b_{n,1} - b_{n,2})^2},$$

¹Орловский филиал Института проблем информатики Российской академии наук, arhipov12@yandex.ru

²Орловский филиал Института проблем информатики Российской академии наук, zykzoia@yandex.ru

где $(L_{n,1}, a_{n,1}, b_{n,1})$ — это Lab-координаты, соответствующие RGB-координатам $\varphi_n(x_1)$, а $(L_{n,2}, a_{n,2}, b_{n,2})$ — $\varphi_n(x_2)$.

Обозначим через x'_1, x'_2 значения пикселей x_1, x_2 после некоторого преобразования. Если Lab-контраст цифрового описания отображений новых значений пикселей значительно превосходит Lab-контраст цифрового описания отображений пикселей x_1 и x_2 :

$$E(\varphi_n(x'_1), \varphi_n(x'_2)) \gg E(\varphi_n(x_1), \varphi_n(x_2)),$$

то, как правило, выполняется следующее соотношение:

$$C_n(\Psi_n(x'_1), \Psi_n(x'_2)) \geq C_n(\Psi_n(x_1), \Psi_n(x_2)).$$

Из соотношения

$$E(\varphi_n(x'_1), \varphi_n(x'_2)) \ll E(\varphi_n(x_1), \varphi_n(x_2))$$

следует

$$C_n(\Psi_n(x'_1), \Psi_n(x'_2)) \leq C_n(\Psi_n(x_1), \Psi_n(x_2)).$$

Приведенные зависимости позволяют определить подход к получению нужного распределения контрастов на отображениях шкал, состоящий в подборе подходящего распределения Lab-контрастов. В [6–9] предлагается подбирать распределение Lab-контрастов по образцу. Для задания образцов используются неотрицательные числовые функции от пикселей шкал.

Процедура управления контрастами отображений такова:

- (1) задается начальный образец распределения Lab-контрастов;
- (2) определяется функция Lab-контрастного градационного преобразования цифрового описания отображений ступенчатых тоновых шкал по образцу;
- (3) вычисляются координаты пикселей преобразованных ступенчатых тоновых шкал и цифровое описание их отображений;
- (4) вычисляется распределение Lab-контрастов на цифровом описании отображений ступенчатых тоновых шкал;
- (5) преобразованные ступенчатые шкалы выводятся на используемое периферийное устройство;
- (6) пользователем проводится визуальный анализ детализации — распределения контрастов на отображениях шкал в сопоставлении с графиками образца и функции Lab-контрастов;

- (7) до установления допустимости детализации отображений шкал проводится необходимая модификация образца и повторяются шаги 2–6;
- (8) после нахождения подходящего образца определяется Lab-контрастное градационное преобразование всех пикселей RGB-куба;
- (9) выполняется предварительное Lab-контрастное градационное преобразование произвольного изображения;
- (10) преобразованное изображение выводится на используемое периферийное устройство.

В результате в отображениях произвольных изображений обеспечивается установленное выбранным образцом распределение контрастов. В качестве начального образца распределения Lab-контрастов можно выбрать, например, равноконтрастный образец, который определяется функцией расстояния между RGB-пикселями, или образец подобия оригиналу, который определяется функцией Lab-контрастов исходных тоновых шкал.

Модификация образца приводит к модификации распределения Lab-контрастов. Модификация Lab-контрастов приводит к модификации распределения реальных контрастов. Таким образом, нужного распределения реальных контрастов на отображениях преобразованных ступенчатых тоновых шкал можно добиться, если модифицировать образец, ориентируясь на соотношение графиков функций образца и Lab-контрастов и распределение контрастов в отображениях преобразованных тоновых шкал.

Из-за дискретности RGB-координат функции Lab-контрастов лишь приближенно соответствуют образцам при использовании любой функции Lab-контрастного градационного преобразования. Кроме того, погрешность вычисления функции Lab-контрастов зависит от значений параметров применяемых функций преобразования, от значений координат векторов образцов, а также от характеристик и цифрового описания цветового пространства в восприятии пользователя.

Очевидно, что чем точнее вычисляется функция Lab-контрастов по отношению к образцу, тем быстрее можно получить нужный результат. В связи с этим в рамках данной работы рассматривается задача оптимизации функции Lab-контрастного градационного преобразования — выбора такой функции преобразования, при которой распределение Lab-контрастов было бы максимально близким к образцу при произвольных образцах и пространствах цветовосприятия.

2 Обозначения и определения

В рамках данной работы рассматривается несколько цифровых описаний цветовосприятия виртуальных наблюдателей. Обозначим через φ_n , $n \in 0, 1, \dots, 5$, RGB-функции, описывающие цветовосприятие некоторых типичных наблюдателей с аномалией цветового зрения в соответствии с моделями цветовосприятия из [10, 11]: φ_0 — для дейтеранопов [10]; φ_1 — для дейтеранопов [11]; φ_2 — для протанопов [10]; φ_3 — для протанопов [11]; φ_4 — для тританопов [10]; φ_5 — для тританопов [11].

Пусть M — множество RGB-пикселей, координаты которых кратны семнадцати. Тогда

$$\varphi_n(M) \subset M, \quad n \in 0, 1, \dots, 5.$$

Обозначим через φ_6 RGB-функцию, описывающую идеальное цветовосприятие при идеальном цветовоспроизведении, при котором

$$\varphi_6(M) = M.$$

Обозначим через φ_n , $n \in 7, 8, \dots, 13$, функции RGB-характеристики цветовосприятия, вычисленные по результатам визуального тестирования различения цветных пикселей из множеств $\varphi_{n-6}(M)$ в соответствии с [1–5].

Рассматривается следующая совокупность тоновых шкал:

$$S_i = \{s_{i,j}\}, \quad i = 0, 1, \dots, 6, \quad j = 0, 1, \dots, J_i^S,$$

где

$$\begin{aligned} S_0 &= \{s_{0,j}\}, \quad s_{0,j} = (j, j, j); \\ S_1 &= \{s_{1,j}\} \cup \{s_{1,j+15}\}, \quad s_{1,j} = (j, 0, 0), \\ &\quad s_{1,j+15} = (255, j, j); \\ S_2 &= \{s_{2,j}\} \cup \{s_{2,j+15}\}, \quad s_{2,j} = (j, j, 0), \\ &\quad s_{2,j+15} = (255, 255, j); \\ S_3 &= \{s_{3,j}\} \cup \{s_{3,j+15}\}, \quad s_{3,j} = (0, j, 0), \\ &\quad s_{3,j+15} = (j, 255, j); \\ S_4 &= \{s_{4,j}\} \cup \{s_{4,j+15}\}, \quad s_{4,j} = (0, j, j), \\ &\quad s_{4,j+15} = (j, 255, 255); \\ S_5 &= \{s_{5,j}\} \cup \{s_{5,j+15}\}, \quad s_{5,j} = (0, 0, j), \\ &\quad s_{5,j+15} = (j, j, 255); \\ S_6 &= \{s_{6,j}\} \cup \{s_{6,j+15}\}, \quad s_{6,j} = (j, 0, j), \\ &\quad s_{6,j+15} = (255, j, 255), \quad j = 0, 1, \dots, 255. \end{aligned}$$

Последовательность S_i будем называть носителем ступенчатой тоновой шкалы $T_i = \{t_{i,j}\}$, если T_i является подпоследовательностью S_i , состоящей

из всех таких пикселей S_i , координаты которых кратны семнадцати.

Как и ранее, для описания Lab-контраста RGB-пикселей x' и x'' будем использовать $E(x', x'')$ — значение расстояния между образами RGB-пикселей x' и x'' в Lab-пространстве. Исходное распределение Lab-контрастов на $\{t_{i,j}\}$ описывается последовательностью $\{E(t_{i,j}, t_{i,j+1})\}$, на $\{\varphi_n(t_{i,j})\} - \{E(\varphi_n(t_{i,j}), \varphi_n(t_{i,j+1}))\}$.

В качестве образцов $\tau_l(t_{i,j})$ будут использоваться произвольные функции, обладающие следующими свойствами:

$$\begin{aligned} \tau_l(t_{i,j}) &\geq 0, \quad \sum_{j=0}^{J_i^T-1} \tau_l(t_{i,j}) > 0, \\ l &= 0, 1, \dots; \quad i = 0, 1, \dots, 6; \quad j = 0, 1, \dots, J_i^T. \end{aligned}$$

3 Lab-контрастное градиционное преобразование по образцу

Надо найти функцию $F(t_{i,j})$, при которой распределение контрастов на преобразованной шкале соответствовало бы образцу τ_l :

$$\begin{aligned} \frac{E(\varphi_n(F(t_{i,j})), \varphi_n(F(t_{i,j+1})))}{J_i^T-1} &= \\ \frac{\sum_{m=0}^{J_i^T-1} E(\varphi_n(F(t_{i,m})), \varphi_n(F(t_{i,m+1})))}{\sum_{m=0}^{J_i^T-1} \tau_l(t_{i,m})} &= \\ = \frac{\tau_l(t_{i,j})}{\sum_{m=0}^{J_i^T-1} \tau_l(t_{i,m})}, \quad j < J_i^T. \quad (1) \end{aligned}$$

Очевидно, что в поле вещественных чисел решение уравнений (1) существует. Но в данном случае из-за дискретности координат RGB-пикселей возможно только приближенное вычисление функции F .

Пусть на носителях тоновых шкал выбраны градации — некоторые подпоследовательности

$$\begin{aligned} X_k &= \{x_{k,i,j}\}, \quad i \in 0, 1, \dots, 6, \\ &\quad j \in 0, 1, \dots, J_{k,i}^X; \quad k \in 0, 1, \dots, K, \end{aligned}$$

причем

$$x_{k,l,0} = s_{i,0}, \quad x_{k,i,J_i^X} = s_{i,J_i^S}.$$

По заданным градациям найдем последовательности

$$Y_{k,l,n} = \{y_{k,l,n,i,j}\} \subset S_i,$$

удовлетворяющие соотношениям:

$$\begin{aligned} & \frac{E(\varphi_n(x_{k,i,j}), \varphi_n(x_{k,i,j+1}))}{\sum_{m=0}^{J_{k,i}^X - 1} E(\varphi_n(x_{k,i,m}), \varphi_n(x_{k,i,m+1}))} = \\ & = \frac{\tau_l(y_{k,l,n,i,j})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(y_{k,l,n,i,m})}, \quad j < J_{k,i}^X. \quad (2) \end{aligned}$$

Определим значения функции τ_l на всех пикселах носителя $s_{i,j}$ с помощью линейной интерполяции по двум ближайшим к ним пикселям ступенчатых тоновых шкал (пусть $t_{i,j'}$ и $t_{i,j'+1}$):

$$\begin{aligned} \tau_l(s_{i,j}) &= \frac{1}{17} (17(j'+1) - j) \tau_l(t_{i,j'}) + \\ &+ (j - 17j') \tau_l(t_{i,j'+1}), \quad 17j' \leq j \leq 17(j'+1). \end{aligned}$$

Заметим, что уравнения из (2) в общем случае не имеют точного решения. Перепишем их в эквивалентном виде

$$\begin{aligned} & \frac{\sum_{m=0}^j E(\varphi_n(x_{k,i,m}), \varphi_n(x_{k,i,m+1}))}{\sum_{m=0}^{J_{k,i}^X - 1} E(\varphi_n(x_{k,i,m}), \varphi_n(x_{k,i,m+1}))} = \\ & = \frac{\sum_{m=0}^j \tau_l(y_{k,l,n,i,m})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(y_{k,l,n,i,m})}, \quad j < J_{k,i}^X, \end{aligned}$$

и на этой основе перейдем к условиям, из которых найдем приближенные значения компонентов искомым последовательностей $Y_{k,l,n}$:

$$\begin{aligned} & \frac{\sum_{m=0}^{j'(k,l,n,i,j)-1} \tau_l(s_{i,m})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(s_{i,m})} \leq \\ & \leq \frac{\sum_{m=0}^j E(\varphi_n(x_{k,i,m}), \varphi_n(x_{k,i,m+1}))}{\sum_{m=0}^{J_{k,i}^X - 1} E(\varphi_n(x_{k,i,m}), \varphi_n(x_{k,i,m+1}))} \leq \\ & \leq \frac{\sum_{m=0}^{j'(k,l,n,i,j)} \tau_l(s_{i,m})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(s_{i,m})}, \quad j < J_{k,i}^X, \quad j'(k,l,n,i,j) < J_i^S. \end{aligned}$$

Полагая

$$\begin{aligned} y_{k,l,n,i,j} &\approx s_{i,j'(k,l,n,i,j)}; \\ x_{k,i,j} &= f_{k,l,n}(s_{i,j'(k,l,n,i,j)}) \approx f_{k,l,n}(y_{k,l,n,i,j}), \end{aligned}$$

получим:

$$\begin{aligned} & \frac{E(\varphi_n(f_{k,l,n}(y_{k,l,n,i,j})), \varphi_n(f_{k,l,n}(y_{k,l,n,i,j+1})))}{\sum_{m=0}^{J_{k,i}^X - 1} E(\varphi_n(f_{k,l,n}(y_{k,l,n,i,m})), \varphi_n(f_{k,l,n}(y_{k,l,n,i,m+1})))} \approx \\ & \approx \frac{\tau_l(y_{k,l,n,i,j})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(y_{k,l,n,i,m})}, \quad j < J_{k,i}^X. \end{aligned}$$

Теперь можно с помощью интерполяции определить значения функции $f_{k,l,n}$ во всех точках носителей, в том числе и в точках ступенчатых тоновых шкал

$$\begin{aligned} & \frac{E(\varphi_k(f_{k,l,n}(t_{i,j})), \varphi_k(f_{k,l,n}(t_{i,j+1})))}{\sum_{m=0}^{J_{k,i}^X - 1} E(\varphi_k(f_{k,l,n}(t_{i,m})), \varphi_k(f_{k,l,n}(t_{i,m+1})))} \approx \\ & \approx \frac{\tau_l(t_{i,j})}{\sum_{m=0}^{J_{k,i}^X - 1} \tau_l(y_{k,l,n,i,m})}, \quad j < J_i^T. \end{aligned}$$

4 Погрешность вычисления функции F

Обозначим через $\varepsilon_{k,l,n}$ погрешность вычисления функции F в случае, когда функция $f_{k,l,n}$ используется в качестве ее приближенного значения. В качестве значения $\varepsilon_{k,l,n}$ используем меру близости между векторами на основе косинуса:

$$\varepsilon_{k,l,n} = \rho(\vec{a}_{k,l,n}, \vec{b}_l) = 1 - \cos(\vec{a}_{k,l,n}, \vec{b}_l),$$

где

$$\begin{aligned} \vec{a}_{k,l,n} &= \{a_{k,l,n,i,j}\}; \\ a_{k,l,n,i,j} &= E(\varphi_n(f_{k,l,n}(t_{i,j})), \varphi_n(f_{k,l,n}(t_{i,j+1}))); \\ \vec{b}_l &= \{b_{l,i,j}\}; \quad b_{l,i,j} = \tau_l(t_{i,j}), \\ & i = 0, 1, \dots, 6, \quad j = 0, 1, \dots, J_i^T - 1; \end{aligned}$$

$$\cos(\vec{a}_{k,l,n}, \vec{b}_l) = \frac{\sum_{i=0}^6 \sum_{j=0}^{J_i^T - 1} a_{k,l,n,i,j} b_{l,i,j}}{\sum_{i=0}^6 \sum_{j=0}^{J_i^T - 1} a_{k,l,n,i,j}^2 \sum_{i=0}^6 \sum_{j=0}^{J_i^T - 1} b_{l,i,j}^2}.$$

Авторами была написана специальная программа, с помощью которой были вычислены значения $\varepsilon_{k,l,n}$ при X_k , $k = 1, 2, \dots, 36$, τ_l , $l = 0, 1$, и φ_n , $n = 0, 1, \dots, 13$.

При выборе градаций носителей ступенчатых тоновых шкал X_k , $k = 1, 2, \dots, 36$, было использовано два способа. При $k \leq 18$ в качестве X_k выбирались подпоследовательности пикселей носителей ступенчатых тоновых шкал с шагом по индексу, равным k , а при $19 \leq k \leq 36$ — с шагом по контрасту, равным числу $1 + 0,5(k - 18)$.

При $l = 0$ значения функции $\tau_l(t_{i,j})$ определялись величиной расстояния между компонентами ступенчатых тоновых шкал, а при $l = 1$ — величиной Lab-контраста между ними.

5 Выбор функции Lab-контрастного градационного преобразования

Обозначим через $\varepsilon_{37,l,n}$ погрешности вида

$$\varepsilon_{37,l,n} = \min_{1 \leq k \leq 36} \varepsilon_{k,l,n},$$

а через \vec{V}_k — векторы вида

$$\vec{V}_k = \{v_{k,l+2n}\} = \{\varepsilon_{k,l,n}\}, \quad 0 \leq l + 2n \leq 27.$$

Необходимо определить, при каком значении j' выполнено соотношение

$$\rho(\vec{V}_{37}, \vec{V}_{j'}) = \min_{1 \leq j \leq 36} \rho(\vec{V}_{37}, \vec{V}_j).$$

Обозначим через $\vec{P}_i = \{p_{i,j}\}$ векторы, составленные из расстояний между векторами \vec{V}_i и \vec{V}_j :

$$p_{i,j} = \rho(\vec{V}_i, \vec{V}_j), \quad i = 1, 2, \dots, 37, \quad j = 1, 2, \dots, 36.$$

Было установлено, что наименьшей координатой вектора \vec{P}_{37} является координата с индексом, равным 24. Следовательно, наиболее близким к вектору \vec{V}_{37} является вектор \vec{V}_{24} (рис. 1). Это означает оптимальность функции $f_{24,l,k}$ и соответствующего метода преобразования ступенчатых тоновых шкал. При определении $f_{24,l,k}$ градации носителей ступенчатых тоновых шкал выбирались с шагом по контрасту, равным четырем.

Среди функций, при определении которых градации выбирались с шагом по индексу, наилучшей оказалась функция $f_{15,l,k}$, поскольку наиболее близким к вектору \vec{V}_{37} среди соответствующего множества векторов является вектор \vec{V}_{15} .

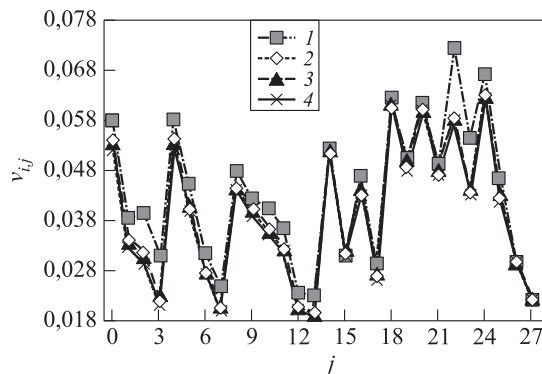


Рис. 1 Соотношение координат векторов: 1 — \vec{V}_1 ; 2 — \vec{V}_{15} ; 3 — \vec{V}_{24} ; 4 — \vec{V}_{37}

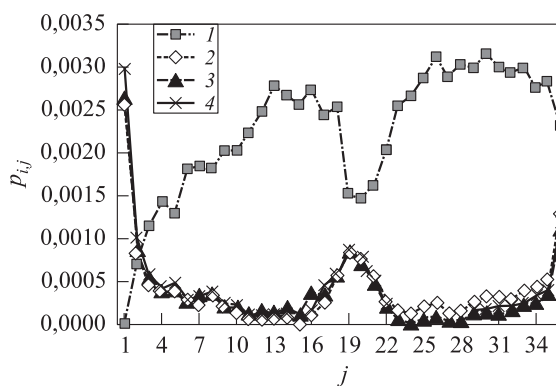


Рис. 2 Соотношение координат векторов: 1 — \vec{P}_1 ; 2 — \vec{P}_{15} ; 3 — \vec{P}_{24} ; 4 — \vec{P}_{37}

Наиболее далеким от вектора \vec{V}_{37} является вектор \vec{V}_1 ; следовательно, $f_{1,l,k}$ является наихудшей среди рассмотренных функций.

Близость векторов \vec{V}_{15} , \vec{V}_{24} , \vec{V}_{37} и их отличие от вектора \vec{V}_1 проявляется и в близости векторов \vec{P}_{15} , \vec{P}_{24} , \vec{P}_{37} и их различии с вектором \vec{P}_1 (рис. 2).

6 Сравнение результатов применения функций $f_{1,l,k}$, $f_{15,l,k}$, $f_{24,l,k}$ на примере

Серая шкала T_0 , исходно имеющая вид, отраженный на рис. 3, а, в восприятии пользователя, цветовосприятие которого описывается функцией φ_{11} ($k = 11$), приобретает зеленоватый оттенок и выглядит как на рис. 3, б. Можно видеть, что в области теней теряется контрастность пикселей.

После равноконтрастного градационного преобразования (рис. 3, в — 3, д) исходный контраст на

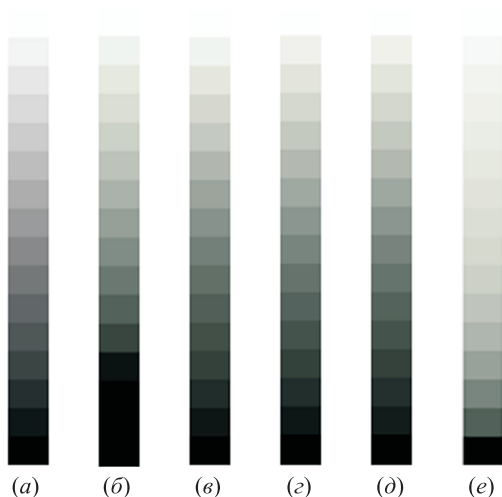


Рис. 3 Вид пикселей шкал: (а) T_0 ; (б) $\varphi_{11}(T_0)$; (в) $f_{1,0,11}(T_0)$; (г) $f_{15,0,11}(T_0)$; (д) $f_{24,0,11}(T_0)$; (е) $f_{24,2,11}(T_0)$

серой шкале практически восстанавливается. При этом после преобразований с помощью функций $f_{1,0,11}$, $f_{15,0,11}$ и $f_{24,0,11}$ (рис. 4) наиболее точное соответствие распределения контрастов образцу соответствует функции $f_{24,0,11}$.

Выбор образца τ_2 (рис. 5) позволяет значительно увеличить контраст между первыми двумя пикселями ступенчатой тоновой шкалы (см. рис. 3, е).

Рассмотрим изображение Img (рис. 6, а). В восприятии пользователя, цветовосприятие которого описывается функцией φ_{11} , Img принимает вид, показанный на рис. 6, б. Как и у отображения серой шкалы, у отображения Img детализация искажается, поскольку уменьшается контрастность пикселей в области теней.

Предварительное Lab-контрастное градиционное преобразование с помощью функции $f_{24,0,11}$ восстанавливает детализацию изображения в восприятии рассматриваемого пользователя (рис. 7, а).

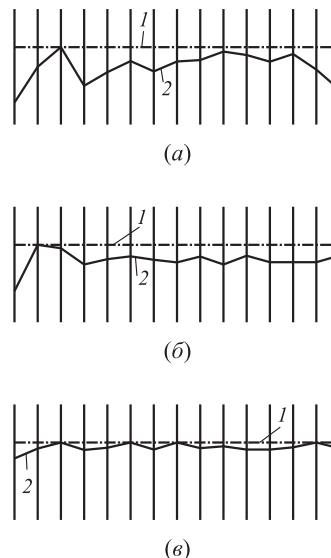


Рис. 4 Соотношение образцов τ_0 (1) и контрастов пикселей преобразованных шкал (2): (а) $f_{1,0,11}(T_0)$; (б) $f_{15,0,11}(T_0)$; (в) $f_{24,0,11}(T_0)$

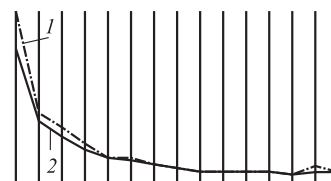


Рис. 5 Соотношение образцов τ_2 (1) и контрастов пикселей преобразованных шкал (2) $f_{24,2,11}(T_0)$

Предварительное Lab-контрастное градиционное преобразование с помощью функции $f_{24,2,11}$ улучшает детализацию изображения в восприятии рассматриваемого пользователя (рис. 7, б), поскольку позволяет выявить новые детали в области теней.



Рис. 6 Изображения: (а) Img ; (б) $\varphi_{11}(\text{Img})$



Рис. 7 Изображения: (а) $\varphi_{11}(f_{24,0,11}(\text{Img}))$; (б) $\varphi_{11}(f_{24,2,11}(\text{Img}))$

7 Заключение

Рассмотрена задача оптимизации функции Lab-контрастного градационного преобразования ступенчатых тоновых шкал.

Задача решена путем сравнения погрешности вычисления функции распределения Lab-контрастов на ступенчатых тоновых шкалах на типичных примерах.

Установлено, что наименьшую погрешность обеспечивает функция Lab-контрастного градационного преобразования ступенчатых тоновых шкал, основанная на выборе градаций носителей ступенчатых тоновых шкал с шагом по контрасту, равным четырем. Эта функция рекомендуется для применения на практике.

Приведен пример, демонстрирующий эффективность применения функции Lab-контрастного градационного преобразования ступенчатых тоновых шкал. Показано, что при выборе подходящего образца можно не только избежать искажения детализации, но и добиться ее улучшения.

Литература

1. Архипов О. П., Зыкова З. П. Допечатное тестирование индивидуального зрительного восприятия // Вестник компьютерных и информационных технологий, 2008. № 12. С. 2–8.
2. Архипов О. П., Зыкова З. П. Интеграция гетерогенной информации о цветных пикселях и их цветовоспри-

ятии // Информатика и её применения, 2010. Т. 4. Вып. 4. С. 14–25.

3. Архипов О. П., Зыкова З. П. Функциональное описание индивидуального цветовосприятия // Информационные системы и технологии, 2010. № 5. С. 5–12.
4. Архипов О. П., Зыкова З. П. RGB-характеризация пространства цветовосприятия // Системы и средства информатики, 2010. Вып. 20. № 1. С. 73–90.
5. Архипов О. П., Зыкова З. П. Многокритериальный выбор тестового множества при исследовании цветовосприятия // Информационные технологии, 2011. № 2. С. 67–73.
6. Архипов О. П., Зыкова З. П. Равноконтрастные градационные преобразования ступенчатых тоновых шкал // Информационные системы и технологии, 2011. № 4. С. 39–46.
7. Архипов О. П., Зыкова З. П. Персонализированное преобразование представлений цветных изображений на мониторе ПЭВМ // Системы и средства информатики, 2012. Т. 22. № 1. С. 22–37.
8. Архипов О. П., Зыкова З. П. Метод улучшения детализации цветных изображений // Вестник компьютерных и информационных технологий, 2013. № 2. С. 6–11.
9. Архипов О. П., Зыкова З. П. Коррекция детализации представлений RGB-изображений на периферийных устройствах ПЭВМ // Информационные технологии, 2013. № 2. С. 56–60.
10. Color Oracle/Institute of Cartography. — Zurich: ETH, 2008. <http://www.colororacle.org>.
11. Vischeck. <http://www.vischeck.com>.

Поступила в редакцию 25.03.13

FUNCTIONS OPTIMIZATION OF LAB-CONTRAST GRADED TRANSFORMATION

O. P. Arkhipov and Z. P. Zykova

Oryol Branch, Institute of Informatics Problems, Russian Academy of Sciences, Oryol 302025, Russian Federation

Abstract: The purpose of this study is to create algorithm for custom conversion of contrast distribution on stepped-tone scales. The idea is to find suitable Lab-contrast distribution, keeping in mind that scaled distortion of digital representation of pixels would keep the real perception of color contrast all the same, if only less bright. In order to approximately calculate the function of Lab-contrast gradual conversion, this study considers two families of parametric algorithms, the first one uses subset of pixels on the scale as parameters, while the another one uses many gradations as they are. The problem of selecting the set of optimal parameters is solved by comparison of range of calculation errors achieved on a set of typical examples by the function of Lab-contrast distribution on stepped-tone scales. The study shows the function of Lab-contrast distribution on stepped-tone scales that gives the least error and provides example that proves that with optimal set of parameters chosen, it is not only possible to avoid visual distortion but the details can also be improved. These findings may be useful in management of maps of RGB-color images for PC peripherals in order to improve perceptive quality of details.

Keywords: color reproduction; color perception; Lab-coordinate; contrast; gradation

DOI: 10.14357/19922264130405

References

1. Arkhipov, O. P., and Z. P. Zykova. 2008. Dopechatnoe testirovanie individual'nogo zritel'nogo vospriyatiya [Preprinting test of individual visual perception]. *Vestnik Komp'yuternykh i Informatsionnykh Tekhnologiy — Herald of Computer and Information Technologies* 12:2–8.
2. Arkhipov, O. P., and Z. P. Zykova. 2010. Integratsiya geterogennoy informatsii o tsvetnykh piksel'yakh i ikh tsvetovospriyatii [Integration of heterogeneous information about color pixels and their color perception]. *Informatika i ee Primeneniya — Inform. Appl.* 4(4):14–25.
3. Arkhipov, O. P., and Z. P. Zykova. 2010. Funktsional'noe opisanie individual'nogo tsvetovospriyatiya [Characteristics of color perceptual space]. *Informatsionnye Sistemy i Tekhnologii [Information Systems and Technologies]* 5:5–12.
4. Arkhipov, O. P., and Z. P. Zykova. 2010. RGB-kharakterizatsiya prostranstva tsvetovospriyatiya [RGB-characterization of color perception space]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 20(1):73–90.
5. Arkhipov, O. P., and Z. P. Zykova. 2011. Mnogokriterial'nyy vybor testovogo mnozhestva pri issledovanii tsvetovospriyatiya [Multicriterion choice of test set when studying the color perception]. *Inform. Tekhnologii [Information Technologies]* 2:67–73.
6. Arkhipov, O. P., and Z. P. Zykova. 2011. Ravnokontrastnyye gradatsionnyye preobrazovaniya stupenchatykh tonovykh shkal [Equal contrast graded transformation of step tinted scales]. *Informatsionnye Sistemy i Tekhnologii [Information Systems and Technologies]* 4:39–46.
7. Arkhipov, O. P., and Z. P. Zykova. 2012. Personifitsirovanoe preobrazovanie predstavleniy tsvetnykh izobrazheniy na monitore PEVM [Personified transformation of color images presentations on a PC monitor]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):22–37.
8. Arkhipov, O. P., and Z. P. Zykova. 2013. Metod uluchsheniya detalizatsii tsvetnykh izobrazheniy [Method of improving the detail of color images]. *Vestnik Komp'yuternykh i Informatsionnykh Tekhnologiy — Herald of Computer and Information Technologies* 2:6–11.
9. Arkhipov, O. P., and Z. P. Zykova. 2013. Korrektsiya detalizatsii predstavleniy RGB-izobrazheniy na periferiynnykh ustroystvakh PEVM [Correcting of detail presentations of RGB-images on peripherals of PC]. *Informatsionnye Tekhnologii [Information Technologies]* 2:56–60.
10. ETH Zurich. 2008. Color Oracle/Institute of Cartography. Available at: <http://www.colororacle.org> (accessed November 5, 2013).
11. Vischeck. 2008. Available at: <http://www.vischeck.com> (accessed November 3, 2013).

Received March 25, 2013

Contributors

Arkhipov Oleg P. (b. 1948) — Candidate of Science (PhD) in technology, Director, Oryol Branch of the Institute of Informatics Problems, Russian Academy of Sciences, Oryol 302025, Russian Federation; arkhipov12@yandex.ru

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, Oryol 302025, Russian Federation; zykzoya@yandex.ru

МЕТОД ИЗВЛЕЧЕНИЯ БИБЛИОГРАФИЧЕСКОЙ ИНФОРМАЦИИ ИЗ ПОЛНОТЕКСТОВЫХ ОПИСАНИЙ ИЗОБРЕТЕНИЙ*

И. М. Зацман¹, В. А. Хавансков², С. К. Шубников³

Аннотация: Предложен метод извлечения библиографической информации из полных описаний патентов, которая необходима для исследования тематических взаимосвязей науки и технологий. Цель исследования заключается в разработке принципов создания отечественных информационных систем для вычисления индикаторов тематических взаимосвязей. Этот вид информационных систем является новым для российской научно-технической сферы. Их создание необходимо для мониторинга и оценивания программ научных исследований и принятия решений на всех этапах программной деятельности. Предлагаемый метод извлечения библиографической информации из текстов на естественном языке (ЕЯ) обладает рядом принципиальных отличий от имеющихся зарубежных и отечественных аналогов. Во-первых, этот метод учитывает тот факт, что в патентных документах библиографическая информация может размещаться внутри текста на ЕЯ. Во-вторых, библиографическая информация является структурированным информационным объектом, который в общем случае является многоязычным.

Ключевые слова: взаимосвязи науки и технологий; методология определения индикаторов; информационные системы; архитектурные решения; библиографическая информация; патентные документы

DOI: 10.14357/19922264130406

1 Введение

На протяжении более 20 лет в ИПИ РАН развивается направление исследований, связанное с автоматической обработкой текстов на ЕЯ с целью извлечения информационных объектов. Основной задачей этих исследований является создание нового класса интеллектуальных систем, основанных на автоматической формализации ЕЯ-текстов с формированием структур знаний для решения логико-аналитических задач [1–6].

Исследование тематических взаимосвязей науки и технологий также базируется на автоматической обработке текстов патентных документов на ЕЯ с целью извлечения библиографической информации, но при этом обладает рядом отличительных черт. Во-первых, библиографическая информация является структурированным информационным объектом, который состоит из нескольких полей и который может размещаться внутри неструктурированного текста. Во-вторых, разные поля библиографической информации могут быть на разных языках, отличных от языка патентного документа на ЕЯ. Другими словами, извлекаемый библиографический информационный объект в общем случае является многоязычным.

До описания предлагаемого метода извлечения библиографической информации из полных описаний патентов, учитывающего отмеченные особенности ее извлечения, рассмотрим кратко ответы на два вопроса, характеризующих актуальность проблематики статьи:

- (1) Каким образом процесс исследования тематических взаимосвязей науки и технологий связан с извлечением библиографической информации из описаний изобретений?
- (2) Как сведения о тематических взаимосвязях науки и технологий могут повлиять на процесс принятия решений о финансировании науки?

Ответ на первый вопрос был получен еще в 1985 г. когда в процессе сопоставительного анализа статей по биологии и библиографических ссылок на статьи в описаниях изобретений по биотехнологиям были экспериментально зафиксированы взаимосвязи между цитируемыми научными публикациями и развитием биотехнологий. В процессе анализа различались патентные ссылки, т. е. ссылки на ранее выданные патенты, и непатентные ссылки, среди которых выделялись ссылки на статьи из журналов. В результате сопоставительного анализа было обработано более 6500 патентных и непатент-

* Работа выполнена при частичной поддержке РГНФ (грант № 12-02-12019в).

¹ Институт проблем информатики Российской академии наук, iz_ipi@i170.ipi.ac.ru

² Институт проблем информатики Российской академии наук, havanskov@i170.ipi.ac.ru

³ Институт проблем информатики Российской академии наук, sergeysh50@yandex.ru

ных ссылок, извлеченных из описаний 399 патентов [7].

Второй вопрос рассматривался в рамках широкомасштабного исследования взаимосвязей науки и технологий с использованием более 1 млн непатентных ссылок, извлеченных из десятилетнего массива описаний изобретений США и Европейского патентного ведомства (ЕПВ). В результате проведенного исследования было экспериментально определено, что 75% научных статей, цитируемых в этих массивах описаний изобретений по широкому спектру технологий, были подготовлены по результатам, полученным именно в некоммерческом секторе научной сферы. Это дало возможность по результатам исследования сделать вывод о сильной зависимости технологического развития от степени государственной поддержки науки в США и европейских странах. Авторы этого исследования приводят пример существенного увеличения государственного финансирования научных проектов в Японии, которое было мотивировано именно выявленной зависимостью технологического развития от степени государственной поддержки науки [8].

При исследовании тематических взаимосвязей науки и технологий один из самых сложных вопросов заключается в том, как зафиксировать факт передачи и использования в технологической сфере результатов научных исследований [9]. Согласно результатам работы [7], один из предложенных и опробованных на практике подходов заключается в том, что процессы передачи знаний от науки к технологиям отслеживаются с помощью научных публикаций, цитируемых экспертами в отчетах о патентном поиске и/или авторами изобретений в их описаниях. Если публикации, цитируемые экспертами, структурно выделены в патентных документах, то публикации, цитируемые авторами патентов, могут быть включены в полнотекстовые описания изобретений. Поэтому исследования тематических взаимосвязей науки и технологий, как правило, основаны на извлеченной из них библиографической информации.

Основная цель статьи заключается в описании метода извлечения библиографической информации из полных описаний патентов. В следующем разделе статьи кратко рассматриваются основные идеи методологии определения тематических взаимосвязей науки и технологий. Подробное описание методологии и реализующей ее технологии, включающей извлечение библиографической информации, дано в работах [10, 11]. Раздел 3 посвящен методу извлечения библиографической информации из полнотекстовых описаний изобретений, за которым следует заключение.

2 Методология определения тематических взаимосвязей науки и технологий

Проведенный в ИПИ РАН анализ зарубежного опыта показал, что вычисление индикаторов тематических взаимосвязей науки и технологий требует формирования, автоматизированной обработки полнотекстовых описаний изобретений с извлечением библиографической информации и сопоставления этой информации как с названиями источников научных публикаций (журналов или материалов конференций), так и с названиями самих статей, хранящихся в научных электронных библиотеках. Это сопоставление дает возможность определить тематику тех научных направлений, к которым относятся научные публикации, цитируемые в описаниях изобретений [10–14].

Разработанный в ИПИ РАН вариант методологии определения значений индикаторов взаимосвязей позволяет вычислять их значения как для целых областей знаний, так и для отдельных направлений исследований [10]. На его основе в настоящее время в ИПИ РАН создается аналитико-информационная система, предназначенная для индикаторного оценивания взаимосвязей науки и одного из направлений технологического развития — информационных технологий (ИТ).

В настоящее время существует несколько вариантов методологии оценивания взаимосвязей науки и технологий. Перечислим основные из них, отмечая примеры реализованных (реализуемых) на практике вариантов, для которых решены теоретические и методологические вопросы.

1. Методология оценивания в статике взаимосвязей отдельных групп технологий при фиксированной системе классификации областей знаний и составляющих их направлений научных исследований на заданном интервале времени. Этот вариант является одним из наиболее проработанных с теоретической и методологической точек зрения. Он был реализован, например, в РФФИ. С его помощью вычислены индикаторы взаимосвязей областей знаний и критических технологий для того массива проектов РФФИ, руководители которых выявили и описали эти взаимосвязи в отчетных формах по своим проектам.
2. Методология оценивания в статике взаимосвязей по широкому спектру групп технологий при фиксированной системе классификации обла-

стей знаний и составляющих их направлений научных исследований на заданном интервале времени (с использованием *основного индекса* Международной патентной классификации — МПК). Этот подход хорошо проработан теоретически и опробован на практике. Он был реализован в рамках информационной системы, разработанной по заказу Еврокомиссии с целью анализа многолетнего массива европейских и американских запатентованных изобретений и оценивания взаимосвязей науки и технологий. Этот вариант используется в настоящее время в директоратах Еврокомиссии.

3. Методология оценивания в статике взаимосвязей по всем группам технологий при фиксированной системе классификации областей знаний и составляющих их направлений научных исследований на заданном интервале времени (с использованием основного и дополнительных индексов МПК). Этот вариант хорошо проработан теоретически. Продемонстрирована его реализуемость в лабораторных условиях. Он был разработан в ИПИ РАН в рамках проекта по гранту РГНФ № 06-02-04043а «Методы мониторинга и оценки инновационного потенциала и результативности направлений научных исследований».

Наибольший интерес при оценивании взаимосвязей науки и технологий представляют не отдельные технологии, а группы технологий по приоритетным направлениям научно-технологического развития. Тематика каждого научного направления, результаты которого используются при разработке технологий, задается в виде одной или нескольких рубрик выбранной системы классификации областей знаний. Тематика каждой анализируемой группы технологий задается в виде списка рубрик МПК. Наиболее часто используются списки рубрик МПК из номенклатуры, разработанной Фраунгоферовским институтом системотехники и инновационных исследований (Fraunhofer Gesellschaft-Institute für Systemtechnik und Innovationsforschung — FhG-ISI) [15].

В номенклатуре FhG-ISI группа ИТ описывается следующими тремя рубриками МПК¹:

- (1) класс G06 — «Вычисление; счет» (эта рубрика МПК включает оптические вычислительные устройства, обработку цифровых данных с помощью компьютеров, аналоговые и гибридные компьютеры);

- (2) подкласс G11C — «Запоминающие устройства статического типа»;

- (3) подкласс G10L — «Анализирование или синтезирование речи; распознавание речи».

Как отмечалось в начале этого раздела, для определения индикаторов взаимосвязей сначала должен быть сформирован массив библиографических описаний научных публикаций, цитируемых в описаниях изобретений. Именно здесь возникает необходимость формирования структурированных библиографических данных на основе анализа слабоструктурированных описаний изобретений. Затем необходимо провести рубрицирование этих данных в рамках заданной классификации научных направлений. Далее необходимо выполнить автоматическое сопоставление частот появления рубрик МПК, проставленных в описаниях изобретений, и частот появления рубрик цитируемых в этих изобретениях научных публикаций (рубрик заданной классификации научных направлений).

Значения частот рубрик научных документов одного направления исследований, цитируемых в изобретениях с рубриками МПК заданной (исследуемой) группы технологий, представляют собой значения индикаторов взаимосвязей этого направления с рубриками МПК заданной группы технологий. Отметим, что значения индикаторов существенно зависят от выбранной классификации научных направлений.

Предлагаемый вариант методологии оценивания индикаторов взаимосвязей научных исследований и технологий предполагает в качестве входных ресурсов использование информационных ресурсов Роспатента. В упрощенном виде входные и выходные информационные ресурсы и их взаимодействие на разных этапах методологии представлены на рис. 1. Используемые патентные ресурсы публикуются Роспатентом на CD/DVD носителях и доступны на его сайте. Выходные информационные ресурсы представляют собой базу данных (БД) структурированных библиографических ссылок на публикации в описаниях изобретений, каждая из которых привязана, с одной стороны, к индексу МПК, с другой стороны, к одной или нескольким рубрикам направлений научных исследований.

В разработанной методологии можно выделить два основных этапа. Первый (подготовительный) этап включает формирование БД структурированных ссылок на публикации, включенных в описания изобретений. Он предполагает обработку

¹ Все рубрики МПК образуют иерархическую структуру и делятся на 5 категорий: разделы (верхний уровень иерархии), классы, подклассы, группы и подгруппы (нижний уровень иерархии).

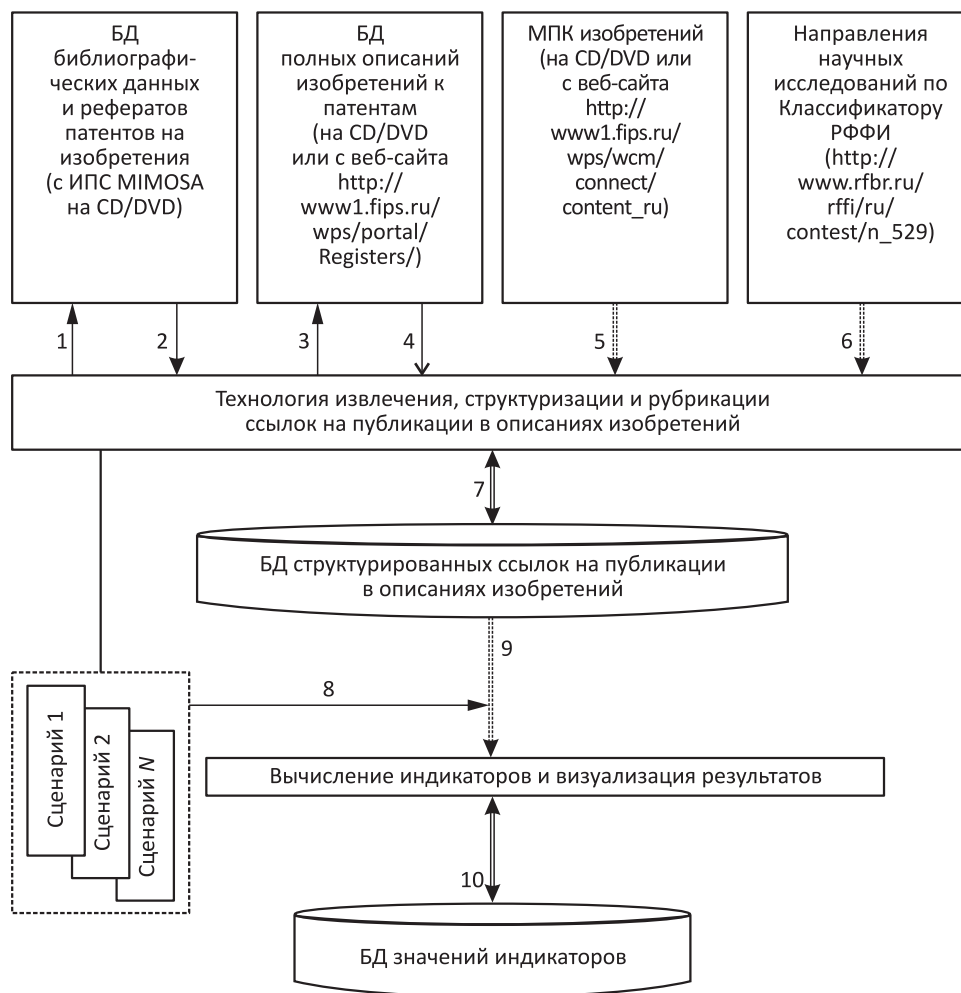


Рис. 1 Входные и выходные информационные ресурсы (описание пронумерованных информационных потоков дано в работе [11])

больших объемов слабоструктурированных описаний изобретений, что трудно поддается полной автоматизации и поэтому требует привлечения операторов. Этот этап может повторяться итерационно в целях пополнения БД. Второй этап включает вычисление значений количественных индикаторов и получение на их основе экспертных оценок взаимосвязей научных исследований и технологий, т. е. на этом этапе предполагается привлечение экспертов.

Задача поиска ссылок на публикации решается с помощью метода автоматического выделения фрагментов неструктурированного текста по заданным признакам с использованием шаблонов, разработанных для разных видов ссылок на публикации. Каждый шаблон строится на основе признаков, характерных для одного вида ссылок на публикации.

Шаблоны формируются на основе данных анализа тестового массива полнотекстовых описаний

изобретений. Все ссылки на публикации в тексте описания изобретения, найденные с помощью шаблонов, выделяются цветом. Подобное выделение дает возможность оператору контролировать точность поиска ссылок на публикации. В ходе контроля оператором точности и полноты поиска ссылок на публикации накапливается статистика по определению успешно выделенных фрагментов текста, которые действительно оказались ссылками цитирования. Такая статистика позволяет увеличить результативность автоматического выделения ссылок цитирования, что, в свою очередь, служит предпосылкой для повышения результативности процесса автоматизированного поиска библиографических ссылок в описаниях изобретений.

Однако ключевым фактором, влияющим на результативность автоматического выделения ссылок цитирования, является используемый метод извлечения, структуризации и нормализации библио-

графической информации, описание которого и является основной целью этой статьи.

3 Метод извлечения библиографической информации

Вычисление значений количественных индикаторов тематических взаимосвязей научных исследований и технологий базируется на предварительной обработке патентных документов, содержащих слабо структурированный или полностью неструктурированный текст. Этими документами являются полнотекстовые описания изобретений. Массив отобранных документов может достигать сотен тысяч полнотекстовых описаний изобретений.

Особенности технологий формирования и использования такого массива наглядно иллюстрирует пример из работы [16], в которой описывается процесс обработки массива из 656 695 патентов на изобретения, выданных Патентным ведомством США. Сначала из описаний изобретений были выделены 1 147 160 непатентных ссылок (т. е. ссылки на патенты были исключены). Затем из них для дальнейшей обработки были отобраны только те ссылки на журнальные статьи, для которых удалось

идентифицировать название журнала и соотнести его с нормативным списком названий журналов, в котором каждому названию присвоена одна или несколько рубрик научных направлений исследований. Таким образом было отобрано 106 636 ссылок, т. е. менее 10% от непатентных ссылок. В результате обработки этих ссылок были вычислены значения индикаторов тематических взаимосвязей групп технологий и научных направлений.

От используемого метода извлечения, структуризации и нормализации зависит доля библиографических данных, для которых удастся идентифицировать название журнала и соотнести его с нормативным списком. Только после такого соотнесения появляется возможность присвоить одну или несколько рубрик научных направлений исследований структурированным ссылкам на публикации, извлеченным из описаний изобретений.

Схема формирования массива патентных документов, используемых для поиска и выделения структурированных ссылок, а также для проведения расчетов значений индикаторов, показана на рис. 2. Этот массив представляет собой список патентов на изобретения и их библиографических описаний, отобранных на основе признаков, описывающих некоторую группу технологий. В качестве базового инструмента для формирования массива используется ИПС (информационно-поисковая система)

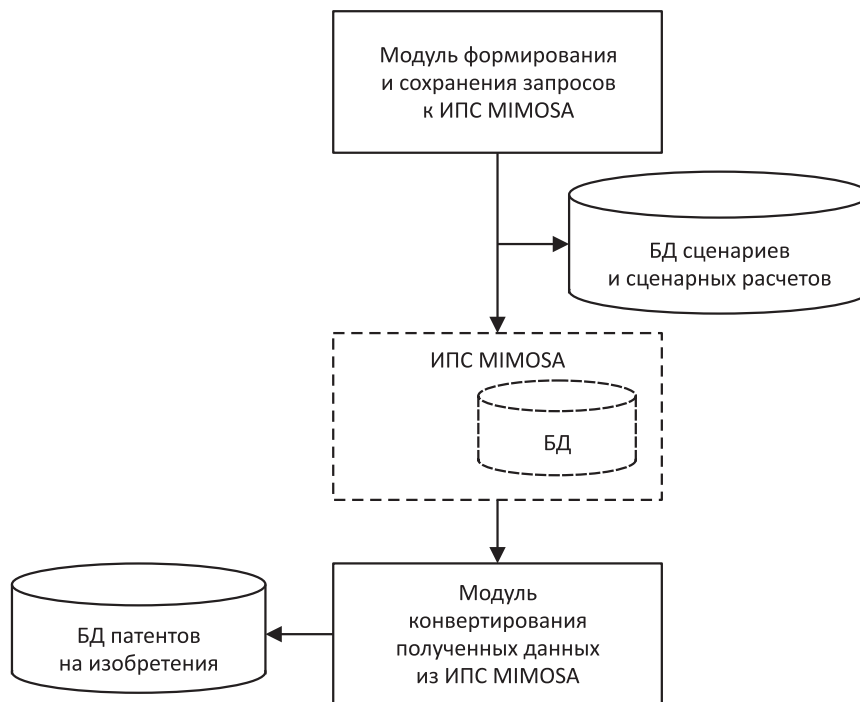


Рис. 2 Схема формирования массива патентных документов

MIMOSA, разработанная в Федеральном институте промышленной собственности (ФИПС).

Для поиска документов в БД библиографических данных и рефератов патентов на изобретения с помощью ИПС MIMOSA существует специально разработанный язык запросов. Индексы МПК в БД библиографических данных содержатся в разных полях, что необходимо учитывать в запросах на поиск. Ниже приводится пример запроса:

(ICA = G06* OR ICAA = G06* OR ICAI = G06*)
AND DP = 2012* ,

где знак «*» служит знакозамениателем. Данный запрос интерпретируется следующим образом: *выбрать все документы, которые содержат код МПК G06 в полях «Классификация – расширенный уровень», «Классификация – дополнительная информация», «Классификация – изобретательская информация» и 2012 год в поле «Дата публикации».*

Информационно-поисковая система MIMOSA состоит из четырех частей. Эти части представляют собой разные программы:

- (1) MIMOSA: это основная часть, предлагающая различные функции для работы с БД: построение запросов, выполнение этих запросов, просмотр документов и т. д.
- (2) MIMOBatch: эта программа позволяет выполнить пакет запросов к одной или нескольким БД за одну операцию; можно просматривать, выгружать или печатать найденные документы.
- (3) MIMOViever: эта программа предназначена в основном для просмотра электронных документов, ранее выгруженных из БД с помощью программ MIMOSA или MIMOBatch; эти документы выгружаются в форме SGML-файлов и изображений.
- (4) JKTools: эта программа запускается только под Windows 95. Она используется для работы с ро-

ботизированными библиотеками в локальном режиме.

Для формирования массива патентных документов, далее используемых для поиска и выделения структурированных ссылок, была спроектирована подсистема подготовки данных, которая использует программу MIMOBatch. С помощью параметров настройки этой программы можно заранее загружать файлы пакетов запросов, выполнять пакетный поиск документов и формировать структурированный текстовый файл библиографических описаний патентов, отвечающих выбранному критерию поиска (рис. 3).

Трудности, возникающие в процессе извлечения ссылок на публикации, цитируемые в полнотекстовых описаниях изобретений, подробно описаны в работе [12]. В числе основных причин этих трудностей отметим:

1. Отсутствие в «Административном регламенте. . .» [17] требований к структурированию ссылок на цитируемые публикации (см. п. (12) раздела регламента 10.11 **Требования к оформлению заявки**).
2. Отсутствие в опубликованных электронных версиях полнотекстовых описаний изобретений групп меток, выделяющих ссылки на цитируемые публикации согласно рекомендациям стандарта ВОИС ST.14 [18].
3. Отсутствие списка нормализованных и сокращенных названий журналов, используемых в ссылках на цитируемые публикации.

Таким образом, при исследовании тематических взаимосвязей технологий и направлений научных исследований возникает задача анализа десятков и сотен тысяч полнотекстовых описаний изобретений и поиска в тексте на ЕЯ ссылок на публикации с последующей их структуризацией и привязкой ссылок к направлениям научных исследований. Следовательно, необходима автоматизация данного процесса. При этом важно учитывать, что биб-

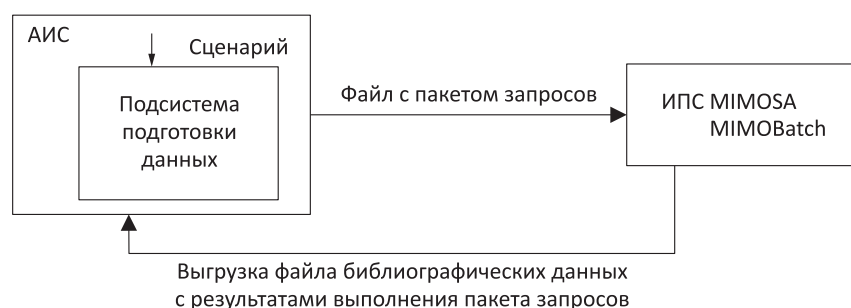


Рис. 3 Взаимодействие подсистемы подготовки данных и ИПС MIMOSA

Примеры ссылок на публикации в текстах описаний изобретений

1. Белоцерковский Г. Б. Основы радиолокации и радиолокационные устройства, М.: Сов. радио, 1975 г.
2. Г. Корн, Т. Корн. Электронные аналоговые и аналого-цифровые вычислительные машины. 1 Теория и основные функциональные блоки. изд-во Москва.: Мир, 1967.
3. Б. М. Каган и В. В. Сташин «Основы проектирования микропроцессорных устройств автоматики», стр.11, М.: Энергоатомиздат, 1987.
4. Шахнович И. Век нынешний и век грядущий. // Электроника: Наука, технология, бизнес, N 6,: 1999, стр.8–11.
5. RODECKU <i>et al.</i> , Tumo growth modulation by monoclonal antibody to the epidermal growth factor receptor immunologically mediated and effector cell independent effects., Cancer Research, 1987, v.47, №14, pp.3692–3696.
6. Высоцкого Б. Ф. Цифровые фильтры и устройства обработки сигналов на интегральных микросхемах. — М.: Радио и связь. 1984 г., стр.46, 54, 55.
7. А. В. Рыжкова, В. Н. Попова «Синтезаторы частот в технике радиосвязи» (М.: Радио и связь, 1991, с.137–142, рис.5.3–5.6)
8. Holden, Science 291: 967(2001).
9. «Радио» №9, 2004, с.47.

лиографическая информация является структурированным информационным объектом, который состоит из нескольких полей и может размещаться внутри неструктурированного текста, а разные поля библиографической информации могут быть в общем случае на разных языках.

Рассмотрим примеры структур объектов поиска в тексте на ЕЯ, т. е. ссылок на публикации, которые собраны в таблицу. Приведенные примеры ссылок на публикации были выделены из реального массива, который был получен в результате обработки порядка 1300 полнотекстовых описаний изобретений.

Из представленной таблицы видно, что в обобщенном виде структура ссылок на публикации может быть представлена следующей схемой:

$$[\text{автор } \{S_1\}][\text{название публикации}] \\ \{ \{S_2\} \text{ название источника} \} \\ \{S_3\} \text{ атрибуты публикации}$$

Наличие квадратных скобок говорит о необязательности присутствия данного элемента схемы в реальной ссылке. $\{S_i\}$ обозначает множество возможных знаков-разделителей, которые стоят между элементами этой схемы. В отличие от элементов [название публикации] и $\{ \{S_2\} \text{ название источника} \}$, которые могут рассматриваться как простые множества слов и знаков пунктуации, элементы [автор $\{S_1\}$] и атрибуты публикации имеют собственные структурные особенности.

Элемент [автор $\{S_1\}$] структурно может иметь следующие варианты представления:

- (а) *Фамилия И.[О.]* — см. ссылки на публикации 1 и 4 из таблицы;
- (б) *И.[О.] Фамилия* — см. ссылки на публикации 2, 3 и 7;

(в) [*Имя*] *Фамилия* или *Фамилия* [*Имя*] — см. ссылку на публикацию 5.

Причем *Фамилия* во всех вариантах представления и *Имя* могут быть набраны как прописными, так и строчными буквами, за исключением первых букв, которые, как правило, всегда являются прописными. Кроме того, в ссылке этот элемент может повторяться в том варианте представления, который принят автором изобретения. В качестве разделителя, как правило, используется запятая.

Элемент $\{S_3\}$ атрибуты публикации имеет произвольную структуру, которая может включать данные об издательстве, томе, номере, страницах и годе публикации. Причем произвольность структуры выражается в том, что порядок следования, наличие и вид представления этих данных могут быть любыми. Кроме того, сам данный элемент может быть заключен в скобки (см. ссылки на публикации 7 и 8 из таблицы). Единственные данные, которые, как правило, всегда присутствуют в элементе $\{S_3\}$ атрибуты публикации, — это данные о годе публикации.

Описанная структура ссылки на публикации использовалась в процессе разработки метода извлечения, структуризации и нормализации библиографической информации полнотекстовых описаний изобретений.

Разработанный метод основан на процессе последовательного перебора символов в линейной последовательности текста на ЕЯ, в которую предварительно преобразуется текст описания изобретения. Лингвистические методы, основанные на автоматической формализации текстов ЕЯ с формированием структур знаний [1–6], не используются.

В процессе перебора выбирается такая последовательность символов, которая удовлетворяет опи-

Приводится в статье Chedmail, Damay et Le Roy, озаглавленной «Réalité virtuelle, maquette numérique du produit, outils de distribution et partage de la conception», Journées Priméca, La Plagne, 7-9 avril 1999.

Рис. 4 Фрагмент текста описания изобретения с частично выделенной ссылкой на публикацию

санию первого элемента структуры ссылки. Далее в случае обнаружения этой последовательности символов ищется такая последовательность символов, которая соответствует описанию последнего элемента структуры. Затем выделяется фрагмент текста, имеющий признаки ссылки на публикацию.

Анализ точности разработанного метода поиска и выделения ссылок дал в целом удовлетворительный результат. Из общего числа ссылок на публикацию в обработанном тестовом массиве описаний изобретений программа выделила около 88% ссылок. Еще 9% ссылок было выделено не полностью, как, например, во фрагменте текста описания изобретения на рис. 4.

Как видно из приведенного фрагмента текста, алгоритм не смог правильно интерпретировать союз ‘et’ в перечислении авторов в элементе [автор {S₁}] описанной выше структуры ссылки на публикацию. Поэтому к знаку разделения перечисляемых авторов (запятой) были добавлены буквенные разделители ‘et’ и ‘u’ (см. ссылку 3 из таблицы). Кроме того, к множеству {S₁}, содержащему, как правило, точку и пробел, добавлено сочетание знаков ‘et al.’ (‘и др.’ в русском варианте) (см. ссылку на публикацию 5) для более точного выделения элемента [автор {S₁}].

Из анализа приведенного примера следует, что для повышения результативности автоматического выделения ссылок цитирования требуется предусмотреть возможность внесения изменений по мере накопления и анализа новых вариантов представления ссылок на публикации, используемых авторами в описаниях изобретений патентов. Для реализации этого подхода предлагается использовать библиотеку шаблонов, описывающих различные варианты представления ссылок, при сохранении единого алгоритма их поиска и выделения в описаниях изобретений патентов.

Для описания шаблонов поиска требуемой последовательности символов при разборе текста описания изобретений используются регулярные выражения, которые широко применяются в различных программных средах (см., например, [19]). Приведем три примера представления элемента [автор {S₁}]:

$$(a) \backslash p\{Lu\} + (\backslash p\{Ll\}+)?[\](\backslash p\{Lu\}\backslash p\{Ll\}?[\])?(\backslash p\{Lu\}[\])[,];$$

$$(b) \backslash p\{Lu\}\backslash p\{Ll\}?[\]?(\backslash p\{Lu\}[\])?[\]?\backslash p\{Lu\} + \backslash p\{Ll\} + ([, \]);$$

$$(v) (\backslash p\{Lu\}\backslash p\{Ll\}+)[\]\backslash p\{Lu\} + \backslash p\{Ll\} + [,][\]$$

Аналогичным образом можно описать варианты представления элемента {S₃} атрибуты публикации. Все сочетания используемых описаний обоих элементов структуры ссылки на публикацию образуют библиотеку шаблонов {R} для поиска ссылок на публикации в описаниях изобретений патентов.

Таким образом, в процессе поиска ссылок в описании изобретения к его тексту применяется библиотека шаблонов {R}. Использование регулярных выражений обеспечивает возможность получения непересекающихся фрагментов текста, содержащих признаки ссылки на публикацию, для отдельно взятого шаблона. Но в то же время разные шаблоны коллекции могут формировать пересекающиеся фрагменты текста. Например, шаблоны (a) и (б) (см. выше) элемента [автор {S₁}] по-разному выделяют в одном и том же фрагменте текста ссылку на публикацию:

(a) *действительному техническому состоянию объекта [Кузнецов В. Е., Лихачев А. М., Парашук И. Б., Присяжнюк С. П. Телекоммуникации. Толковый словарь основных терминов и сокращений. Под редакцией А. М. Лихачева, С. П. Присяжнюка. — СПб.: Издательство МО РФ, 2001.]*

(б) *действительному техническому состоянию объекта [Кузнецов В. Е., Лихачев А. М., Парашук И. Б., Присяжнюк С. П. Телекоммуникации. Толковый словарь основных терминов и сокращений. Под редакцией А. М. Лихачева, С. П. Присяжнюка. — СПб.: Издательство МО РФ, 2001.]*

После применения к тексту библиотечных шаблонов выполняется процедура интеграции выделенных каждым шаблоном фрагментов текста (рис. 5).

На этом рисунке отрезок TT₁ обозначает фрагмент текста описания изобретения; r₁, r₂, ..., r_n — шаблоны библиотеки {R}; s_{i,j} — позиция начала найденного фрагмента текста в соответствии с шаблоном r_i; e_{i,j} — позиция окончания найденного фрагмента текста в соответствии с шаблоном r_i;

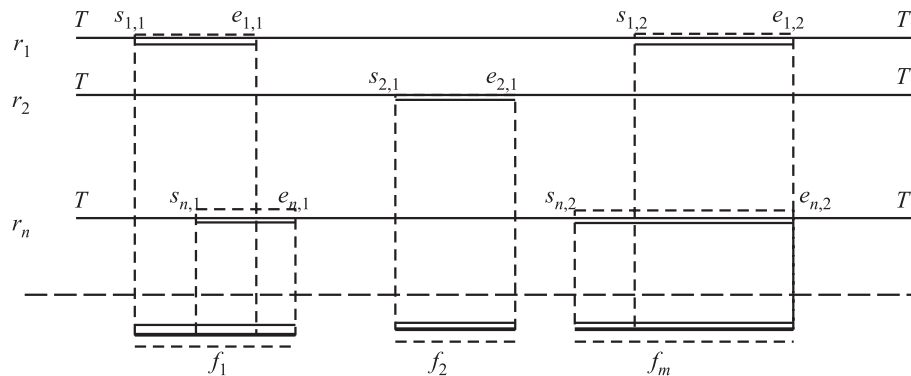


Рис. 5 Интеграция выделенных каждым шаблоном фрагментов текста

f_m — выделенный фрагмент текста, имеющий признаки ссылки на публикацию и являющийся элементом множества $\{F\}$ ссылок на публикации, которые найдены в обрабатываемом фрагменте текста описания изобретения.

Как видно из рис. 5, при нахождении выделенных разными шаблонами фрагментов текста применяется следующее правило: из нескольких значений начала выделенных фрагментов $s_{i,j}$ выбирается минимальное, а для значений окончания выделенных фрагментов $e_{i,j}$ — максимальное.

Функциональная схема подсистемы, обеспечивающей реализацию описанного метода поиска и структуризации ссылок на публикации, представлена на рис. 6. После формирования массива данных для проведения расчетов (см. рис. 2) на вход данной подсистемы поступает список отобранных номеров патентов. По каждому номеру производится запрос в открытый реестр описаний изобретений Роспатента РФ. По запросу в подсистему возвращается полнотекстовое описание изобретения в формате HTML. Далее текст описания программно обрабатывается в соответствии с методом, который описан выше (см. рис. 5).

После завершения процесса поиска и выделения ссылок на публикации программа анализирует число ссылок в данном описании изобретения. При нулевом значении описание изобретения передается на рассмотрение оператору.

В случае обнаружения оператором пропущенных ссылок он формирует множество $\{F\}$ для данного описания изобретения. Кроме того, тексты данных ссылок передаются как образцы на технологическую операцию ведения библиотеки шаблонов. Там эти образцы ссылок на публикации анализируются и при необходимости либо корректируются существующие описания шаблонов, либо формируется новый шаблон для поиска, который добавляется в библиотеку $\{R\}$.

В методологии поиска ссылок цитирования публикаций и их привязки к направлениям научных исследований следующим шагом является структуризация выделенных ссылок и их рубрицирование по областям знаний и научным дисциплинам. Для того чтобы определить научную область знаний статьи, которая цитируется, используется подход, примененный в работе [16]. В этом подходе учитывается то, что научные журналы имеют тематическую направленность и соотнесены с некоторой областью знаний или научной дисциплиной, которая может быть выражена одной или совокупностью рубрик некоторого классификатора научных направлений исследований, например классификатора ГРНТИ или РФФИ. Таким образом, как правило, можно сделать вывод, что статья, опубликованная в журнале, имеет ту же рубрику классификатора, которая указана в его библиографических данных.

Отсюда следует, что процедура рубрицирования ссылки цитируемой публикации в случае журнальной публикации или публикации в сборнике трудов конференции сводится к соотнесению выделенной ссылки цитируемой публикации с названием научного журнала или сборника, указанного в ссылке на публикацию, из нормативного списка журналов.

Это означает, что необходимо выполнить структуризацию выделенной ссылки с целью поиска названия журнала с последующим извлечением из нормативного списка журналов его рубрик по классификатору научных направлений исследований. Иначе говоря, задачей этого этапа методологии является нахождение в ссылке на публикацию названия журнала и года публикации. Первый параметр будет использоваться для рубрицирования ссылки, а второй важен при построении ряда индикаторов тематических взаимосвязей направлений научных исследований и технологий.

При технологическом решении данной задачи используется подход, основанный на описанной

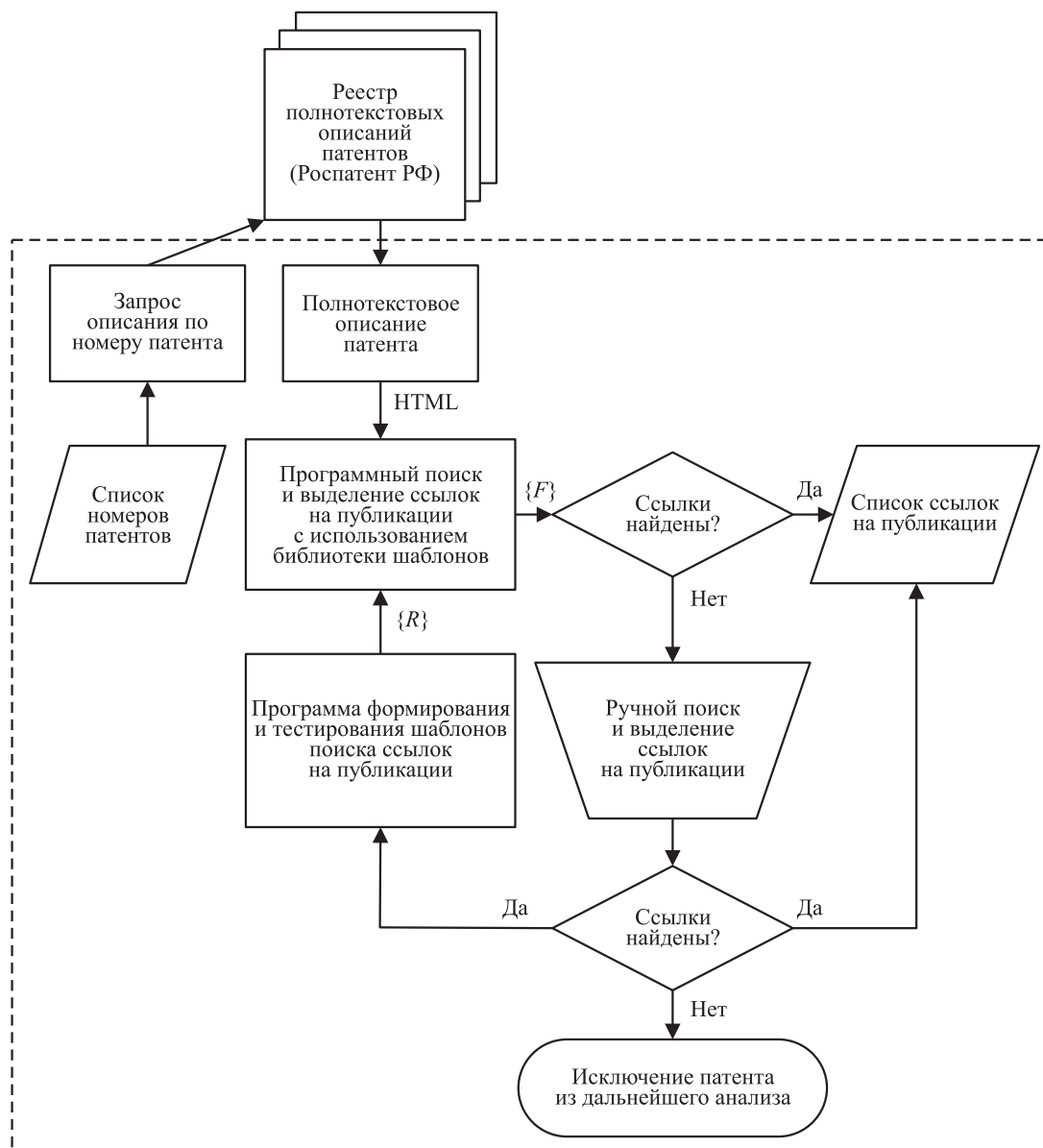


Рис. 6 Функциональная схема поиска ссылок на публикации

в предыдущем разделе схеме структур ссылок на публикации и использовании описания элементов структур и их разделителей с помощью регулярных выражений. При этом для эксперта обеспечивается возможность просмотра, анализа и редактирования результатов рубрицирования ссылок на публикации.

Таким образом, в результате накопления и обработки полнотекстовых описаний патентов на изобретения получается массив записей, в каждой из которых указаны: номер патента, индексы МПК, год публикации патента, ссылки на выявленные публикации с указанием рубрик(и) научных на-

правлений исследований и года публикации статьи. Данный массив может быть использован для вычисления и построения различных индикаторов, описывающих тематические взаимосвязи научных исследований и технологий.

4 Заключение

В работе [10] была описана методология вычисления значений количественных индикаторов тематических взаимосвязей науки и технологий как для целых областей знаний, так и для отдельных

направлений исследований в целях идентификации технологически ориентированных научных направлений. Она обладает рядом принципиальных отличий от зарубежных аналогов:

- выбор именно тех систем классификации областей знаний, которые используются в процессе принятия решений;
- использование научных публикаций, цитируемых экспертами в отчетах о поиске и авторами в описаниях изобретений, для определения значений индикаторов взаимосвязей науки и технологий;
- использование ключевых слов из названий публикаций для уточнения рубрики публикации в тех случаях, когда одному источнику публикации приписано несколько рубрик системы классификации областей знаний.

Использование ключевых слов из названий публикаций для уточнения рубрики публикации предполагает, что предварительно уже выполнено базовое рубрицирование этих публикаций. Разработанный метод извлечения библиографической информации позволяет решить задачу базового рубрицирования в рамках технологии автоматизированной обработки полнотекстовых описаний изобретений и сопоставления извлеченной информации с названиями источников научных публикаций. Это сопоставление дает возможность определить тематику тех научных направлений, к которым относятся научные публикации, цитируемые в описаниях изобретений.

Разрабатываемая в ИПИ РАН информационная система индикаторного оценивания тематических взаимосвязей науки и технологий не имеет аналогов в российской научно-технической сфере. Ее создание необходимо для проведения мониторинга [20–23], многоаспектного оценивания программ научных исследований и прогнозирования научно-технологического развития страны.

Литература

1. *Kuznetsov I., Kozerenko E.* The system for extracting semantic information from natural language texts // Conference (International) on Machine Learning (MLMTA-03) Proceedings. — Las Vegas, 2003. P. 75–80.
2. *Кузнецов И. П.* Семантико-ориентированная система обработки неформализованной информации с выдачей результатов на естественном языке // Системы и средства информатики, 2006. Вып. 16. С. 235–253.
3. *Кузнецов И. П., Мацкевич А. Г.* Семантико-ориентированные системы на основе баз знаний. — М.: МТУСИ, 2007. 173 с.
4. *Кузнецов И. П.* Объектно-ориентированная система, основанная на знаниях в виде XML-представлений // Системы и средства информатики. — М.: Наука, 2008. Вып. 18. С. 96–118.
5. *Kuznetsov I. P., Kozerenko E. B.* Linguistic processor Semantix for knowledge extraction from natural texts in Russian and English // Conference (International) on Artificial Intelligence (ICAI 2008) Proceedings. — Las Vegas: CSREA Press, 2008. P. 835–841.
6. *Кузнецов И. П., Солин Н. В.* Выявление имплицитной информации из текстов на естественном языке: проблемы и методы // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 49–58.
7. *Narin F., Noma E.* Is technology becoming science? // *Scientometrics*, 1985. Vol. 7. No. 3-6. P. 369–381.
8. *Narin F., Olivastro D.* Linkage between patents and papers: An interim EPO/US comparison // *Scientometrics*, 1998. Vol. 41. No. 1-2. P. 51–59.
9. *Schmoch U.* Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. P. 193–211.
10. *Минин В. А., Зацман И. М., Кружков М. Г., Норек-ян Т. П.* Методологические основы создания информационных систем для вычисления индикаторов тематических взаимосвязей науки и технологий // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 70–81.
11. *Минин В. А., Зацман И. М., Хавансков В. А., Шубников С. К.* Архитектурные решения для систем вычисления индикаторов тематических взаимосвязей науки и технологий // Системы и средства информатики, 2013. Т. 23. № 2. С. 260–283.
12. *Зацман И. М., Шубников С. К.* Принципы обработки информационных ресурсов для оценки инновационного потенциала направлений научных исследований // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды 9-й Всеросс. научн. конф. RCDL'2007. — Переславль: Университет города Переславля, 2007. С. 35–44.
13. *Зацман И. М., Курчавова О. А., Галина И. В.* Информационные ресурсы и индикаторы для оценки инновационного потенциала направлений научных исследований // Системы и средства информатики, 2008. Вып. 18 (доп.). С. 159–175.
14. *Кожунова О. С.* Цитирование документов в патентах как индикатор взаимосвязи областей науки и технологий // Системы и средства информатики, 2012. Т. 22. № 2. С. 106–128.
15. *Van Looy B., Zimmermann E., Veugelers R., Verbeek A., Mello J., Debackere K.* Do science–technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains // *Scientometrics*, 2003. Vol. 57. No. 3. P. 355–367.
16. *Verbeek A., Debackere K., Luwel M., Andries P., Zimmermann E., Deleus D.* Linking science to technology: Using bibliographic references in patents to build linkage schemes // *Scientometrics*, 2002. Vol. 54. No. 3. P. 399–420.

17. Административный регламент исполнения Роспатентом приема заявок на изобретение, их рассмотрения и экспертизы. — ФИПС, 2008. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/order_minobr/administrative_regulations/test_8/.
18. Рекомендации по включению ссылок, цитируемых в патентных документах: Стандарт ВОИС ST.14. <http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03.14.01.pdf>.
19. Регулярные выражения в .NET Framework. <http://msdn.microsoft.com/ru-ru/library/hs600312.aspx>.
20. Зацман И. М., Веревкин Г. Ф. Информационный мониторинг сферы науки в задачах программно-целевого управления // Системы и средства информатики, 2006. Вып. 16. С. 164–189.
21. Зацман И. М., Веревкин Г. Ф., Дрынова И. В., Курчавова О. А., Ларин Н. В., Норежан Т. П. Моделирование систем информационного мониторинга как проблема информатики // Системы и средства информатики. Спец. вып. Научно-методологические проблемы информатики, 2006. С. 112–139.
22. Зацман И. М., Кожунова О. С. Семантический словарь системы информационного мониторинга в сфере науки: задачи и функции // Системы и средства информатики, 2007. Вып. 17. С. 124–141.
23. Zatsman I., Kozhunova O. Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators // 2009 Atlanta Conference on Science and Innovation Policy Proceedings / Eds. S. E. Cozzens, P. Catalan. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.

Поступила в редакцию 07.10.13

METHOD OF BIBLIOGRAPHIC INFORMATION EXTRACTION FROM FULL-TEXT DESCRIPTIONS OF INVENTIONS

I. M. Zatsman, V. A. Havansky, and S. K. Shubnikov

Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: The method of bibliographic information extraction from full-text descriptions of inventions, which is necessary for analysis of thematic linkages between science and technologies, is considered. The research objective consists in the development of principles for creation of domestic information systems for calculation of indicators of thematic linkages. This type of information systems is new to the Russian scientific and technical sphere. Their creation is necessary for monitoring and evaluation of research and development programs and decision-making at all stages of program activities. The suggested method of bibliographic information extraction from the texts in natural language differs a lot from available foreign and domestic analogs. First, this method considers the fact that bibliographic information can be found inside the natural language text of descriptions of inventions. Second, paper bibliographic information is the structured information object, which is generally multilingual.

Keywords: linkages between science and technologies; methodology of indicator calculation; information systems; architectural decisions; bibliographic information; patent documents

DOI: 10.14357/19922264130406

Acknowledgments

The research was partially supported by the Russian Foundation for Humanities (grant No. 12-02-12019v).

References

1. Kuznetsov, I., and E. Kozerenko. 2003. The system for extracting semantic information from natural language texts. *Conference (International) on Machine Learning (MLMTA-03) Proceedings*. Las Vegas. 75–80.
2. Kuznetsov, I. P. 2006. Semantiko-orientirovannaya sistema obrabotki neformalizovannoy informatsii s vydachey rezul'tatov na estestvennom yazyke [Semantic-oriented system for processing of nonformalized information with outcomes in a natural language]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:235–253.
3. Kuznetsov, I. P., and A. G. Matskevich. 2007. *Semantiko-orientirovannye sistemy na osnove baz znaniy* [Semantic-oriented systems based on knowledge bases]. Moscow: MTUSI. 173 p.
4. Kuznetsov, I. P. 2008. Ob'ektno-orientirovannaya sistema, osnovannaya na znaniyakh v vide XML-predstavleniy [The object-oriented system based on knowledge in the form of XML representations]. *Sistemy*

- i Sredstva Informatiki — Systems and Means of Informatics* 18:96–118.
5. Kuznetsov, I. P., and E. B. Kozerenko. 2008. Linguistic processor Semantix for knowledge extraction from natural texts in Russian and English. *Conference (International) on Artificial Intelligence (ICAI 2008) Proceedings*. Las Vegas: CSREA Press. 835–841.
 6. Kuznetsov, I. P., and N. V. Somin. 2012. Vyyavlenie implitsitnoy informatsii iz tekstov na estestvennom yazyke: Problemy i metody [Retrieval of implicit information from natural language texts: Problems and methods]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):49–58.
 7. Narin, F., and E. Noma. 1985. Is technology becoming science? *Scientometrics* 7(3-6):369–381.
 8. Narin, F., and D. Olivastro. 1998. Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics* 41(1-2):51–59.
 9. Schmoch, U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* 26:193–211.
 10. Minin, V. A., I. M. Zatsman, M. G. Kruzhkov, and T. P. Norekjan. 2013. Metodologicheskie osnovy sozdaniya informatsionnykh sistem dlya vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologii [Methodological basis for the creation of information systems for the calculation of indicators of thematic linkages between science and technology]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):70–81.
 11. Minin, V. A., I. M. Zatsman, V. A. Havanskov, and S. K. Shubnikov. 2013. Arkhitekturnye resheniya dlya sistem vychisleniya indikatorov tematicheskikh vzaimosvyazey nauki i tekhnologii [Information system conceptual decisions for assessment of linkages between science and technologies]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):260–283.
 12. Zatsman, I., and S. Shubnikov. 2007. Printsipy obrabotki informatsionnykh resurov dlya otsenki innovatsionnogo potentsiala napravleniy nauchnykh issledovaniy [Processing principles of information resources for an assessment of innovation potential of the scientific domains]. *Trudy 9-y Vserossiyskoy nauchnoy konferentsii “Elektronnye biblioteki” [9th All-Russian Scientific Conference on Digital Libraries Proceedings]*. Pereslavl’: Publishing House of Pereslavl’ University. 35–44.
 13. Zatsman, I., O. Kurchavova, and I. Galina. 2008. Informatsionnye resursy i indikatory dlya otsenki innovatsionnogo potentsiala napravleniy nauchnykh issledovaniy [Information resources and indicators for an assessment of innovation potential of the scientific domains]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 18 (add.):159–175.
 14. Kozhunova, O. 2012. Tsitirovanie dokumentov v patentakh kak indikator vzaimosvyazi oblastey nauki i tekhnologii [Citing documents in patents as an indicator for science and technologies linkages]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(2):106–128.
 15. Van Looy, B., E. Zimmermann, R. Veugelers, A. Verbeek, J. Mello, and K. Debackere. 2003. Do science-technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains. *Scientometrics* 57(3):355–367.
 16. Verbeek, A., K. Debackere, M. Luwel, P. Andries, E. Zimmermann, and D. Deleus. 2002. Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(3):399–420.
 17. FIPS. 2008. Administrativnyy reglament ispolneniya Rospatentom priema zayavok na izobreteniye, ikh rassmotreniya i ekspertizy [Rospatent administrative regulations for filing invention applications, their considerations and examination]. http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/order_minobr/administrative_regulations/test_8/ (accessed December 10, 2013).
 18. Standart VOIS ST.14 “Rekomendatsii po vkluyчениyu ssylok, citiruemykh v patentnykh dokumentakh” [WIPO Standard ST.14 “Recommendation for the inclusion of references cited in patent documents”]. http://www.rupto.ru/rupto/nfile/52b8dfc1-1049-11e1-a520-9c8e9921fb2c/03_14_01.pdf (accessed December 10, 2013).
 19. Regulyarnye vyrazheniya v .NET Framework [NET Framework Regular Expressions]. <http://msdn.microsoft.com/ru-ru/library/hs600312.aspx> (accessed December 10, 2013).
 20. Zatsman, I. M., and G. F. Verevkin. 2006. Informatsionnyy monitoring sfery nauki v zadachakh programmno-tselevogo upravleniya [Information monitoring in the science sphere in problems of goal-oriented program management]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:164–189.
 21. Zatsman, I. M. G. F. Verevkin, I. V. Drynova, O. A. Kurchavova, N. V. Larin, and T. P. Norekjan. 2006. Modelirovanie sistem informatsionnogo monitoringa kak problema informatiki [Modeling of systems of information monitoring as informatics problem]. *Sistemy i sredstva informatiki. Nauchno-metodologicheskie problemy informatiki [Systems and means of informatics. Scientific and methodological problems of informatics]*. Moscow: IPI RAN. 112–139.
 22. Zatsman, I., and O. Kozhunova. 2007. Semanticheskii slovar’ sistemy informatsionnogo monitoringa v sfere nauki: Zadachi i funktsii [Semantic vocabulary of the system of information monitoring in scientific sphere: The tasks and functions]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 17:124–141.
 23. Zatsman, I., and O. Kozhunova. 2009. Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators. *2009 Atlantica Conference on Science and Innovation Policy Proceedings*. Eds. S. E. Cozzens, and P. Catalan. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.

Received October 7, 2013

Contributors

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; iz_lipi@a170.ipi.ac.ru

Havanskov Valerij A. (b. 1950) — scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; havanskov@a170.ipi.ac.ru

Shubnikov Sergej K. (b. 1955) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; sergeysh50@yandex.ru

О СХОДИМОСТИ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ СУММ К СКОШЕННЫМ ЭКСПОНЕНЦИАЛЬНО-СТЕПЕННЫМ ЗАКОНАМ*

М. Е. Григорьева¹, В. Ю. Королев²

Аннотация: Предложено обобщение класса экспоненциально-степенных распределений (обобщенных распределений Лапласа) на несимметричный случай. Класс скошенных экспоненциально-степенных распределений (скошенных обобщенных распределений Лапласа) вводится как семейство специальных дисперсионно-сдвиговых смесей нормальных законов. Найдены выражения для моментов скошенных экспоненциально-степенных распределений. Показано, что скошенные экспоненциально-степенные распределения могут использоваться в качестве асимптотических аппроксимаций. С этой целью доказывается теорема о необходимых и достаточных условиях сходимости распределений сумм случайного числа независимых одинаково распределенных случайных величин к скошенным экспоненциально-степенным распределениям. Для частного случая — специальных случайных блужданий с непрерывным временем, порожденных обобщенными дважды стохастическими пуассоновскими процессами, — приводятся оценки скорости этой сходимости.

Ключевые слова: случайная сумма; обобщенное распределение Лапласа; скошенное обобщенное распределение Лапласа; экспоненциально-степенное распределение; симметричное устойчивое распределение; одностороннее устойчивое распределение; дисперсионно-сдвиговая смесь нормальных законов; смешанное пуассоновское распределение; смесь распределений вероятностей; идентифицируемые смеси; аддитивно замкнутое семейство; оценка скорости сходимости

DOI: 10.14357/19922264130407

1 Экспоненциально-степенные распределения

Пусть $0 < \alpha \leq 2$. Экспоненциально-степенным распределением (обобщенным распределением Лапласа) называется абсолютно непрерывное распределение вероятностей, задаваемое левобеговой плотностью

$$l_{\alpha}(x) = \frac{\alpha}{2\Gamma(1/\alpha)} e^{-|x|^{\alpha}}, \quad -\infty < x < \infty. \quad (1)$$

Чтобы упростить обозначения и вычисления, здесь и далее в представлении (1) используется единственный параметр α , поскольку именно этот параметр является в определенном смысле характеристическим и определяет форму экспоненциально-степенного распределения. При $\alpha = 1$ соотношение (1) определяет классическое распределение Лапласа с нулевым средним и дисперсией 2. При $\alpha = 2$ соотношение (1) определяет нормальное (гауссово) распределение с нулевым средним и дисперсией 1/2.

Класс распределений (1) был предложен и изучен М. Ф. Субботиным в 1923 г. [1]. Наряду с термином *обобщенное распределение Лапласа*, восходящим к оригинальной работе [1], для распределения вида (1) используются по крайней мере четыре других термина. В частности, в [2] это распределение названо *экспоненциально-степенным*, в [3] и [4] оно названо *обобщенным распределением ошибок*, в [5] используется термин *обобщенное показательное распределение*, тогда как в работах [6] и [7] это распределение соответственно названо *обобщенным гауссовым* и *обобщенным нормальным*. Распределения типа (1) широко используются в байесовских моделях и разнообразных прикладных задачах от астрономии до обработки сигналов и изображений.

В работе [8] было замечено, что при $0 < \alpha \leq 2$ распределения вида (1) представимы в виде масштабных смесей нормальных законов (этот результат также цитируется в [9]). Пусть $G_{\alpha, \theta}(x)$ и $g_{\alpha, \theta}(x)$ — соответственно функция распределения и плотность строго устойчивого закона с характеристическим показателем α и параметром θ , определяемые характеристической функцией

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515-а, 11-07-00112-а, 12-07-00115-а).

¹ Parexel International, maria-grigoryeva@yandex.ru

² Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, victoryukorolev@yandex.ru

$$g_{\alpha,\theta}(t) = \exp \left\{ -|t|^\alpha \exp \left\{ -\frac{i\pi\theta\alpha}{2} \operatorname{sign} t \right\} \right\}, \quad t \in \mathbb{R},$$

где $0 < \alpha \leq 2$, $|\theta| \leq \theta_\alpha = \min\{1, 2/\alpha - 1\}$ (см., например, [10]). Положим

$$\begin{aligned} h_{\alpha/2}(z) &= \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \cdot \frac{g_{\alpha/2,1}(z)}{\sqrt{z}}, \quad z \geq 0; \\ w_{\alpha/2}(z) &= \frac{h_{\alpha/2}(z^{-1})}{z^2} = \\ &= \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \frac{g_{\alpha/2,1}(z^{-1})}{z^{3/2}}, \quad z \geq 0. \end{aligned}$$

Несложно убедиться, что $h_{\alpha/2}(z)$ и $w_{\alpha/2}(z)$ — плотности распределения. Будем считать, что все случайные величины, упоминаемые в данной статье, заданы на одном и том же достаточно богатом вероятностном пространстве $(\Omega, \mathfrak{A}, P)$. Символ $\stackrel{d}{=}$ обозначает совпадение распределений. Если $V_{\alpha/2}$ и $U_{\alpha/2}$ — неотрицательные абсолютно непрерывные случайные величины с плотностями $h_{\alpha/2}(z)$ и $w_{\alpha/2}(z)$ соответственно, то легко видеть, что

$$U_{\alpha/2} \stackrel{d}{=} \frac{1}{V_{\alpha/2}}. \quad (2)$$

Хорошо известно, что если $\zeta_{\alpha,\theta}$ — случайная величина с устойчивым распределением, соответствующим характеристической функции $g_{\alpha,\theta}$, то $E|\zeta_{\alpha,\theta}|^p < \infty$ для любого $p < \alpha$ (см., например, [10]). Поэтому из определения плотности $h_{\alpha/2}(z)$ вытекает, что $EV_{\alpha/2}^p < \infty$ для любого $p < (\alpha + 1)/2$, а следовательно, из (2) вытекает, что $EU_{\alpha/2}^q < \infty$ для любого $q > -(\alpha + 1)/2$.

Функции распределения, соответствующие плотностям $h_{\alpha/2}(z)$ и $w_{\alpha/2}(z)$, будут обозначаться соответствующими заглавными буквами: $L_{\alpha/2}(x)$, $H_{\alpha/2}(z)$ и $W_{\alpha/2}(z)$. Стандартная нормальная функция распределения ($\alpha = 2$) и ее плотность будут соответственно обозначаться $\Phi(x)$ и $\varphi(x)$:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad \Phi(x) = \int_{-\infty}^x \varphi(z) dz.$$

Лемма 1. При $0 < \alpha \leq 2$ экспоненциально-степенное распределение (1) является масштабной смесью нормальных законов:

$$\begin{aligned} L_{\alpha}(x) &= \int_0^{\infty} \Phi(x\sqrt{z}) dH_{\alpha/2}(z), \quad x \in \mathbb{R}; \\ L_{\alpha}(x) &= \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{z}}\right) dW_{\alpha/2}(z), \quad x \in \mathbb{R}. \end{aligned} \quad (3)$$

Доказательство см. в [8] или [11].

В работе [11] показано, что распределения вида (1) могут выступать в качестве предельных законов в довольно простых предельных теоремах для сумм случайного числа независимых одинаково распределенных случайных величин и статистик, построенных по выборкам случайного объема, и, как следствие, обосновано их применение в качестве асимптотических аппроксимаций для статистических закономерностей в прикладных задачах. В данной статье вводится несимметричное расширение класса распределений (1) и доказываются теоремы, устанавливающие критерии сходимости распределений случайных сумм к несимметричным экспоненциально-степенным законам и скорость этой сходимости.

2 Скошенные экспоненциально-степенные распределения

Все распределения вида (1) симметричны. Рассмотрим естественное несимметричное расширение класса распределений вида (1). С этой целью используем подход, аналогичный тому, с помощью которого в 1977 г. О. Барндорфф-Нильсен ввел класс обобщенных гиперболических распределений как класс специальных дисперсионно-сдвиговых смесей нормальных законов [12]. Основой для соответствующих рассуждений является представление (3).

Пусть $\alpha \in (0, 2]$, $\mu \in \mathbb{R}$. Скошенным экспоненциально-степенным распределением (скошенным обобщенным распределением Лапласа) с параметром формы α и параметром скошенности (асимметрии) μ назовем распределение вероятностей, функция распределения которого имеет вид

$$L_{\alpha,\mu}(x) = \int_0^{\infty} \Phi\left(\frac{x - \mu z}{\sqrt{z}}\right) dW_{\alpha/2}(z), \quad x \in \mathbb{R}. \quad (4)$$

Формально в смеси (4) смешивание производится по обоим параметрам нормального распределения. Однако за счет того, что эти параметры в представлении (4) жестко связаны и математические ожидания (сдвиги) смешиваемых нормальных законов в (4) оказываются пропорциональными дисперсиям, фактически в (4) смешивание производится по одному параметру. В силу этого обстоятельства Барндорфф-Нильсен и его коллеги назвали смеси вида (4) дисперсионно-сдвиговыми (variance-mean mixtures) [13].

Если X — случайная величина со стандартным нормальным распределением, независимая от случайной величины $U_{\alpha/2}$, введенной ранее, то, как несложно видеть, функция распределения $L_{\alpha,\mu}(x)$ (см. (4)) соответствует случайной величине

$$Z_{\alpha,\mu} = X\sqrt{U_{\alpha/2}} + \mu U_{\alpha/2}.$$

Найдем моменты случайной величины $Z_{\alpha,\mu}$. С этой целью сначала вычислим моменты случайной величины $U_{\alpha/2}$. При $\mu = 0$ имеет место соотношение

$$Z_{\alpha,0} \stackrel{d}{=} X\sqrt{U_{\alpha/2}}.$$

При этом из независимости случайных величин X и $U_{\alpha/2}$ вытекает, что для каждого $k \in \mathbb{N}$

$$EZ_{\alpha,0}^{2k} = EX^{2k} \cdot EU_{\alpha/2}^k.$$

Из (1) несложно видеть, что

$$\begin{aligned} EZ_{\alpha,0}^{2k} &= \frac{\alpha}{\Gamma(1/\alpha)} \int_0^\infty x^{2k} e^{-|x|^\alpha} dx = \\ &= \frac{1}{\Gamma(1/\alpha)} \int_0^\infty x^{(2k+1)/\alpha-1} e^{-|x|^\alpha} dx = \frac{\Gamma((2k+1)/\alpha)}{\Gamma(1/\alpha)}. \end{aligned}$$

С другой стороны, как известно, $EX^{2k} = (2k-1)!!$. Поэтому для любого $k \in \mathbb{N}$

$$EU_{\alpha/2}^k = \frac{EZ_{\alpha,0}^{2k}}{EX^{2k}} = \frac{\Gamma((2k+1)/\alpha)}{(2k-1)!!\Gamma(1/\alpha)}.$$

Таким образом,

$$\begin{aligned} EZ_{\alpha,\mu} &= \mu EU_{\alpha/2} = \frac{\mu\Gamma(3/\alpha)}{\Gamma(1/\alpha)}; \\ EZ_{\alpha,\mu}^2 &= EU_{\alpha/2} + \mu^2 EU_{\alpha/2}^2 = \frac{3\Gamma(3/\alpha) + \mu^2\Gamma(5/\alpha)}{3\Gamma(1/\alpha)}; \\ EZ_{\alpha,\mu}^3 &= 3\mu EU_{\alpha/2}^2 + \mu^3 EU_{\alpha/2}^3 = \\ &= \frac{15\mu\Gamma(5/\alpha) + \mu^3\Gamma(7/\alpha)}{15\Gamma(1/\alpha)}; \\ EZ_{\alpha,\mu}^4 &= 3EU_{\alpha/2}^2 + 6\mu^2 EU_{\alpha/2}^3 + \mu^4 EU_{\alpha/2}^4 = \\ &= \frac{105\Gamma(5/\alpha) + 42\mu^2\Gamma(7/\alpha) + \mu^4\Gamma(9/\alpha)}{105\Gamma(1/\alpha)}. \end{aligned}$$

Обозначим

$$\Delta U_{\alpha/2} = \frac{EU_{\alpha/2}}{DU_{\alpha/2}} = \frac{3\Gamma(3/\alpha)\Gamma(1/\alpha)}{\Gamma(1/\alpha)\Gamma(5/\alpha) - 3[\Gamma(3/\alpha)]^2}.$$

Тогда коэффициенты асимметрии (скошенности) $\kappa_3(Z)$ и эксцесса $\kappa_4(Z)$ (островершинности) случайной величины Z принимают вид:

$$\begin{aligned} \kappa_3(Z) &= \frac{E(Z - EZ)^3}{(DZ)^{3/2}} = \\ &= \mu \frac{3DU_{\alpha/2} + \mu^2 E(U_{\alpha/2} - EU_{\alpha/2})^3}{(EU_{\alpha/2} + \mu^2 DU_{\alpha/2})^{3/2}} = \\ &= \frac{\mu\kappa_3(U_{\alpha/2})}{(\mu^2 + \Delta U_{\alpha/2})^{3/2}}; \\ \kappa_4(Z) &= \frac{E(Z - EZ)^4}{(DZ)^2} = \\ &= \frac{1}{(\mu^2 + \Delta U_{\alpha/2})^2} \left[3 \frac{EU_{\alpha/2}^2}{(EU_{\alpha/2})^2} + \right. \\ &\quad \left. + \mu^2 \left(\frac{\kappa_3(U_{\alpha/2})}{\sqrt{DU_{\alpha/2}}} + \mu^2 \kappa_4(U_{\alpha/2}) \right) \right], \quad (5) \end{aligned}$$

где $\kappa_3(U_{\alpha/2})$ и $\kappa_4(U_{\alpha/2})$ — соответственно коэффициенты асимметрии и эксцесса случайной величины $U_{\alpha/2}$:

$$\begin{aligned} \kappa_3(U_{\alpha/2}) &= \frac{E(U_{\alpha/2} - EU_{\alpha/2})^3}{(DU_{\alpha/2})^{3/2}} = \\ &= \left\{ \left[\Gamma\left(\frac{1}{\alpha}\right) \right]^2 \Gamma\left(\frac{7}{\alpha}\right) - 15\Gamma\left(\frac{1}{\alpha}\right) \Gamma\left(\frac{3}{\alpha}\right) \Gamma\left(\frac{5}{\alpha}\right) + \right. \\ &\quad \left. + 30 \left[\Gamma\left(\frac{3}{\alpha}\right) \right]^3 \right\} / \left(5 \left\{ \Gamma\left(\frac{1}{\alpha}\right) \Gamma\left(\frac{5}{\alpha}\right) - \right. \right. \\ &\quad \left. \left. - 3 \left[\Gamma\left(\frac{3}{\alpha}\right) \right]^2 \right\}^{3/2} \right); \\ \kappa_4(U_{\alpha/2}) &= \frac{E(U_{\alpha/2} - EU_{\alpha/2})^4}{(DU_{\alpha/2})^2} = \\ &= \left(\Gamma\left(\frac{9}{\alpha}\right) \left[\Gamma\left(\frac{1}{\alpha}\right) \right]^3 - \right. \\ &\quad - 28 \left[\Gamma\left(\frac{1}{\alpha}\right) \right]^2 \Gamma\left(\frac{3}{\alpha}\right) \Gamma\left(\frac{7}{\alpha}\right) + \\ &\quad + 210\Gamma\left(\frac{1}{\alpha}\right) \left[\Gamma\left(\frac{3}{\alpha}\right) \right]^2 \Gamma\left(\frac{5}{\alpha}\right) - \\ &\quad \left. - 315 \left[\Gamma\left(\frac{3}{\alpha}\right) \right]^4 \right) / \left(105 \left\{ \Gamma\left(\frac{1}{\alpha}\right) \Gamma\left(\frac{5}{\alpha}\right) - \right. \right. \\ &\quad \left. \left. - 3 \left[\Gamma\left(\frac{3}{\alpha}\right) \right]^2 \right\}^2 \right). \end{aligned}$$

Если $\mu = 0$, то смесь (4) является чисто масштабной. Из (5) видно, что в таком случае $\kappa_4(Z) \geq 3$, т.е. при $0 < \alpha < 2$ экспоненциально-степенные распределения вида (1) всегда имеют более острую вершину и, соответственно, более тяжелые хвосты, нежели нормальное распределение.

3 Критерий сходимости распределений случайных сумм к скошенным экспоненциально-степенным распределениям

Пусть $\{X_{n,j}\}_{j \geq 1}, n = 1, 2, \dots$ — семейство последовательностей одинаково распределенных в каждой последовательности (при каждом фиксированном n) случайных величин. Пусть $\{N_n\}_{n \geq 1}$ — последовательность целочисленных неотрицательных случайных величин таких, что при каждом $n \geq 1$ случайные величины $N_n, X_{n,1}, X_{n,2}, \dots$ независимы. Положим

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}.$$

Во избежание недоразумений полагаем $\sum_{j=1}^0 = 0$. Символ \implies будет обозначать слабую сходимость (сходимость по распределению). Везде далее сходимость будет подразумеваться при $n \rightarrow \infty$.

Как показано в [14], дисперсионно-сдвиговые смеси нормальных законов вида (3) оказываются идентифицируемыми, поскольку при каждом фиксированном $\mu \in \mathbb{R}$ однопараметрическое семейство распределений $\{\Phi((x - \mu z)/\sqrt{z}) : z \geq 0\}$ является аддитивно-замкнутым. В статье [14] доказано следующее общее утверждение (см. также [15]).

Теорема 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\mu \in \mathbb{R}$ такие, что*

$$P(S_{n,k_n} < x) \implies \Phi(x - \mu). \quad (6)$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к некоторой функции распределения $F(x)$:

$$P(S_{n,N_n} < x) \implies F(x),$$

необходимо и достаточно, чтобы существовала функция распределения $Q(x)$ такая, что

$$Q(0) = 0; \quad F(x) = \int_0^\infty \Phi\left(\frac{x - \mu z}{\sqrt{z}}\right) dQ(z);$$

$$P(N_n < x k_n) \implies Q(x).$$

Замечание 1. Условие (6) выполняется в следующей довольно общей ситуации. Предположим, что случайные величины $X_{n,j}$ имеют конечные дисперсии. Также предположим, что величины $X_{n,j}$ могут быть представлены в виде

$$X_{n,j} = X_{n,j}^* + \mu_n,$$

где $\mu_n \in \mathbb{R}$, а $X_{n,j}^*$ — случайная величина с $EX_{n,j}^* = 0$, $DX_{n,j}^* = \sigma_n^2 < \infty$, так что $EX_{n,1} = \mu_n$ и $DX_{n,1} = \sigma_n^2$. Предположим, что $\mu_n k_n \rightarrow a$ и $k_n \sigma_n^2 \rightarrow 1$ при $n \rightarrow \infty$. Тогда вследствие хорошо известного результата о необходимых и достаточных условиях сходимости к нормальному закону распределений сумм независимых случайных величин с конечными дисперсиями в схеме серий (см., например, [16]), можно заметить, что соотношение (6) имеет место тогда и только тогда, когда выполнено условие Линдбергга: для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} k_n E(X_{n,1}^*)^2 \mathbb{I}(|X_{n,1}^*| \geq \varepsilon) = 0$$

(здесь $\mathbb{I}(A)$ — индикаторная функция множества (события) A), т.е. квадратичные хвосты распределений слагаемых должны убывать достаточно быстро.

Следствие 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\mu \in \mathbb{R}$ такие, что имеет место сходимость (6). Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость*

$$P(S_{n,N_n} < x) \implies L_{\alpha,\mu}(x), \quad (7)$$

необходимо и достаточно, чтобы

$$P(N_n < x k_n) \implies W_{\alpha/2}(x). \quad (8)$$

Замечание 2. В соотношениях (6), (7) и (8) предельные функции распределения непрерывны. Поэтому в этих соотношениях сходимость по распределению эквивалентна равномерной сходимости функций распределения.

4 Оценки скорости сходимости распределений случайных сумм к скошенным экспоненциально-степенным законам

В данном разделе будет использоваться специальная и довольно естественная конструкция случайных блужданий, удовлетворяющая комплексу условий, указанному в замечании 1.

Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины с $E\xi_1 = 0, 0 < D\xi_1 = 1, \beta^3 = E|\xi_1|^3 < \infty, a \in \mathbb{R}, n$ — натуральное число.

Положим

$$X_{n,j} = \frac{\xi_j}{\sqrt{n}} + \frac{\mu}{n}. \tag{9}$$

В терминах случайных блужданий случайные величины $X_{n,j}$, определенные соотношением (9), могут быть интерпретированы как элементарные приращения процесса, при этом рассматриваемая их конструкция (9) предполагает *одинаковый порядок малости* элементарных трендов и *дисперсий*, что характерно, например, для приращений винеровского процесса со сносом. Обозначим

$$S_n = \sum_{j=1}^n X_{n,j} \left(= \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j + \mu \right).$$

В силу классической центральной предельной теоремы имеем

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |P(S_n < x) - \Phi(x - \mu)| = 0,$$

т.е. так определенные случайные величины $X_{n,j}$ удовлетворяют условию (6) с $k_n = n$.

В книге [17] и статьях [18, 19] предложено моделировать эволюцию неоднородных хаотических стохастических процессов, в частности динамику цен финансовых активов, с помощью обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса). Этот подход получил дополнительное обоснование и развитие в [20–23]. В работах [23, 24] этот подход успешно применен к моделированию процессов плазменной турбулентности. В соответствии с указанным подходом поток информативных событий, в результате каждого из которых появляется очередное «наблюденное» значение рассматриваемой характеристики, описывается с помощью точечного случайного процесса вида $M(\Lambda(t))$, где $M(t), t \geq 0$, — однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t), t \geq 0$, — независимый от $M(t)$ случайный процесс, обладающий следующими свойствами: $\Lambda(0) = 0, P(\Lambda(t) < \infty) = 1$ для любого $t > 0$, траектории $\Lambda(t)$ не убывают и непрерывны справа. Процесс $M(\Lambda(t)), t \geq 0$, называется дважды стохастическим пуассоновским процессом (процессом Кокса). В частности, если процесс $\Lambda(t)$ допускает представление

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0,$$

в котором $\lambda(t)$ — положительный случайный процесс с интегрируемыми траекториями, то $\lambda(t)$ можно интерпретировать как мгновенную стохастическую интенсивность процесса Кокса.

В соответствии с такой моделью в каждый момент времени t распределение случайной величины $M(\Lambda(t))$ является смешанным пуассоновским. С практической точки зрения для описания статистических закономерностей поведения интенсивности потока информативных событий удобно использовать такую гибкую модель, как обобщенное обратное гауссовское распределение. Для большей наглядности рассмотрим случай, когда в рассматриваемой модели время t остается фиксированным, а $\Lambda(t) = nU_{\alpha/2}$, где n — вспомогательный параметр; $U_{\alpha/2}$ — введенная выше случайная величина, имеющая плотность распределения $w_{\alpha/2}(x)$, независимая от стандартного пуассоновского процесса $M(t), t \geq 0$. При этом асимптотика $n \rightarrow \infty$ может интерпретироваться как то, что (случайная) интенсивность потока информативных событий считается очень большой, а при использовании подобных моделей в задачах финансовой математики распределение случайной величины $U_{\alpha/2}$ довольно естественно отождествляется со статистическими закономерностями поведения (случайной) волатильности. Для каждого натурального n положим

$$N_n = M(nU_{\alpha/2}).$$

Очевидно, что так определенная случайная величина N_n имеет смешанное пуассоновское распределение:

$$P(N_n = k) = P(M(nU_{\alpha/2}) = k) = \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} w_{\alpha/2}(z) dz, \quad k = 0, 1, \dots \tag{10}$$

Обозначим $A_n(z) \equiv A_n(z; \nu, \mu, \lambda) = P(N_n < nz), z \geq 0$ ($A_n(z) = 0$ при $z < 0$). Несложно видеть, что $A_n(z) \implies W_{\alpha/2}(z)$ (см., например, [15]), т.е. так определенные случайные величины N_n удовлетворяют условию (8) с $k_n = n$. Впредь будем считать, что при каждом $n \geq 1$ случайная величина N_n независима от последовательности $\{\xi_j\}_{j \geq 1}$, что гарантирует независимость случайных величин $N_n, X_{n,1}, X_{n,2}, \dots$.

Таким образом, в силу непрерывности функции распределения $W_{\alpha/2}(x)$ из следствия 1 вытекает, что

$$D_n \equiv \sup_{x \in \mathbb{R}} \left| P \left(\sum_{j=1}^{N_n} X_{n,j} < x \right) - L_{\alpha,\mu}(x) \right| \longrightarrow 0 \quad (n \rightarrow \infty)$$

(см. замечание 2).

Скорость стремления D_n к нулю описывается следующим утверждением.

Теорема 2. Для любого $n \geq 1$ справедлива оценка

$$D_n \leq 0,5681 \frac{\alpha}{\Gamma(1/\alpha)} \frac{\beta^3}{\sqrt{n}} + 0,1210 \frac{\mu^2}{n}.$$

Доказательство. Как уже было показано, распределение случайной величины N_n является смешанным пуассоновским (см. (10)). Следовательно, по теореме Фубини

$$\begin{aligned} P\left(\sum_{j=1}^{N_n} X_{n,j} < x\right) &= P\left(\sum_{j=1}^{M(nU_{\alpha/2})} X_{n,j} < x\right) = \\ &= \int_0^\infty P\left(\sum_{j=1}^{M(nz)} X_{n,j} < x\right) w_{\alpha/2}(z) dz. \end{aligned}$$

При этом

$$EX_{n,j} = \frac{\mu}{n}; \quad DX_{n,j} = \frac{1}{n}; \quad E|X_{n,j} - EX_{n,j}|^3 = \frac{\beta^3}{n^{3/2}}.$$

Таким образом, при каждом $z \in (0, \infty)$

$$E \sum_{j=1}^{M(nz)} X_{n,j} = \mu z;$$

$$D \sum_{j=1}^{M(nz)} X_{n,j} = nz \left(\frac{\mu^2}{n^2} + \frac{1}{n} \right) = z \left(1 + \frac{\mu^2}{n} \right).$$

Из (10) вытекает, что

$$\begin{aligned} D_n = \sup_x \left| \int_0^\infty w_{\alpha/2}(z) \left[P\left(\sum_{j=1}^{M(nz)} X_{n,j} < x\right) - \right. \right. \\ \left. \left. - \Phi\left(\frac{x - \mu z}{\sqrt{z(1 + \mu^2/n)}}\right) + \Phi\left(\frac{x - \mu z}{\sqrt{z(1 + \mu^2/n)}}\right) - \right. \right. \\ \left. \left. - \Phi\left(\frac{x - \mu z}{\sqrt{z}}\right) \right] dz \right| \leq I_1 + I_2, \quad (11) \end{aligned}$$

где

$$I_1 = \int_0^\infty w_{\alpha/2}(z) \sup_x \left| P\left(\sum_{j=1}^{M(nz)} X_{n,j} < x\right) - \Phi\left(\frac{x - \mu z}{\sqrt{z(1 + \mu^2/n)}}\right) \right| dz;$$

$$I_2 = \int_0^\infty w_{\alpha/2}(z) \sup_x \left| \Phi\left(\frac{x - \mu z}{\sqrt{z(1 + \mu^2/n)}}\right) - \Phi\left(\frac{x - \mu z}{\sqrt{z}}\right) \right| dz.$$

В дальнейшем понадобится следующее утверждение.

Лемма 2. Пусть случайные величины X_1, X_2, \dots одинаково распределены. Пусть N_λ — пуассоновская случайная величина с параметром $\lambda > 0$. Предположим, что случайные величины $N_\lambda, X_1, X_2, \dots$ независимы в совокупности. Обозначим

$$S_\lambda = X_1 + \dots + X_{N_\lambda}.$$

Тогда

$$\begin{aligned} \sup_x \left| P(S_\lambda < x) - \Phi\left(\frac{x - ES_\lambda}{\sqrt{DS_\lambda}}\right) \right| &\leq \\ &\leq \frac{0,4532}{\sqrt{\lambda}} \cdot \frac{E|X_1 - EX_1|^3}{(DX_1)^{3/2}}. \end{aligned}$$

Доказательство леммы 2 приведено в работе [25] (см. также [9, с. 144]).

Продолжим доказательство теоремы 2. Рассмотрим I_1 . Применяя лемму 2, получаем

$$\begin{aligned} I_1 &\leq 0,4532 \frac{\beta^3}{\sqrt{n}} \int_0^\infty \frac{w_{\alpha/2}(z)}{\sqrt{z}} dz = \\ &= 0,4532 \frac{\beta^3}{\sqrt{n}} EU_{\alpha/2}^{-1/2}. \quad (12) \end{aligned}$$

Лемма 3. Для любого $\alpha \in (0, 2]$

$$EU_{\alpha/2}^{-1/2} = \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}}.$$

Доказательство см. в [11].

Продолжив (12) с учетом леммы 3, получим

$$I_1 \leq 0,5681 \frac{\alpha}{\Gamma(1/\alpha)} \frac{\beta^3}{\sqrt{n}}.$$

Рассмотрим I_2 . В дальнейшем понадобится еще одно вспомогательное утверждение.

Лемма 4. Пусть $b \in \mathbb{R}$, $0 < c < \infty$, $0 < d < \infty$. Тогда

$$\sup_y |\Phi(y) - \Phi(cy)| \leq \frac{1}{\sqrt{2\pi e}} \left| \max\left\{c, \frac{1}{c}\right\} - 1 \right|; \quad (13)$$

$$\sqrt{1+d} - 1 \leq \frac{d}{2}. \quad (14)$$

Элементарное доказательство неравенств (13) и (14) можно получить, например, с помощью формулы Лагранжа.

Продолжим доказательство теоремы 2. В лемме 4 положим

$$y = \frac{x - \mu z}{\sqrt{z(1 + \mu^2/n)}}; \quad c = \sqrt{1 + \frac{\mu^2}{n}}.$$

Тогда $c \geq 1$ и в силу утверждения (13) леммы 4 имеем:

$$I_2 \leq \frac{1}{\sqrt{2\pi e}} \left(\sqrt{1 + \frac{\mu^2}{n}} - 1 \right).$$

При этом в силу утверждения (14) леммы 4

$$\sqrt{1 + \frac{\mu^2}{n}} - 1 \leq \frac{\mu^2}{2n}.$$

Окончательно получаем

$$I_2 \leq \frac{\mu^2}{2\sqrt{2\pi e} \cdot n}. \quad (15)$$

Подставляя (12) и (15) в (11), получаем утверждение теоремы. Теорема доказана.

Для случая симметричных экспоненциально-степенных распределений (т. е. $\mu = 0$) оценку, представленную в теореме 2, можно уточнить не только за счет того, что в таком случае обнуляется второе слагаемое в правой части, но и за счет уменьшения коэффициента при первом слагаемом. В работе [11] с использованием оценки точности нормальной аппроксимации для пуассоновских сумм, полученной в [26] в терминах начальных моментов, доказано следующее утверждение.

Теорема 3. Пусть в дополнение к условиям теоремы 1 $\mu = 0$. Тогда

$$D_n \leq 0,3812 \frac{\alpha}{\Gamma(1/\alpha)} \frac{\beta^3}{\sqrt{n}}.$$

Литература

1. *Subbotin M. T.* On the law of frequency of error // Матем. сб., 1923. Т. 31. № 2. С. 296–301.
2. *Box G., Tiao G.* Bayesian inference in statistical analysis. — Reading, MA: Addison–Wesley, 1973. 608 с.
3. *Evans M., Hastings N., Peacock J. B.* Statistical distributions. — 3rd ed. — N.Y.: John Wiley & Sons, 2000. 170 p.
4. *Leemis L. M., McQueston J. T.* Univariate distribution relationships // Amer. Stat., 2008. Vol. 62. No. 1. P. 45–53.
5. RiskMetrics Technical Document. — N.Y.: RiskMetric Group, J. P. Morgan, 1996.
6. *Varanasi M. K., Aazhang B.* Parametric generalized Gaussian density estimation // J. Acoust. Soc. Am., 1989. Vol. 86. No. 4. P. 1404–1415.
7. *Nadaraja S.* A generalized normal distribution // J. Appl. Stat., 2005. Vol. 32. No. 7. P. 685–694.
8. *West M.* On scale mixtures of normal distributions // Biometrika, 1987. Vol. 74. No. 3. P. 646–648.
9. *Choy S. T. B., Smith A. F. F.* Hierarchical models with scale mixtures of normal distributions // Test, 1997. Vol. 6. P. 205–221.
10. *Золотарев В. М.* Одномерные устойчивые распределения. — М.: Наука, 1983. 304 с.
11. *Korolev V. Yu., Bening V. E., Zaks L. M., Zeifman A. I.* Exponential power distributions as asymptotic approximations in applied probability and statistics // Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems (APTP + MS'2012): Book of Abstracts of the 6th Workshop (International) (Autumn Session). — М.: ИПИ РАН, 2012. P. 60–71.
12. *Barndorff-Nielsen O. E.* Exponentially decreasing distributions for the logarithm of particle size // Proc. R. Soc. Lond. Ser. A, 1977. Vol. 353. P. 401–419.
13. *Barndorff-Nielsen O. E., Kent J., Sørensen M.* Normal variance-mean mixtures and z -distributions // Int. Stat. Rev., 1982. Vol. 50. No. 2. P. 145–159.
14. *Королев В. Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // Теория вероятностей и ее применения, 2013. Т. 58. Вып. 1. С. 117–132.
15. *Закс Л. М., Королев В. Ю.* Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 105–115.
16. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. — М.–Л.: ГИТТЛ, 1949. 264 с.
17. *Gnedenko B. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996. 275 p.
18. *Королев В. Ю.* Построение моделей распределений биржевых цен при помощи методов асимптотической теории случайного суммирования // Обзорные промышленной и прикладной математики. Сер. Финансовая и страховая математика, 1997. Т. 4. Вып. 1. С. 86–102.
19. *Королев В. Ю.* Асимптотические свойства экстремумов обобщенных процессов Кокса и их применение к некоторым задачам финансовой математики // Теория вероятностей и ее применения, 2000. Т. 45. Вып. 1. С. 182–194.
20. *Bening V., Korolev V.* Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002. 434 p.
21. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008. 200 с.

22. Королев В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
23. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011. 510 с.
24. Stochastic models of structural plasma turbulence / Eds. V. Korolev, N. Skvortsova. — Utrecht: VSP, 2006. 400 p.
25. Королев В. Ю., Шевицова 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.
26. Korolev V., Shevtsova I. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // *Scand. Actuar. J.*, 2012. No. 2. P. 81–105. Available online since June 4, 2010. DOI:10.1080/03461238.2010.485370.

Поступила в редакцию 10.01.13

ON CONVERGENCE OF THE DISTRIBUTIONS OF RANDOM SUMS TO SKEW EXPONENTIAL POWER LAWS

M. E. Grigor'eva¹ and V. Yu. Korolev^{2,3}

¹Parexel International, Moscow 121609, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation

³Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: An extension of the class of exponential power distributions (also known as generalized Laplace distributions) to the nonsymmetric case is proposed. The class of skew exponential power distributions (skew generalized Laplace distributions) is introduced as a family of special variance-mean normal mixtures. Expressions for the moments of skew exponential power distributions are given. It is demonstrated that skew exponential power distributions can be used as asymptotic approximations. For this purpose, a theorem is proved establishing necessary and sufficient conditions for the convergence of the distributions of sums of a random number of independent identically distributed random variables to skew exponential power distributions. Convergence rate estimates are presented for a special case of random walks generated by compound doubly stochastic Poisson processes.

Keywords: random sum; generalized Laplace distribution; skew generalized Laplace distribution; exponential power distribution; symmetric stable distribution; one-sided stable distribution; variance-mean normal mixture; mixed Poisson distribution; mixture of probability distributions; identifiable mixtures; additively closed family; convergence rate estimate

DOI: 10.14357/19922264130407

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (projects Nos. 11-01-00515-a, 11-07-00112-a, and 12-07-00115-a).

References

- Subbotin, M. T. 1923. On the law of frequency of error. *Mat. Sb.* 31(2):296–301.
- Box, G., and G. Tiao. 1973. *Bayesian inference in statistical analysis*. Reading, MA: Addison–Wesley. 608 p.
- Evans, M., N. Hastings, and J. B. Peacock. 2000. *Statistical distributions*. 3rd. ed. N.Y.: John Wiley & Sons. 170 p.
- Leemis, L. M., and J. T. McQueston. 2008. Univariate distribution relationships. *Amer. Stat.* 62(1):45–53.
- RiskMetric Group, J. P. Morgan. 1996. RiskMetrics Technical Document. N.Y.
- Varanasi, M. K., and B. Aazhang. 1989. Parametric generalized Gaussian density estimation. *J. Acoust. Soc. Am.* 86(4):1404–1415.
- Nadaraja, S. 2005. A generalized normal distribution. *J. Appl. Stat.* 32(7):685–694.
- West, M. 1987. On scale mixtures of normal distributions. *Biometrika* 74(3):646–648.
- Choy, S. T. B., and A. F. F. Smith. 1997. Hierarchical models with scale mixtures of normal distributions. *Test* 6:205–221.
- Zolotarev, V. M. 1983. *Odnomernye ustoychivye raspredeleniya [One-dimensional stable distributions]*. Moscow: Nauka. 304 p.

11. Korolev, V. Yu., V. E. Bening, L. M. Zaks, and A. I. Zeifman. 2012. Exponential power distributions as asymptotic approximations in applied probability and statistics. *Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems (APTP + MS'2012). Book of abstracts of the 6th Workshop (International) (Autumn Session)*. Moscow: IPI RAN. 60–71.
12. Barndorff-Nielsen, O. E. 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. Ser. A* 353:401–419.
13. Barndorff-Nielsen, O. E., J. Kent, and Sørensen M. 1982. Normal variance-mean mixtures and z -distributions. *Int. Stat. Rev.* 50(2):145–159.
14. Korolev, V. Yu. 2013. Obobshchennye giperbolicheskie raspredeleniya kak predel'nye dlya sluchaynykh summ [Generalized hyperbolic distributions as limit distributions for random sums] *Teoriya Veroyatnostey i ee Primeneniya — Theory Probab. Appl.* 58(1):117–132.
15. Zaks, L. M., and V. Yu. Korolev. 2013. Obobshchennye dispersionnyye gamma-raspredeleniya kak predel'nye dlya sluchaynykh summ [Variance-generalized-gamma-distributions as limit laws for random sums]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):105–115.
16. Gnedenko, B. V., and A. N. Kolmogorov. 1949. *Predel'nye raspredeleniya dlya summ nezavisimykh sluchaynykh velichin* [Limit distributions for sums of independent random variables]. Moscow–Leningrad: GITTL. 264 p.
17. Gnedenko, B. V., and V. Yu. Korolev. 1996. Random summation: Limit theorems and applications. Boca Raton: CRC Press. 275 p.
18. Korolev, V. Yu. 1997. Postroenie modeley raspredeleniy birzhevykh tsen pri pomoshchi metodov asimptoticheskoy teorii sluchaynogo summirovaniya [Construction of models for stock prices by methods of the asymptotic theory of random summation]. *Obozrenie Promyshlennoy i Prikladnoy Matematiki* [Surveys in Applied and Industrial Mathematics] 4(1):86–102.
19. Korolev, V. Yu. 2000. Asimptoticheskie svoystva ekstremumov obobshchennykh protsessov Koksya i ikh primeneniye k nekotorym zadacham finansovoy matematiki [Asymptotic properties of extrema of compound Cox processes and their applications to some problems of financial mathematics]. *Teoriya Veroyatnostey i ee Primeneniya — Theory Probab. Appl.* 45(1):182–194.
20. Bening, V., and V. Korolev. 2002. *Generalized Poisson models and their applications in insurance and finance*. Utrecht: VSP. 434 p.
21. Korolev, V. Yu., and I. A. Sokolov. 2008. *Matematicheskie modeli neodnorodnykh potokov ekstremal'nykh sobytii* [Mathematical models of nonhomogeneous flows of extremal events]. Moscow: TORUS PRESS. 200 p.
22. Korolev, V. Yu., V. E. Bening, and S. Ya. Shorgin. 2011. *Matematicheskie osnovy teorii riska* [Mathematical foundations of risk theory]. 2nd ed. Moscow: Fizmatlit. 620 p.
23. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh processov* [Probabilistic and statistical methods for the decomposition of the volatility of chaotic processes]. Moscow: Moscow University Press. 510 p.
24. Korolev, V., and N. Skvortsova, eds. 2006. *Stochastic models of structural plasma turbulence*. Utrecht: VSP. 400 p.
25. Korolev, V. Yu., I. G. Shevtsova, and S. Ya. Shorgin. 2011. O neravenstvakh tipa Berri–Esseena dlya puassonovskikh sluchaynykh summ [On the Berry–Esseen-type inequalities for Poisson random sums]. *Informatika i ee Primeneniya — Inform. Appl.* 5(3):64–66.
26. Korolev, V., and I. Shevtsova. 2012. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scand. Actuar. J.* 2:81–105. Available online since June 4, 2010. DOI:10.1080/03461238.2010.485370.

Received January 10, 2013

Contributors

Grigorieva Maria E. (b. 1986) — biostatistician II, Parexel International, Moscow 121609, Russian Federation; maria-grigoryeva@yandex.ru

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; Moscow 119991, Russian Federation; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; victoryukorolev@yandex.ru

ОБРАЩЕНИЕ СФЕРИЧЕСКОГО ПРЕОБРАЗОВАНИЯ РАДОНА В КЛАССЕ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ФУНКЦИЙ*

О. В. Шестаков¹, М. Г. Кузнецова², И. А. Садовой³

Аннотация: Рассмотрена задача восстановления вероятностных распределений случайных функций по распределениям сферических проекций, описывающих данные в некоторых видах томографических экспериментов, включая термоакустическую томографию, оптоакустическую томографию и радиолокацию. Задачи подобного рода возникают в ситуациях, когда исследуемый объект может случайным образом менять свою структуру в процессе регистрации проекционных данных. Это приводит к невозможности использования традиционных методов вычислительной томографии. Предполагается, что случайный объект может иметь не более счетного числа структурных состояний, каждое из которых описывается интегрируемой функцией с компактным носителем. Для такого дискретного класса случайных функций доказывается взаимно однозначное соответствие между распределением случайной функции и распределениями ее проекций и разрабатывается метод восстановления, основанный на свойствах так называемых моментов проекций. Также исследована устойчивость разработанного метода к погрешностям и показано, что метод дает адекватные результаты в случаях, когда проекционные данные содержат шум.

Ключевые слова: случайные функции; сферическое преобразование Радона; стохастическая томография

DOI: 10.14357/19922264130408

1 Введение

Сферическое преобразование Радона возникает в различных приложениях, включая термоакустическую томографию, оптоакустическую томографию, обработку сигналов, получаемых радарными и сонарами, и т. п. [1–3]. Оно также используется в различных областях теории аппроксимации и математической физики [4].

Пусть $f(x, y)$ — непрерывная функция с компактным носителем. Далее, не ограничивая общности, будем считать что носителем функции $f(x, y)$ является круг U единичного радиуса с центром в начале координат. Определим сферическое преобразование Радона на плоскости следующим образом:

$$Rf(p, r) = \int_{S(p, r)} f(x, y) ds. \quad (1)$$

Здесь $p \in \mathbf{R}^2$; $r \in (0, \infty)$, $S(p, r)$ — окружность с центром в точке p и радиусом r ; ds — мера на $S(p, r)$. В данной работе будет рассмотрен важный для приложений (в частности, томографических) случай, когда центры окружностей $S(p, r)$ располагаются на границе носителя $f(x, y)$, т. е. $p = (\cos \theta, \sin \theta)$,

$\theta \in [0, 2\pi)$. По аналогии с классическим преобразованием Радона интегральные преобразования вида (1) называются сферическими проекциями, т. е. сферическая проекция представляет собой функцию от $r \in (0, \infty)$ при фиксированном p (или в данном случае θ). Такие сферические проекции будем обозначать через $Rf(\theta, r)$.

В ряде томографических приложений (см., например, [5]) функцию $f(x, y)$, описывающую изображение изучаемого объекта, необходимо считать случайной. При этом основной особенностью является то обстоятельство, что состояния (реализации) функции меняются случайным образом во время процесса получения проекций. Это приводит к тому, что восстановление даже одной реализации случайной функции обычными томографическими методами невозможно. Основной интерес в такого рода задачах представляют вероятностные характеристики функции $f(x, y)$.

В работах [6, 7] рассмотрена задача определения вероятностных характеристик двумерных случайных функций в модели классического преобразования Радона. Показано, что в общем случае эта задача характеризуется сильной неоднозначностью и содержательные результаты удается получить лишь

*Работа выполнена при финансовой поддержке Министерства образования и науки РФ (государственный контракт № 14.740.11.0996).

¹Московский государственный университет им. М. В. Ломоносова, факультет ВМК; Институт проблем информатики Российской академии наук, oshestakov@cs.msu.ru

²Московский государственный университет им. М. В. Ломоносова, факультет ВМК, m.g.kuznetsova@gmail.com

³Московский государственный университет им. М. В. Ломоносова, факультет ВМК, isadovoy@gmail.com

в том случае, когда случайная функция имеет не более чем счетное число состояний. В работе [7] для класса таких функций разрабатывается метод восстановления распределений двумерных случайных функций.

В данной работе для класса случайных функций, имеющих не более чем счетное число состояний, разрабатывается метод восстановления распределений случайной функции по распределениям ее сферических проекций.

2 Постановка задачи

Постановка задачи, рассматриваемой в настоящей работе, следующая. Имеется двумерная случайная функция $\xi(x, y)$ — стохастический объект вида $\xi(x, y) = f_\nu(x, y)$, где $f_1(x, y), f_2(x, y), \dots$ — последовательность интегрируемых функций, определенных в единичном круге U , а ν — случайная величина, принимающая целые положительные значения. Вероятностная структура $\xi(x, y)$ полностью определяется набором $(f_1(x, y), f_2(x, y), \dots; p_1, p_2, \dots)$, где $p_i = P(\xi(x, y) = f_i(x, y))$, $i = 1, 2, \dots, \sum_{i=1}^{\infty} p_i = 1$. Распределение $\xi(x, y)$ будем обозначать через P_ξ , а распределения сферических проекций (при фиксированном $\theta \in [0, 2\pi)$) — через $P_{\xi(\theta)}$. Требуется по распределениям $P_{\xi(\theta)}$ для $\theta \in \Lambda$, где Λ — некоторое подмножество $[0, 2\pi)$, определить распределение P_ξ .

3 Теорема единственности

Проблема описания множества точек p , для которого задача обращения преобразования (1) имеет единственное решение, рассматривалась многими авторами (см., например, [1, 2, 4]). В \mathbb{R}^2 эта проблема полностью решена в работе [4]. В частности, в этой работе доказана следующая теорема.

Теорема 1. Пусть P — множество всех точек p , для которых известны сферические проекции вида (1). Функцию $f(x, y)$ с компактным носителем невозможно восстановить единственным образом по преобразованиям $Rf(p, r)$, $p \in P$, $r \in (0, \infty)$, тогда и только тогда, когда

$$P \subset M(\Sigma_N) \cup \Phi.$$

Здесь M — оператор движения (перемещения и/или поворота), Φ — конечное множество; Σ_N — система прямых Коксетера, определяемая следующим образом: для любого натурального N система Σ_N состоит из прямых L_k ($k = 0, \dots, N - 1$), проходящих через начало координат под углом $\pi k/N$ к оси OX .

Кроме того, в работе [2] также доказана еще одна теорема единственности.

Теорема 2. Пусть D — ограниченное открытое множество в \mathbb{R}^2 с гладкой границей ∂D , и пусть \overline{D} (замыкание множества D) представляет собой строго выпуклое множество. Обозначим через Λ любое открытое подмножество ∂D . Если носитель гладкой функции $f(x, y)$ содержится в \overline{D} и $Rf(p, r) = 0$ для всех $p \in \Lambda$ и всех r , тогда $f(x, y) \equiv 0$.

Из этих теорем следует, что в рамках рассматриваемой модели функцию $f(x, y)$ можно однозначно восстановить по значениям сферических проекций $Rf(\theta, r)$, заданным при $\theta \in \Lambda$, где Λ — любое подмножество $[0, 2\pi)$, имеющее положительную меру Лебега.

Используя этот факт, покажем, что распределение двумерной случайной функции полностью определяется распределениями проекций, а именно имеет место следующая теорема.

Теорема 3. Пусть случайные функции $\xi(x, y)$ и $\eta(x, y)$ имеют описанный выше вид и

$$P_{\xi(\theta)} = P_{\eta(\theta)}$$

для всех $\theta \in \Lambda$, где Λ — любое подмножество $[0, 2\pi)$ положительной меры. Тогда

$$P_\xi = P_\eta.$$

Доказательство. Идея доказательства в основном повторяет идеи аналогичной теоремы в работе [6]. Предположим, что $P_\xi \neq P_\eta$. Это означает, что существует функция $f(x, y)$ с носителем в круге U такая, что

$$P(\xi(x, y) = f(x, y)) \neq P(\eta(x, y) = f(x, y)).$$

Обозначим через $f_1(x, y), f_2(x, y), \dots$ значения случайной функции $\xi(x, y)$, отличные от $f(x, y)$, а через $g_1(x, y), g_2(x, y), \dots$ обозначим аналогичные значения $\eta(x, y)$ (таким образом, $f(x, y) \neq f_i(x, y)$ и $f(x, y) \neq g_i(x, y)$, $i = 1, 2, \dots$). Для каждого фиксированного $i = 1, 2, \dots$ пусть A_i обозначает множество всех $\theta \in \Lambda$, для которых

$$Rf(\theta, r) \equiv Rf_i(\theta, r)$$

и, соответственно, B_i — множество всех $\theta \in \Lambda$, для которых

$$Rf(\theta, r) \equiv Rg_i(\theta, r).$$

Каждое из множеств A_i и B_i имеет меру нуль. Если для некоторого i это было бы не так, то $Rf(\theta, r)$ совпадала бы с $Rf_i(\theta, r)$ или с $Rg_i(\theta, r)$ на множестве положительной меры и в силу приведенных

выше утверждений это влекло бы за собой совпадение функции $f(x, y)$ с функцией $f_i(x, y)$ или с функцией $g_i(x, y)$.

Таким образом, множество

$$C = \bigcup_{i=1}^{\infty} (A_i \cup B_i)$$

имеет меру нуль. Следовательно, множество $\Lambda \setminus C$ непусто. Возьмем произвольное $\theta \in \Lambda \setminus C$. Поскольку для этого θ выполнено $Rf(\theta, r) \neq Rf_i(\theta, r)$, $i = 1, 2, \dots$,

$$P(R\xi(\theta, r) = Rf(\theta, r)) = P(\xi(x, y) = f(x, y))$$

и аналогично

$$P(R\eta(\theta, r) = Rf(\theta, r)) = P(\eta(x, y) = f(x, y)).$$

Полученное противоречие доказывает теорему.

4 Группировка проекций

Как следует из предыдущего раздела, в рамках описанной модели возможно восстановить распределение двумерной случайной функции, зная распределения ее сферических проекций на множестве Λ , имеющем положительную меру. В этом разделе будет предложен метод, позволяющий разделить множество зарегистрированных сферических проекций на группы, соответствующие различным состояниям случайной функции. Для удобства будем полагать, что Λ совпадает с $[0, 2\pi)$.

Для простоты изложения в данной работе будет рассматриваться класс функций, имеющих всего два состояния. Обобщение на любое конечное число состояний очевидно, а для случая счетного числа состояний можно произвести процедуру «усечения» распределений сферических проекций, так же как это делается в работе [7].

Итак, пусть случайная функция $\xi(x, y)$ принимает значения $f_1(x, y)$ и $f_2(x, y)$ с вероятностями p_1 и p_2 . Предполагается, что известны распределения сферических проекций для всех $\theta \in [0, 2\pi)$, т. е. для каждого $\theta \in [0, 2\pi)$ известны функции $Rf_i(\theta, r)$, $i = 1, 2$, являющиеся проекциями функций $f_i(x, y)$, $i = 1, 2$, и реализующиеся с вероятностями p_1 и p_2 соответственно. Причем, вообще говоря, заранее не известно, какому состоянию случайной функции соответствует состояние проекции, т. е. может оказаться так, что $Rf_1(\theta, r)$ является проекцией $f_2(x, y)$, а $Rf_2(\theta, r)$ — проекцией $f_1(x, y)$. Необходимо разделить функции $Rf_i(\theta, r)$, $i = 1, 2$, для всех $\theta \in [0, 2\pi)$ на группы так, чтобы каждая группа состояний проекций относилась к одному состоянию случайной функции.

Если $p_1 \neq p_2$, то такое разделение можно произвести по вероятностям состояний сферических проекций, т. е. для всех $\theta \in [0, 2\pi)$ то значение $Rf_i(\theta, r)$, которое реализуется с вероятностью p_1 , относится к первой группе, а значение $Rf_j(\theta, r)$, которое реализуется с вероятностью p_2 , — ко второй.

В случае, когда $p_1 = p_2 = 1/2$, метод группировки основан на использовании некоторых свойств сферических проекций.

Лемма. Пусть $Rf(\theta, r)$ — сферическая проекция функции $f(x, y)$, имеющей носитель в круге U . Тогда интеграл

$$J^{(k)}(\theta) = \int_0^{\infty} r^{2k} Rf(\theta, r) dr, \quad k = 0, 1, \dots, \quad (2)$$

который называется $2k$ -м моментом сферической проекции $Rf(\theta, r)$, представляет собой тригонометрический многочлен от θ степени не выше чем k .

Доказательство.

$$\begin{aligned} & \int_0^{\infty} r^{2k} Rf(\theta, r) dr = \\ &= \int_0^{\infty} r^{2k} \int_0^{2\pi} f(\cos \theta + r \cos \varphi, \sin \theta + r \sin \varphi) r d\varphi dr = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x - \cos \theta)^2 + (y - \sin \theta)^2)^k f(x, y) dx dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2 - 2x \cos \theta - 2y \sin \theta + 1)^k f(x, y) dx dy. \end{aligned}$$

Указанные интегралы существуют, так как функция $f(x, y)$ имеет носитель в круге U . Последний интеграл, очевидно, представляет собой тригонометрический многочлен от $\cos \theta$ и $\sin \theta$ степени не выше чем k . Лемма доказана.

Замечание. Нетрудно убедиться, что при данном $\theta \in [0, 2\pi)$

$$Rf(\theta, r) \equiv 0$$

при $r \in (0, \infty)$ тогда и только тогда, когда

$$\int_0^{\infty} r^{2k} Rf(\theta, r) dr = 0$$

для всех $k = 0, 1, \dots$ (интеграл на самом деле имеет конечные пределы по $r \in (0, 2)$ в силу того,

что функция $f(x, y)$ имеет носитель в круге U . Действительно, условие $Rf(\theta, r) \equiv 0$ эквивалентно тому, что

$$\int_0^{\infty} g(r^2)Rf(\theta, r) dr = 0$$

для любой непрерывной функции $g(r)$, имеющей компактный носитель. В свою очередь, любую такую функцию можно равномерно приблизить многочленом.

Используя приведенные свойства сферических проекций, можно построить метод группировки.

Сначала вычисляются интегралы от $Rf_i(\theta, r)$, $i = 1, 2$, по r для некоторого $\theta \in [0, 2\pi)$. Если эти интегралы отличны друг от друга, то, поскольку их значения не зависят от θ (они представляют собой тригонометрические многочлены нулевой степени, т.е. константы), можно произвести группировку, основываясь на этих значениях. Для этого, вычисляя для каждого $\theta \in [0, 2\pi)$ интегралы по r от $Rf_i(\theta, r)$, $i = 1, 2$, сферические проекции $Rf_i(\theta, r)$ следует относить к той или иной группе в зависимости от того, чему равен вычисленный интеграл. В результате в каждой группе окажутся сферические проекции $Rf_i(\theta, r)$, $\theta \in [0, 2\pi)$, интегралы от которых по r равны одному и тому же значению.

В случае, когда интегралы от $Rf_i(\theta, r)$, $i = 1, 2$, по r совпадают, рассматриваются моменты $J_i^{(k)}(\theta)$. В силу сделанного выше замечания, если все моменты $J_i^{(k)}(\theta)$ двух функций совпадают между собой, то эти функции эквивалентны. Значит, если функции $Rf_i(\theta, r)$, $i = 1, 2$, различны, то найдется номер m , для которого моменты сферических проекций, $J_i^{(m)}(\theta)$, $i = 1, 2$, различаются. Поскольку моменты являются тригонометрическими многочленами от θ , они либо тождественно равны, либо пересекаются в конечном числе точек. Пусть δ задано так, что найдется точка θ , в окрестности которой разница между функциями m -х моментов сферических проекций $Rf_i(\theta, r)$, $i = 1, 2$, по модулю больше δ (т.е. в этой окрестности одна функция момента превышает другую больше чем на δ). Посчитав значения $J_i^{(m)}(\theta_j)$, $i = 1, 2$, в $2m+1$ близких точках θ_j из этой окрестности, можно разделить эти значения на группы, соответствующие каждому состоянию случайной функции (если разности $J_1^{(m)}(\theta_j) - J_2^{(m)}(\theta_j)$ и $J_1^{(m)}(\theta_{j+1}) - J_2^{(m)}(\theta_{j+1})$ имеют разный знак, то значения $J_1^{(m)}(\theta_j)$ и $J_1^{(m)}(\theta_{j+1})$ соответствуют разным состояниям случайной функции и следует поменять местами $Rf_1(\theta_{j+1}, r)$ и $Rf_2(\theta_{j+1}, r)$), а затем найти явный вид функций моментов сферических проекций $J_i^{(m)}(\theta)$, $i = 1, 2$, от переменной θ . Для

этого нужно решить следующие две линейные системы уравнений:

$$a_{m,0}^i + \sum_{n=1}^m a_{m,n}^i \cos(n\theta_j) + \sum_{n=1}^m b_{m,n}^i \sin(n\theta_j) = J_i^{(m)}(\theta_j), \quad j = 0, \dots, 2m, \quad i = 1, 2, \quad (3)$$

и найти коэффициенты $a_{m,n}^i$ и $b_{m,n}^i$ тригонометрических многочленов, которыми представляются функции $J_i^{(m)}(\theta)$.

Вывод явного вида функций $J_i^{(m)}(\theta)$ аналогичен выводу интерполяционного многочлена Лагранжа (подробности можно найти в [8]). Решая (3), имеем

$$J_i^{(m)}(\theta) = \sum_{j=0}^{2m} J_i^{(m)}(\theta_j) \prod_{k=0, k \neq j}^{2m} \frac{\sin((\theta - \theta_k)/2)}{\sin((\theta_j - \theta_k)/2)}, \quad i = 1, 2. \quad (4)$$

Далее при вычислении для каждого θ значения $J_i^{(m)}(\theta)$ по формуле (2) сферическая проекция относится к той или иной группе в зависимости от того, со значением какой из найденных функций (4) в точке θ совпадает это вычисленное значение.

После того как сферические проекции распределены по группам, можно восстановить каждое состояние случайной функции, а значит, и ее распределение, с помощью обычных формул обращения (см., например, [9–11]).

5 Устойчивый метод группировки при наличии погрешностей

На практике имеется конечный набор сферических проекций для $\theta_1, \dots, \theta_N$. Причем, как правило, проекции задаются не точно, а с некоторой погрешностью. Погрешности возникают вследствие несовершенства оборудования, регистрирующего проекции, случайных помех при измерении, ошибок интерполяции и других причин.

Предположим, что сферические проекции каждого состояния случайной функции $\xi(x, y)$ заданы с погрешностью, не превышающей заданного уровня $\varepsilon > 0$:

$$|Rf_i(\theta_l, r) - Rf_i^e(\theta_l, r)| < \varepsilon, \quad r \in [0, 2], \quad i = 1, 2, \quad l = 1, \dots, N,$$

где $Rf_i^e(\theta_l, r)$ — проекции, измеренные с ошибкой.

Тогда значения моментов сферических проекций заданы с погрешностью

$$\begin{aligned} & \left| \int_0^\infty r^{2m} (Rf_i(\theta_l, r) - Rf_i^e(\theta_l, r)) dr \right| = \\ & = \left| \int_0^2 r^{2m} (Rf_i(\theta_l, r) - Rf_i^e(\theta_l, r)) dr \right| \leq \\ & \leq \int_0^2 r^{2m} |Rf_i(\theta_l, r) - Rf_i^e(\theta_l, r)| dr < \\ & < \int_0^2 \varepsilon |r^{2m}| dr < \frac{2^{2m+1}}{2m+1} \varepsilon. \quad (5) \end{aligned}$$

Следовательно, можно считать, что интегралы от $Rf_i(\theta_l, r)$, $i = 1, 2$, $l = 1, \dots, N$, не совпадают, и производить группировку проекций на основании значений этих интегралов, если выполнено условие

$$\left| \int_0^2 (Rf_i(\theta_j, r) - Rf_i^e(\theta_j, r)) dr \right| > 4\varepsilon$$

для некоторого $1 \leq j \leq N$. Если же это условие не выполнено, то можно считать, что интегралы совпадают и разница между ними возникает за счет погрешностей.

Для оценки погрешности, с которой вычисляются функции моментов сферических проекций, воспользуемся известной оценкой погрешности интерполяции тригонометрическими многочленами [8]. Вычислим по формуле (2) моменты $J_i^{(m)}(\theta_n)$ в точках $\theta_n = 2n\pi/(2m+1)$, $i = 1, 2$, $n = 0, \dots, 2m$. Всего существует 2^{2m+1} способов распределить значения $J_l^{(m)}(\theta_n)$, $l = 1, 2$, по двум группам (на практике число $2m+1$, как правило, невелико). Обозначим через H множество всех возможных распределений. Решим системы уравнений

$$\begin{aligned} & a_{m,0}^h + \sum_{j=1}^m a_{m,j}^h \cos(j\theta_n) + \sum_{j=1}^m b_{m,j}^h \sin(j\theta_n) = \\ & = J_{i_n}^{(m)}(\theta_n), \quad i_n^h = 1 \text{ или } 2 \text{ в зависимости от } h, \\ & \quad \quad \quad n = 0, \dots, 2m, \end{aligned}$$

для всех возможных распределений h из H . В результате получится 2^{2m+1} функций (обозначим их через $I_h^{(m)}(\theta)$, $h = 1, \dots, 2^{2m+1}$), претендующих на роль функций моментов $J_i^{(m)}(\theta)$, $i = 1, 2$. Если предположить отсутствие погрешностей, то достаточно, перебирая $I_h^{(m)}(\theta)$, проверять, равно ли значение $I_h^{(m)}(\theta^*)$ какому-либо из значений $J_i^{(m)}(\theta^*)$, $i = 1, 2$, вычисленных по формуле (2) в произвольно выбранной точке θ^* , отличной от точек θ_n . При

том выборе h^* , при котором равенство имеет место, функция $I_{h^*}^{(m)}(\theta)$ совпадает с одной из $J_i^{(m)}(\theta)$, $i = 1, 2$, для всех $\theta \in [0, 2\pi)$, поскольку тригонометрические многочлены степени m , совпадающие более чем в $2m+1$ точках, тождественно равны.

В случае же наличия погрешностей для функции $I_{h^*}^{(m)}(\theta)$, претендующей на роль функции момента $J_i^{(m)}(\theta)$, $i = 1$ или 2 , в силу (5) должно выполняться

$$\left| I_{h^*}^{(m)}(\theta_l) - J_1^{(m)}(\theta_l) \right| \leq \varepsilon \frac{2^{2m+1}}{2m+1} \left(8 + \frac{4}{\pi} \ln m \right)$$

или

$$\left| I_{h^*}^{(m)}(\theta_l) - J_2^{(m)}(\theta_l) \right| \leq \varepsilon \frac{2^{2m+1}}{2m+1} \left(8 + \frac{4}{\pi} \ln m \right)$$

для всех $l = 1, \dots, N$ (см., например, [8]). Поэтому практический алгоритм поиска такой функции $I_{h^*}^{(m)}(\theta_i)$ описывается следующим образом.

1. Выбираем распределение h из множества возможных распределений H и находим функцию $I_h^{(m)}(\theta_l)$, $l = 1, \dots, N$.
2. Полагаем $l = 1$.
3. Проверяем, выполняется ли условие

$$\left| I_h^{(m)}(\theta_l) - J_1^{(m)}(\theta_l) \right| \leq \varepsilon \frac{2^{2m+1}}{2m+1} \left(8 + \frac{4}{\pi} \ln m \right)$$

или

$$\left| I_h^{(m)}(\theta_l) - J_2^{(m)}(\theta_l) \right| \leq \varepsilon \frac{2^{2m+1}}{2m+1} \left(8 + \frac{4}{\pi} \ln m \right),$$

где $J_1^{(m)}(\theta_l)$ и $J_2^{(m)}(\theta_l)$ вычисляются по формуле (2).

4. Если условие не выполнено, то исключаем распределение h из множества возможных распределений H и переходим к шагу 1. Если условие выполнено и $l \neq N$, то полагаем $l = l + 1$ и переходим к шагу 3. Если же условие выполнено и $l = N$, то алгоритм завершает работу и полагаем $I_{h^*}^{(m)}(\theta_l) = I_h^{(m)}(\theta_l)$, $l = 1, \dots, N$. При этом распределение h^* и определяет группы сферических проекций, соответствующие каждому состоянию случайной функции.

Описанный метод является более трудоемким, чем метод, предложенный в предыдущем разделе, но при этом в случае наличия погрешностей он дает более точное приближение функций моментов, что позволяет надеяться на более точное восстановление состояний случайной функции.

Литература

1. *Louis A. K., Quinto E. T.* Local tomographic methods in Sonar // *Surveys on solution methods for inverse problems.* — Vienna: Springer, 2000. P. 147–154.
2. *Finch D., Patch S., Rakesh.* Determining a function from its mean values over a family of spheres // *SIAM J. Math. Anal.*, 2004. Vol. 35. No. 5. P. 1213–1240.
3. *Ambartsoumian G., Kuchment P.* On the injectivity of the circular Radon transform arising in thermoacoustic tomography // *Inverse Probl.*, 2005. Vol. 21. P. 473–485.
4. *Agranovsky M. L., Quinto E. T.* Injectivity sets for the Radon transform over circles and complete systems of radial functions // *J. Funct. Anal.*, 1996. Vol. 139. P. 383–413.
5. *Liu W., Frank J.* Estimation of variance distribution in three-dimensional reconstruction. I. Theory // *J. Opt. Soc. Am. A*, 1995. Vol. 12. P. 2615–2627.
6. *Ушаков В. Г., Ушаков Н. Г.* Восстановление вероятностных характеристик многомерных случайных функций по проекциям // *Вестн. Моск. ун-та. Сер. 15: Вычисл. матем. и киберн.*, 2001. № 4. С. 32–39.
7. *Shestakov O. V.* An algorithm to reconstruct probabilistic distributions of multivariate random functions from the distributions of their projections // *J. Math. Sci.*, 2002. Vol. 112. No. 2. P. 4198–4204.
8. *Намансон И. П.* Конструктивная теория функций. — М.-Л.: ГИТТЛ, 1949. 684 с.
9. *Norton S. J.* Reconstruction of a two-dimensional reflecting medium over a circular domain: Exact solution // *J. Acoust. Soc. Amer.*, 1980. Vol. 67. P. 1266–1273.
10. *Kunyansky L.* Explicit inversion formulas for the spherical mean Radon transform // *Inverse Probl.*, 2007. Vol. 23. P. 373–383.
11. *Finch D., Haltmeier M., Rakesh.* Inversion of spherical means and the wave equation in even dimensions // *SIAM J. Appl. Math.*, 2007. Vol. 68. No. 2. P. 392–412.

Поступила в редакцию 03.03.13

INVERSION OF SPHERICAL RADON TRANSFORM IN THE CLASS OF DISCRETE RANDOM FUNCTIONS

O. V. Shestakov^{1,2}, M. G. Kuznetsova¹, and I. A. Sadovoy¹

¹Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

Abstract: The paper deals with the problem of reconstructing the probabilistic distributions of random functions from distribution of spherical projections that describe the images in certain types of tomographic experiments, including optoacoustic tomography, thermoacoustic tomography, and radiolocation. The problems of this kind arise when the object under study may randomly change its structure during the registration of the projection data and the time within which its structure changes radically is considerably smaller than the time of registration of a required number of projections. In such cases, the conventional tomographic approach cannot be used directly. The authors assume that a random object may have at most countable set of structural states which are described by integrable functions with compact support. For such discrete class of random functions, the uniqueness of solution is proved and the reconstruction method is developed which is based on the properties of the so-called moments of projections. It is shown that the developed method is stable and gives adequate results when the projection data are corrupted by noise.

Keywords: random functions; spherical Radon transform; stochastic tomography

DOI: 10.14357/19922264130408

Acknowledgments

The research was financially supported by the Ministry of Education and Science of the Russian Federation (State Contract No. 14.740.11.0996).

References

1. Louis, A. K., and E. T. Quinto. 2000. Local tomographic methods in Sonar. *Surveys on solution methods for inverse problems*. Vienna: Springer. 147–154.
2. Finch, D., S. Patch, Rakesh. 2004. Determining a function from its mean values over a family of spheres. *SIAM J. Math. Anal.* 35(5):1213–1240.
3. Ambartsoumian, G., and P. Kuchment. 2005. On the injectivity of the circular Radon transform arising in thermoacoustic tomography. *Inverse Probl.* 21:473–485.
4. Agranovsky, M. L., and E. T. Quinto. 1996. Injectivity sets for the Radon transform over circles and complete systems of radial functions. *J. Funct. Anal.* 139:383–413.
5. Liu, W., and J. Frank. 1995. Estimation of variance distribution in three-dimensional reconstruction. I. Theory. *J. Opt. Soc. Am. A* 12:2615–2627.
6. Ushakov, V. G., and N. G. Ushakov. 2001. Vosstanovlenie veroyatnostnykh kharakteristik mnogomernykh sluchaynykh funktsiy po proekciyam [Reconstruction of probabilistic characteristics of multivariate random functions from projections]. *Vestn. Mosk. un-ta. Ser. 15. Vychisl. matem. i kibern.* [Bulletin of Moscow University. Computational mathematics and cybernetics ser.] 4:32–39.
7. Shestakov, O. V. 2002. An algorithm to reconstruct probabilistic distributions of multivariate random functions from the distributions of their projections. *J. Math. Sci.* 112(2):4198–4204.
8. Natanson, I. P. 1949. *Konstruktivnaya teoriya funktsiy* [Constructive theory of functions]. Moscow–Leningrad: GITTL. 684 p.
9. Norton, S. J. 1980. Reconstruction of a two-dimensional reflecting medium over a circular domain: Exact solution. *J. Acoust. Soc. Amer.* 67:1266–1273.
10. Kunyansky, L. 2007. Explicit inversion formulas for the spherical mean Radon transform. *Inverse Probl.* 23:373–383.
11. Finch, D., M. Haltmeier, Rakesh. 2007. Inversion of spherical means and the wave equation in even dimensions. *SIAM J. Appl. Math.* 68(2):392–412.

Received March 3, 2013

Contributors

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, assistant professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; oshestakov@cs.msu.su

Kuznetsova Maria G. (b. 1991) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation; m.g.kuznetsova@gmail.com

Sadovoy Ivan A. (b. 1990) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation; isadovoy@gmail.com

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ АВТОМАТИЗИРОВАННАЯ СИСТЕМА «МЕГАЛИТ» В ОПТИМИЗАЦИИ ДИАГНОСТИКИ И ЛЕЧЕНИЯ МОЧЕКАМЕННОЙ БОЛЕЗНИ

М. П. Кривенко¹, С. А. Голованов², П. А. Савченко³, А. В. Сивков⁴, А. П. Сучков⁵

Аннотация: В статье, первой из предполагаемой серии научных публикаций, рассматриваются результаты исследований по автоматизации информационных и аналитических процессов обследования, диагностики и лечения мочекаменной болезни (МКБ). Существенную роль в создании систем диагностики МКБ играет разработка информационных технологий сбора клинических данных и формирования специализированных баз данных (БД). Изучена возможность создания и способы реализации информационно-аналитической автоматизированной системы (ИААС) по сбору, хранению и обработке клинических данных обследования больных, а также алгоритмизации процессов принятия решений при диагностике МКБ и выборе схем лечения и профилактики этого заболевания. Предложенные математические методы и алгоритмы могут найти применение при дальнейшем развитии фундаментальных научных исследований в области разработки математических методов моделирования медико-биологических систем, а также при создании необходимого математического инструментария.

Ключевые слова: информационно-аналитическая система; урология; компьютерная диагностика; схема лечения; схема профилактики

DOI: 10.14357/19922264130409

1 Введение

В настоящее время доля людей, у которых на протяжении их жизни диагностируется МКБ, довольно значительна и составляет в странах Западной Европы 5%–9%, в Канаде и США — 7%–12%, в странах Азии — 1%–5% [1–4]. Эпидемиологические исследования, проводимые в ряде индустриально развитых стран, указывают на сохранение тенденции к росту частоты возникновения МКБ среди населения. Так, число впервые выявленных случаев МКБ на 100 000 населения за последние десятилетия возросло в США с 58,7 (1950–1954 гг.) до 85,1 (2000 г.) [2, 4], в Японии — с 43,7 (1965 г.) до 134 (2005 г.) [5, 6], в России — со 123,3 (2002 г.) до 138,6 (2010 г.) [7, 8].

По данным исследований [9] с использованием БД Pediatric Health Information System (национальная БД, в которую включены данные об амбулаторных визитах, срочных госпитализациях и стационарном лечении детей из 42 детских больниц США) по сравнению с общим количеством

госпитализированных пациентов число пациентов с МКБ увеличилось с 18,4 на 100 000 населения в 1999 г. до 57,0 в 2008 г., годовой прирост составил 10,6% ($p < 0,0001$).

В основе развития МКБ лежат характерные нарушения обмена веществ, приводящие к образованию камней в мочевых путях. Эти литогенные (камнеобразующие) нарушения обмена веществ характеризуются большим многообразием и проявляются различными патологическими изменениями биохимического состава крови и мочи пациента.

Необходимым условием для выбора правильной тактики консервативного лечения с целью предупреждения повторного камнеобразования является исследование всего комплекса метаболических факторов риска (МФР), ответственных за развитие МКБ.

В этой связи большое внимание уделяется изучению особенностей физико-химических параметров мочи, во многом определяющих вероятность образования мочевых камней [10]. Кроме того, литогенные нарушения метаболизма зачастую

¹Институт проблем информатики Российской академии наук, mkrivenko@ipiran.ru

²Научно-исследовательский институт урологии, sergeygol124@mail.ru

³Институт проблем информатики Российской академии наук, psavchenko@ipiran.ru

⁴Научно-исследовательский институт урологии, uroinfo@yandex.ru

⁵Институт проблем информатики Российской академии наук, asuchkov@ipiran.ru

имеют сложный многофакторный характер воздействия на процесс формирования камня. Это создает особые трудности для врача в полной и объективной оценке всех влияющих литогенных факторов обмена веществ, а также в принятии решения по диагностике и выбору лечебной тактики для конкретного больного. Отсюда возникает необходимость формирования БД анкетных и лабораторных исследований, систем, связанных с диагностикой МКБ и формирования базы знаний по профилактике и лечению этого заболевания.

2 Системы компьютерной диагностики в области урологии

Существенную роль в создании систем диагностики МКБ является разработка информационных технологий сбора клинических данных и формирования специализированных БД. К ним относится упомянутая Pediatric Health Information System. В ряде медицинских работ упоминается реестр по уролитиазу (БД по больным и результатам лечения) Юго-западного медицинского центра Техасского университета: «Retrospective data from the University of Texas Southwestern Medical Center *Nephrolithiasis Registry* from 17 studies that dealt with physiologic and physicochemical effects of various magnesium and potassium salts were categorized into three groups and analyzed. . . » [11]. Однако подробного описания данного реестра не приведено.

Задачи диагностики, дифференциальной диагностики, прогнозирования, выбора стратегии и тактики лечения позволяют решать экспертные медицинские системы [12].

Ряд работ посвящен использованию в урологии компьютерных диагностических систем на основе методов искусственных нейронных сетей (ИНС) [13]. Так, в онкоурологии смогли прогнозировать 5-летнюю выживаемость пациентов, перенесших радикальную цистэктомию по поводу рака мочевого пузыря [14]. Искусственные нейронные сети применили также для автоматизированного анализа показаний к биопсии предстательной железы [15]. Методика основывалась на выявлении общего простат-специфического антигена (ПСА) и определении доли свободного ПСА. Чувствительность составила 95%, специфичность — 34%. При дополнении нейросети моделью логистической регрессии специфичность возросла до 95%. Искусственная нейронная сеть использовалась для выявления группы риска рака предстательной железы в сравнении с моделью логисти-

ческой регрессии [15]. Искусственная нейронная сеть также продемонстрировала более точные прогностические возможности. Компьютерных систем диагностики именно МКБ по литературным данным не выявлено.

Отсюда ясно, что имеется настоятельная необходимость разработки аналитической системы диагностики и лечения больных МКБ в процессе их динамического наблюдения (мониторинге) для предупреждения повторного камнеобразования. Отсутствие подобных аналитических систем для мониторинга больных МКБ послужило основанием для разработки опытного образца ИААС «Мегалит». Создание системы осуществляется ИПИ РАН совместно с НИИ урологии Минздравсоцразвития России в рамках серии совместных научно-исследовательских работ.

Основные цели и задачи создания информационно-аналитической автоматизированной системы «Мегалит»

1. Создание БД по результатам обследования пациента, включающей:
 - формализованные данные опроса пациента при первом и последующих визитах, содержащие информацию о факторах, способных оказывать влияние на возникновение и особенности клинического течения МКБ (lifestyle-факторы индивида, факторы среды, питания, профессии и проч.);
 - данные лабораторного обследования (результатов простого или расширенного лабораторного обследования).
2. Создание аналитической подсистемы, обеспечивающей решение следующих задач:
 - на основании данных первичного опроса выявление наличия или отсутствия, а также степень риска развития МКБ и определение объема предполагаемого лабораторного обследования пациента (простое или расширенное обследование);
 - на основе анализа входных данных лабораторного обследования осуществление выбора дальнейшей тактики ведения больного — дополнительные виды исследования, выбор лечебных мероприятий (тип хирургического лечения, схема медикаментозной терапии, коррекция диеты и проч.);
 - реализация методов оптимального выбора (с учетом показаний и противопоказаний) вида хирургического лечения или схемы проведения профилактического лечения

(включая прием специальных фармпрепаратов, рекомендации по модификации диеты и образа жизни).

При создании опытного образца ИААС «Мегалит» учитывались следующие требования.

1. Опытный образец аналитической системы «Мегалит» должен иметь возможность ведения распределенной БД пациентов, содержащей результаты обследований, профили МФР и относительный индекс перенасыщенности мочи (ОИП) как исходные, так и измененные в результате назначенного лечения, и включать набор подсистем, включающих программную реализацию разработанных методов диагностирования и лечения.
2. Данные простого лабораторного обследования пациента должны включать:
 - исследование химического состава мочевого камня;
 - биохимическое исследование крови и мочи по различным параметрам;
 - клинический анализ мочи с посевом на микрофлору;
 - обзорный рентгеновский снимок, сонограмму и другие виды инструментального обследования пациента.

Данные биохимического исследования представлены величинами содержания в крови и моче ионов и соединений, способных приводить к образованию мочевых камней. При наличии патологических отклонений в биохимических исследованиях проводится расширенное лабораторное обследование.

3. Данные расширенного лабораторного обследования включают протокол диагностики типа гиперкальциемии (ПД-ГКУ) (при выявлении повышенной суточной экскреции кальция у пациента). Выполняется поэтапно, с помощью модифицированной по кальцию диеты. В расширенное лабораторное обследование входит также полный диагностический протокол (ПДП) больного МКБ.
4. Полный диагностический протокол представляет собой выраженное в графическом виде исходное состояние обмена веществ у пациента с МКБ с выявленными МФР и динамику изменения показателей обмена веществ в результате проводимого лечения. Графическое отображение МФР и ОИП больного МКБ позволяет оценить степень выявленных нарушений и их динамику в процессе профилактиче-

ского лечения и вносить в лечебную схему необходимые коррекции, также выбираемые по особому алгоритму. Выявленные при первичном обследовании МФР и ОИП служат основой для программного выбора схем коррекции метаболических нарушений и предупреждения рецидивов МКБ. Коррекция включает в себя лечебные мероприятия, прием специальных фармпрепаратов, рекомендации по модификации диеты и образа жизни.

5. В аналитической системе «Мегалит» предусматривается возможность ее обучения и настройки на основе получаемых новых данных о результатах лечения пациента (пациентов) на каждом этапе наблюдения.
6. Предусмотреть в программной реализации алгоритмов экспертного модуля:
 - алгоритм оценки эффективности выбранной схемы лечения;
 - алгоритм принятия решения по дальнейшему лечению;
 - алгоритм поиска и выбора рациональной схемы профилактического лечения.
7. Оценить возможности разработки метода корректировки параметров подсистемы диагностирования и лечения на основе анализа вновь поступающих данных (обратная связь).
8. Разработанные аналитические методы и алгоритмы, реализованные в составе опытного образца аналитической системы «Мегалит», должны пройти апробацию и тестирование в реальных клинических условиях. По результатам применения опытного образца должны быть сформулированы рекомендации по его совершенствованию и развитию.

3 Основные подходы к созданию информационно-аналитической автоматизированной системы «Мегалит» и их реализация

Основные функции ИААС:

- сбор и формализация данных, включая ведение реестра пациентов, системы словарей и справочников;
- поддержка принятия решения по назначению и сбор данных диагностических исследований;
- первичный и ретроспективный анализ тестов;

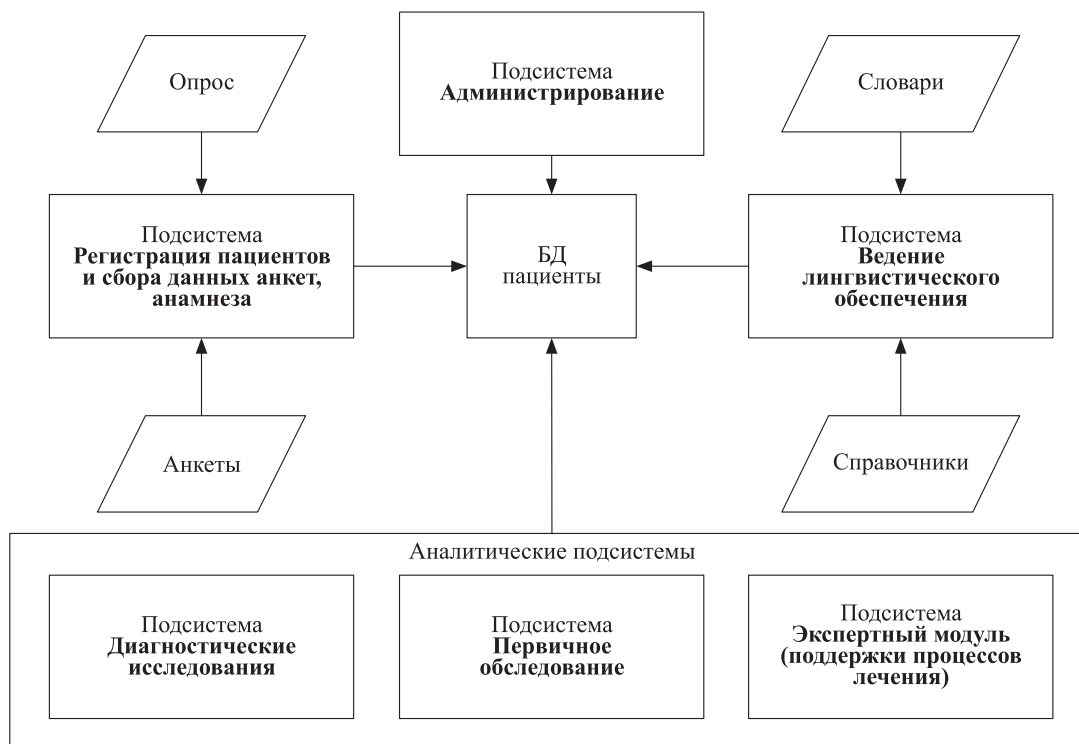


Рис. 1 Структура ИААС

- поддержка принятия решения по выбору схемы лечения, оценка эффективности схемы лечения;
- поддержка принятия решения по дальнейшему лечению;
- поиск и поддержка принятия решения по выбору рациональной схемы профилактического лечения.

Информационно-аналитическая автоматизированная система «Мегалит» включает в себя подсистемы:

- администрирования;
- регистрации пациентов и сбора данных анкет, анамнеза;
- ведения лингвистического обеспечения;
- первичного обследования;
- диагностических исследований;
- экспертный модуль (поддержки процессов лечения).

Структурная схема ИААС «Мегалит» представлена на рис. 1.

Для обеспечения возможности коллективной работы по формированию БД системы и многопользовательского режима работы с ее аналитическим модулем она проектируется в виде веб-сайта,

доступного авторизованным пользователям в сети Интернет. Основные базовые функции информационного сайта должны быть реализованы общесистемным функционалом его платформы. Таким образом, в программно-технологической платформе должны быть заложены следующие функции:

- (1) выполнение приложений — позволяет легко разрабатывать, развертывать различные приложения и управлять ими;
- (2) возможность совместной работы — позволяет отдельным пользователям и крупным организациям объединить свои ресурсы и работать вместе через Интернет;
- (3) управление содержимым — придает гибкость производству и управлению отдельными веб-узлами, позволяя поставлять конечному пользователю приспособленное под него (персонализированное) содержимое сайта;
- (4) управление пользователями — позволяет организации управлять пользователями, ресурсами и безопасностью внутри и вне системы сетевой защиты, а также предоставлять канал для внешних связей и проведения электронных транзакций;
- (5) контроль и управление производительностью — позволяет улучшать качество пользовательского интерфейса, обеспечивая:

- управление знаниями — помогает объединять внутреннюю и внешнюю информацию и предоставлять информацию, основанную на контекстной концепции;
- поддержку поиска — обеспечивает клиента доступом к широкому спектру источников информации как внутри, так и вне сайта;
- безопасность — защиту данных, приложений и транзакций;
- стандартный www-доступ к сайту — для технического обеспечения функционирования его содержимого.

Определяющими характеристиками веб-решений являются масштабируемость, доступность, надежность, защита данных от несанкционированного доступа, транзакционная целостность и распространение.

Важной особенностью платформы является то, что она объединяет все необходимые модули, которые позволяют выполнять практически любую работу, связанную с созданием и обновлением сайта специалистом предметной области.

В качестве языка программирования выбран один из самых современных языков — C#. ASP.NET — технология, которая является частью .NET и используется для разработки интернет-ориентированного программного обеспечения и интернет-сайтов.

Для работы интернет-сайтов используется связка: операционная система Windows Server 2008 + интернет-сервер IIS 7.0 + СУБД Ms SQL Server 2008.

4 Концепция экспертного модуля системы

4.1 Основные подходы к использованию статистических методов анализа данных в урологии

Клинические БД содержат большое количество информации о пациентах и их заболеваниях. Скрытые (латентные) связи и структуры в этих данных могут быть источником новых медицинских знаний. К сожалению, немногие из существующих технологий анализа данных оказываются непосредственно применимыми и действенными при обнаружении и описании этих латентных знаний, но, безусловно, универсальной из них является технология на принципах *Data Mining* — извлечение скрытой информации из уже накопленных и пополняемых сведений об объекте исследования. Эта и ряд других сформировавшихся технологий, ориентированных на анализ массивов данных, терминологически пересекаются или оказываются взглядом на одном и том же, но с разных точек зрения (краткое освещение данного вопроса дано в [17, разд. 1]). В первую очередь речь идет о следующих подходах:

- разведочный анализ данных — Exploratory Data Analysis (EDA);
- извлечение скрытой информации из данных — Data Mining (DM);
- обнаружение знаний в данных — Knowledge Discovery in Databases (KDD);
- машинное обучение — Machine Learning (ML).

Таким образом, обнаружение в данных ранее не известных, нетривиальных, практически полез-



Рис. 2 Пирамида анализа данных

ных и доступных интерпретации знаний, необходимых для поддержки принятия решений в различных сферах человеческой деятельности, составляет суть DM. Говоря далее об анализе данных, будем понимать при этом цели, задачи, технологии, методы и алгоритмы, присущие DM.

На рис. 2 схематично изображена иерархия содержательной стороны анализа данных. Вертикальная стрелка слева показывает направление роста отдельных характеристик задач анализа данных в зависимости от их уровня. Примеры постановок практических задач приведены справа. Надо понимать, что «восхождение» по пирамиде анализа данных должно обеспечиваться обязательным существенным ростом объема используемой информации (данных и предположений об объектах исследования), а также глубиной проработки вопроса о качестве предлагаемых решений.

Основные принципы анализа данных. Среди методов, которые использовались при решении проблемы обучения в ML, те, которые представляют наибольший интерес при анализе данных (снижение размерности, оценивание распределения данных, регрессионный анализ, классификация, кластеризация), теперь все вместе упоминаются как статистическое обучение. Проблема обучения делится на различные категории: две из них, наиболее близкие к статистике, суть контролируемое обучение или обучение с «учителем» и неконтролируемое, без «учителя».

Одна из самых важных задач в анализе данных состоит в том, чтобы оценить качество полученных решений, в частности точность предложенного прогноза (например, качество построенного классификатора). В качестве меры точности прогноза обычно используется ошибка прогноза. Простейшая оценка ошибки прогноза строится с помощью тех же данных, которые используются для построения модели (такой вариант оценки называют самооценкой, оценкой переподстановки). Понятно, что в результате сформируется чрезмерно оптимистический взгляд на точность прогноза.

Очевидный способ улучшения состоит в обобщении: оценивать точность прогноза с помощью данных, независимых от тех, которые использовались для подгонки модели. Получить подобные независимые данные можно путем сбора новых данных. Если это невозможно, то имеет смысл разделить исходные данные на части и воспользоваться ими для решения самостоятельных задач. Обычная практика заключается в следующем: если набор данных достаточно велик, то необходимо использовать случайный механизм для деления данных на два непересекающихся и независимых набора:

- (1) данные для обучения, которые можно использовать для предварительного контроля данных, для формирования моделей;
- (2) тестирующие данные, которые будут использоваться для оценки качества построенной модели.

Альтернативные методы расщепления данных для того, чтобы оценить тестовую ошибку, основаны на перепроверке [16] и бутстреп-методе [17].

Суть вероятностной модели бутстреп-метода в данном случае состоит в следующем. Предположим, что по выборке $x = (x_1, \dots, x_N)$ данных лабораторных исследований из распределения $F(u)$ оценивается значение $\vartheta = \vartheta(F)$ некоторого функционала (например, классификатора заболеваний), заданного на семействе \mathbf{F} . Качество оценки $\vartheta^*(X)$ характеризуется величиной

$$R(\vartheta^*(X), \vartheta(F)) = E_F\{L(\vartheta^*(X), \vartheta(F))\},$$

где $L(\vartheta^*(X), \vartheta(F))$ — потери от принятия оценки $\vartheta^*(X)$ вместо неизвестного значения $\vartheta(F)$. Бутстреп-метод позволяет оценить $R(\vartheta^*(X), \vartheta(F))$ с помощью замены распределения F его некоторой оценкой F^B и вычисления статистики ϑ^* по выборке x^B объемом N из F^B . Совокупность x^B называется бутстреп-выборкой, статистика $\vartheta^*(x^B)$ — бутстреп-реализацией ϑ^* .

Условное распределение

$$\begin{aligned} \Pr\{\vartheta^*(X^B) < u | x_1, \dots, x_N\} &= \\ &= \int_{\{y: \vartheta^*(y) < u\}} dF^B(y_1) \dots dF^B(y_N) \end{aligned}$$

является бутстреп-оценкой функции распределения $\Pr\{\vartheta^*(X) < u\}$ статистики ϑ^* .

Процедура выбора оценки F^B для F мотивируется наличием априорной информации. В параметрической ситуации, когда $\mathbf{F} = \{F_\lambda, \lambda \in \Lambda\}$, оценка F^B часто оказывается результатом подстановки вместо λ некоторой оценки λ^* , т. е. $F^B = F_{\lambda^*}$.

Другая ситуация относится к области непараметрической статистики. Здесь F^B обычно оказывается эмпирической функцией распределения, т. е. каждому наблюдаемому значению (элементу исходной выборки) приписывается вероятность $1/N$. Бутстреп-выборки тогда подчиняются условному полиномиальному распределению, сосредоточенному на x_1, \dots, x_N .

Наиболее трудную часть бутстреп-метода составляет нахождение распределения $\vartheta^*(X^B)$, для чего применяются три приема:

- (1) прямое теоретическое вычисление;
- (2) аппроксимация с помощью метода статистических испытаний;
- (3) аппроксимация с помощью аналитических методов (например, используя разложение в ряд Тейлора).

Прямое теоретическое вычисление распределения $\vartheta^*(X^B)$ может осуществляться либо аналитическим путем, либо путем непосредственного перечисления бутстреп-выборок и подсчета соответствующих вероятностей. Если оба приема недоступны (первый из-за аналитических сложностей, второй из-за вычислительных), то приходится прибегать к методу статистических испытаний, т. е. к повторению экспериментов по случайному формированию бутстреп-выборки x^B и подсчету значения $\vartheta^*(x^B)$.

Следует обратить внимание на реальные возможности бутстреп-метода: он не позволяет получить новую информацию о наблюдаемых объектах, его назначение — сформировать объективное представление о свойствах использованных процедур анализа данных.

В аналитической системе «Мегалит» накапливаются данные следующих типов:

1. Неформализованные (неструктурированные), представленные в виде текста (например, текст назначения врача).
2. Формализованные:
 - 2.1. Качественные:
 - 2.1.1. Измеренные по шкале наименований (например, пол пациента).
 - 2.1.2. Измеренные по порядковой шкале (например, порядковый номер сезона, когда обследовался пациент).
 - 2.2. Количественные:
 - 2.2.1. Измеренные по одной из соответствующих шкал и принимающие значения из небольшого набора числовых значений (например, дата взятия анализов или количество обнаруженных у пациента камней).
 - 2.2.2. Измеренные по одной из соответствующих шкал и принимающие значения в виде действительных чисел (например, уровень кальция в анализе крови пациента).

Приведенный систематизированный перечень встречающихся типов данных требует привлечения

разнообразного арсенала средств, таких как лингвистический анализ (п. 1), статистический анализ категориальных данных (п. 2.1.1), ранговые процедуры (п. 2.1.2), статистический анализ на основе моделей дискретных и непрерывных распределений (пп. 2.2.1 и 2.2.2).

Ошибки есть во всех видах БД; к сожалению, встречаются они и в данном случае. В различных прикладных областях накоплен опыт (см., в частности, [18]), позволяющий привести типичный перечень источников ошибок: фальсификация, неполнота, несогласованность, дублирование.

Те ошибки, которые легко обнаружить, вероятнее всего можно найти на стадии «очистки» данных, более же скрытые, неочевидные могут быть обнаружены только при анализе данных. «Очистка» данных обычно происходит, когда данные получены и прежде, чем они сохраняются в формате только для чтения в хранилище данных. В частности, должны быть исключены ошибки, при которых переменные принимают значения, противоречащие естественным ограничениям (например, при описании химического состава камней значения отдельных переменных не могут превосходить 100%). Доля подобных грубых ошибок в медицинских исследованиях может превышать 10%.

Ошибки недопустимости значений должны быть описаны с помощью логических выражений, истинность которых проверяется на этапе «очистки». В случаях, когда их не удастся исправить автоматически или автоматизированно, результат должен помечаться специальным образом.

Для данных, уже хранящихся в БД и являющихся объектом анализа, могут быть характерны следующие проблемы:

- наличие аномальных наблюдений (значения, которые существенно отличаются от основной массы наблюдений);
- пропуски в данных;
- малочисленность данных (ситуация, когда количество переменных превышает число наблюдений).

Таким образом, статистический анализ конкретных данных является многоэтапным процессом, включающим планирование статистического исследования, организацию сбора необходимых статистических данных, первичное описание данных, оценивание характеристик данных, проверку статистических гипотез, анализ полученных решений, формулировку выводов, составление итоговых документов.

В основе принципов построения статистического вывода относительно данных лежат следующие положения:

- при выборе семейства вероятностных распределений, описывающих данные, существенную роль играет предварительный анализ данных; последующий итеративный процесс уточнения априорных предположений направлен на построение модели, являющейся достаточно реалистичной и позволяющей строить содержательные выводы;
- при построении методов анализа наряду с постановкой задач разработки оптимальных процедур и попыткой их решения следует не пренебрегать разумными подходами к созданию каких-либо процедур с последующим обязательным анализом предлагаемых решений;
- завершающим этапом построения методов анализа должен быть количественный или качественный анализ влияния на предлагаемые решения отклонений от априорных предположений, при этом исследование качества полученных решений реальнее всего проводить с помощью бутстреп-метода.

4.2 Принципиальные возможности создания экспертного модуля системы «Мегалит»

Повседневная деятельность врача требует решения задач интерпретации, диагностики, контроля и прогнозирования, т. е. таких задач, которые могут быть решены с помощью систем поддержки принятия решений. Медицина представляет одну из областей человеческой деятельности, где знания специалистов трудно формализуемы, однако разработка диагностических медицинских систем в настоящее время является актуальной задачей.

При создании экспертного модуля системы «Мегалит», предназначенного для поддержки принятия решения по диагностике заболевания, предполагается:

- разработать систему представления медицинских знаний (с использованием данных анкет, анамнеза, данных инструментальных и лабораторных методов исследования);
- разработать алгоритм механизма логического вывода (выполнение диагностики типа литогенного нарушения обмена веществ; выбора адекватной схемы лечения, оценки эффективности заданной схемы лечения с возможностью ее коррекции при дальнейшем мониторинге пациента).

Система представления медицинских знаний позволяет выделять значимые для принятия врачебного решения или постановки медицинского диа-

гноза данные (качественные или количественные). Так, анализ качественных данных анамнеза, анкетных данных позволяет сделать заключение о силе влияния наследственных, средовых и социальных факторов риска развития МКБ; потенциальной активности процесса камнеобразования. Этой же цели служат качественные и количественные данные, полученные при инструментальном/лабораторном (рентгенологическом, микробиологическом, ультразвуковом или антропометрическом) обследовании пациента.

Большой массив количественных данных в виде числовых значений показателей получают при биохимическом исследовании. Именно он является основным объектом алгоритмизации при разработке экспертного модуля системы «Мегалит». Применение этого модуля предназначено для объективной и более точной диагностики метаболического литогенного синдрома, выбора адекватной схемы лечения, качественной оценки результатов лечения, коррекции лечебной схемы на основе полученных данных в целях выбора оптимального лечебного воздействия на нарушенный обмен веществ у пациента.

Учитывая, что указанные процессы являются алгоритмируемыми, можно полагать, что принципиальные возможности создания экспертного модуля для системы «Мегалит» имеются.

4.3 Качественное описание алгоритмической базы экспертного модуля

Качественное описание алгоритмической базы включает: постановку задачи, описание входных и выходных данных, описание входных/выходных форм пользовательского интерфейса, описание событий и реакций системы в рамках поддержки процесса оценки эффективности.

Алгоритм оценки эффективности заданной схемы лечения

Постановка задачи. Качественно и количественно оценить эффективность применения выбранной схемы лечения (с выводом о продолжении ее использования в лечении; ее модификации в той или иной степени; замены на другую схему лечения).

Входные данные. Входными данными служат те количественно измененные биохимические показатели, которые на предыдущем этапе программного анализа были определены (диагностированы) как характерные для данного метаболического синдрома.

Выходные данные. Выходными данными служат количественные значения биохимических признаков пролеченного метаболического синдрома, полученные при лабораторном исследовании после курса лечения. Эти данные должны быть программно проанализированы в сравнении с их исходными (до начала лечения) значениями. Используется алгоритм «Оценка качества лечебного эффекта», который предполагает следующие варианты вывода о качестве лечения:

- «отсутствие эффекта»;
- «слабо выраженный положительный эффект»;
- «выраженный положительный эффект»;
- «слабо выраженный отрицательный эффект»;
- «выраженный отрицательный эффект».

Выходные формы пользовательского интерфейса. Выходные формы пользовательского интерфейса при этом могут быть представлены в виде таблицы со значениями биохимических признаков метаболического синдрома до и после лечения. Возможна опция отображения исходных данных в виде диаграммы или графика.

Алгоритм принятия решения по дальнейшему лечению

Постановка задачи. Построить алгоритмические правила, позволяющие пользователю сделать вывод и принять решение об использовании данной схемы в дальнейшем лечении; модификации схемы в той или иной степени; замены данной схемы на другую схему лечения.

Входные данные:

- данные сравнительного анализа численных биохимических величин до и после лечения;
- данные качественной оценки лечебного эффекта, получаемые в результате обработки данных сравнительного анализа численных биохимических величин до и после лечения.

При формировании и сборе данных первого типа выполняется процедура сравнения достигнутых в результате лечения величин значимых для данного метаболического синдрома показателей с их исходными значениями, диагностированными до начала лечения в ходе биохимического лабораторного исследования.

Данные второго типа являются качественными, производными от данных первого типа. Эти данные представляют собой возможные варианты вывода о качестве лечения.

Выходные данные. Выходными данными алгоритма принятия решения по дальнейшему лечению служат установленные типы рекомендаций по применявшейся схеме лечения:

- сохранение схемы без изменений и продолжение лечения;
- модификация схемы трех степеней выраженности (незначительная, умеренная, существенная);
- отказ от применения данной схемы и выбор новой схемы лечения.

5 Перспективы развития информационно-аналитической автоматизированной системы «Мегалит»

Таким образом, разработана методологическая и техническая база для экспертной системы комплексной диагностики и профилактического лечения пациентов с МКБ. Учитывая не только медицинскую, но и социальную актуальность проблемы МКБ, а также трудности принятия врачебного решения в выборе адекватной тактики противоречивого лечения этого заболевания, следует считать целесообразным создание технологий расширения и совершенствования функционала экспертного модуля системы «Мегалит» на основе анализа вновь поступающих данных с использованием принципа обратной связи.

На следующих этапах планируется работа по оптимизации практического применения ИААС «Мегалит» в клинической урологии.

Предложенные математические методы и алгоритмы найдут применение при дальнейшем развитии фундаментальных научных исследований в области разработки математических методов моделирования медико-биологических систем, а также при создании необходимого математического инструментария. В первую очередь это касается постановки развернутого диагноза, наиболее полно отражающего особенности метаболического типа конкретного больного (обследуемого), специфику функционального состояния почек и мочевых путей пациента. При этом также учитывается влияние различных модифицирующих факторов, таких как наличие или отсутствие инфекции и степени ее выраженности, воздействие социальных факторов, факторов питания, наследственности и др.

Внедрение и практическое использование системы «Мегалит» позволит сформировать представительный набор данных, на основе которого разработать методы выбора оптимальной лечебной тактики. Таким образом, конечные пользователи получат возможность дистанционной диагностики метаболических литогенных синдромов у пациента, оценки степени риска развития МКБ, выбора адекватных терапевтических схем лечения МКБ и/или профилактики рецидивов камнеобразования, ввода данных о пациенте в единый банк данных для последующего мониторинга и др.

Литература

1. *Ramello A., Vitale C., Marangella D.* Epidemiology of nephrolithiasis // *J. Nephrol.*, 2000. Vol. 13. Suppl. 3. P. 45–50.
2. *Trinchieri A., Coppi F., Montanari E., Del Nero A., Zanetti G., Pisani E.* Increase in the prevalence of symptomatic upper urinary tract stones during the last ten years // *Eur. Urol.*, 2000. Vol. 37. P. 23–25.
3. *Pearle M. S., Calhoun E. A., Curhan G. C.* Urologic diseases in America project: Urolithiasis // *J. Urology*, 2005. Vol. 173. P. 848–857.
4. *Lieske J. C., Pena de la Vega L. S., Slezak J. M., Bergstralh E. J., Leibson C. L., Ho K. L., Gettman M. T.* Renal stone epidemiology in Rochester, Minnesota: An update // *Kidney Int.*, 2006. Vol. 69. No. 4. P. 760–764.
5. *Johnson C. M., Wilson D. M., O'Fallon W. M., Malek R. S., Kurland L. T.* Renal stone epidemiology: A 25-year study in Rochester, Minnesota // *Kidney Int.*, 1979. Vol. 16. No. 5. P. 624–631.
6. *Yasui T., Iguchi M., Suzuki S., Kohri K.* Prevalence and epidemiological characteristics of urolithiasis in Japan: National trends between 1965 and 2005 // *Urology*, 2008. Vol. 71. No. 2. P. 209–213.
7. Заболеваемость населения России в 2003 году: Статистические материалы. — М., 2004 (электронная версия МЗ и СР РФ и ЦНИИ организации и информатизации здравоохранения МЗ и СР РФ). <http://www.minzdravsoc.ru/docs/mzsr/stat/17>.
8. *Аполихин О. И., Сивков А. В., Солнцева Т. В., Комарова В. А.* Анализ урологической заболеваемости в Российской Федерации в 2005–2010 годах // *Экспериментальная и клиническая урология*, 2012. № 2. С. 4–12. <http://ecuro.ru/article/analiz-urologicheskoi-zabolevaemosti-v-rossiiskoi-federatsii-v-2005-2010-godakh>.
9. *Routh J. C., Graham D. A., Nelson C. P.* Epidemiological trends in pediatric urolithiasis at United States freestanding pediatric hospitals // *J. Urology*, 2010. Vol. 184. No. 3. P. 1100–1104.
10. *Голованов С. А., Дрожжева В. В.* Кристаллообразующая активность мочи при оксалатном уролитиазе // *Экспериментальная и клиническая урология*, 2010. № 2. С. 24–29. <http://ecuro.ru/article/kristalloobrazuyushchaya-aktivnost-mochi-pri-oksalatnom-urolitiaze>.
11. *Bonny O., Rubin A., Huang Ch.-L., Frawley W. H., Pak C. Y. C., Moe O. W.* Mechanism of urinary calcium regulation by urinary magnesium and pH // *J. Am. Soc. Nephrol.*, 2008. Vol. 19. No. 8. P. 1530–1537.
12. *Дюк В. А., Эмануэль В. Л.* Информационные технологии в медико-биологических исследованиях. — СПб.: Питер, 2003. 525 с.
13. *Liew P. L., Lee Y. C., Lin Y. C., et al.* Comparison of artificial neural networks with logistic regression in prediction of gallbladder disease among obese patients // *Digest. Liver Dis.*, 2007. Vol. 39. No. 4. P. 356–362.
14. *Bassi P., Sacco E., De Marco V., et al.* Prognostic accuracy of an artificial neural network in patients undergoing radical cystectomy for bladder cancer: A comparison with logistic regression analysis // *BJU Int.*, 2007. Vol. 99. No. 5. P. 1007–1012.
15. *Stephan C., Xu C., Finne P., et al.* Comparison of two different artificial neural networks for prostate biopsy indication in two different patient populations // *J. Urology*, 2007. Vol. 70. No. 3. P. 596–601.
16. *Stone M.* Cross-validated choice and assessment of statistical predictions (with discussion) // *J. Roy. Stat. Soc. B*, 1974. Vol. 36. P. 111–147.
17. *Efron B.* Bootstrap methods: Another look at the jackknife // *Ann. Stat.*, 1979. Vol. 7. P. 1–26.
18. *Izenman A. J.* Modern multivariate statistical techniques. — Springer, 2008. 731 p.

Поступила в редакцию 17.04.13

THE INFORMATION-ANALYTICAL COMPUTER SYSTEM “MEGALITH” IN OPTIMIZATION OF THE DIAGNOSIS AND TREATMENT OF UROLITHIASIS

M. P. Krivenko¹, S. A. Golovanov², P. A. Savchenko¹, A. V. Sivkov², and A. P. Suchkov¹

¹Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

²Research Institute of Urology, Moscow 105425, Russian Federation

Abstract: In this article, that is the first of an expected series of scientific publications, the results of research on automation of the information and analytical processes of the urolithic disease (ULD) survey, diagnosis, and treatment are discussed. A significant role in creating the systems of ULD diagnostics has the development of information technologies for clinical data collection and formation of specialized databases. The possibility of creation and the ways of realization of information-analytical computer system of collection, storage, and processing of the clinical data of patients examination, as well as programming decision-making processes in the diagnosis ULD and the choice of schemes of treatment and prevention of this disease has been studied. The developed mathematical methods and algorithms may be applied to the further fundamental scientific researches in the field of development of mathematical methods of medical and biological systems modeling; besides, they may be applied for necessary mathematical tools creation.

Keywords: informational-analytical system; urology; computer diagnostics; treatment scheme; scheme of prevention

DOI: 10.14357/19922264130409

References

1. Ramello, A., C. Vitale, and D. Marangella. 2000. Epidemiology of nephrolithiasis. *J. Nephrol.* 13(Suppl. 3):45–50.
2. Trinchieri, A., F. Coppi, E. Montanari, A. Del Nero, G. Zanetti, and E. Pisani. 2000. Increase in the prevalence of symptomatic upper urinary tract stones during the last ten years. *Eur. Urol.* 37:23–25.
3. Pearle, M. S., E. A. Calhoun, and G. C. Curhan. 2005. Urologic diseases in America project: Urolithiasis. *J. Urology* 173:848–857.
4. Lieske, J. C., L. S. Pena de la Vega, J. M. Slezak, E. J. Bergstralh, C. L. Leibson, K. L. Ho, and M. T. Gettman. 2006. Renal stone epidemiology in Rochester, Minnesota: An update. *Kidney Int.* 69(4):760–768.
5. Johnson, C. M., D. M. Wilson, W. M. O’Fallon, R. S. Malek, and L. T. Kurland. 1979. Renal stone epidemiology: A 25-year study in Rochester, Minnesota. *Kidney Int.* 16(5):624–631.
6. Yasui, T, M. Iguchi, S. Suzuki, and K. Kohri. 2008. Prevalence and epidemiological characteristics of urolithiasis in Japan: National trends between 1965 and 2005. *Urology* 71(2):209–213.
7. Russian Ministry of Health: Central Research Institute of Organization and Informatization of Population. 2004. Zaboлеваemost’ naseleniya Rossii v 2003 godu: Statisticheskie materialy [Morbidity of population of Russia in 2003: Statistical materials]. Moscow. Electronic version. <http://www.minzdravsoc.ru/docs/mzsr/stat/17>.
8. Apolikhin, O. I., A. V. Sivkov, T. V. Solntseva, and V. A. Komarova. 2012. Analysis of urological morbidity in the Russian Federation within the period of 2005–2010. *Experimental and Clinical Urology* 2:4–12. <http://ecuro.ru/en/article/analysis-urological-morbidity-russian-federation-within-period-2005-2010>.
9. Routh, J. C., D. A. Graham, and C. P. Nelson. 2100. Epidemiological trends in pediatric urolithiasis at United States freestanding pediatric hospitals. *J. Urology* 184(3):1100–1104.
10. Golovanov, S. A., and V. V. Drozhzheva. 2010. Crystal formation activity of urine in oxalate urolithiasis. *Experimental and Clinical Urology* 2:24–29. <http://ecuro.ru/en/article/crystal-formation-activity-urine-oxalate-urolithiasis>.
11. Bonny, O., A. Rubin, Ch.-L. Huang, W. H. Frawley, C. Y. C. Pak, and O. W. Moe. 2008. Mechanism of urinary calcium regulation by urinary magnesium and pH. *J. Am. Soc. Nephrol.* 19(8):1530–1537.
12. Djuk, V. A., and V. L. Jemanujel’. 2003. *Informatsionnye tekhnologii v mediko-biologicheskikh issledovaniyakh [Information technologies in medical and biological researches]*. St. Petersburg, Russia: Piter, 2003. 525 p.
13. Liew, P. L., Y. C. Lee, Y. C. Lin, et al. 2007. Comparison of artificial neural networks with logistic regression

- in prediction of gallbladder disease among obese patients. *Digest. Liver Dis.* 39(4):356–362.
14. Bassi, P., E. Sacco, V. De Marco, *et al.* 2007. Prognostic accuracy of an artificial neural network in patients undergoing radical cystectomy for bladder cancer: A comparison with logistic regression analysis. *BJU Int.* 99(5):1007–1012.
 15. Stephan, C., C. Xu, P. Finne, *et al.* 2007. Comparison of two different artificial neural networks for prostate biopsy indication in two different patient populations. *J. Urology* 70(3):596–601.
 16. Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Stat. Soc. B* 36:111–147.
 17. Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* 7:1–26.
 18. Izenman, A. J. 2008. *Modern multivariate statistical techniques*. Springer. 731 p.

Received April 17, 2013

Contributors

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

Golovanov Sergey A. (b. 1950) — Doctor of Science in medicine, Head of Laboratory, Research Institute of Urology, Moscow 105425, Russian Federation; sergeygol124@mail.ru

Savchenko Pavel A. (b. 1967) — software engineer, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; psavchenko@ipiran.ru

Sivkov Andrey V. (b. 1957) — Doctor of Science in medicine, Deputy director, Research Institute of Urology, Moscow 105425, Russian Federation; uroinfo@yandex.ru

Suchkov Alexander P. (b. 1954) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; asuchkov@ipiran.ru

АНАЛИЗ ОДНОРОДНОСТИ ДАННЫХ О ХИМИЧЕСКОМ СОСТАВЕ КАМНЕЙ ПРИ УРОЛИТИАЗЕ

М. П. Кривенко¹, С. А. Голованов², А. В. Сивков³

Аннотация: Рассмотрены методы исследования однородности статистических данных о химическом составе камней при уролитиазе. Многомерность данных, бедный спектр различных значений показателей, невозможность использовать простые адекватные вероятностные модели для них, а также отсутствие накопленного опыта подобного моделирования в области урологии придают специфику рассматриваемым задачам. Предложен и проанализирован критерий значимости, основанный на расстоянии между множествами точек в евклидовом пространстве. Он применен для анализа зависимости состава камней от пола пациента. При анализе изменения состава камней от времени предложенный критерий использован как элемент парных сравнений данных для различных временных фрагментов. В этом случае впервые применены идеи аппроксимации возникающих связей между фрагментами данных с помощью отношения эквивалентности. Это позволяет обоснованно проводить стратификацию исходных данных. В ходе экспериментов получены статистически обоснованные результаты как об особенностях мочекаменной болезни в зависимости от пола пациента, так и об их изменении с течением времени.

Ключевые слова: критерии однородности; стратификация; бутстреп-метод; мочекаменная болезнь

DOI: 10.14357/19922264130410

1 Введение

Заболеваемость уролитиазом существенно варьируется в различных странах мира. В терминах частоты встречаемости этого заболевания на протяжении жизни индивида в исследуемой популяции (lifetime prevalence) распространенность мочекаменной болезни (МКБ) составляет в странах Азии 1%–5%, Европы — 5%–9%, Северной Америки — 13%, достигая 20% среди населения Саудовской Аравии [1].

Несмотря на то что эпидемиологические исследования, проведенные в 10 странах, указывают на определенное сходство распределения типов уролитиаза, определяемых по химическому составу конкрементов [2], химический состав мочевых камней у больных в разных странах все же имеет свои особенности. Кроме того, замечено, что в одном и том же регионе клинические и метаболические характеристики МКБ могут существенно изменяться с течением времени [1, 3–5].

Существующие различия, как полагают, имеют тесную связь с инфекцией мочевых путей, нарушением уродинамики, изменениями обмена веществ и физико-химических свойств мочи, а также с факторами окружающей среды, характером питания населения и социально-экономическими условиями [6].

В связи с этим целью настоящей работы было изучение особенностей МКБ и распространенности ее метаболических типов по результатам исследования минерального состава мочевых камней пациентов, проходивших лечение в клинике НИИ урологии МЗ РФ (г. Москва) и городской клинической урологической больнице № 47 г. Москвы в период с 2005 по 2009 гг. Данная работа является частью продолжающегося исследования особенностей распространенности метаболических типов МКБ в московском регионе, начатого в НИИ урологии с 1985 г. [7].

Сравнительный анализ данных по химическому составу мочевых камней за многолетний период представляет как эпидемиологический, так и практический интерес для клинической урологии, позволяя выявить определенные тенденции распространенности типов МКБ за исследуемый период. Например, полученные данные по частоте выявления основных метаболических типов мочевых камней среди мужского и женского населения г. Москвы могут быть использованы при планировании оказания специализированной урологической помощи и разработке соответствующих лечебно-профилактических мероприятий. Результаты проведенной работы могут стать основой для дальнейших аналитических эпидемиологических ис-

¹Институт проблем информатики Российской академии наук, mkrivenko@ipiran.ru

²Научно-исследовательский институт урологии, sergeyGol124@mail.ru

³Научно-исследовательский институт урологии, uroinfo@yandex.ru

следований, направленных на выявление факторов риска и их устранение, в целях снижения заболеваемости МКБ среди населения.

Исследование реальных клинических баз данных играет важную роль для развития теории и практики методов их анализа, давая толчок к детализации постановок задач обработки данных, вынуждая развивать отдельные подходы по выявлению скрытых (латентных) связей и структур в этих данных. И наконец, только проводя прикладные исследования, можно понять истинную значимость предлагаемых наработок в области прикладной математики и информатики.

Материалом для исследования послужили результаты анализа химического состава 4217 мочевых конкрементов, удаленных оперативно, фрагментированных с помощью дистанционной литотрипсии или отошедших самостоятельно. Всего обследовано 2413 мужчин и 1804 женщин в возрасте от 16 до 77 лет, больных МКБ.

В данной работе подробно рассматриваются задачи анализа однородности данных о составе камней в двух конкретных постановках: по имеющимся данным о составе камней требуется исследовать его зависимость от пола пациента и времени. Решение этих задач должно включать как выявление (обнаружение) наличия зависимости, так и формирование предположений, какие особенности состава камней характерны для того или иного пола или промежутка времени.

2 Общая характеристика данных

Объектом анализа стали данные, включающие указание пола пациента (М — мужской и F — женский), результаты измерений состава камней, даты этих измерений. Мочевые камни на 98%–99% состоят из кристаллической фазы, представленной чаще всего оксалатами и фосфатами кальция, мочевой кислотой и ее солями. Кристаллическая фаза может содержать один, два и более минеральных компонентов [8]. Поэтому состав камней представляется либо с помощью первичных признаков, отражающих тип минеральной кристаллической фазы, либо с помощью вторичных или интегральных признаков. Для первичных признаков (переменных) приняты следующие обозначения: WH — вевеллит (кальция оксалат моногидрат); WD — ведделлит (кальция оксалат дигидрат); UA — мочевая кислота; UADH — мочевая кислота дигидрат; AMUR — (моно)аммония урат; NAUR — (моно)натрия урат моногидрат; WTL — витлокит (трикальция фосфат); CA — карбонатапатит; HAP — гидроксиапатит; BRU — брусит; STRU —

струвит; CYS — L-цистин. Вторичные признаки суть: O — оксалаты; U — ураты; P — фосфаты. При этом

$$O = WH + WD ;$$

$$U = UA + UADH + AMUR + NAUR ;$$

$$P = WTL + CA + HAP + BRU + STRU .$$

Все первичные признаки характеризуют долю соответствующего вещества в единице веса отдельного камня, т. е.

$$WH + WD + UA + UADH + AMUR + NAUR + WTL + CA + HAP + BRU + STRU + CYS = 100 .$$

В имеющихся данных обращают на себя внимание следующие моменты:

- высокая повторяемость значений признаков (среди 4217 наблюдений встречается 425 различных значений 12-мерных векторов первичных признаков, а различных значений 3-мерных векторов вторичных признаков всего 114);
- отсутствие даже намека на возможность использовать модель нормального распределения в качестве вероятностной модели данных.

Особое место занимают камни, содержащие CYS. Перечень встречающихся комбинаций различных значений состава камней с ненулевым значением CYS приведен в табл. 1. Из нее видно, что:

- имеется 49 случаев, когда речь идет о пациентах со 100%-ным содержанием CYS в камнях, они могут быть выделены в самостоятельную группу;
- остальные два случая из-за малочисленности не играют при анализе данных существенной роли.

Наличие связей между признаками дает толчок к исследованию возможности снижения размерности. В этом плане эффективность перехода к первым главным компонентам иллюстрируется значениями доли суммарной дисперсии, приведенными в табл. 2 для различных величин сниженных размерностей признакового пространства. Из этой таблицы видна возможность снижения размер-

Таблица 1 Состав камней с ненулевым значением CYS

O	U	P	CYS	Число значений
0	0	0	100	49
0	0	10	90	1
0	0	80	20	1

Таблица 2 Доля суммарной дисперсии (в %) для сниженной размерности данных

Сниженная размерность	Признаки	
	Первичные	Вторичные
1	45	64
2	66	99
3	85	100
4	90	—
5	94	—
6	96	—
7	98	—
8	99	—
...	...	—
12	100	—

ности (например, задавшись желанием обеспечить при снижении размерности не менее 95% суммарной дисперсии, можно перейти к 6-мерным данным в случае первичных признаков и к 2-мерным для вторичных признаков). Такое снижение размерности вряд ли имеет значение для уменьшения алгоритмической сложности обработки данных, а вот с точки зрения повышения качества статистического вывода это обычно дает преимущества.

Довольно бедный спектр различных значений данных, невозможность применить простые адекватные вероятностные модели для них, а также отсутствие накопленного опыта подобного моделирования в области урологии вынуждают обратиться в первую очередь к непараметрическим методам анализа данных.

Задачу анализа зависимости состава камней от пола можно сформулировать как задачу сравнения двух выборок, соответствующих тому или иному полу. Анализ временных зависимостей можно также свести к парному сравнению подвыборок, соответствующих определенным интервалам времени. При этом важно не только проверить гипотезу об однородности двух выборок, но и получить представление о характере отличий, если они есть.

3 Критерии однородности двух выборок

Достаточно богатый арсенал подходов и решений двух сформулированных задач анализа данных (обнаружение и оценивание) имеется в одномерном случае, но это не означает, что их можно считать полностью решенными. Как пример здесь достаточно упомянуть проблему Беренса—Фишера (Behrens—Fisher) о проверке гипотезы равенства средних двух независимых нормальных выборок с неизвестными дисперсиями. Как уже отмечалось, наибольший интерес вызывают непараметрические

постановки задач и соответствующие методы анализа, но при их обобщении на многомерный случай появляются дополнительные трудности: неоднозначность понятия медианы, ранга, зависимость статистик от исходного распределения. Как следствие, возникают и развиваются новые подходы к построению методов анализа однородности данных для многомерных моделей, в частности различные способы обобщения понятий медианы, знака и ранга на многомерный случай [9, 10].

Наиболее известными являются двухвыборочные критерии для параметра положения [9, 11]. К ним относятся критерий Хотеллинга (Hotelling), многомерный Манна—Уитни—Вилкоксона (Mann—Whitney—Wilcoxon) критерий, многомерный критерий Муда (Mood). В качестве параметра положения могут выступать средние или медианы в зависимости от принятой модели.

При использовании критерия Хотеллинга предполагается нормальное распределение данных и осуществляется проверка гипотезы об однородности относительно вектора средних. В построении статистики участвуют выборочные средние и ковариационные матрицы; ее распределение при условии нулевой гипотезы известно. Если предположение о нормальности не действует, то можно обратиться к многомерной версии критерия Манна—Уитни—Вилкоксона. Соответствующая статистика W^2 строится на основе объединения ранговых статистик для отдельных компонент, известно ее асимптотическое распределение. На идею объединения результатов покомпонентного анализа строится и знаковая статистика T^2 медианного критерия Муда. Для нее известно предельное распределение. Для конечных объемов выборок распределения W^2 и T^2 не являются свободными от распределения данных.

Необходимость предположений о непрерывности распределений и того, что различие выборок описывается параметром положения, приводит к тому, что перечисленные статистики малоприменимы при сравнительном анализе данных о составе камней.

С позиций устойчивости статистического вывода конкурентом вектора средних является вектор медиан [10]. Здесь можно рассматривать вектор координатных медиан, пространственную медиану, многомерную медиану Ойя (Oja), многомерную медиану Лиу (Liu), существуют и другие менее известные обобщения. Многообразие вариантов объясняется желанием авторов обеспечить такие свойства характеристик положения, как эффективность, эквивариантность, робастность, вычислительные удобства. К сожалению, для рассматриваемой ситуации с анализом состава камней все эти

обобщения не позволяют освободиться от ярко выраженной повторяемости данных и, как следствие, сводят на нет декларируемые преимущества.

Как новый позиционирует автор [12] критерий, основанный на использовании многомерных квантильных функций и сравнении параметров биномиального распределения. В предложенной редакции этот критерий не представляет интереса по следующим причинам:

- выделение всего одного множества, за вероятностью попадания в которое ведется контроль, приведет к малой мощности критерия, что наиболее ярко должно проявиться при возрастании размерности пространства данных в силу «проклятия размерности»;
- сводить итоговое принятие решения об однородности двух выборок на основе асимптотических результатов о равенстве частоты попадания данных в некоторые множества заданному значению неправильно, так как существует (см. [13, разд. 4.5]) равномерно наиболее мощный несмещенный критерий сравнения двух биномиальных совокупностей, основанный на гипергеометрическом распределении. Для последнего известно, что аппроксимация с помощью нормального распределения крайне неудовлетворительная (вопросы аппроксимации гипергеометрического распределения рассматривались в докладе [14]).

Подводя итог, можно сделать следующий общий вывод: отсутствие сформировавшихся моделей и ярко выраженная дискретность данных приводят к необходимости привлекать при анализе критериев значимости методы управления обработкой выборок, а раз так, то выбор этих методов может быть достаточно произвольным. Важно, чтобы они по возможности были чувствительны к различным отклонениям от нулевой гипотезы об однородности двух выборок и не выходили бы за границы реальных вычислительных возможностей при их воплощении.

В данной работе получил развитие подход, основанный на использовании расстояния между множествами точек. Имеющиеся выборки объемом и представляют собой два множества точек в евклидовом пространстве, поэтому естественно в качестве статистики для проверки гипотезы об однородности взять какое-то расстояние между этими множествами. В частности, это может быть сумма всех расстояний между элементами одного множества и элементами другого. Данное интуитивное представление о мере близости двух множеств укрепляется на основе доказанного в [15] следующего факта: для независимых случайных векторов $X_1,$

$X_2, Y_1, Y_2,$ где X_1, X_2 имеют одно и то же распределение F с конечным значением $E\{\|X_1\|\}$ и Y_1, Y_2 имеют одно и то же распределение G с конечным значением $E\{\|Y_1\|\}$, действует неравенство

$$E\{\|X_1 - Y_1\|\} - \frac{1}{2} E\{\|X_1 - X_2\|\} - \frac{1}{2} E\{\|Y_1 - Y_2\|\} \geq 0, \quad (1)$$

которое становится равенством тогда и только тогда, когда $F = G$. Заменяя левую часть (1) ее выборочным аналогом, помноженным на $mn/(m+n)$, получаем статистику $T_{m,n}$ для проверки нулевой гипотезы $H : F = G$ против конкурирующей $K : F \neq G$. Она имеет следующий вид:

$$T_{m,n} = \frac{mn}{m+n} \left[\frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n \|X_j - Y_k\| - \frac{1}{2m^2} \sum_{j=1}^m \sum_{k=1}^m \|X_j - X_k\| - \frac{1}{2n^2} \sum_{j=1}^n \sum_{k=1}^n \|Y_j - Y_k\| \right]. \quad (2)$$

Здесь (X_1, \dots, X_m) суть элементы одной выборки, (Y_1, \dots, Y_n) — другой. Отклонение нулевой гипотезы происходит при больших значениях статистики $T_{m,n}$. Распределение статистики критерия неизвестно. Поэтому для получения критических уровней значимости приходится прибегать к бутстреп-методу. В этой связи важным результатом является доказательство в [15] того, что распределение статистики при нулевой гипотезе является пределом ее бутстреп-распределения при $m, n \rightarrow \infty$ так, что $m/n \rightarrow \tau \in (0, 1)$.

Отсутствие априорной информации о распределении данных приводит к непараметрическому бутстреп-методу, т.е. к использованию в качестве бутстреп-распределения смеси

$$H_{m,n} = \frac{m}{m+n} F_m + \frac{n}{m+n} G_n,$$

где F_m и G_n — эмпирические функции распределения для F и G соответственно (например, для F_m — это появление одного из X_1, \dots, X_m с вероятностью $1/m$). Для получения критических уровней значимости необходимо описать распределение статистики $T_{m,n}$ для данных из распределения $H_{m,n}$; это можно сделать либо теоретическим путем, либо с помощью метода статистических испытаний. Теоретический подход не годится в основном из-за вычислительных сложностей: достаточно

Таблица 3 Степень сжатия представления данных

Признаки	Общее количество данных	Число различных значений	Коэффициент сжатия
Первичные	4217	425	10
Вторичные		114	37

большие объемы исходных данных полностью исключают прямое перечисление всевозможных бутстреп-выборок, а аналитический вывод распределения $T_{m,n}$ при всей его неочевидности, скорее всего, приведет к громоздким комбинаторным формулам, расчет по которым сопряжен с большими вычислительными погрешностями. Поэтому использовался метод статистических испытаний.

В этом случае многократное вычисление статистики $T_{m,n}$, включающей двойные суммы, создаст проблемы временного характера при реализации бутстреп-метода, решить которые можно, используя особенность данных о составе камней — относительную бедность спектра значений. Таблица 3 демонстрирует возможности сжатия данных при переходе от исходных данных к их представлению в виде совокупности различных значений.

Пусть $(Z_1, \dots, Z_{m+n}) = (X_1, \dots, X_m, Y_1, \dots, Y_n)$, $(\tilde{Z}_1, \dots, \tilde{Z}_k)$ — различные среди (Z_1, \dots, Z_{m+n}) значения, $C(l)$ — индекс \tilde{l} значения Z_l в последовательности $(\tilde{Z}_1, \dots, \tilde{Z}_k)$, т.е. $\tilde{Z}_{C(l)} = Z_l$, где $1 \leq l \leq m+n$ и $1 \leq \tilde{l} \leq k$. Это отображение C строится единожды в процессе формирования $(\tilde{Z}_1, \dots, \tilde{Z}_k)$ из (Z_1, \dots, Z_{m+n}) . Теперь можно заранее вычислить все необходимые значения $\|\tilde{Z}_i - \tilde{Z}_j\|$ и использовать их при нахождении слагаемых в (2), а генерацию бутстреп-выборок осуществлять на множестве $\{1, 2, \dots, m+n\}$

Для обнаружения зависимости состава камней от пола пациента критерий $T_{m,n}$ используется напрямую, в случае же зависимости от времени — как основа для парных сравнений групп (фрагментов) данных, относящихся к определенному промежутку времени (например, в один год).

4 Эксперименты

Для количества бутстреп-экспериментов далее было принято значение 10^4 , что обеспечивало до-

статочно точное представление для единиц процента при оценивании критического уровня \hat{a}^B описанного критерия значимости. Для задачи сравнения составов камней у мужчин и женщин полученные результаты приведены в табл. 4. Из них следует, что есть веские основания для отклонения нулевой гипотезы об однородности. Кроме крайне малых значений оценок критического уровня для наглядности приведены максимальные значения статистики критерия, встретившиеся в ходе бутстреп-экспериментов, которые также показывают существенное отклонение при нулевой гипотезе статистики критерия от обычных значений:

$$T_{mn}^* > \max_B T_{mn}^B.$$

Таким образом, обосновано наличие зависимости состава камней от пола, причем значимость этого вывода весьма высока. Но примененный метод анализа однородности данных ничего не говорит о том, что меняется в составе камней в зависимости от пола пациентов.

Далее речь пойдет о вторичных признаках О, У и Р. Напомним, что для большинства данных действует соотношение: $O + U + P = 100$. Задача состоит в формировании предположений относительно того, в чем проявляется отличие состава камней для мужчин и женщин. Визуальный анализ пар вторичных признаков выявил следующие связи: данные преимущественно сосредотачиваются вдоль отрезков на осях координат и вдоль отрезков, принадлежащих прямым вида:

$$O + U = 100; O + P = 100; U + P = 100.$$

В плоскости $O + U + P = 100$ были выделены области (для гистограммной оценки распределения их также называют ячейками), описанные в табл. 5.

Для визуализации все данные были спроецированы на плоскость $O + U + P = 100$ (рис. 1). Для

Таблица 4 Результаты проверки нулевой гипотезы об однородности данных для различных полов

Признаки	Сниженная размерность	T_{mn}^*	$\max_B T_{mn}^B$	Вывод для нулевой гипотезы	\hat{a}^B
Первичные	6	763,2	323,8	Отвергается	0%
Вторичные	2	1339,1	356,4	Отвергается	0%

Таблица 5 Вид ячеек для группирования данных

№	Состав ячеек	Наличие в составе камней
1	$\Delta_O = \{\text{abs}(O - 100) < \delta\}$	Только оксалаты
2	$\Delta_U = \{\text{abs}(U - 100) < \delta\}$	Только ураты
3	$\Delta_P = \{\text{abs}(P - 100) < \delta\}$	Только фосфаты
4	$\Delta_{OU} = \{\text{abs}(O + U - 100) < \delta\} - \Delta_O - \Delta_U$	Только оксалаты и ураты
5	$\Delta_{OP} = \{\text{abs}(O + P - 100) < \delta\} - \Delta_O - \Delta_P$	Только оксалаты и фосфаты
6	$\Delta_{UP} = \{\text{abs}(U + P - 100) < \delta\} - \Delta_U - \Delta_P$	Только ураты и фосфаты
7	$\Delta_C = \mathcal{R}^3 - \Delta_O - \Delta_U - \Delta_P - \Delta_{OU} - \Delta_{OP} - \Delta_{UP}$	Все остальные

этого достаточно воспользоваться преобразованием поворота, матрица которого имеет вид:

$$T = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{pmatrix}.$$

Далее находились частоты попадания данных в перечисленные ячейки отдельно для мужчин и женщин. Общая картина распределения данных приведена на рис. 2. На нем из общего объема данных в количестве 4217 рассматривались лишь 114 различных вариантов состава камней.

Теперь каждая полученная пара частот попадания в ячейку может быть формально исследована с целью проверки гипотезы о равенстве частот и, если они не равны, для выяснения направления изменения. Полученные результаты сведены в табл. 6.

Анализ полученных и представленных в табл. 6 результатов позволяет достаточно обоснованно сделать следующие выводы:

- нет камней по составу явно соответствующих мужчинам или женщинам, т.е., опираясь на

вторичные признаки, нельзя по составу камня определять с малыми ошибками пол пациента;

- состав камней у мужчин и женщин отличаются по частоте встречаемости O и U (уменьшение вероятности встречаемости у женщин по сравнению с мужчинами), а также по встречаемости P и O + P (увеличение вероятности встречаемости у женщин по сравнению с мужчинами);
- ячейка Δ_C достаточно хорошо представляет тип камней, в составе которых присутствует только CYS; в нее попали 49 наблюдений с CYS = 100, только 2 наблюдения с CYS = 20 и CYS = 90, а также только 2 наблюдения со следующими значениями: O = 70, U = 20, P = 10; суммарное количество наблюдений — 53 (см. в табл. 6 значения абсолютных частот 32 и 21).

В задаче анализа зависимости состава камней от времени все данные были разбиты на фрагменты, соответствующие одному году обследования (всего 5 фрагментов). Далее для каждого пола в отдельности и каждого набора признаков (первичных или вторичных) строилась матрица критических уровней значимости критерия $T_{m,n}$ для сравнения i -го

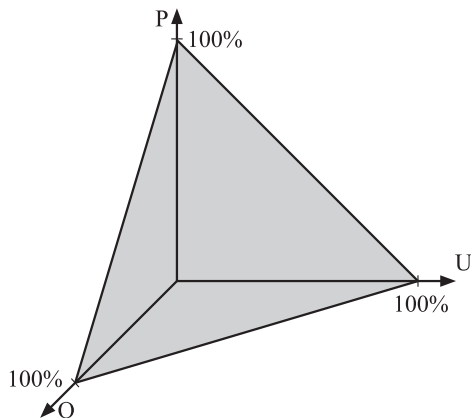


Рис. 1 Область возможных значений состава камней на плоскости $O + U + P = 100$

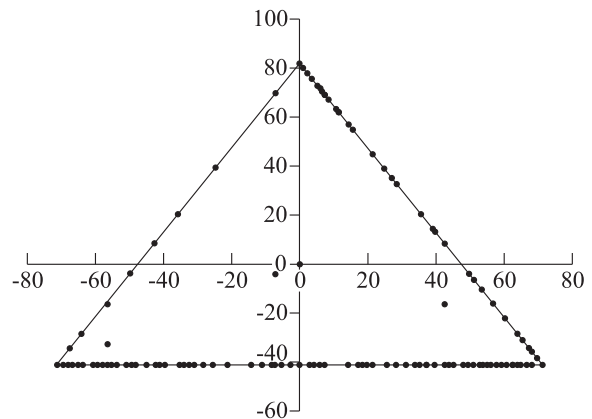


Рис. 2 Распределение данных о составе камней

Таблица 6 Сравнительный анализ состава камней для различных полов

Тип камней	Абсолютная частота		Относительная частота, %		\hat{a} , %	Направление изменения
	Пол М	Пол F	Пол М	Пол F		
Δ_O	1144	717	47,4	39,7	0,0	Уменьшение
Δ_U	273	160	11,3	8,9	0,5	Уменьшение
Δ_P	127	213	5,3	11,8	0,0	Увеличение
Δ_{OU}	164	106	6,8	5,9	12,6	Нет
Δ_{OP}	669	579	27,7	32,1	0,1	Увеличение
Δ_{UP}	4	8	0,2	0,4	2,5	Нет
Δ_C	32	21	1,3	1,2	37,4	Нет
Сумма	2413	1804				

и j -го фрагментов, $j > i$. Примеры двух подобных матриц для первичных и вторичных признаков даны в табл. 7, затемненные клетки определяют очевидным образом: на диагонали стоят значения 100% (принятие нулевой гипотезы), таблица для применяемого критерия является симметричной относительно диагонали.

Все четыре матрицы (два пола и два типа признаков) свидетельствуют о следующем:

- 5-й фрагмент (данные за 2009 г.) значительно отличается от остальных,
- первые четыре фрагмента в совокупности дают не совсем ясную картину, требующую привлечения дополнительных приемов для формирования конкретных предположений.

Установим для пары фрагментов отношение совпадения распределений данных для этих фрагментов,

оно будет выполняться при условии, если соответствующий критический уровень значимости будет превосходить некоторый порог. Здесь в качестве порогового было принято значение 1%, оно является привычным в практике анализа данных и служит, по мнению авторов работы, компромиссом между ошибками первого и второго рода при принятии гипотезы.

Матрицы данного отношения для двух рассмотренных примеров представлены в табл. 8.

Введенное отношение необязательно является транзитивным (например, для табл. 8 в случае первичных признаков условие транзитивности нарушается, а в случае вторичных — нет). Но в рассматриваемом случае анализа однородности фрагментов свойство транзитивности оказывается крайне нецелесообразным: в дополнение к имеющимся рефлексивности и симметрии (см. ранее сделанное

Таблица 7 Критические уровни значимости (в %) критерия $T_{m,n}$ в случае мужского пола

i	j									
	Первичные признаки					Вторичные признаки				
	1	2	3	4	5	1	2	3	4	5
1		1,3	2,2	0,3	0,0		18,9	49,4	81,7	0,0
2			36,0	8,3	0,0			36,8	5,0	0,0
3				46,9	0,0				25,4	0,0
4					0,0					0,0
5										0,0

Таблица 8 Матрицы отношения совпадения в случае мужского пола

i	j									
	Первичные признаки					Вторичные признаки				
	1	2	3	4	5	1	2	3	4	5
1		1	1	0	0		1	1	1	0
2			1	1	0			1	1	0
3				1	0				1	0
4					0					0
5										0

замечание о затемненных клетках таблиц) транзитивность приводит к отношению эквивалентности и, следовательно, к возможности получить разбиение для исходного множества фрагментов. Так, в случае вторичных признаков из табл. 8 получаем, что таковым разбиением является $\{\{1, 2, 3, 4\}, \{5\}\}$, что означает следующее: состав камней по вторичным признакам у мужчин в 2005–2008 г. не изменялся, а 2009 г. в этом смысле отличается от предыдущих.

Поставим задачу нахождения транзитивного приближения — такого транзитивного отношения, которое было бы наиболее близким к заданному. Если R_1 и R_2 — матрицы рефлексивных, симметричных отношений, то для определения их меры близости используем расстояние Хемминга, а именно:

$$\rho(R_1, R_2) = |\{i, j : 1 < i < j < n, R_1(i, j) \neq R_2(i, j)\}|.$$

Тогда если R_0 — матрица заданного отношения (например, матрица из табл. 8 для первичных признаков), то искомой матрицей $R_T^*(R_0)$ из множества \mathfrak{R}_T матриц всех рефлексивных, симметричных, транзитивных отношений будет следующая:

$$R_T^*(R_0) = \arg(\min_{R \in \mathfrak{R}_T} \rho(R_0, R)).$$

Полученное решение $R_T^*(R_0)$ необязательно единственное. При этом $R_T^*(R_0)$ необязательно является $\text{TrCl}(R_0)$ — транзитивным замыканием для R_0 . Более того, число случаев s , когда $\text{TrCl}(R_0)$ вообще не попадает в множество найденных $R_T^*(R_0)$, увеличивается с ростом n . Как иллюстрация этого, табл. 9 содержит первые значения отношения s/r , где r — общее число нетранзитивных отношений среди всех рефлексивных и симметричных отношений. Для $n > 7$ подобные значения получить за приемлемое время не удастся. Следует отметить, что для проведенных в рамках работы экспериментов пришлось реализовывать алгоритмы перечисления всех рефлексивных и симметричных отношений и перечисления всех отношений эквивалентности (через построение всех разбиений множества).

Для поиска решений $R_T^*(R_0)$ в данной работе был применен алгоритм простого перебора: для всех рефлексивных, симметричных и транзитивных отношений подсчитывались расстояния $\rho(R_0, R_T)$, минимум ρ^* из которых давал решение искомой задачи. Подобная процедура нахождения $R_T^*(R_0)$ реально срабатывает при $n \leq 14$; причина проста: быстрый рост мощности перебираемого множества

Таблица 9 Доля случаев, когда транзитивное замыкание не попадает в множество приближающих транзитивных отношений

n	s/r
3	0%
4	57%
5	72%
6	91%
7	94%
> 7	Не доступно

Таблица 10 Разбиения данных по годам

Пол	Тип признаков	ρ^*	R_T^*
М	Первичные	1	$\{\{1, 2, 3, 4\}, \{5\}\}$
	Вторичные	0	$\{\{1, 2, 3, 4\}, \{5\}\}$
Ф	Первичные	1	$\{\{1, 2, 3, 4\}, \{5\}\}$
	Вторичные	2	$\{\{1, 2, 3, 4\}, \{5\}\}$ $\{1, 2, 3, 4, 5\}$

отношений эквивалентности (это суть числа Белла, а именно: при $n = 2$ — это 2, при $n = 5$ — это 52, при $n = 10$ — это 115 975, при $n = 14$ — это 190 899 322, при $n = 15$ — это 1 382 958 545).

Напомним, что минимум может быть не единственный, поэтому при выборе соответствующего разбиения возникает некоторый произвол. Результаты анализа всех исследуемых случаев сведены в табл. 10. Для женского пола и вторичных признаков было отобрано разбиение $\{\{1, 2, 3, 4\}, \{5\}\}$ как наиболее отвечающее ранее сделанному выводу о том, что 5-й фрагмент (данные за 2009 г.) значимо отличается от остальных.

Полученные разбиения позволяют высказать предположения о содержании изменений в составе камней на рубеже 4-го и 5-го фрагментов (переход от 2008 к 2009 г.). Для выделенных типов камней (см. табл. 6) исследовалось изменение частот их встречаемости для периодов времени 2005–2008 гг. и 2009 г. Формальное сравнение частот встречаемости за эти периоды позволило выделить часть типов камней, для которых изменения были значимыми. Для мужского пола речь идет о типах камней, по составу входящих в ячейки Δ_O , Δ_{OP} , Δ_{UP} и Δ_C (напомним, что в Δ_C оказываются камни фактически только с CYS), для женского пола — это Δ_O и Δ_{OP} . Соответствующие данные о доле камней, входящих в перечисленные ячейки, приведены на рис. 3; на них штриховыми линиями обозначены оценки частот f для разбиения $\{1, 2, 3, 4, 5\}$, а сплошными линиями выделены оценки частот f^* , полученные для разбиения $\{\{1, 2, 3, 4\}, \{5\}\}$.

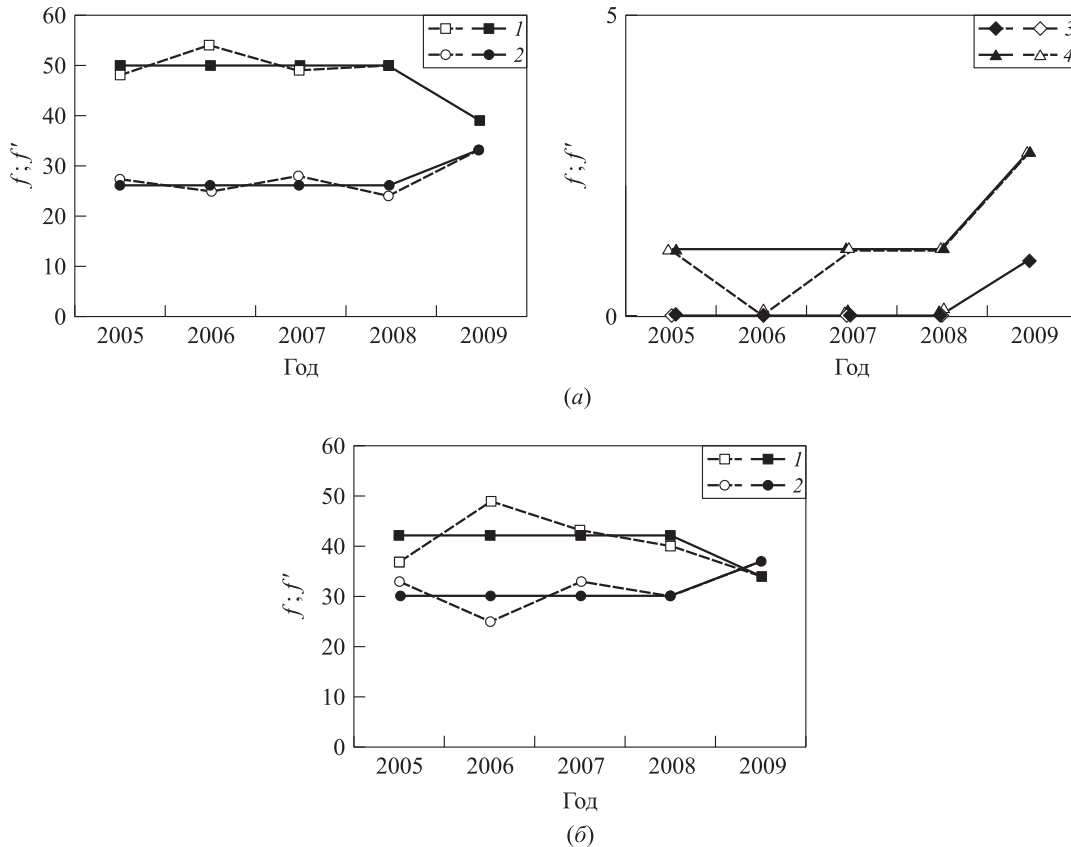


Рис. 3 Изменения состава камней в зависимости от года: (а) пол мужской; (б) пол женский; 1 — О; 2 — ОР; 3 — УР; 4 — С; штриховые линии — f , разбиение $\{1, 2, 3, 4, 5\}$; сплошные линии — f' , разбиение $\{\{1, 2, 3, 4\}, \{5\}\}$

Приведенные результаты свидетельствуют о реалистичности следующих выводов:

- для обоих полов изменения в первую очередь касаются камней типа О и О + Р;
- для мужского пола также характерен рост доли камней, включающих U + Р и С.

5 Заключение

В статье рассмотрены методы исследования однородности данных, учитывающие специфику предметной области: многомерность данных, бедный спектр различных значений показателей, невозможность применить простые адекватные вероятностные модели для них.

Предлагается и анализируется критерий значимости, основанный на расстоянии между множествами точек в евклидовом пространстве, строится эффективная оригинальная реализация бутстреп-метода. При множественном использовании данного критерия впервые применяются идеи аппроксимации возникающих связей между фрагментами данных с помощью отношения эквивалентности,

что позволяет обоснованно проводить разбиение исходных данных.

В ходе экспериментов получены статистически обоснованные результаты как об особенностях распространенности метаболических типов мочекаменной болезни в зависимости от пола пациента, так и о характерных изменениях в соотношении этих типов с течением времени.

Литература

1. Ramello A., Vitale C., Marangella D. Epidemiology of nephrolithiasis // J. Nephrol., 2000. Vol. 13. Suppl. 3. P. 45–50.
2. Pak C. Y., Resnick M. I., Preminger G. M. Ethnic and geographic diversity of stone disease // Urology, 1997. Vol. 50. No. 4. P. 504–507.
3. Takasaki E. Chronological variation in the chemical composition of upper urinary tract calculi // J. Urology, 1986. Vol. 136. No. 1. P. 5–9.
4. Trinchieri A., Coppi F., Montanari E., Del Nero A., Zanetti G., Pisani E. Increase in the prevalence of symptomatic upper urinary tract stones during the last ten years // Eur. Urol., 2000. Vol. 37. P. 23–25.

5. *Arias Funez F., Garcia Cuerpo E., Lovaco Castellanos F., Escudero Barrilero A., Avila Padilla S., Villar Palasi J.* Epidemiologia de la litiasis urinaria en nuestra Unidad. Evolucion en el tiempo y factores predictivos [Epidemiology of urinary lithiasis in our unit. Clinical course in time and predictive factors] // *Arch. Esp. Urol.*, 2000. Vol. 53. No. 4. P. 343–347.
6. *Тиктинский О. Л., Александров В. П.* Мочекаменная болезнь. — СПб.: Питер, 2000. 379 с.
7. *Шуберт Г., Чудновская М. В., Тыналиев М. Т., Поповкин Н. Н., Тимин А. Р.* Особенности химического состава и структуры мочевых камней и их распространенность в городах Москве, Берлине и Киргизской ССР // *Урология и нефрология*, 1990. № 5. С. 49–54.
8. *Фрейтаг Д., Хруска К.* Патофизиология нефролитиаза // *Почки и гомеостаз в норме и при патологии / Под ред. С. Клар; пер. с англ. Е. И. Дайхина.* — М.: Медицина, 1987. С. 390–420. (*Klahr S.* The kidney and body fluids in health and disease. — N.Y., L.: Plenum Medical Books, 1983.)
9. *Hettmansperger T. P.* Multivariate location tests // *Encyclopedia of statistical sciences.* — N.Y.: John Wiley & Sons, 2006. P. 5249–5252.
10. *Minimaa A., Oja H.* Multivariate median // *Encyclopedia of statistical sciences.* — N.Y.: John Wiley & Sons, 2006. P. 5258–5266.
11. *Marden J. I.* Multivariate rank tests // *Multivariate analysis, design of experiments and survey sampling / Ed. S. Ghosh.* — N.Y.: Marcel Dekker, 1999. P. 401–432.
12. *Чистяков С. П.* О новом многомерном статистическом критерии однородности двух выборок // *Тр. Карельского научного центра РАН*, 2010. № 3. С. 93–97.
13. *Леман Э.* Проверка статистических гипотез. — М.: Наука, 1979. 408 с.
14. *Кривенко М. П.* Задачи выборочного контроля при досмотре лиц, багажа и транспорта // *Обозрение прикладной и промышленной математики*, 2011. Т. 18. Вып. 1. С. 125–126.
15. *Baringhaus L., Franz C.* On a new multivariate two-sample test // *J. Multivariate Anal.*, 2004. Vol. 88. P. 190–206.

Поступила в редакцию 05.02.13

ANALYSIS OF DATA HOMOGENEITY OF THE CHEMICAL COMPOSITIONS OF STONES IN CASE OF UROLITHIASIS

M. P. Krivenko¹, S. A. Golovanov², and A. V. Sivkov²

¹Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

²Research Institute of Urology, Moscow 105425, Russian Federation

Abstract: The article considers the methods for researching homogeneity of the data on chemical composition of urinary stones. Multidimensionality of the data, scarce spectrum of various values of indicators, impossibility to apply simple adequate likelihood models to them, and, also, absence of experience in such type of modeling in the field of urology are the cause of the considered problems specificity. The test of significance based on distance between sets of points in Euclidean space is suggested and analyzed. It is used to analyze a composition of stones dependence on the sex of a patient. In the case of analysis of changes in the composition of stones with time (as time goes), the proposed criterion is used as an element of pairwise comparisons of data for different time intervals. That is the first time when the idea of approximation of emerging relationships between groups of data with the help of equivalence relations is used. It allows reasonably carrying out the stratification of the original data. The experiments produced statistically valid results about the features of urolithiasis depending both on the sex of the patient and the changes over time.

Keywords: tests of homogeneity; stratification; bootstrap method; urolithiasis

DOI: 10.14357/19922264130410

References

1. Ramello, A., C. Vitale, and D. Marangella. 2000. Epidemiology of nephrolithiasis. *J. Nephrol.* 13(3):45–50.
2. Pak, C. Y., M. I. Resnick, and G. M. Preminger. 1997. Ethnic and geographic diversity of stone disease. *Urology* 50(4):504–507.
3. Takasaki, E. 1986. Chronological variation in the chemical composition of upper urinary tract calculi. *J. Urology* 136(1):5–9.
4. Trinchieri, A., F. Coppi, E. Montanari, A. Del Nero, G. Zanetti, and E. Pisani. 2000. Increase in the prevalence of symptomatic upper urinary tract stones during the last ten years. *Eur. Urol.* 37:23–25.

5. Arias Funez, F., E. Garcia Cuerpo, F. Lovaco Castellanos, A. Escudero Barrilero, S. Avila Padilla, and J. Villar Palasi. 2000. Epidemiologia de la litiasis urinaria en nuestra Unidad. Evolucion en el tiempo y factores predictivos. [Epidemiology of urinary lithiasis in our unit. Clinical course in time and predictive factors]. *Arch. Esp. Urol.* 53(4):343–347.
6. Tiktinskij, O. L., and V. P. Aleksandrov. 2000. *Mochekamennaya bolezni' [Urolithiasis]*. St. Petersburg, Russia: Piter. 379 p.
7. Shubert, G., M. V. Chudnovskaja, M. T. Tynaliev, N. N. Popovkin, and A. R. Timin. 1990. Osobennosti khimicheskogo sostava i struktury mochevykh kamney i ikh rasprostranennost' v gorodakh Moskve, Berline i Kirgizskoy SSR [Peculiarities of chemical composition and structure of urinary stones and their prevalence in the cities of Moscow, Berlin, and the Kirghiz SSR]. *Urologiya i Nefrologiya [Urology and Nefrology]* 5:49–54.
8. Freitag, D., and K. Hruska. 1983. Pathophysiology of nephrolithiasis. *The kidney and body fluids in health and disease*. Ed. S. Klahr. N.Y., L.: Plenum Medical Books.
9. Hettmansperger, T. P. 2006. Multivariate location tests. *Encyclopedia of statistical sciences*. N.Y.: John Wiley & Sons. 5249–5252.
10. Minimaa, A., and H. Oja. 2006. Multivariate median. *Encyclopedia of statistical sciences*. N.Y.: John Wiley & Sons. 5258–5266.
11. Marden, J. I. 1999. Multivariate rank tests. *Multivariate analysis, design of experiments and survey sampling*. Ed. S. Ghosh. N.Y.: Marcel Dekker. 401–432.
12. Chistjakov, S. P. 2010. O novom mnogomernom statisticheskom kriterii odnorodnosti dvukh vyborok [On a new multidimensional statistical test of homogeneity of two samples]. *Trudy Karel'skogo Nauchnogo Centra RAN* 3:93–97.
13. Leman, Je. 1979. *Proverka statisticheskikh gipotez* [Testing of statistical hypothesis]. Moscow: Nauka. 408 p.
14. Krivenko, M. P. 2011. Zadachi vyborochnogo kontrolya pri dosmotre lits, bagazha i transporta [The tasks of sampling during the inspection of persons, baggage, and vehicles]. *Obozrenie Prikladnoy i Promyshlennoy Matematiki* 18(1):125–126.
15. Baringhaus, L., and C. Franz. 2004. On a new multivariate two-sample test. *J. Multivariate Anal.* 88:190–206.

Received February 5, 2013

Contributors

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

Golovanov Sergey A. (b. 1950) — Doctor of Science in medicine, Head of Laboratory, Research Institute of Urology, Moscow 105425, Russian Federation; sergeygol124@mail.ru

Sivkov Andrey V. (b. 1957) — Doctor of Science in medicine, Deputy director, Research Institute of Urology, Moscow 105425, Russian Federation; uroinfo@yandex.ru

О МЕТОДЕ ПРОГНОЗИРОВАНИЯ И КЛАССИФИКАЦИИ ДЛЯ ЦЕНЗУРИРОВАННЫХ ДАННЫХ

Т. В. Захарова¹, Е. М. Абрамова²

Аннотация: В данной работе решалась задача классификации пациентов с инсулинонезависимым сахарным диабетом (ИНСД) и выявления признаков, по которым у пациентов можно предположить наличие данного диагноза. В первоначальном виде медицинские данные не удалось классифицировать. Основной помехой для использования классического метода дискриминации стала недостаточность и неполнота данных. Для обработки таких данных предлагается метод классификации, в котором определенным образом выбираются различные наборы дискриминантных признаков и строятся классификационные функции для каждого набора. Число таких наборов зависит от степени неполноты данных: чем больше потерянных данных, тем больше число различных наборов. Пациент классифицируется каждой дискриминантной системой и относится к той группе, для которой было получено максимальное число совпадений классификации. Такая многоступенчатая классификация компенсирует малый объем выборки и максимально использует информацию о каждом пациенте.

Ключевые слова: гипотеза; цензурированные данные; дискриминантные переменные; классификационные функции; прогнозирование

DOI: 10.14357/19922264130411

1 Введение

Задачи классификации и прогнозирования являются достаточно важными и актуальными проблемами, решаемыми методами математической статистики. Задачи подобного рода в медицинских исследованиях имеют дополнительную значимость. Правильно поставленный диагноз на начальных стадиях болезни может значительно облегчить само лечение и продлить жизнь человека, тем более если это опасное заболевание.

ФГУ РЦСМЭ Росздрава (Российский центр судебно-медицинской экспертизы, г. Москва) предоставил авторам статьи данные, характеризующие такое тяжелое заболевание, как сахарный диабет. Число больных сахарным диабетом в России составляет около 10 млн человек. Из них почти 80% страдают ИНСД, или диабетом 2-го типа. Приблизительно у половины всех больных ИНСД заболевание не распознается вовремя. Перед авторами была поставлена задача выявления пациентов с подозрением на заболевание.

Существенной особенностью анализируемых данных было отсутствие у каких-то пациентов одних данных, у каких-то — других, т. е. налицо была различная неполная информация о каждом пациенте. В таких случаях данные принято называть

многократно цензурированными. Как строить выборку? Естественно, не хотелось терять никакую полезную информацию о пациентах, тем более что число наблюдаемых пациентов было невелико.

Задачи оценивания, связанные с медициной, часто приходится решать по неполной выборке. Для таких задач разработаны специальные методы, наиболее известными из них являются методы анализа выживаемости, например построение таблиц времен жизни. Отметим, что анализ выживаемости применяется к многомерным непрерывным величинам, в данном же случае большинство переменных наблюдения были дискретными. Специфика всех применяемых при этом методов определяется наличием цензурированных наблюдений.

Исследование, проведенное авторами в сотрудничестве с РЦСМЭ Росздрава, было основано на методах дискриминантного анализа. Дискриминантный анализ позволяет решать многие прикладные статистические задачи, связанные с распознаванием [1]. Его методы широко применяются и в медицинских исследованиях [2–4].

В данной работе решалась задача классификации пациентов с ИНСД (далее — СД) и выявления признаков, по которым у человека можно предположить наличие данного диагноза. Авторам были

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, isa@cs.msu.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, houselake@gmail.com

предоставлены данные по группам людей, так называемые «обучающие выборки»:

- группа 1 — с недиагностированным СД (при жизни диагноз СД не выставлялся);
- группа 2 — с выявленным СД;
- группа 3 — здоровые.

В первоначальном виде медицинские данные не удалось классифицировать. Основной помехой для использования классического метода дискриминации служила недостаточность и неполнота данных. Также работа осложнялась малым объемом выборки, некоторыми неточностями заполнения медицинских карт и различием регистрируемых параметров.

Для обработки таких данных предлагается эвристический метод классификации, в котором определенным образом выбираются различные наборы дискриминантных признаков и строятся классификационные функции для каждого набора. Число таких наборов зависит от степени неполноты данных: чем больше потерянных данных, тем больше число различных наборов. Каждый набор определяет свои классификационные функции для каждой из трех групп (такую совокупность назовем дискриминантной системой). Новый пациент со своим набором признаков классифицируется каждой дискриминантной системой и относится к той группе, для которой было получено максимальное число совпадений классификации. Такая многоступенчатая классификация в некотором смысле компенсирует малый объем выборки и максимально использует информацию о каждом пациенте. Подробное описание метода приведено ниже.

В сходной постановке задачи классификации решались в [5–7].

Исследование, проведенное в данной работе, позволяет установить некоторые совокупности медицинских измерений, которые являются информативными и впоследствии смогут помочь при постановке и выявлении диагноза СД.

2 Постановка задачи

Классификация больных проводилась на основе медицинских показателей при помощи дискриминантного анализа в пакете STATISTICA.

Пусть задано пространство измерений (в рассматриваемой задаче это патологоанатомические измерения), обладающее размерностью $m > 1$. Элементами данного пространства являются векторы $x = (x_1, \dots, x_i, \dots, x_m)$, где x_i — значение признака номер i для конкретного пациента. Кроме того, заданы n групп (классов). Для каждого

больного необходимо указать группу, к которой он принадлежит.

В пространстве измерений можно выделить множество векторов обучающей выборки, т. е. векторов, для которых известна принадлежность к конкретной группе. Символом α_k обозначим число измерений обучающей выборки, принадлежащих группе k , пусть $x^{(j,k)}$ — вектор из обучающей выборки, соответствующий j -му пациенту из группы k .

В основе линейного дискриминантного анализа лежит нахождение такой линейной комбинации переменных, которая бы оптимально разделила рассматриваемые группы. Семейство линейных функций

$$d^{(k)} = a_0^{(k)} + a_1^{(k)} * x_1 + a_2^{(k)} * x_2 + \dots + a_m^{(k)} * x_m, \\ k = \overline{1, n}$$

называется дискриминантной функцией. Ее аргументом служат компоненты вектора x , принадлежащего пространству измерений. Коэффициенты $a_j^{(i)}$ подлежат определению на основе обучающей выборки. В задаче, рассматриваемой в настоящей работе, число групп равно трем. Поэтому необходимо рассчитать 3 линейные дискриминантные функции.

Для расчета коэффициентов дискриминантных функций нужен статистический критерий, оценивающий различия между группами. Задачу поиска наилучшей дискриминации данных можно свести к поиску таких дискриминантных функций $d^{(k)}$, которые были бы основаны на максимуме отношения межгрупповой вариации к внутригрупповой. Предположим, что все наблюдения являются реализацией некоторой случайной величины X , имеющей многомерное нормальное распределение. Такое предположение является общим для большинства методов дискриминантного анализа.

Введем некоторые обозначения.

Символом $\overline{x_i^{(k)}}$ обозначим среднее значение признака i , рассчитанного для представителей группы k , символом $\overline{\overline{x_i}}$ — среднее значение признака i по всем пациентам сразу из всех групп. Через $x_i^{(l,k)}$ обозначим i -й признак у l -го пациента в группе k .

Используя введенные осреднения, определим квадратные матрицы W и T :

$$W_{ij} = \sum_{k=1}^n \sum_{l=1}^{\alpha_k} (x_i^{(l,k)} - \overline{x_i^{(k)}})(x_j^{(l,k)} - \overline{x_j^{(k)}}); \\ T_{ij} = \sum_{k=1}^n \sum_{l=1}^{\alpha_k} (x_i^{(l,k)} - \overline{\overline{x_i}})(x_j^{(l,k)} - \overline{\overline{x_j}}).$$

Матрица W называется внутригрупповой выборочной вариационной матрицей признаков i и j . Матрица T называется матрицей рассеяния объединенных данных.

Пусть $B = T - W$. Величины элементов B по отношению к величинам элементов W дают меру различия между группами.

Коэффициенты разделяющих функций могут быть найдены по методу дискриминантного анализа Фишера как элементы матрицы, обратной к W , что соответствует общей вычислительной процедуре множественной линейной регрессии.

Для нахождения p наборов коэффициентов канонических дискриминантных функций необходимо решить обобщенную задачу на собственные значения:

$$B\nu^{(k)} = \lambda_k W\nu^{(k)}, \quad \|\nu^{(k)}\| = 1, \quad j = \overline{1, k}.$$

Найденные $\nu^{(k)}$ используются для подсчета коэффициентов дискриминантной функции:

$$a_i^{(k)} = \nu^{(k)} \sqrt{\alpha_k - n}.$$

Для проверки «качества» найденных коэффициентов возьмем статистическую гипотезу $H_0 : m_1 = m_2 = \dots = m_p$, заключающуюся в проверке отсутствия различий между групповыми средними.

Для принятия или отклонения гипотезы H_0 используется критерий, основанный на U -статистике Уилкса, которая вычисляется как отношение определителей (детерминантов) матрицы внутригрупповой вариации W и полной вариационной матрицы T :

$$U = \frac{\det(W)}{\det(T)}.$$

Если значения статистики U близки к единице, то вероятность ошибки дискриминации близка к 1.

После того как функции построены, можно провести классификацию любого произвольного измерения. Алгоритм классификации следующий.

В каждую функцию в качестве аргументов подставляются компоненты вектора наблюдений x и вычисляются значения функции. За номер группы, к которой принадлежит наблюдение, берется номер той функции, значение которой оказалось наибольшим.

3 Решение задачи

3.1 Первичная обработка данных.

Разведочный статистический анализ

На первых этапах работы делались попытки представить имеющиеся исходные данные в наи-

более удобной и наглядной форме. Все данные были собраны в таблицу.

Изначально данные состояли из наблюдений над 272 пациентами по 132 признакам. Кроме того, что объем выборки был небольшим, обработка осложнялась неполнотой данных: практически у каждого пациента отсутствовал ряд наблюдений вследствие утери данных, которые не подлежали восстановлению. Таким образом, имелись наблюдения над пациентами со своими наборами признаков у каждого. Отметим, что исходная матрица данных была заполнена менее чем на четверть, по большинству признаков наблюдения имелись не более чем у 10 пациентов.

Очевидно, что входные данные должны быть различимы в каждой группе. Но часть предоставленных признаков равнялась константе (в основном это были бинарные признаки) и не могла служить материалом для исследования. Часть признаков относилась к типу качественных и была переведена в числовой вид путем разумного упорядочивания и последующей кодировки цветов.

После первичной обработки было отобрано 90 пациентов из первой группы и 62 пациента из второй. Затем при помощи модуля Discriminant Analysis пакета STATISTICA с использованием обучающих выборок была проведена классификация данных, результаты которой представлены на рис. 1.

Из рис. 1 видно, что данные не удалось разделить по совокупности всех предоставленных признаков. Требовалось снизить размерность пространства признаков, оставив наиболее информативные. Поэтому следующим этапом стал подбор значимых признаков из имеющихся 132 признаков и отсеивание пациентов, не имеющих данных признаков. Выборка признаков производилась так: измерения должны были присутствовать не менее чем у половины пациентов, признаки внутри групп должны были различаться, пациенты с данными

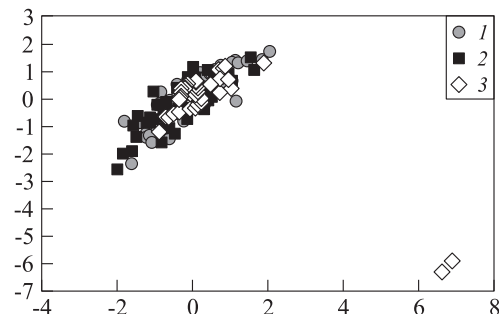


Рис. 1 Деление начальных данных на 3 группы: 1 — недиагностированные; 2 — диагностированные; 3 — отсутствие диагноза

менее чем по 10 признакам не вносились в таблицу.

3.2 Алгоритм классификации

Для построения адекватной модели были выбраны 32 признака, которые считались наиболее значимыми по медицинским показаниям. Следующим этапом было нахождение дискриминантных функций. Результаты, полученные в подразд. 3.1, показали, что при использовании всех 32 признаков теряется существенная часть наблюдений. Пациенты с не полностью заполненными векторами наблюдений не учитываются при нахождении коэффициентов дискриминантных функций, это становится причиной потери точности расчетов. Вследствие отсутствия у большинства пациентов наблюдений по всем 32 признакам и желания сохранить как можно больше наблюдений, было принято решение о построении нескольких систем дискриминантных функций. Построение каждой из 32 комбинаций происходило следующим образом: сначала из матрицы наблюдений удалялись все пациенты, не имеющие наблюдений по данному признаку, затем удалялись признаки, не заполненные целиком для каждого из оставшихся пациентов. Проведя данную операцию 32 раза, авторы получили 32 комбинации признаков (по одной комбинации на каждый признак). Были учтены все пациенты и не потеряно ни одно измерение. При сохранении максимального объема данных было получено 32 системы дискриминантных функций, что позволило в дальнейшем проводить более точную дискриминацию пациентов, увеличивая размер выборки. При помощи 32 дискриминантных систем стало возможным 32 раза проводить дискриминацию пациентов и более точно определять номер группы пациента.

Для определения возможности использования дискриминантного анализа необходимо было проверить данные на принадлежность к нормальному распределению.

В процессе проверки на нормальность были произведены следующие измерения:

- проверка матрицы данных для дискриминации групп 1 и 2;
- проверка матрицы данных по совокупности групп 1 и 2 с группой 3;
- проверка матрицы данных для каждой из совокупностей признаков, найденных в предыдущем пункте.

Результаты в каждом случае оказались разными. Многие признаки не были распределены нормально, что видно из табл. 1. Буквой N отмечался

Таблица 1 Типы распределения признаков

Признак	Распределение
НО-02А	N
ОБ-02	
ССС-06	
ССС-07С	
ССС-14	
П-3	N
ПЖ-2	N
ПЖ-6а	
НО-47А	N
ОБ-03	
ССС-07	
ССС - 07D	
ССС-15	
П-5а*	N
ПЖ-3	
ПЖ-7	
ССС-01А	
ССС-04А	
ССС-07Аа*	N
ССС-12*	N
П-1а	N
П-6А	
ПЖ-4	N
ПЖ-7А	N
ССС-02	
ССС-05А	
ССС-07Ва	N
ССС-13*	
П-2а	N
ПЖ-1	N
ПЖ-5	N
ПЖ-8	N

признак, для которого выполнялась гипотеза о нормальном распределении. Символом * отмечены численные признаки.

В табл. 1 приведены данные для общей матрицы данных по трем группам. Оставшиеся признаки были перекодированы из качественных. При проверке данных для каждой подгруппы на нормальность результаты оказались следующими: в подгруппах, построенных для дискриминации групп 1 и 2, около 80% признаков были распределены нормально. В подгруппах, построенных для дискриминации совокупностей групп 1 и 2 с группой 3, около 40% признаков обладали нормальным распределением. Следует отметить, что нормальность признаков не является критической для правильной работы дискриминантного анализа [1, 2].

3.3 Исследование различий пациентов групп 1 и 2

После обработки данных было получено пространство из 32 признаков, в котором была про-

Таблица 2 Разделение 1-й и 2-й групп по наборам параметров

Комбинация параметров	Значение U -статистики
$X_1 X_8 X_{17} X_{18}$	0,945466
$X_1 X_2 X_8 X_{17} X_{18}$	0,940606
$X_1 X_3 X_4 X_8 X_{17} X_{18}$	0,911782
$X_1 X_7 X_8 X_{17} X_{18}$	0,945990
$X_1 X_8 X_{10} X_{17} X_{18}$	0,948152
$X_1 X_8 X_{11} X_{17} X_{18}$	0,942741
$X_1 X_8 X_{12} X_{14} X_{17} X_{18}$	0,930141
$X_1 X_8 X_{12} X_{13} X_{14} X_{17} X_{18}$	0,929915
$X_1 X_8 X_{12} X_{14} X_{17} X_{18}$	0,930141
$X_1 X_8 X_{15} X_{16} X_{17} X_{18}$	0,924349
$X_1 X_8 X_{16} X_{17} X_{18}$	0,939543
$X_1 X_8 X_{17} X_{18} X_{19} X_{22}$	0,958094
$X_1 X_8 X_{17} X_{18} X_{20} X_{22}$	0,944039
$X_1 X_8 X_{17} X_{18} X_{22}$	0,945476
$X_1 X_8 X_{17} X_{18} X_{23}$	0,938776
$X_1 X_8 X_{17} X_{18} X_{28}$	0,912175
$X_1 X_8 X_{17} X_{18} X_{22} X_{29}$	0,945256
$X_1 X_8 X_{17} X_{18} X_{31}$	0,929865
$X_1 X_8 X_{17} X_{18} X_{22} X_{32}$	0,946031

ведена дискриминация по методу, описанному в подразд. 3.2. Было видно, что полученные группы не разделяются статистически. Приведем средние значения статистики Уилкса для некоторых комбинаций признаков (табл. 2). В табл. 2 признаки перекодированы.

Видно, что все значения статистики превосходят 0,9. Таким образом, группы 1 и 2 по выбранным

признакам оказались схожи. Это соответствовало действительности: исследуемые категории лиц были больны СД.

3.4 Решение задачи разделения совокупностей групп 1 и 2 и группы 3

Для проверки предлагаемого метода разделения были использованы данные пациентов, не имеющих СД. Алгоритм классификации был описан в подразд. 3.2.

В табл. 3 приведены значения статистики Уилкса для различных комбинаций параметров.

На рис. 2 приведена дискриминация данных на 3 группы. Из рисунка можно видеть, что группы 1 и 2 не отличимы, но заметно разделение на больных СД и здоровых.

3.5 Пример классификационных функций и верификация предложенного метода

Проверка результатов проходила при помощи нового массива данных пациентов 1-й, 2-й и 3-й групп. Процент верного прогноза составил 75%–80%. Таким образом, полученная классификация при помощи обучающей выборки на реальных данных показала хорошие результаты. Была существенно снижена размерность дискриминантных

Таблица 3 Разделение между группами по наборам параметров

Комбинация параметров	Значение U -статистики
X_1, X_8, X_{17}, X_{18}	0,74961
$X_1, X_2, X_8, X_{17}, X_{18}$	0,732198
$X_1, X_3, X_4, X_8, X_{17}, X_{18}$	0,726553
$X_1, X_5, X_6, X_7, X_8, X_{12} - X_{20}, X_{22}$	0,210394
$X_1, X_6, X_7, X_8, X_{12} - X_{18}, X_{20}, X_{22}$	0,349879
$X_1, X_7, X_8, X_{17}, X_{18}$	0,669331
$X_1, X_8, X_9, X_{12}, X_{14}, X_{17}, X_{18}$	0,527898
$X_1, X_8, X_{10}, X_{17}, X_{18}$	0,630852
$X_1, X_8, X_{12}, X_{13}, X_{14}, X_{17}, X_{18}$	0,708013
$X_1, X_8, X_{17}, X_{18}, X_{19}, X_{22}$	0,688674
$X_1, X_8, X_{12}, X_{13}, X_{14}, X_{17}, X_{18}, X_{21}, X_{22}$	0,604662
$X_1, X_8, X_{17}, X_{18}, X_{23}$	0,684530
$X_1, X_8, X_{12}, X_{13}, X_{14}, X_{17}, X_{18}, X_{24}$	0,548464
$X_1, X_8, X_{12}, X_{13}, X_{14}, X_{17}, X_{18}, X_{24}, X_{25}, X_{26}$	0,518108
$X_1, X_8, X_{12}, X_{13}, X_{14}, X_{17}, X_{18}, X_{26}$	0,525807
$X_1, X_3, X_4, X_8, X_{11} - X_{18}, X_{20}, X_{22}, X_{27}, X_{30}, X_{31}$	0,128547
$X_1, X_8, X_{17}, X_{18}, X_{22}, X_{29}$	0,674781
$X_8, X_{14}, X_{17}, X_{18}, X_{22}, X_{30}, X_{31}$	0,668601
$X_1, X_3, X_4, X_8, X_{11} - X_{14}, X_{17}, X_{18}, X_{22}, X_{30}, X_{31}$	0,646518
$X_3, X_4, X_8, X_{13}, X_{14}, X_{17}, X_{18}, X_{22}, X_{30}, X_{31}$	0,652206

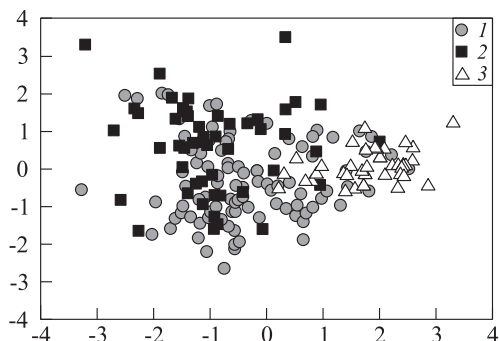


Рис. 2 Конечное деление данных на 3 группы: 1 — недиагностированные; 2 — диагностированные; 3 — отсутствие диагноза

признаков и получена высокая достоверность прогнозирования.

Для некоторого набора признаков ниже приведена система классификационных функций:

$$\begin{aligned}
 f_1 &= 5,91X_1 + 12,03X_3 - 31,42X_4 + 7,94X_8 - \\
 &- 0,02X_{11} - 1,8X_{12} - 6,1X_{13} + 16,77X_{14} + 137,72X_{15} + \\
 &+ 13,9X_{16} + 1,89X_{17} - 2,35X_{18} - 11,34X_{20} - 0,02X_{22} + \\
 &+ 42,23X_{27} + 3,96X_{30} - 0,48X_{31} - 96,71; \\
 f_2 &= 5,53X_1 + 24,2X_3 - 46,25X_4 + 5X_8 - 0,01X_{11} + \\
 &+ 17,87X_{12} + 23,5X_{13} - 14,41X_{14} + 93,54X_{15} + \\
 &+ 23,43X_{16} - 2,4X_{17} - 7,93X_{18} - 3,63X_{20} - 0,02X_{22} + \\
 &+ 21,74X_{27} + 4,84X_{30} - 0,19X_{31} - 65,22.
 \end{aligned}$$

Литература

1. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ / Пер с англ. И.С. Енюкова, И.Д. Новикова. — М.: Мир, 1982. 488 с. (Afifi A. A., Azen S. P. Statistical analysis. A computer oriented approach. — 2nd ed. — New York — San Francisco — London: A Subsidiary of Harcourt Brace Jovanovich Pubs., 1979. 442 p.)
2. Truett J., Cornfield J., Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham // J. Chronic Diseases, 1967. Vol. 20. P. 511–524.
3. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. — М.: Медиа Сфера, 2002. 312 с.
4. Халафян А. А. Современные статистические методы медицинских исследований. — М.: Эдиториал УРСС, 2008. 320 с.
5. Захарова Т. В., Золоева М. В. Прогнозирование состояния пациентов // Обозрение прикладной и промышленной математики, 2007. Т. 14. Вып. 2. С. 298–299.
6. Драницына М. А., Захарова Т. В. Классификация состояний пациентов с целью прогнозирования результатов лечения // Обозрение прикладной и промышленной математики, 2009. Т. 16. Вып. 5. С. 840–841.
7. Драницына М. А., Захарова Т. В. Дискриминантный анализ для классификации и прогнозирования результатов лечения // Системы и средства информатики, 2013. Т. 23. № 2. С. 89–95.

Поступила в редакцию 10.01.13

PREDICTION AND CLASSIFICATION METHOD FOR CENSORED DATA

T. V. Zakharova and E. M. Abramova

Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation

Abstract: The classification method for noninsulin-dependent diabetes mellitus patients cohort is presented and the technique for identification of diabetes mellitus indicators is described. The basic medical data we dealt with turned out to be unfit for classification. The main obstacle for applying classical discrimination approaches was insufficiency and incompleteness of original data. For data processing, the authors suggest to select different sets of discriminant characteristics and to obtain classification functions for each set. The number of these sets depends on data incompleteness degree. The more data are omitted, the more different sets are needed. Each patient finally refers to the group, for which he gets the greater number of matches in classification. This multistep procedure reimburses small sample size and insufficiency and incompleteness of original data.

Keywords: hypothesis; censored data; discriminant variables; classification functions; forecasting

DOI: 10.14357/19922264130411

References

1. Afifi, A. A., and S. P. Azen. 1979. *Statistical analysis. A computer oriented approach*. 2nd ed. New York – San Francisco – London: A Subsidiary of Harcourt Brace Jovanovich Publ. 442 p.
2. Truett, J., J. Cornfield, and W. Kannel. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. *J. Chronic Diseases* 20:511–524.
3. Rebrova, O. Ju. 2002. *Statisticheskii analiz meditsinskikh dannykh. Primenenie paketa prikladnykh programm STATISTICA [Statistical analysis of medical data. Application software package STATISTICA]*. Moscow: Media Sfera. 312 p.
4. Halafjan, A. A. 2008. *Sovremennye statisticheskie metody meditsinskikh issledovaniy [Advanced statistical methods for medical research]*. Moscow: Editorial URSS. 320 p.
5. Zakharova, T. V., and M. V. Zoloeva. 2007. Prognozirovaniye sostoyaniya patsientov [The patients' conditions forecast]. *Obozrenie Prikladnoi i Promyshlennoy Matematiki [Review of Applied Industrial Mathematics]* 14:298–299.
6. Dranitsyna, M. A., and T. V. Zakharova. 2009. Klassifikatsiya sostoyaniy patsientov s tsel'yu prognozirovaniya rezul'tatov lecheniya [Classification of patients' conditions to forecast the outcomes of the treatment]. *Obozrenie Prikladnoy i Promyshlennoy Matematiki [Review of Applied Industrial Mathematics]* 16:840–841.
7. Dranitsyna, M. A., and T. V. Zakharova. 2013. Diskriminantnyy analiz dlya klassifikatsii i prognozirovaniya rezul'tatov lecheniya [Discriminant analysis for classification and forecasting outcomes of the treatment]. *Systemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):89–95.

Received January 10, 2013

Contributors

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, Moscow 119991, Russian Federation; isa@cs.msu.ru

Abramova Ekaterina M. (b. 1989) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation; houselake@gmail.com

CONCEPTUAL DECLARATIVE PROBLEM SPECIFICATION AND SOLVING IN DATA INTENSIVE DOMAINS*

L. Kalinichenko¹, S. Stupnikov¹, A. Vovchenko¹, and D. Kovalev¹

Abstract: Various notations aimed at defining the semantics of a computation in terms of the application domains have been experienced for conceptual modeling. For example, entity-relationship (ER) approach and UML (Unified Modeling Language) diagrams allow one to specify the semantics informally. Ontology languages based on description logic (DL) have been developed to formalize the semantics of data. However, it is now generally acknowledged that data semantics alone are insufficient and still representation of data analysis algorithms is necessary to specify data and behavior semantics in one paradigm. Moreover, the curse of ever increasing diversity of multistructured data models gave rise to a need for their unified, integrated abstraction to make specifications independent of real data in data intensive domains (DID). To overcome these disadvantages, a novel approach for applying a combination of the semantically different declarative rule-based languages (dialects) for interoperable conceptual specifications over various rule-based systems (RSs) relying on the logic program transformation technique recommended by the W3C (World Wide Web Consortium) Rule Interchange Format (RIF) has been investigated. Such approach is coherently combined with the specification facilities aimed at the semantic rule-based mediation intended for the heterogeneous data base integration. The infrastructure implementing the multidialect conceptual specifications by the interoperable RSs and mediating systems (MSs) is introduced. The proof-of-concept prototype of the infrastructure based on the SYNTHESIS MS and RIF standard is presented. The approach for multidialect conceptualization of a problem domain, rule delegation, rule-based programs and mediators interoperability is explained in detail and illustrated on a real nondeterministic polynomial time (NP) complete use-case in the finance domain. The research results are promising for the usability of the approach and of the infrastructure for conceptual, declarative, resource independent and reusable data analysis in various application domains.

Keywords: conceptual specification; W3C RIF; logic rule languages; SYNTHESIS; database integration; mediators; RIF-BLD; RIF-CASPD; multidialect infrastructure; rule delegation

DOI: 10.14357/19922264130412

1 Introduction

For decades, informatics is investigating proper facilities for conceptualization of the diverse application domains, descriptive and predictive data analysis to extract information from data, decision making, and generally, for the problem solving based on such insight. Many times a point was reached when the level of maturity of the conceptual modeling techniques could have been evaluated as satisfactory from the theoretical and practical points of view. But every time new technological developments and research results disproved such optimistic opinions. Now, in the epoch of the data intensive research² and system development, a lack of the proper declarative conceptual modeling facilities is felt again. For instance, the call for such facilities has been expressed in the ACM (Association for Computing Machinery) publica-

tion “Challenges and Opportunities with Big Data” [2]. In particular, the call considers important to examine a combination including

- facilities for support of variety of data (heterogeneity of data types, representation, and semantic interpretation) and semantic data integration; and
- declarative specifications required to meet the programmability and composition of complex analytical pipelines in an understandable form applying appropriate high-level languages to express the analytics.

This paper is related to such statements investigating a novel methodology and infrastructure supporting conceptually-driven problem specifications and solving in the DID. This research is motivated by aiming at the

*This research has been done under the support of the RFBR (project 11-07-00402-a) and the Program for Basic Research of the Presidium of the Russian Academy of Sciences.

¹Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

²Data Intensive Research (DIR) has emerged as a new paradigm for inferring information from data (known as the 4th paradigm for scientific domains [1]) emphasizing the increasing value of observational, experimental, and computer-generated data in virtually all domains. Similarly, “Big Data” is developing as a recognition of the increasing significance of massive multistructured data in almost each area of the human activity.

specification reusability in various domains over different sets of information resources (IRs), widely diverse data, and knowledge semantic integration capability as well as at the accumulation of reproducible data analysis and problem solving methods, related specifications, and experience in various DID.

One of the intended outcomes of the work is to expose current practically reachable limit of declarative conceptual specification construction applying the wealth of various available facilities of logic programming, knowledge representation, semantic Web, and heterogeneous database mediation in a coherent, cooperative way.

Conceptual specifications should be expressed in terms of an application domain independently of design and implementation concerns. Usually, the domain specification understandable and acceptable in the respective community includes a set of coherently tied ontologies together with a conceptual schema of the domain defined on the basis of such ontologies. Conceptual specifications of the applications in this domain are defined in terms of the conceptual schema to express abstract representations of information structures and algorithms of their analysis for the application problem solving.

This research reflects an intention to define and implement facilities for conceptually-driven problems specification and solving in DID to create the perspectives ensuring the following capabilities for expressing the specifications:

- declarative nature of the language facilities involved;
- an ability to provide complete and precise specification of the abstract structure and behavior of the domain entities, their consistency, relationship, and interaction;
- well-grounded diversity of semantics of the modeling facilities providing for proper expressiveness, compactness, and precision of the definition of behavior of the domain entities and problem solving algorithm specifications;
- arrangements for the probable extensions of the modeling facilities for conceptual specifications satisfying the changing technological and practical needs;
- provisions for specification independence from implementation languages and systems;
- provisions for specification independence from specific IRs combined with facilities for their semantic integration and interoperability coping with the increasing diversity of the techniques used for implementation of programs, databases, services, ontologies, and other resources;

- built-in methodologies for creation of unifying specification languages providing for construction of semantic preserving mappings of conceptual specifications into their implementations in specific languages and systems; and
- combination of ease of a human perception with a sufficient efficiency and expressiveness of the implementation for publication in communities intended for specification reuse.

The brief vision of the state-of-the-art in conceptual modeling (not claimed to be exhaustive) is as follows. Various notations aimed at defining the semantics of a computation in terms of the application problems to be solved have been experienced for conceptual modeling for many years. Structured and object-oriented specification notations for conceptual modeling in graphical (diagram) form dominate. Entity-relationship approach models the domain in terms of entities, attributes, and relationships, extended sometimes with cardinality constraints, aggregation, sets, temporality, taxonomic structures. Many parts of UML are the diagrams for specific purposes (e.g., class diagrams, activity diagrams, use cases, state charts, object interaction diagrams) used for the domain modeling and the design phase in software engineering. Such notations often allow one to specify the semantics informally, in terms of natural language; or the semantics is hard-coded using the tools supporting the notation [3]. Various logic specification languages, such as ontology languages based on DL, have been developed to describe the semantics of data. With certain limitations, the ontology languages allow one to define formal models of domain concepts, their taxonomic relations, and constraints over them. They facilitate automated reasoning (inference) over taxonomic structures. However, it is now generally acknowledged that data semantics alone are insufficient and representation of behavior is necessary to specify data and behavior semantics, algorithms for data analysis in one paradigm.

This is just what paper is focused on, emphasizing the importance of declarative nature of the specification facilities. In particular, the approach proposed is aimed at conceptual modeling related to DID applying rule-based declarative languages possessing different, complementary semantics and capabilities combined with the methods for heterogeneous data mediation and integration.

In the work presented, the issues of interoperability of declarative programs defined in the languages having different semantics combined with an ability to integrate various IRs (such as data and knowledge bases, software services, ontologies) for problem solving are investigated on the basis of two fundamental techniques:

- (1) constructing of the unifying extensible language providing for semantic preserving representation in it of various IR specification languages (e. g., such as data definition (DDL) and manipulation (DML) languages for databases); and
- (2) creation of the unified extensible family of rule-based languages (dialects) and a model of interoperability of the programs expressed in such dialects with different semantics.

The first technique is based on the experience obtained in course of the SYNTHESES language development [4, 5] accompanied by the methods and facilities for constructing its extensions. The kernel of the SYNTHESES language is based on the object-frame data model used together with the declarative rule-based facilities in the logic language similar to a stratified Datalog with functions and negation. The extensions of the SYNTHESES kernel are constructed in such a way that each extension together with the kernel is a result of semantic preserving mapping of some IR language (mainly, a data model language considered to be a combination of the respective DDL and DML) into the SYNTHESES [6, 7]. The canonical information model (CIM) is constructed as a union of the kernel with such extensions constructed for various IR languages. Canonical information model is used for development of mediators positioned between the users, conceptually formulating problems in terms of the mediators, and various distributed IRs (such as databases and services) needed for a specific application. A schema of a subject mediator for a class of problems includes the specification of the domain concepts defined by the respective ontologies.

Another, multidialect technique for rule-based programs interoperability applied in the current work is based on the RIF standard [8] of W3C. Rule interchange format introduces a unified family of rule-based languages (dialects) together with a methodology for constructing semantic preserving mappings of specific languages used in various RSs¹ into such dialects. Besides, a methodology for extension of the family with the definition of new dialects is also defined as a part of RIF. From the RIF point of view, an IR is a program developed in a specific language of some RS. For the logic dialects, the RIF family should include declarative, high-level languages reflecting in a systematic way the basic semantics of rule-based languages of various existing logic-based RSs. These RSs and their declarative rule-based languages form now a mature technology with decades of theoretical development, practical and

commercial use based on logic programming and non-monotonic reasoning.

The approach proposed here is built on the techniques briefly introduced above aiming at construction of declarative specifications containing definitions of the concepts and their relationships in the DID area (conceptual schema), conceptual definitions of problems, and their respective algorithms in declarative form simplifying their reuse. The *conceptually-driven* problem specifications of an application can be defined applying advanced declarative languages, modularization, and composition facilities. Thus, the problem specifications can be considered also as domain-driven, that is, they are defined in terms of the domain independently of IRs (databases, services, workflows, mediators) and implementation notations.

The paper presents the results obtained including the description of an approach and an infrastructure supporting:

- application domain conceptual specification and problem solving algorithms² definitions based on the combination of the heterogeneous database mediation technique and rule-based multidialect facilities;
- interoperability of distributed multidialect rule-based programs and mediators integrating heterogeneous databases; and
- rule delegation approach for the peer interactions in the multidialect environment.

The proof-of-concept prototype of the infrastructure based on the SYNTHESES environment and RIF standards has been implemented. The approach for multidialect conceptualization of a problem domain, rule delegation, rule-based programs, and mediators interoperability is explained in detail and illustrated on a real NP-complete use-case in the finance domain. For the conceptual definition of the use-case problem, Web Ontology Language (OWL) is used for the domain concepts definition and programs in two dialects RIF-BLD (Basic Logic Dialect) and RIF-CASPD (Core Answer Set Programming Dialect [9]) mapped into the SYNTHESES formula language and answer set programming (ASP) based DLV [10] language, respectively.

The paper³ is structured as follows. After introduction, the sections containing an overview of the approach and an infrastructure for distributed multidialect rule-based programs support are given. In section 4, the lessons learned during the use-case specification and implementation are summarized. A related work

¹Examples of RSs include SILK, OntoBroker, DLV, IBM ILOG, IRIS, etc. More examples can be found at <http://www.w3.org/2005/rules/wiki/Implementations>.

²In the approach proposed, the logic programming model of computation is assumed.

³The approach presented in this paper has been reported for the first time at the ADBIS'2013 conference and annotated in a short publication [9].

section, the future plans section that outlines the intentions for the future research, and the conclusion which summarizes the results end the paper. The detailed description of the use-case and its implementation is given in an appendix. Such detailed exposition of the results is justified by the novelty of the approach and a necessity to show and make readable all the steps required for the specification and implementation of the use-case applying the languages and the infrastructure proposed.

2 An Approach for Conceptual Problem Specification and Solving Based on the Techniques for Heterogeneous Database Mediation and Multidialect Program Interoperability

According to the above, the problems of integration of various IRs and interoperability of rule-based programs arising during application development are proposed to be resolved on the combination of two techniques:

- (1) construction of the unifying extensible language providing for representation in it of various IR languages preserving their semantics and for creation of mediators for integration of heterogeneous IRs; and
- (2) creation of the unified family of languages (dialects) for expressing the problem solving algorithms and the model of interoperability of programs represented in such dialects.

2.1 Heterogeneous Database Mediation

In this research, the first technique is based on the results of the SYNTHESIS language project [4] including the methods and means for the development of this language extensions and for its use for database integration.

The *Unifier of information models* [11] is the main facility created for the development of the SYNTHESIS language extensions and the mappings of specific data models into such extensions. Preserving semantics of a specific data model by its mapping into SYNTHESIS in the Unifier is based on the specification *refinement* principle (it is said that specification A refines specification B if it is possible to use A instead of B so that a user of B does not notice such substitution) developed by Abrial [12]. Thus, it is possible to construct provably correct extensions of the SYNTHESIS language and mappings of specific languages into them based on the

proof of refinement. Such methodology has been applied for mapping into SYNTHESIS of diverse database languages, service and process specification languages, ontological languages [5–7, 11]. Now, a possibility of mapping into SYNTHESIS of various multistructured data models typical for the Big Data epoch (such as, e. g., various NoSQL models [13], graph-based models [14], triplet data model [15], and array data model [16]) is studied.

The main characteristic properties of the SYNTHESIS language use as the kernel of the extensible language are as follows:

- (1) the merge of the kernel extensions constructed for diverse data models is treated as a single language known as the CIM; and
- (2) CIM is considered as a basis for the development of mediators positioned between the programs providing for problem solving in terms of mediators (the application domain terms) and specific distributed IRs (such as database, services) selected for the problem solving. The mediator specification for a class of problems includes definitions of the application domain concepts expressed by the respective declarative specifications, specifications of the classes of the domain entities, specifications of types of the instances of the classes mentioned and of their methods, and specification of functions.

A mediator in SYNTHESIS can be defined as a triple (T, S, M) where T is the mediator (target) schema; S is the IR schema; and M is the set of assertions (rules) relating elements of the target schema with elements of the resource schema. It is assumed that IR data model (that is, the respective DDL and DML) has been mapped into the canonical one; thus, both resource and target schemas are defined in the canonical data model. Both resource and canonical data models may allow for the expression of various constraints Σ . In this generality, a data integration setting is $(S, T, \Sigma_r, \Sigma_{rt}, \Sigma_t)$ in which S and Σ_r constitute the IR schema, T and Σ_t form the target schema, and the resource-to-target dependencies included into Σ_{rt} are the assertions of the schema mappings in the mediator. Σ_{rt} should be defined as a set of assertions of weakly-acyclic class [7, 17]. Note that Σ_r includes not only data dependencies corresponding to the IR database in the original data model, but also the dependencies generated in accordance with the extension of the kernel of the canonical data model determining an information-preserving mapping of the resource data model into the canonical one (some categories of such dependencies are shown in [6, 7]).

Specifications of mediators are expressed in CIM. Generally, a specification of M in the approach oriented on the definition of the mediator schema T in terms

of the application domain independently of schemas of IRs involved is based on the GLAV (Global and Local As View) [18, 19] schema mapping technique. In GLAV, two schema mapping techniques are combined: LAV (Local as View) according to which a schema of IR is defined as a view over the mediator schema and GAV (Global as View) according to which the mediator schema is defined as a global view over the IRs schemas. In LAV, the schemas of IR classes are considered as materialized views over the virtual classes of the mediator. Usually, GAV views are used for resolving various conflicts between the specifications of IRs and of the mediator. In such cases, the GAVs provide the rules for transformation of the results of a query over IRs to their representation in the mediator. The GLAV technique provides for stability of the mediator specification in time of the specific IRs change or the change of the actual IR presence (remove of an IR, adding of new resources, and so on). Also, the GLAV technique provides for scalability of mediators with respect to the number of IRs involved in them. It worth mentioning that the mediators, in their turn, can play the role of IRs towards the mediators of the higher level.

The following characterization of the GLAV setting is applied in the SYNTHESIS project. The sound GLAVs are assumed [18]. The GLAV mapping is constructed as a composition of two schema mappings — GAV mapping $M_{rg} = (S_r, S_g, \Sigma_{rg})$ (from resource schemas to GAV schema) and LAV mapping $M_{gt} = (S_g, T_{rt}, \Sigma_{gt})$ (from GAV schema to mediator (target) schema). GLAV is formed by a definition of the composition $M_{rg} \circ M_{gt}$ by means of the GAV and LAV dependencies (Σ_{rg} and Σ_{gt} , respectively). The GLAV mappings are formed showing how two parts of the schema (resource and target) are linked together by the Σ_{rt} dependencies.

Queries and programs over the mediator are expressed declaratively in the SYNTHESIS formula language (resembling Datalog) in terms of the mediator schema T [4]. An example of the mediator definition can be found in appendix (see subsections A3.1 and A3.2).

2.2 Diverse Rule-Based Programs Interoperability

Another fundamental principle used in this research is based on a multidialect technique intended for providing diverse rule-based programs interoperability and known as the W3C RIF standard [8]. The range of intended applications of the RIF standard (besides the Semantic Web) includes also development of the intelligent information systems as well as knowledge representation in various application domains. The family of dialects proposed in RIF has a common core (the *Core dialect*)

and an ensemble of the dialects extending the core and forming the directed acyclic graph. An arc of the graph is directed from the dialect being extended to the dialect extending it.

According to RIF, any specific RS can act in two independent roles — the role of *provider* and the role of *consumer*. The first one indicates that RS is capable to transform its native rule-based programs into the programs in the respective dialect; the second one indicates that RS is capable to accept a program defined in a proper dialect and transform it into the program in the native RS language. Thus, to include any rule-based language into the family of interoperable dialects, it is required to provide the respective RS with two semantic preserving transformers — from its own language into the proper dialect (the role of provider) and from the proper dialect into its own language (the role of consumer). In contrast to the development of the SYNTHESIS language extensions, each dialect extending other dialects may have its own, different of others, semantics. For example, for logic rule-based languages, each dialect can use its own semantics (e.g., the classic first-order semantics), well-founded semantics, stable model semantics, etc. [14]. Unlike the SYNTHESIS language, in which for development of the language extensions and respective semantic preserving language mappings, it is required to establish the refinement relation between the specifications, in RIF for the logic rule-based languages, it is required to establish the *entailment* relation between the specifications. The entailment relation between formulas f and g , $f \models g$ (f entails g) is valid if and only if g is true for all models where f is true.

In the RIF standard, only the basic, simple dialects have been fixed. Thus, RIF-BLD [20] corresponds to Horn logic with various syntactic and semantic extensions. RIF-PRD (Production Rules Dialect) [21] specifies a production rules dialect to enable the interchange of production rules. RIF-PRD captures the main aspects of various production rule systems. Production rules are defined using *ad hoc* computational mechanisms, which are not based on a logic. The condition language of RIF-PRD is defined as a common subset of RIF-BLD and RIF-PRD. RIF-Core (Core Dialect) is a subset of both RIF-BLD and RIF-PRD thus enabling limited rule exchange between logic rule dialects and production rules.

Significant contribution of RIF consists in inclusion into the standard of the Framework for Logic Dialects RIF-FLD [22]. RIF-FLD represents quite general logic language including the broad set of syntactic and semantic constructs frequently used in logic. On the basis of the framework, concrete dialects having specific semantics can be defined. Definitions of all dialects in the standard are strictly formalized. Rule interchange format defines the general conception of the unified

language construction as a family of dialects intended for provision of the interoperability of the numerous RSs. On the basis of the RIF-FLD framework, several dialects have been designed in the world. They need still to pass through the procedure of approval by W3C. Among these dialects are the following. RIF-CLPWD (Core Logic Programming Well-founded Dialect [23]) uses WFS [24] with the default negation and functions. RIF-CASPD [25] uses the semantics of ASP, known also as the stable model semantics [26]. Difference of semantics and application areas of WFS and ASP is a typical example showing different capabilities of the rule-based programs expressible in the languages with such semantics. The WFS and ASP language classes correspond to two philosophically different approaches in logic programming: WFS keeps the idea of defining a single (well-founded) model for a program while ASP provides for the computation of the set of preferred models that is more than query answering. Thus, the ASP-based systems are specifically oriented on solving the complex, combinatorial (NP-complete) problems. The ASP application areas include: diagnosis, information integration, constraint satisfaction, reasoning about actions (including planning), routing, and scheduling, health care, biomedicine and biology, text mining and classification, question answering, etc. In contrast to the ASP systems, the WFS-based systems incorporate three-valued logic are computationally complete and can be used as the universal logic programming facilities. The joint usage of programs with different semantics for problem solving is obviously reasonable.

The RIF standard has also defined the necessary concepts to ensure compatibility of RIF with resource description framework (RDF) and OWL [27], in spite of dissimilarity of their syntaxes and semantics. Rule interchange format uses its frame syntax to communicate with RDF/OWL. These frames are mapped onto RDF triples and a joint semantics is defined for the combination. The basic idea is that rules and OWL will view each other as “black boxes” with well-defined interfaces through exported predicates. The OWL-based ontologies will export some of their classes and properties, while rule-based knowledge bases will export some of the predicates that they define. Each type of the knowledge base will be able to refer to the predicates defined in the other knowledge bases and treat them extensionally as collections of facts. Ontologies are assumed to provide commonly shared conceptualization of a domain. There may be different application-specific rule programs for different applications in the domain.

In the use-case considered in the paper, the following languages for conceptual specifications are used: OWL 2, RIF-BLD and RIF-CASPD (the examples of specifications in these languages can be found in the appendix, section A2).

2.3 Basic Ideas of the Conceptual Problem Solving Approach Built on the Symbiosis of the Above

The idea of the conceptual problem solving approach considered in the paper consists in developing a modular infrastructure in which alongside with the mediators integrating data and services in the SYNTHESIS-like setting, the modules representing knowledge and declarative rule-based programs over various resources will be introduced. The infrastructure is based on the following principles:

- *the multidialect construction of the conceptually-driven problem specifications combining the SYNTHESIS and RIF approaches.* The specifications are represented as a functional composition of declarative modules, each based on its own language (dialect) with an appropriate semantics. Semantics of a conceptual definition in such setting becomes a multidialect one. To the SYNTHESIS language, one of the existing RIF dialects (or a new, specific dialect) should be put into correspondence (currently, OWL is used for the conceptual schema specification together with the RIF-BLD for the mediator rules to be interpreted in SYNTHESIS);
- *the specification modules are treated as peers.* Rule-based program modules are included into the specification alongside with the mediators that can be supported by various MSs. Interoperability of such combination of modules is based on P2P (peer-to-peer) and W3C RIF interoperability techniques;
- *combination of integration and interoperability.* The IR integration can be provided in the scope of individual mediator modules. Interoperability is provided between the modules supporting different dialects according to the RIF methodology; and
- *rule-based specifications on different levels of the infrastructure.* Rule-based, inference providing modules can be used for declarative programming over the mediators. In this case, the mediators support schema mapping for semantic integration of the IRs supported by various MSs.

In the paper, the ideas of construction of the infrastructure providing for conceptual declarative programming and heterogeneous IR mediation interoperating in the multidialect environment are presented.

The infrastructure generally outlined in Fig. 1 provides for implementation of conceptual specifications on the basis of the CIM-like kernel and its extensions used in the mediators combined with declarative languages of the rule-based programming systems possessing diverse semantics and capabilities. Such environment provides for

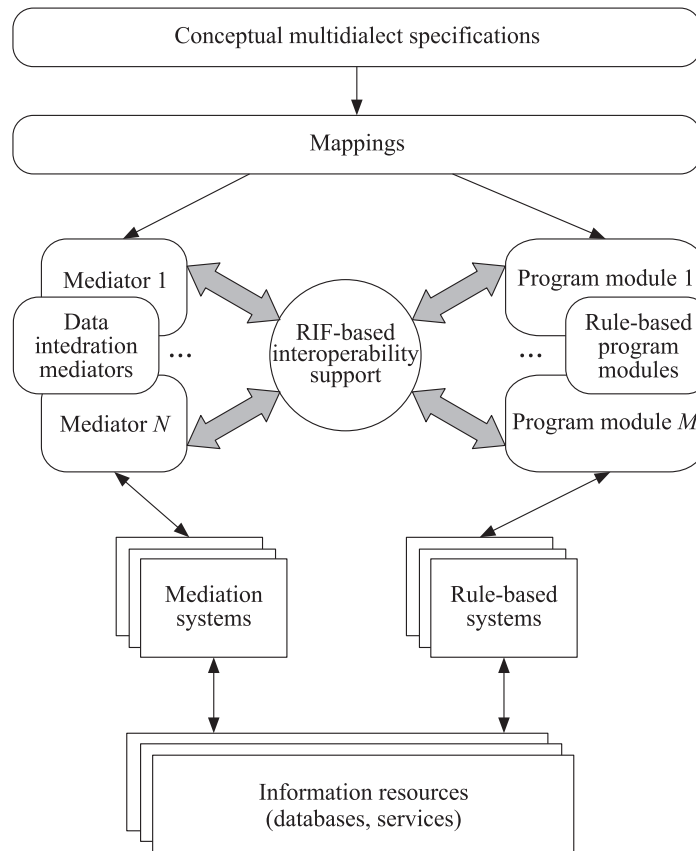


Figure 1 Infrastructure for conceptual problem solving

programming over heterogeneous resources (e. g., over databases and services) accessed through the mapping of specific resource schemas into the target schemas of the mediators. Multidialect program is implemented by means of transformation of conceptual specifications into modular, component-based P2P program represented on the languages of the mediation and rule-based systems. Interaction of components of such distributed program is carried out by means of the delegation technique (see subsection 3.3).

An approach for transformation of conceptual specifications into the distributed mediators and modules interpreted in RSs with different semantics interoperating in P2P style by the rule delegation in the multidialect W3C RIF-based environment is presented in section 3.

Practical value of the results obtained is demonstrated with a use-case in the finance domain. The important result of the research conducted consists also in accumulation of the experience of usage of the multidialect specification facilities for *conceptually-driven problem specifications* in specific domains and implementation of such specifications in the prototype that has been developed.

3 Infrastructure of the Multidialect Environment for Distributed Rule-based Programs and Mediators Interoperability

3.1 Conceptual Programming over the Conceptual Schema

The aim of the novel infrastructure proposed is a conceptual programming of problems in the RIF dialects and an implementation of conceptual specifications using declarative languages of the RSs and MSs. Conceptual multidialect logic programs specify the algorithms for problem solving in a subject domain. They are implemented using their transformation into RS or MS programs.

Conceptual specification of a problem (class of problems) is defined in the context of a subject domain and consists of a set of *RIF-documents* (document is a specification unit of RIF). Each document contains the groups of rules. Conceptual specification of a problem is an abstract program of a problem solving.

The subject domain conceptualization is performed using OWL 2 [28] ontologies containing entities of the domain and their relationships (Fig. 2, right-hand part). Thus, the *conceptual schema* of the domain can be formed. Conceptual specification is defined over conceptual schema: names of the entities (classes and attributes) and only they are used in the rules of the RIF-documents as extensional predicates.

Ontologies are imported into the RIF-documents specifying an import profile, for instance, *OWL Direct*. A profile defines a semantics of an OWL. The profiles are formally defined in [27].

Modular construction of the conceptual specification is based on the techniques of document import and link. Each RIF-document is defined using a dialect which is a *specialization* of the RIF Framework for Logic Dialects (RIF-FLD) [22]. Documents *import* other documents having the same semantics (the *Import* directive) or *link* documents defined using other dialects and having different semantics (remote module directive *Module*).

Documents refer to predicates from imported documents (using the name of an imported module *module:predicate*) as well as to predicates from remote modules using *remote terms f@r* [22]. Here, *f* means a term and *r* means a reference to a remote module.

Retrieving results of problem solving is performed by querying the virtual knowledge/database formed by combined infrastructure of mediators and rule-based systems.

Queries are formulated in terms of the conceptual schema and the intentional predicates of the conceptual specification of a problem. The result is a Boolean value (if the formula is closed) or a collection of tuples of values of free variables of the formula.

3.2 Mapping of the Peer Schemas into the Conceptual Specification

In the proposed infrastructure, (1) the logic programs implementing RIF-documents of the conceptual schema in specific RSs and (2) the mediators supporting virtual collections of facts as the result of heterogeneous databases integration are considered as peers.

A schema S_R of a peer R is a set of entities (classes or relations and their attributes) corresponding to extensional and intensional predicates of the resources.

The RS or the MS of each peer R should be a *conformant* D_R consumer where D_R is the respective RIF dialect (Fig. 2, left-hand part). Conformance is formally defined using formula entailment and language mappings [22].

The peer R is *relevant* to a RIF-document d of a conceptual specification of a problem (see Fig. 2, right-hand part) if:

- D_R is a subdialect of the document d dialect; and
- entities of the peer schema S_R (if they exist) are *ontologically relevant*¹ to entities of the conceptual schema the names of which are used in d for extensional predicates.

The schema of a relevant peer is mapped into the conceptual schema. The mapping establishes the correspondence of the conceptual entities referenced in the document d to their expressions in terms of entities of the schema S_R using logic rules of the D_R dialect. These schema mapping rules constitute separate RIF-document (Fig. 2, middle part). An example of mapping is given in appendix (see subsection A3.1).

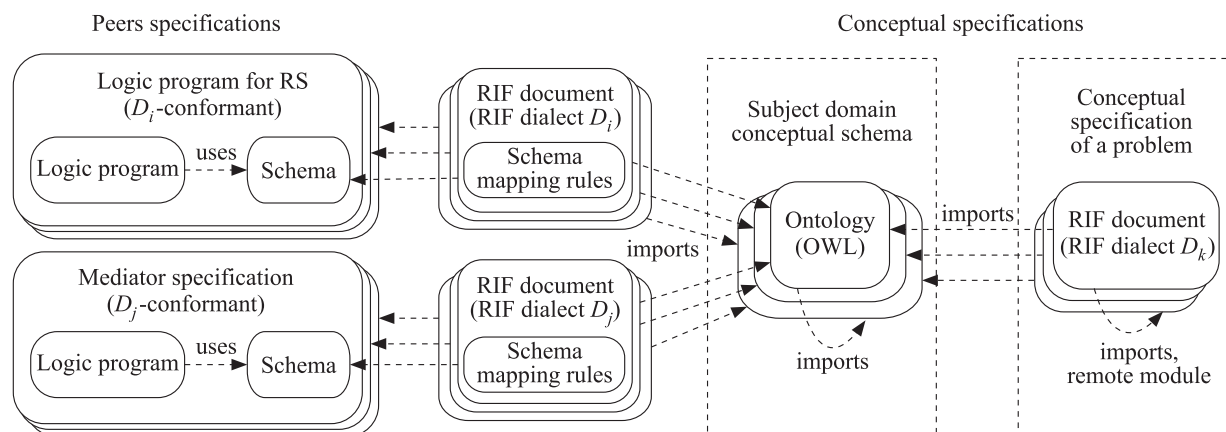


Figure 2 Conceptual and peer specifications

¹In this paper, the methods for ontological matching of the schemas [15] are not considered.

3.3 Implementation of the Conceptual Specification

Programs defined in the conceptual specification are implemented in P2P environment formed by relevant peers which are related to conceptual specification by mapping rules (see Fig. 2, middle part).

Peers communicate using a technique for distributed execution of the logic programs. The basic notion of the technique is *delegation* — transferring facts and rules from one peer to another. A peer is installed on a *node* of the multidialect infrastructure. A node is a combination of a wrapper, an RS or an MS, and a peer (Fig. 3). A wrapper transforms programs and facts from the specific RIF dialect into the language of the RS or MS and *vice versa*. A wrapper also implements the delegation mechanism. A definition of the delegation is given in the latter part of this subsection. Transferring facts and rules among peers is performed in RIF dialects. The wrappers implement an interface *RIFNodeWrapper* (see Fig. 3).

A special component (*Supervisor*) of the proposed architecture stores shared information of the environment, i. e., conceptual specifications related to the domain and to the problem, a list of the relevant resources, RIF-documents combining logic rules for the conceptual specification, and a resource schema mapping.

Implementation of the conceptual specification includes the following steps:

- rewriting the conceptual documents into the RIF-programs of the peers performed by the *Supervisor*. A rewriting includes also (1) replacing the document identifiers (used to mark predicates) by peer identifiers; and (2) adding schema mapping rules to programs (see Fig. 2, middle part);
- a transfer of the rewritten programs to nodes containing peers relevant to the respective conceptual documents. The transfer is performed by the *Supervisor* by calling the method *loadRules* of the respective node wrappers;

- a transformation of the RIF-programs into the concrete RS or MS languages. The transformation is performed by the *NodeWrapper* or by the RS or MS itself (if the RS or MS supports the respective RIF dialect); and
- an execution of the produced programs in P2P environment.

During the process of rewriting the conceptual schema into the resource programs, the relationships between RIF-documents of the conceptual schema defined by remote or imported terms are replaced by relationships between peers also defined by remote or imported terms. To implement remote and imported terms, a *rule delegation* mechanism is used — transferring facts and rules from one peer to another. The details of rule delegation approach implemented in the current prototype of the multidialect infrastructure are provided below.

In general case, the programs transferred to some peer may include *nonlocal* rules. These rules contain remote or imported terms. On the contrary, *local* rules do not contain such terms. For simplicity, only remote terms are mentioned in the latter part of this subsection. To make possible an execution of a program on a node, the program should be *normalized*, i. e., transformed into an equivalent program including only local rules and *delegation rules*. There are two kinds of delegation rules for a peer *n*:

- (1) *fact delegation rule* like $p@m(X) : -q@n(X)$ where *p* and *q* are the predicate names; *X* is the variable list; and *m* is the peer different from *n*. The rule means that all the facts turning *q* into *true* have to be transferred to the peer *m* as facts turning *p* into *true*. A fact is a term $p(v_1, \dots, v_n)$ such that

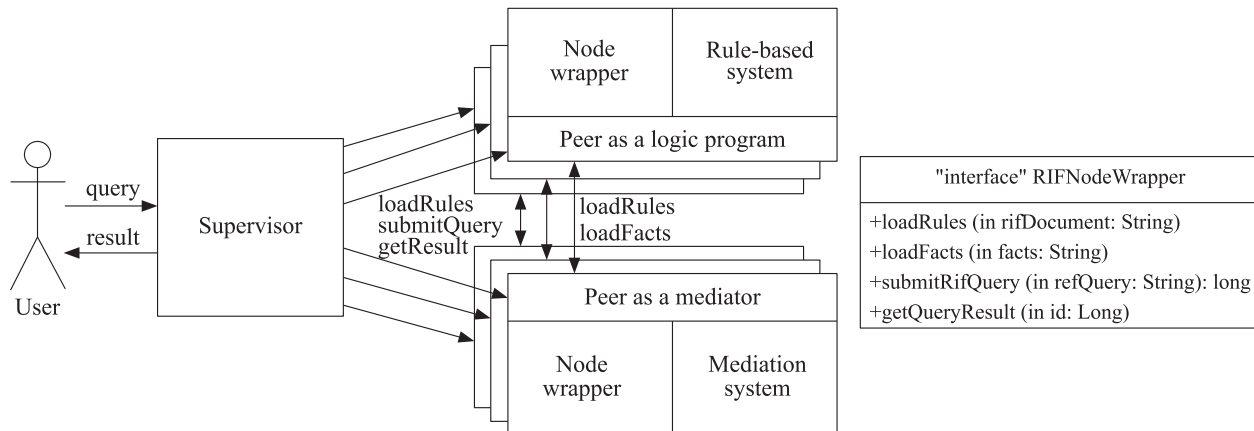


Figure 3 Peer-to-peer multidialect architecture

$p(v_1, \dots, v_n) = true$ where v_1, \dots, v_n are the values of variables $X = x_1, \dots, x_n$; and

- (2) *delegation rule* like $q@n(X) :- p@m(X)$. The rule has to be transferred to the peer m where it becomes a fact delegation rule.

The procedure of *normalization* (Algorithm 1) of a program pr on a peer n is as follows. For each occurrence of a remote term in a rule r , a fact delegation rule is added to pr aimed at importing facts from another peer to the peer n or at exporting facts from n to another peer. The rule includes a new local predicate created for storing imported or exported facts. Each occurrence of a remote term in a rule r is replaced by an occurrence of the respective local predicate. In that way, the rule r is transformed into a local rule.

Algorithm 1. *Normalization*(pr, n)

Input: pr is a program — a set of rules; n is a peer name

Output: normalized program pr

for each rule $r \in pr$ like $head(r) :- body(r)$. **do**

if the $head(r)$ of r is a remote term $p@m(X)$ **then**

// remove r from pr

$pr \leftarrow pr \setminus \{r\}$

// add new fact delegation rule,

// here, p_m is a new local predicate

$pr \leftarrow pr \cup \{p@m(X) :- p_m(X).\}$

// add new local rule

$pr \leftarrow pr \cup \{p_m(X) :- body(r)\}$

elseif the $head(r)$ is a local predicate **and**

the $body(r)$ contains an occurrence

of a remote term $q@l(Y)$

then

for each remote term $q@l(Y)$

contained in the $body(r)$

do

// add new fact delegation rule

// here, q_l is a new local predicate

$pr \leftarrow pr \cup \{q_l(Y) :- q@l(Y).\}$

// replace the occurrence of the remote term in r

// by a local term $q_l(Y)$

$r \leftarrow head(r) :- body(r)[q@l(Y) \rightarrow q_l(Y)].$

end for

endif

end for

return pr

The algorithm of execution of a program on a node containing a peer n looks as follows (Algorithm 2). The program stored on the peer is normalized. Delegation rules produced during normalization are transferred to the respective peers. After that, the algorithm waits for the facts from all the peers to which the delegation rules were transferred. During waiting, some delegation rules can be obtained from other peers (no mutual recursion between programs sent to different peers is presumed).

When all facts are got, they become a part of the peer. Then, local RIF-rules of the peer are transformed into the RS or MS rule language and executed in the RS or MS. The last step consists in transferring the required facts to other peers in accordance with the fact delegation rules.

Algorithm 2. *Execute*(n)

Input: n is a peer name;

$prog(n)$ is a program stored at n at the moment

(additional rules can be loaded into $prog(n)$)

during execution of *Execute*)

Output: $result$ — a set of tuples which are the result of the executing program

Local variables

prn — a variable to store a normalized program

pre — a variable to store local normalized rules

$pre1$ — a variable to store a program transformed into the RS or MS rule language

$prn \leftarrow Normalization(prog(n), n)$

for each fact delegation rule $r \in prn$

like $q_l(Y) :- q@l(Y)$

do

// transfer rule $q_l@n(Y) :- q(Y)$ to peer l

$l.loadRule("q_l@n(Y) :- q(Y)")$

$pre \leftarrow pre \setminus \{r\}$

end for

wait until for all fact delegation rules, $r \in prn$

like $q_l(Y) :- q@l(Y)$

a respective peer l has transferred facts

like $q(v_1, \dots, v_n)$ to peer n as facts

like $q_l(v_1, \dots, v_n)$

for each fact delegation rule $r \in prn$

like $q_l(Y) :- q@l(Y)$

do

load facts from predicate q_l into the peer n

end for

// put all normalized local rules into pre

$pre \leftarrow \{r | r \in prn \vee r \text{ is local}\}$

transform the program pre into the resource

rule language program $pre1$

$result \leftarrow execute$ the program $pre1$ on the resource

(mediator or RS) connected to n

for each fact delegation rule $r \in prog(n)$

like $p@m(X) :- q(X)$

do

// transfer facts from local predicate q into remote

// predicate p at the peer m

$m.loadFacts(\{p(x) | x \in q(x)\})$

end for

return $result$

A rule delegation mechanism described in this section resembles one proposed in WebdamLog [30].

4 Lessons Learned During the Use Case Implementation

To provide a proof of the multidialect infrastructure concept, a use case in the financial domain has been implemented. The problem to be solved in the use case is called *investment portfolio diversification problem*. The detailed description of the use case is included in the appendix. In this section, the important lessons learned during the implementation of the use case are summarized.

During the implementation of the *investment portfolio* use case in the multidialect infrastructure, the conceptual specifications were defined: a conceptual schema of the financial domain (using OWL 2) and conceptual specifications of the problem (using RIF dialects BLD and CASPD). Conceptual specification was defined over conceptual schema: names of the entities (classes and attributes) of conceptual schema and only they were used in the rules of the conceptual specification as extensional predicates. Conceptual specifications were implemented using a multidialect infrastructure which includes a mediation system integrating the financial data and an ASP [26] RS.

Application specification in the multidialect infrastructure is started with the informal definition of a problem (see appendix, section A1). Then, a conceptual schema of a subject domain is defined using OWL 2 (see section A2). Due to the fact that OWL 2 does not provide facilities to specify functions, it is necessary to model the functions related to the collections of facts of the domain as OWL classes (e. g., *correlation*):

```
Class(correlation)
  DataProperty(series1 domain(correlation)
    range(DatedValue))
  DataProperty(series2 domain(correlation)
    range(DatedValue))
  DataProperty(corr domain(correlation)
    range(xsd:double))
```

Input and output parameters of the functions are modeled as data or object properties of the respective classes (for instance, *series1*, *series2*, and *corr*). The types of the parameters are modeled as range of the respective properties.

Lesson 1. Functions related to the collections of facts in conceptual schema of the domain have to be modeled as OWL classes.

The classes of the conceptual schema are used as predicates in rules which constitute documents of the conceptual specification of the problem. Logic rules over taxonomic structures of the schema may require existential quantifier in their heads (see section A2):

```
Exists ?ts( And(?ts#gex:tickers
?ts[symbol->?symbol] ) :-
  And(?t#srt:stockRates ?t[ticker->?symbol]
  ?t[isInTop500->?inTop500]
  External(pred:boolean-equal(?inTop500
  true))
)
```

The rule states that for each object *?t* of *stockRates* class there exists an object *?ts* of a class *tickers*. A predicate *e#s* denotes that the element *e* belongs to the set *s*.

Built-in predicates (like Boolean equality *pred:boolean-equal*) and functions are used as *external terms* of RIF. The symbol *External* indicates that an atomic formula or a function term is defined externally (neither in the current RIF-document nor in the remote RIF-document). Built-in predicates are defined in a special part of the RIF standard.

Lesson 2. Conceptual specification of a problem using RIF logic rules over taxonomic structures may require existential quantifier in the rule heads. The RIF-BLD dialect intended for serving as the basic logic dialect does not provide such facility. Respective extension of RIF-BLD is suggested.

Lesson 3. Conceptual specification of a problem often requires usage of built-in predicates and functions as external terms.

During the application of the RIF-CASPD dialect [31] for conceptual specification, it was found out that the ASP programs require a specific facility — so-called *weak constraints* [10]. Such constraints should be satisfied if it is possible, but their violation does not invalidate the models. For instance, a weak constraint used in the specification of the portfolio problem looks as follows (see section A2):

```
Forall ?X( :~ prt:nonPortfolio(?X))
```

The rule states that only such stable models that minimize the truth set of the predicate *nonPortfolio* are considered as a result of the logic program (that includes the mentioned rule). The difference of the rule with a strong constraint like

```
Forall ?X( :- prt:nonPortfolio(?X))
```

consists in that in a program with the strong constraint the predicate *nonPortfolio(?X)* has to be turned to *false* for all *?X*. Weak constraint requires the set of all possible values of *?X* that turns *nonPortfolio(?X)* into *true* to be of minimum size.

Lesson 4. Conceptual specification using ASP-like dialects requires a specific kind of logic rules — *weak constraint*. The RIF-CASPD dialect intended for serving as the dialect supporting ASP does not provide such

facility. Respective extension of RIF-CASPD is suggested.

To implement the conceptual schema, a mediator intended to provide the collection of facts implied by the schema was created. The mediator integrates the respective financial data. The schema of the mediator (see subsection A3.2) was created by mapping the conceptual schema into the CIM (the SYNTHESIS language [32]). The classes of the conceptual schema were mapped into types and/or classes of the mediator schema, including their attributes. It appeared that some of the attributes (e. g., *isInTop500*) of the conceptual schema taking into account implementation reasons better to interpret by the methods of the mediator schema. It appeared also that the definition of *isInTop500* as a method on the mediator layer made it simpler to establish semantic mappings (GLAV views) between the mediator and the resources.

Another peculiarity of the conceptual schema implementation worth mentioning here is that the classes of the conceptual schema (namely, *correlation*, mentioned earlier in this section) can be mapped into a function of the mediator schema:

```
{ correlation; in: function;
  params: {+s1/{set; type_of_element:
    DatedValue;},
    +s2/{set; type_of_element:
    DatedValue;}, -corr/real };
};
```

Lesson 5. The attributes of the conceptual schema can be mapped into attributes or functions of the schemas of the mediators implementing the conceptual specifications. The classes of the conceptual schema can be mapped into functions of the schemas of the mediators.

The mapping of conceptual schema into the mediator schema is formally specified using schema mapping rules constituting a separate RIF-document as shown in Fig. 3, middle part. These rules are similar to GAV views which represent classes of the conceptual schema as views over the mediator schema. This allows one to consider the schema mapping rules just as a consistent part of the logic program that have to be executed in a peer (e. g., in a mediator).

The document containing schema mapping rules refers to the conceptual schema and to the mediator schema. Both schemas are identified by URIs:

```
Import(<http://synthesis.ipi.ac.ru/
  optimalSecurityPortfolio>
  <http://www.w3.org/ns/entailment/
  OWL-Direct>)
Prefix(srt <http://synthesis.ipi.ac.ru/
  optimalSecurityPortfolio#>)
```

```
Prefix(fs <http://synthesis.ipi.ac.ru/
  FinanceServices#>)
```

Entities of the mediator schema are referred in the schema mapping rules using external terms of the RIF, for instance, `External(?t#fs:stockRates)` (see subsection A3.1). This means that the mediator schema is neither OWL-ontology or RDF-document nor RIF-document and possesses its own semantics.

Mapping of the *correlation* class of the conceptual schema into the correlation function of the mediator schema is defined using a rule with a frame predicate in the head and functional predicate in the body (see subsection A3.1):

```
And(?c#correlation
  ?c[?series1->series1 ?series2->series2
  ?corr->corr]) :-
External(pred:numeric-equal(
  fs:correlation(?series1 ?series2) ?corr))
```

Lesson 6. Entities of the peer (mediator) schemas have to be referred in schema mapping rules using external terms of the RIF.

Lesson 7. Logic programs defined in the native languages of RS or MS look more elegant and concise than the respective programs defined using RIF-dialects. The present authors consider this as a price paid for providing the interoperability of multidialect programs with different semantics. More elegant notation to apply RIF for conceptual specifications should be investigated.

It appeared that the mediation system used does not support all the features of the RIF-BLD (terms with named arguments, conjunctions of atoms in the head of a rule, and some others). DLV system also does not support all the features of the CASPD dialect (terms with named arguments, conjunctions of atoms in the head of a rule, frame terms, subclass terms, membership terms, and some others). These systems can be the consumers only for the subsets of the respective dialects.

Lesson 8. According to RIF, one could have expected that the rule-based programs represented in a specific RIF-dialect should be independent on the RS (or MS) consumer languages at least in the class of RSs (or MSs) conforming to this dialect. For instance, any consumer in the class of the ASP systems should be considered as conformant to the CASPD dialect. Noncompliance with such condition makes the conceptual modeling in the RIF dialects difficult: it is required to choose a specific RS (or MS) in advance and to adjust the dialect to be used appropriately (as it was shown in Lessons 2 and 4 or in the observation above showing that it is unlikely that an RS or MS would support all the features of some standard dialect). Practically, it means that at least the RIF-dialect standards should include instructions for the dialect adjustment making possible to produce

appropriate subdialects or superdialects at the early stage of the conceptual design.

5 Related Work

Among various future research issues predicted in [33] to obtain reasonable solutions in the next decades, the conceptual modeling has been positioned as the basis for interoperability and for shareable information services. The solution proposed in the present paper is tightly related to the issues listed in [33].

Ambiguity and incompleteness of structured and object-oriented notations for conceptual modeling in graphical form causes the unceasing process of generating various contributions suggesting formal meaning to the diverse graphical constructions, enhancing their expressiveness, e. g., as in several recent papers [34–36]. The present paper suggests more radical formalization decisions.

Diverse ontology centered conceptual modeling tools have been developed. Just a couple of them to mention.

KAON2¹ is an infrastructure for managing OWL-DL, SWRL (Semantic Web Rule Language), and F-Logic ontologies. Its main feature is its own inference engine. KAON2 supports the SHIQ(D) subset of OWL-DL, DL-safe subset of the SWRL, making reasoning decidable [32], as well as the function-free subset of F-Logic. This approach is related to the multidialect architecture considered here due to a provision of using several different languages treated as ontological notations. ICOM 3.0 tool [37] for conceptual modeling is aimed at the support of the design of multiple extended ER or UML diagrams with inter- and intramodel constraints. Reasoning support via DL to help validate the models is provided. The system is helpful for validating integrated models (support for the integration of models organized in several ontologies is provided). But it is difficult to improve such notations as ER or UML completely; they allow only to specify the semantics informally, in terms of natural language; or the semantics are hard-coded using the tools supporting the notation.

Rule-based modeling is applied in various DIS. For example, in biology, rule-based modeling has increasingly attracted attention due to enabling a concise and compact description of biochemical systems. Proteins, individual cells, and cell populations denote different levels of an organizational hierarchy, each of them with its own dynamics. Multilevel modeling is concerned with describing a system at these different levels and relating their dynamics. The approach proposed facilitates developing and maintaining multilevel models that, for

instance, interrelate intracellular and intercellular dynamics [38].

The idea used in the infrastructure prototype for ontology-based access (using OWL 2) to the mediators integrating heterogeneous databases resembles the idea of the MASTRO-I system [39] in which QuOnto engine supporting inference in the DL-Lite ontologies and conjunctive query answering is positioned above the federated database. The latter is the result of relational databases integration.

The rule exchange using RIF-PRD is discussed in [40, 41]. In such case, the production rule systems share the same operational semantics opposed to the present approach studying the problem of rule exchange between systems with different semantics.

Several papers on the use of database query languages for specifying declarative distributed programs and managing data in distributed environment have been published [30, 42–44]. In contrast to multidialect approach, a single declarative language is used in each of the proposed systems. Usually, it is a conventional Datalog extended with the notion of localization and possibly other nondatalog constructs [42]. In the multidialect approach, localization is specified with the RIF remote and imported terms.

Conceptual notion of *delegation* applied in the present approach is similar to the notion of delegation in Webdamlog defined as the “possibility of installing a rule at another peer. In its simplest form, delegation is essentially a remote materialized view. In its general form, it allows peers to exchange rules, i. e., knowledge beyond simple facts, and thereby provides the means for a peer to delegate work to other peers” [16]. Actually, the current implementation supports delegation as a remote materialized view. Extending the approach for delegation of knowledge is a future work. The idea of program normalization is similar to the rule localization rewriting the step described in [42].

6 Future Plans

A number of related issues remain open. Some of them that are planned to be investigated are listed below.

Conceptual specification of processes. How to define compositions of algorithmic modules in a process structure (like Petri nets, UML activity diagrams, the Business Process Modeling) can be found in [45] where process modeling semantics, service discovery, mediation, and composition are considered. Such notations as WSDL-S, SAWSDL, OWL-S, and WSMO (Web Service Modeling Ontology) can be applied for that. For example, OWL-S is a semantic web service description language that borrows ontological operators from

¹<http://kaon2.semanticweb.org/>.

OWL and provides a service upper ontology to describe web services and service processes in a standard manner. The WSMO and the combination of UML and Object Constraints Language (OCL) are viewed as promising alternatives to OWL-S. Notation WSMO, like OWL, is also based on the first-order logic but employs an object-oriented style of modeling based on frames. Unlike OWL-S, WSMO includes service and ontology specification formalisms with well-defined formal semantics, whereas OWL-S relies on OWL for ontology specifications and on languages such as SWRL and KIF (Knowledge Interchange Format) for formal semantics. The present authors plan to investigate an approach for reusable process specifications in conceptual terms compliant with the infrastructure proposed.

Multiple layers of specifications. Metamodeling as a kind of modeling across multiple layers uses an “instance-of” relationships between layers. UML/MOF/MDA by OMG is an example of a strict metamodeling hierarchy in which metamodels introduce higher levels of abstractions for specification and language concepts. Similar hierarchy can be introduced also in ontologies (e.g., upper ontologies above the domain ontologies in one and the same language). The models can be used in a descriptive or prescriptive manner [46]. In this paper, the metamodeling issues are not considered though in future plans, the authors are closer to the Ontology-based software Development Environment (ODE) approach of [47].

Hybrid specifications. Conceptual modeling should likely explicitly involve hybrid aspects of information (i.e., the coexistence of formal reasoning and “informal” human interactions). Correspondingly, “hybrid ontologies” will need to be modeled and deployed within their supporting social implementation environments [48].

Modeling hypotheses in the DIS experiments. Evaluating the conceptual model and infrastructure proposed in the paper, the following far-reaching DIS related ideas are planned to be taken into account [49]. As the DIS research evolves, references to previous results and specifications are needed as a source of provenance data. *In-silico*¹ experiments must be supported by a hypotheses model that describes the elements involved in a scientific exploration and supports hypotheses assessment. Adopting a conceptual specification perspective to represent hypotheses would allow high-level references to experiments and provides support for hypotheses evolution. Such hypotheses model would support scientists in describing, running simulations and interpreting their results [49]. In conjunction with an axiom set specified as rules that model known facts over the same universe

and experimental data, the knowledge base may contradict or validate some of the sentences in the hypotheses. In the case of contradictions, the rules that caused the problems must be identified and eliminated from the theory formed by the hypotheses [50]. The conceptual specification model proposed in the present paper looks suitable to such intention providing for a possibility of hypotheses representation as a set of first-order predicate calculus sentences applying the multiplicity of the dialects required.

7 Concluding Remarks

The approach presented is an attempt of introducing the multidialect interoperable conceptual programming over various semantically different rule-based programming systems relying on the logic program transformation technique recommended by W3C RIF. It is also shown how to coherently combine such approach with the heterogeneous data bases integration applying the semantic mediation. Thus, the data independence of conceptual specifications is provided. The results obtained so far are quite encouraging for future work aimed at reaching the conceptual specifications reusability in various applications over different sets of data, as well as for sharing and accumulation of reproducible data analysis and problem solving methods and experience in various DID. At the same time, the RIF-dialect standards are suggested to be made more flexible including into them the instructions for the dialect adjustment. This makes possible to produce appropriate subdialects or superdialects at the early stage of the conceptual design (see section 4, Lesson 8).

APPENDIX A

The Use-case for the Multidialect Infrastructure

A1 Investment Portfolio Diversification Problem

The capabilities of the multidialect architecture are illustrated with the solution of the *investment portfolio diversification problem* [51]. The portfolio is a collection of securities (such as equities or bonds) of companies, and its size is the number of securities in the portfolio. The task is to build a diversified portfolio of maximum size. Diversification means that the prices of the securities in portfolio should be almost independent of each other. If the price of one security falls, it will not significantly affect the prices of other. Thus, the risk of a portfolio sharp decrease is significantly reduced.

¹*In silico* is an expression used to mean “performed on computer or via computer simulation.” The phrase was coined in 1989 as an analogy to the Latin phrases *in vivo*, *in vitro*, and *in situ*, which are commonly used in biology and refer to experiments done in living organisms, outside of living organisms, and where they are found in nature, respectively (refer: Differences between *in vitro*, *in vivo*, and *in silico* studies. 2013. Available at: http://mpkb.org/home/patients/assessing_literature/in_vitro_studies (accessed December 2, 2013)).

The input data for the problem is a set of securities and respective time series of indicators of the security price for each security. Time series for each security is a set of pairs (d, v) where d is a date and v is an indicator of the security price (for instance, closing price).

For a pair of time series $X = \{(d_1, x_1), \dots, (d_n, x_n)\}$ and $Y = \{(d_1, y_1), \dots, (d_n, y_n)\}$, their *correlation* (a measure of similarity) can be calculated. The Pearson correlation¹ r_{XY} was used between X and Y

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where \bar{x} and \bar{y} are the means of x_1, \dots, x_n and y_1, \dots, y_n .

For the diversified portfolio, the securities which time series are noncorrelated should be used. Noncorrelation of the time series means that their correlation is less than some predetermined price correlation value. This predetermined value is a parameter of the portfolio problem. The lower is the correlation — the lower is the risk measure of a sharp decrease of the portfolio value.

So, the predetermined correlation serves as the maximum risk measure. In practice, specified correlation may differ for various types of securities.

The output data for the problem is a subset of securities of the maximum size, for which the pairwise correlation will be less than the predetermined one.

The problem is divided into the following tasks:

- (1) computation of the security pairwise correlations (for specified dates); and
- (2) calculation of the maximum satisfying subset of securities.

To solve the first task, the financial services *Google Finance*² and *Yahoo! Finance*³ are considered, both of which provide

current and historical information about various securities: stock prices, currencies, bonds, stock indexes, etc. In particular, the information provided includes various indicators of the security price for all trading days of last decades. The mediation system is used to solve the problem of resource integration [5].

The second task can be formulated as follows. Let G be a graph where the vertices are the securities, and an edge between two securities exists if absolute value of their correlation is less than a specified number. So, any two securities connected by an edge are considered as noncorrelated. In that case, the problem of finding the portfolio of the maximum size is exactly the problem of finding a maximum clique in an undirected graph. Indeed, a clique in a graph is a subset of its vertices such that every two vertices in the subset are connected by an edge. All the vertices (securities) in a clique are connected by edges (noncorrelated) and any clique is a candidate for a portfolio. A maximal clique is a maximal portfolio.

Finding a maximum clique in an undirected graph is a well-known NP-complete problem⁴.

DLV [10] is one of the ASP logic programming systems well-suited for solving such problems [31]. Answer set programming is based on the stable model (answer set) semantics of logic programming [26]. In ASP, the search problems are reduced to computing stable models, and *answer set solvers* — programs for generating stable models — are used to perform search. With regard to solving the maximum clique problem, ASP methods are applicable nowadays for graphs with thousands of vertices.

A2 Conceptual Specification of the Application Domain and the Problem

Conceptual schema (ontology) of the application domain of historical prices of securities is written in the simplified⁵ OWL functional syntax [28]:

```
Ontology(<http://synthesis.ipi.ac.ru/optimalSecurityPortfolio>
  Class(stockRates)
    DataProperty(ticker domain(stockRates) range(xsd:string))
    DataExactCardinality(1 ticker stockRates)
    ObjectProperty(rates domain(stockRates) range(DatedValue))
    DataProperty(isInTop500 domain(stockRates) range(xsd:boolean))
    DataExactCardinality(1 isInTop500 stockRates)

  Class(DatedValue)
    DataProperty(value domain(DatedValue) range(value xsd:double))
    DataExactCardinality(1 value DatedValue)
    DataProperty(date domain(DatedValue) range(xsd:date))
    DataExactCardinality(1 date DatedValue)
```

¹Refer: De Smith, M.J. 2013. *Statistical analysis handbook*. Available at: http://www.statsref.com/HTML/pearson_product_moment_correla.html (accessed December 2, 2013).

²<https://www.google.com/finance>

³<http://finance.yahoo.com/>

⁴Refer: Cormen, T. H., T. Leiserson, R. L. Rivest, and C. Stein. 2009. NP-complete problems. *Introduction to algorithms*. 3rd ed. Cambridge: The MIT Press. 1292 p.

⁵To save space, “Declaration” keyword is omitted; property, domain and range declarations are combined.


```

Class(correlation)
  DataProperty(series1 domain(correlation) range(DatedValue))
  DataProperty(series2 domain(correlation) range(DatedValue))
  DataProperty(corr domain(correlation) range(xsd:double))
  DataExactCardinality(1 corr correlation)
)

```

The `stockRates` class is used to denote securities, which are characterized by:

- identifier (attribute `ticker`);
- time series of historical prices (attribute `rates`);
- belonging to S&P 500 list of companies (attribute `isInTop500`).

The S&P 500 is a stock market index maintained by the Standard & Poor's, comprising 500 large-cap American companies.

The *correlation* class is the correlation of time series pairs. For each class instance, the value of *corr* attribute equals to the correlation of its attributes *series1* and *series2* (time series).

The conceptual specification of the problem includes two RIF documents that correspond to the specified tasks. The first of the documents (name *gex* is the local prefix of the document) contains a program that calculates the correlation graph of securities (predicate *noncorrelated*) based on the prices in a given period of time. For each pair of securities, a correlation of their time series (defined in section A1) of historical prices is calculated that is why the graph is called “correlation graph.” But as the candidates for the elements of the portfolio, only noncorrelated securities (which correlation is lower than a predefined value) are considered. That is why the predicate is called “*noncorrelated*.” The document is defined in the RIF-BLD¹ dialect [20]:

```

Document( Dialect(RIF-BLD)
  Import(<http://synthesis.ipi.ac.ru/optimalSecurityPortfolio>
    <http://www.w3.org/ns/entailment/OWL-Direct>)
  Prefix(srt <http://synthesis.ipi.ac.ru/optimalSecurityPortfolio#>)
  Prefix(gex <http://synthesis.ipi.ac.ru/graphExtraction#>)
  Group(
    Forall ?t ?symbol ?ticker ?inTop500(
      Exists ?ts( And(?ts#gex:tickers ?ts[symbol -> ?symbol]) ):-
        And(?t#srt:stockRates ?t[ticker->?symbol]
          ?t[isInTop500->?inTop500]
          External(pred:boolean-equal(?inTop500 true))
        )
    )
  )

  Forall ?m ?n ?c ?ticker1 ?ticker2 ?start ?end ?rates1 ?rates2
    ?top1 ?top2 ?dv1 ?dv2 ?date1 ?date2 ?series1 ?series2 ?corr (
    Exists ?e (
      And(?e#gex:noncorrelated ?e[start->?ticker1 end->?ticker2])):-
      ?m#srt:stockRates ?n#srt:stockRates
      ?m[ticker->?ticker1 rates->?rates1 isInTop500->?top1]
      ?n[ticker->?ticker2 rates->?rates2 isInTop500->?top2]
      ?dv1#?rates1 ?dv1[date -> ?date1]
      ?dv2#?rates2 ?dv2[date -> ?date2]
      External(pred:date-greater-than-or-equal(?date1 2012-01-01))
      External(pred:date-less-than-or-equal(?date1 2012-12-31))
      External(pred:date-greater-than-or-equal(?date2 2012-01-01))
      External(pred:date-less-than-or-equal(?date2 2012-12-31))
      ?c#srt:correlation ?c[corr->?corr
        series1->?rates1 series2->?rates2]
      External(pred:numeric-greater-than(?corr -0.25))
      External(pred:numeric-less-than(?corr 0.25))
      External(pred:boolean-equal(?top1 true))
      External(pred:boolean-equal(?top1 true))
      External(pred:string-less-then(?ticker1 ?ticker2))
    ))
  )
)

```

¹The RIF-BLD dialect is extended with the possibility to use the existential quantifier in the head of a rule.

The first rule of the document defines the predicate-collection tickers, in which the attribute symbol element runs through the list of security identifiers. Here, $e \# p$ is a predicate denoting membership of element e in collection p ; predicate $e[a \rightarrow v]$ means that the value of attribute a of object e is v . Securities taken only from S&P 500 list are put into tickers collection (attribute `isInTop500` equals to `true`). This means that only large companies from S&P 500 are considered as candidates for portfolio elements.

The second rule defines predicate `noncorrelated`, which is a noncorrelation relation between securities. Object $?e$ belongs to a relation `noncorrelated` if

- both securities belong to S&P top 500 list;
- the absolute value of correlation between the time series

if historical prices of securities for given attribute values `e.start` and `e.end` is less than 0.25. Here, 0.25 is the value chosen for the use-case as the predetermined maximal correlation value (see section A1); and

- identifiers of securities `id1` and `id2` are in lexicographic order ($id1 < id2$). This is required to prevent equivalent pairs (`id1`, `id2`) and (`id2`, `id1`) to be both presented in a relation corresponding to the predicate `noncorrelated` — only one pair is sufficient.

Dates in time series range from January 1, 2012 to December 31, 2012.

The second document (`prt`) contains a program that computes the maximum clique in a graph of correlations. The document is defined in the RIF-CASPD¹ dialect [25]:

```
Document( Dialect(RIF-CASPD)
  Import(<http://synthesis.ipi.ac.ru/optimalSecurityPortfolio>
    <http://www.w3.org/ns/entailment/OWL-Direct>)
  Module(<http://synthesis.ipi.ac.ru/graphExtraction#>)
  Prefix(prt <http://synthesis.ipi.ac.ru/portfolio#>)
  Prefix(gex <http://synthesis.ipi.ac.ru/graphExtraction#>)
  Group (
    Forall ?X(Or(prt:portfolio(?X) prt:nonPortfolio(?)) :- tickers@gex(?X))
    Forall ?X ?Y( :- And(prt:portfolio(?X) prt:nonPortfolio(?X)))
    Forall ?X ?Y( :- And(prt:portfolio(?X) prt:portfolio(?Y)
      (Naf noncorrelated@gex(?X ?Y))) )
    Forall ?X( :~ prt:nonPortfolio(?X)
      prt:portfolio(?X).
  ) )
```

The program defines a predicate `portfolio`, whose values are the security identifiers in the portfolio, and predicate `nonPortfolio`, whose values include all other securities under consideration.

The first rule states that the only securities considered are the securities which turn to truth the predicate `tickers` in document `gex`.

The second rule states that a security cannot simultaneously belong and not belong to the portfolio. So, the predicates `portfolio` and `nonPortfolio` divide the set of all securities into two nonintersecting sets.

The third rule claims that any two securities in `portfolio` are `noncorrelated` (according to `noncorrelated` predicate in the document `gex`). `Naf` here means a sort of negation — *negation as failure* [52]. This exactly means that securities belonging to `portfolio` form a clique in a graph of `noncorrelated` securities.

The fourth rule is a weak constraint, which minimizes the number of securities not belonging to the portfolio. So, the `portfolio` itself is maximized.

Obviously, these four rules declaratively specify the problem of finding the maximum cliques in a graph of `noncorrelated` securities.

Fifth rule is just a predicate `portfolio(?X)` with a free variable $?X$. This rule forms a result of the program, that

is, a collection of sets of identifiers of securities. Each set corresponds to a particular solution of the problem (a stable ASP-model for the program in the document `prt`) and forms a maximal portfolio.

A3 Peers of the Use-Case Infrastructure

For implementation of the conceptual specification of the problem of the investment portfolio diversification, the two peers should be produced. One of them is the mediator used for integration of data produced by the *Google Finance* and the *Yahoo! Finance* services.

Another one is a program for the rule-based programming system DLV [10]. The mediator produces a virtually integrated collection of facts, and DLV is a system for executing ASP-programs. Initially, the nodes do not contain any logic programs.

A3.1 The mediator schema

The mediator schema implements the conceptual specification of the application domain by expressing its semantics. Schema is written in the SYNTHESIS [4] — the canonical information model used for the mediator specifications:

¹The CASPD dialect is extended with the operation `:~` for a weak constraint. Such constraints should be satisfied if possible, but their violation on some tuples of variables values does not invalidate the models [13].

```

{ FinanceServices; in: module;
  type:
  { DatedValue; in: type;
    date: time;
    value: real;
  },
  { StockRates; in: type;
    ticker: string;
    rates: {set; type_of_element: DatedValue;};
    isInTop500:
    { in: function;
      params: {-returns/boolean};
    };
  };

  class_specification:
  { stockRates; in: class;
    instance_section: StockRates;
  };
}
function:
{ correlation; in: function;
  params: {+s1/{set;
  type_of_element: DatedValue;};
  +s2/{set;
  type_of_element: DatedValue;};, -corr/real };
};
}

```

The schema includes types `DatedValue` and `StockRates`, a class `stockRates`, and a function `correlation`. Semantics of class `stockRates`, type `DatedValue`, and function `correlation` correspond to the semantics of classes `stockRates`, `DatedValue`, and `correlation` belonging to `optimalSecurityPortfolio` ontology (see section A2).

The correspondence between the mediator schema and the conceptual schema is described by the schema mapping rules (see Fig. 3, middle part) constituting a separate RIF-document:

```

Document( Dialect(RIF-BLD)
  Import(<http://synthesis.ipi.ac.ru/optimalSecurityPortfolio>
    <http://www.w3.org/ns/entailment/OWL-Direct>)
  Prefix(srt <http://synthesis.ipi.ac.ru/optimalSecurityPortfolio#>)
  Prefix(fs <http://synthesis.ipi.ac.ru/FinanceServices#>)
  Group(
    And(?t#srt:stockRates
      ?t[?ticker-> srt:ticker ?top-> srt:isInTop500
        ?rates-> srt:rates]) :-
    And(External(?t#fs:stockRates
      External(?t[?ticker->fs:ticker ?top->fs:isInTop500
        ?rates->fs:rates]) )

    And(?dv#srt:DateValue ?dv[?v->srt:value ?d->srt:date]) :-
    And(External(?dv#fs:DateValue
      External(?dv[?v->fs:value ?d->fs:date]) )

    And(?c#correlation
      ?c[?series1->series1 ?series2->series2 ?corr->corr]) :-
    External(pred:numeric-equal(
      fs:correlation(?series1 ?series2) ?corr))
  )
)

```

The document consists of three rules. The first one defines a correspondence between class `stockRates` of the `optimalSecurityPortfolio` ontology (denoted as `srt:stockRates`) and class `stockRates` of the `FinanceServices` mediator module (denoted as `fs:stockRates`) as well as correspondences between their attributes. The second and the third rules define correspondences between ontology class `srt:DatedValue` and mediator type `fs:DatedValue` and between ontology class `srt:correlation` and mediator function `fs:correlation`.

Entities of the mediator schema and conceptual schema are in one-to-one correspondence, the names and the semantics of the relevant entities are the same. Also, the mediator is a conformant RIF-BLD consumer. Thus, the mediator

`FinanceServices` is relevant to the RIF-document `gex` of the conceptual schema.

A3.2 Resource integration by the mediator

The mediator integrates two resources: *Google Finance* and the *Yahoo! Finance services*. For both resources, specific wrappers (the *Google Finance* wrapper and the *Yahoo! Finance* wrapper) were implemented to embed the resources in the mediation system [53].

Schema of the *Google Finance* service presented for uniformity in the canonical information model (the SYNTHESES language) looks as follows:

```

{ GoogleFinance; in: module;
  type:
  { GoogleQuote; in: type;
    symbol: string;
    Date: time;
    Open: real;
    High: real;
    Low: real;
    Close: real;
    Volume: integer;
  };

  function:
  { groupByTicker; in: function;
    params: {+symbol/string,
      +date/time, +close/real,
      -rates/{set; type_of_element:
        DatedValue;}}
  },
  { isInTop500; in: function;
    params: {+ticker/string,
      -returns/boolean};
  };

  class_specification:
  { historicalData; in: class;
    instance_section: GoogleQuote;
  };
}

```

The schema includes a type `GoogleQuote`, which includes attributes that match the identifier of the company (`symbol`), date (`Date`), and different indicators of the security price. Daily closing price of the shares (`Close`) is used to form the portfolio.

The function `top500` reflects the S&P 500 index: if the value of the ticker parameter belongs to S&P 500 list, then the function returns true and false otherwise.

The function `groupByTicker` is aimed to group flat instances of the `GoogleQuote` type by ticker attribute to produce a time series of an indicator of the security price for all securities identified by ticker.

Functions of the schema are implemented by the *Google Finance* wrapper.

Schema of the *Yahoo! Finance* service presented in the canonical model of subject mediators looks as follows:

```

{ YahooFinance; in: module;
  type:
  { YahooQuote; in: type;
    symbol: string;
    Date: time;
    Open: real;
    High: real;
    Low: real;
    Close: real;
    Volume: integer;
    Adj_Close: real;
  };
}

```

```

function:
{ groupByTicker; in: function;
  params: {+symbol/string,
    +date/time, +close/real, -rates/{set;
    type_of_element: DatedValue;}};
},
{ isInTop500; in: function;
  params: {+ticker/string,
    -returns/boolean};
};

class_specification:
{ historicalData; in: class;
  instance_section: YahooQuote;
};
}

```

The schema includes a type `YahooQuote` similar to `GoogleQuote` and functions `groupByTicker` and `isInTop500` similar to respective functions of *Google Finance* schema.

Integration of the resources in the mediator has been provided using GLAV-mappings. The GLAV-mapping is a combination of GAV-mappings and LAV-mappings. The GAV-mapping looks as follows:

```

financeData(x/[ticker, rates]) :-
  group_by({ticker}, [ticker, p1: partition],
    GoogleFinance.historicaldata(
      x/[ticker: symbol, date: Date,
        value: Close]))&
  group_by({ticker}, [ticker, p2: partition],
    YahooFinance.historicaldata(
      x/[ticker: symbol, date: Date,
        value: Close]))&
  is_equal(rates, union(p1, p2))
).

```

An intermediate predicate `financeData` (used to bind GAV and LAV-mappings) is expressed as a view over resource classes `GoogleFinance.historicaldata` and `YahooFinance.historicaldata`. Flat data structures of the resource classes are turned into hierarchical data structures of the mediator class using `group_by` operation. Curly braces in `group_by` operation enclose the grouping attribute (`ticker`). Square brackets in `group_by` operation enclose attributes of objects in a collection formed by `group_by` operation [4]. Partition is a default attribute storing grouped partitions. Here, resource attributes `Date` and `Close` are grouped in the partition attribute. Partitions from `Google` and `Yahoo` resource classes are merged into `rates` attribute of `stockRates` class (union operation). Attributes of the resource classes are renamed properly in order to meet the structure of the mediator class (for instance, `Close` attribute is renamed into `value`).

The LAV mappings look as follows:

```

financeData(x/[ticker, rates]) :-
  stockRates(x/[ticker, rates]).
GoogleFinance.isInTop500(x.ticker, b) ->
  isInTop500(x, b)
YahooFinance.isInTop500(y.ticker, b) ->
  isInTop500(y, b)

```

Here, `financeData` predicate is expressed as a simple view over mediator schema class `stockRates`, and `isInTop500` resource functions are expressed through mediator `isInTop500` function using *functional inverse rules* [19].

A3.3 DLV system

DLV system is a conformant RIF-CASPD consumer, and initially, the system does not contain any logic program or facts. Document `prt` does not contain any occurrence of an extensional predicate corresponding to an entity in the conceptual schema. Thus, the system is relevant to the RIF-document `prt` of the conceptual schema, since the relevance

condition (see subsection 3.2) is reduced to the relation of dialects.

A4 Portfolio Problem Infrastructure

In the provided example, the infrastructure (Fig. 4) includes two nodes corresponding to the mediation system (called `fsv`) and to a rule-based programming system DLV (called `dlv`).

Prefixes of the documents of the conceptual schema in the rules are replaced with the actual node prefixes during rewriting the conceptual programs into the programs over the P2P network nodes. For instance, the rules of the `prt` document were rewritten as follows:

```
Document( Dialect(RIF-CASPD)
  Module(<http://synthesis.ipi.ac.ru/resources/FinanceServices#>)
  Prefix(dlv <http://synthesis.ipi.ac.ru/resources/DLV#>)
  Prefix(fsv <http://synthesis.ipi.ac.ru/resources/FinanceServices#>)
  Group (
    Forall ?X(Or(dlv:portfolio(?X) dlv:nonPortfolio(?)) :-
      tickers@fsv(?X))
    Forall ?X ?Y( :- And(dlv:portfolio(?X) dlv:nonPortfolio(?Y)))
    Forall ?X ?Y( :- And(dlv:portfolio(?X) dlv:portfolio(?Y)
      (Naf noncorrelated@fsv(?X ?Y))) )
    Forall ?X( :~ dlv:nonPortfolio(?X)
      dlv:portfolio(?X).
  )
)
```

Rules from the document `gex` were rewritten similarly. Rewritten documents are sent by the supervisor to the proper nodes: `gex` to the mediation node, and `prt` to the `dlv` node.

After that, the node wrappers automatically execute distributed programs in accordance with the algorithm described in subsection 3.3. First, the program normalization is done. Thus, the nonlocal rule mentioned above

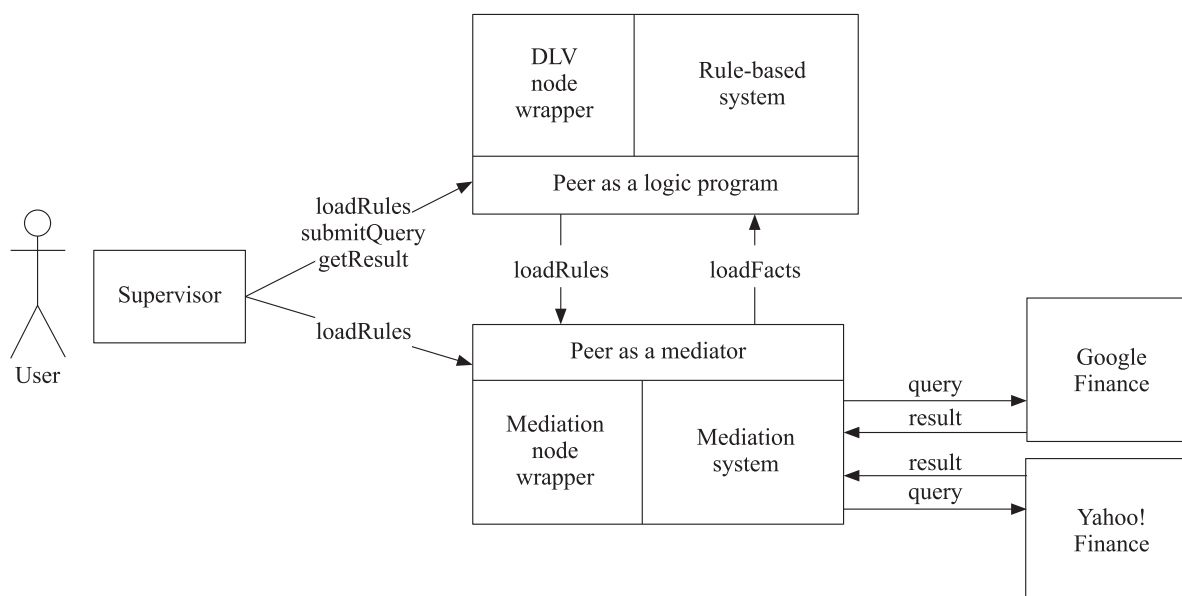


Figure 4 Portfolio problem infrastructure


```

forall ?X ?Y ( :- And(dlv:portfolio(?X)
  dlv:portfolio(?Y)
  (Naf noncorrelated@fsv(?X ?Y))) )

```

is transformed into a delegation rule

```

noncorrelated_fsv(?X ?Y) :-
  noncorrelated@fsv(?X ?Y)

```

and a local rule

```

forall ?X ?Y ( :- And(dlv:portfolio(?X)
  dlv:portfolio(?Y)
  (Naf noncorrelated_fsv(?X ?Y))) )

```

Remaining rules from both documents are normalized in a similar way.

After that, the normalized program is executed. The fact delegation rule, which sends the `noncorrelated` predicate to the `dlv` node, is transmitted to the `fsv` node. The bodies of the rules of the program on the `fsv` node do not contain remote terms; so, the program on the node can be executed without waiting for the facts from the other nodes. In contrast, the `dlv` node has to wait for the arrival of the facts from the `fsv` node, which turns the tickers and the `noncorrelated` predicates to true. After receiving all necessary facts, the program computing portfolio is executed in the `dlv` node.

Before the execution, the normalized programs written in RIF dialects were automatically transformed into the logic languages supported on the nodes — the SYNTHESES and the DLV, respectively. Transformations were implemented using a model transformation language and toolkit ATL (ATLAS Transformation Language) [54].

Toolkit ATL is a hybrid of declarative and imperative languages. An ATL transformation program is composed of rules that define how source model elements are matched and navigated to create and initialize the elements of the target models.

The result of the transformation of the `prt` document into the SYNTHESES language looks as follows:

```

tickers(ts/[symbol]) :-
  stockRates(t/[symbol: ticker]) &
  isInTop500(t, inTop500) & inTop500 = true.

```

```

noncorrelated(e/[start: ticker1,
  end: ticker2]) :-
  stockRates(m/[ticker1: ticker, rates1: rates,
    top1: isInTop500]) &
  stockRates(n/[ticker2: ticker, rates2: rates,

```

```

  top2: isInTop500]) &
  in_set(dv1, rates1) & in_set(dv2, rates2) &
  dv1.date = date1 & dv2.date = date2 &
  date1 >= '2012-01-01' & date1 <= '2012-12-31' &
  date2 >= '2012-01-01' & date2 <= '2012-12-31' &
  correlation(c/[corr, rates1: series1,
    rates2: series2 ]) &
  corr >= -0.25 & corr <= 0.25 &
  top1 = true & top2 = true &
  ticker1 < ticker2.

```

Note that existential quantifier in heads of the rules is discarded during the mapping. According to the semantics of the SYNTHESES logic rules, the variables in the head of a rule which are not presented in the body of the rule are bounded by existential quantifier by default. The RIF predicates of belonging of a variable to a class (like `?t\#stockRates`) are combined with frame predicates (like `?t[ticker->?symbol]`) and mapped into a predicate-collection [4] of the SYNTHESES language (like `stockRates(t/[symbol: ticker])`).

The result of the transformation of the `gex` document into DLV language looks as follows:

```

portfolio(X) v nonPortfolio(X) :- tickers(X).
:- portfolio(X), portfolio(Y),
  not noncorrelated(X, Y).
:~ nonPortfolio(X).
portfolio(?X).

```

A5 Result of the Use Case Program Execution

The maximal models of the `portfolio` predicate, found during the execution of the program on the `dlv` node, are the diversified portfolios of maximal size containing securities of companies from the S&P 500 for 2012. As the result of the program execution, the 11 stable models containing 10 ground atoms each were generated (the maximum size of a diversified portfolio appeared to be equal to 10). The models can be found below. Atoms contain symbols of different companies, e.g., `DUK` and `SBUX` denote *Duke Energy* and *Starbucks Corp.*, respectively. The full list of abbreviations is provided in the table. The models are outlined as the sets of company identifiers (like `SBUX`) related to the security ground terms in `portfolio(SBUX)`. Note that the models have nonempty intersections:

```

Model 1: {CAH, DUK, EL, ETR, HOT, LM, PSA, TJX, TYC, UNH}
Model 2: {CAH, DUK, EL, ETR, HOT, LM, PSA, TJX, UNH, VNO}
Model 3: {CAH, DUK, EL, ETR, HON, LM, PSA, TJX, UNH, VNO}
Model 4: {CAH, DUK, EL, ETR, LM, PSA, TJX, TYC, UNH, WDC}
Model 5: {BMY, DUK, EL, FDX, IPG, KSS, PSA, STT, TJX, VIAB}
Model 6: {BA, BMY, DUK, EL, IPG, KSS, PSA, STT, TJX, VIAB}
Model 7: {BA, BMY, DUK, EL, KSS, LH, PSA, STT, TJX, VIAB}
Model 8: {AGN, BA, BMY, DUK, EL, KSS, LH, PSA, TJX, VIAB}
Model 9: {BA, BMS, BMY, DUK, EL, KSS, LH, PSA, SBUX, TJX}
Model 10: {BMY, BSX, DUK, EL, FDX, GAS, KMI, MDLZ, RRC, TJX}
Model 11: {BA, BMY, DUK, FMC, GIS, KIM, LH, LTD, MNST, SBUX}

```

Tickers symbols and company names

Ticker	Company	Ticker	Company
CAH	Cardinal Health Inc.	STT	State Street Corp.
DUK	Duke Energy	VIAB	Viacom Inc.
EL	Estee Lauder Cos.	BA	Boeing Company
ETR	Entergy Corp.	LH	Laboratory Corp. of America Holding
HOT	Starwood Hotels & Resorts	AGN	Allergan Inc
LM	Legg Mason	BMS	Bemis Company
PSA	Public Storage	SBUX	Starbucks Corp.
TJX	TJX Companies Inc.	BSX	Boston Scientific
TYC	Tyco International	GAS	AGL Resources Inc.
UNH	United Health Group Inc.	KMI	Kinder Morgan
VNO	Vornado Realty Trust	MDLZ	Mondelez International
HON	Honeywell Int'l Inc.	RRC	Range Resources Corp.
WDC	Western Digital	FMC	FMC Corporation
BMJ	Bristol-Myers Squibb	GIS	General Mills
FDX	FedEx Corporation	KIM	Kimco Realty
IPG	Interpublic Group	LTD	L Brands Inc.
KSS	Kohl's Corp.	MNST	Monster Beverage

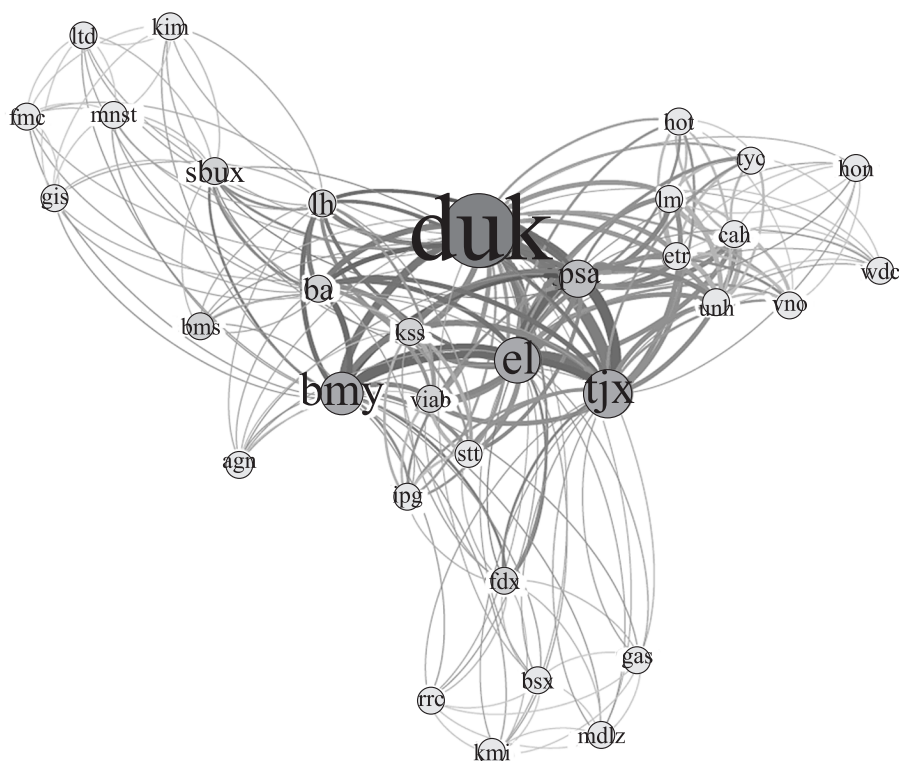


Figure 5 Models and their elements

Models and their elements are illustrated in Fig. 5¹.

The circles in the figure denote companies. The lines connecting two circles denote that two companies belong to the same model. The larger is the circle, the larger is the number of models to which a company belongs. The thicker is the line, the larger is the number of models to which a pair of companies belongs to.

¹Figure was created using a graph visualization platform *Gephi* (<https://gephi.org/>). The *ForceAtlas* layout algorithm was applied.

References

1. Hey, T., S. Tansley, and K. Tolle, eds. 2009. *The fourth paradigm: Data-intensive scientific discovery*. Redmond: Microsoft Research. 252 p.
2. Challenges and opportunities with big data. A community white paper developed by leading re-

- searchers across the United States. 2012. Available at: <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf> (accessed November 21, 2013).
3. Kappel, G., M. Wimmer, W. Retschitzegger, and W. Schwinger. 2011. Leveraging model-based tool integration by conceptual modeling techniques. *The evolution of conceptual modeling*. Eds. R. Kaschek and L. M. L. Delcambre. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6520:254–284.
 4. Kalinichenko, L. A., S. A. Stupnikov, and D. O. Martynov. 2007. SYNTHESIS: A language for canonical information modeling and mediator definition for problem solving in heterogeneous information resource environments. Moscow: IPI RAN. 171 p.
 5. Kalinichenko, L. A., D. O. Briukhov, D. O. Martynov, N. A. Skvortsov, and S. A. Stupnikov. 2007. Mediation framework for enterprise information system infrastructures. *9th Conference (International) on Enterprise Information Systems ICEIS 2007 Proceedings*. Funchal. Vol. Databases and information systems integration. 246–251.
 6. Kalinichenko, L. A. 1990. Methods and tools for equivalent data model mapping construction. *Advances in database technology — EDBT'90*. Eds. F. Buncilhon, C. Thanos, and D. Tsichritzis. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 416:92–119.
 7. Kalinichenko, L. A., and S. A. Stupnikov. 2012. Synthesis of the canonical models for database integration preserving semantics of the value inventive data models. *Advances in database and information systems*. Eds. T. Morzy, T. Härder, R. Wrembel, et al. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 7503:223–239.
 8. Boley, H., and M. Kifer, eds. 2013. RIF overview. W3C working group note. 2nd ed. Available at: <http://www.w3.org/TR/rif-overview/> (accessed November 21, 2013).
 9. Kalinichenko, L., S. Stupnikov, A. Vovchenko, and D. Kovalev. 2013. Rule-based multi-dialect infrastructure for conceptual problem solving over heterogeneous distributed information resources. New trends in databases and information systems. *Selected Papers of the 17th European Conference on Advances in Databases and Information Systems and Associated Satellite Events. Advances in Intelligent Systems and Computing* 241:61–68.
 10. Leone, N., G. Pfeifer, W. Faber, T. Eiter, G. Gottlob, S. Perri, and F. Scarcello. 2006. The DLV system for knowledge representation and reasoning. *ACM Trans. Comput. Log.* 7(3):499–562.
 11. Kalinichenko, L. A., and S. A. Stupnikov. 2008. Constructing of mappings of heterogeneous information models into the canonical models of integrated information systems. *ADBIS 2008 Proceedings*. Pori: Tampere University of Technology. 106–122.
 12. Abrial, J.-R. 1996. *The B-book: Assigning programs to meanings*. Cambridge: Cambridge University Press. 816 p.
 13. Skvortsov, N. A. 2012. Otobrazhenie modeley dannykh NoSQL v ob"ektnye specifikatsii [Mapping of NoSQL data models to object specifications]. *Trudy 14-y Vserossiyskoy nauchnoy konferentsii "Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektzii" RCDL 2012 [14th Russian Conference on Digital Libraries RCDL 2012 Proceedings]*. CEUR Workshop Proceedings 934:53–62.
 14. Stupnikov, S. A. 2013. Otobrazhenie grafovoy modeli dannykh v kanonicheskuyu ob"ektno-freymovuyu informatsionnyuyu model' pri sozdanii sistem integratsii neodnorodnykh informatsionnykh resursov [Mapping of a graph data model into an object-frame canonical information model for the development of heterogeneous information resources integration systems]. *Trudy 15-y Vserossiyskoy nauchnoy konferentsii "Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektzii" RCDL 2013 [15th Russian Conference on Digital Libraries RCDL 2013 Proceedings]*. Yaroslavl: P. G. Demidov Yaroslavl State University. 193–202.
 15. Skvortsov, N. A. 2013. Otobrazhenie modeli dannykh RDF v kanonicheskuyu model' predmetnykh posrednikov [Mapping of RDF data model into the canonical model of subject mediators]. *Trudy 15-y Vserossiyskoy nauchnoy konferentsii "Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektzii" RCDL 2013 [15th Russian Conference on Digital Libraries RCDL 2013 Proceedings]*. Yaroslavl: P. G. Demidov Yaroslavl State University. 202–209.
 16. Stupnikov, S. A. 2013. Verifitsiruemoe otobrazhenie modeli dannykh, osnovannoy na mnogomernykh massivakh, v ob"ektnuyu model' dannykh [A verifiable mapping of a multidimensional array data model into an object data model]. *Informatika i ee Primeneniya — Inform. Appl.* 7(3):22–34.
 17. Fagin, R., P. G. Kolaitis, R. J. Miller, and L. Popa. 2005. Data exchange: Semantics and query answering. *Theor. Comput. Sci.* 336:89–124.
 18. Lenzerini, M. 2002. Data integration: A theoretical perspective. *ACM Symposium on Principles of Database Systems (PODS) Proceedings*. 233–246.
 19. Kalinichenko, L. A., D. O. Martynov, and S. A. Stupnikov. 2004. Query rewriting using views in a typed mediator environment. *Advances in databases and information systems*. Eds. G. Gottlob, A. Benczur, and J. Demetrovics. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 3255:37–53.
 20. Boley, H., and M. Kifer, eds. 2013. RIF basic logic dialect. W3C recommendation. 2nd ed. Available at: <http://www.w3.org/TR/rif-bld/> (accessed November 21, 2013).
 21. De Sainte Marie, C., G. Hallmark, and A. Paschke, eds. 2013. RIF production rule dialect. W3C recommendation. 2nd ed. Available at: <http://www.w3.org/TR/rif-prd/> (accessed November 21, 2013).
 22. Boley, H., and M. Kifer, eds. 2013. RIF framework for logic dialects. W3C recommendation. 2nd ed. Available at: <http://www.w3.org/TR/rif-fld/> (accessed November 21, 2013).
 23. Kifer, M., ed. 2010. RIF core logic programming dialect based on the well-founded semantics. Avail-

- able at: <http://ruleml.org/rif/RIF-CLPWD.html> (accessed November 21, 2013).
24. Van Gelder, A., K. A. Ross, and J. S. Schlipf. 1991. The well-founded semantics for general logic programs. *J. ACM* 38(3):620–650.
 25. Heymans, S., and M. Kifer, eds. 2009. RIF core answer set programming dialect. Available at: <http://ruleml.org/rif/RIF-CASPD.html> (accessed November 21, 2013).
 26. Gelfond, M., and V. Lifschitz. 1988. The stable model semantics for logic programming. *Logic Programming: 5th Conference and Symposium Proceedings*. 1070–1080.
 27. De Bruijn, J., and C. Welty, eds. 2013. RIF RDF and OWL Compatibility. W3C recommendation. 2nd ed. Available at: <http://www.w3.org/TR/rif-rdf-owl/> (accessed November 21, 2013).
 28. Motik, B., P. F. Patel-Schneider, and B. Parsia, eds. 2012. OWL 2 Web ontology language structural specification and functional-style syntax. W3C recommendation. 2nd ed. Available at: <http://www.w3.org/TR/owl2-syntax/> (accessed November 21, 2013).
 29. Shvaiko, P., and J. Euzenat. 2005. A survey of schema-based matching approaches. *J. Data Semantics IV*:146–171.
 30. Abiteboul, S., M. Bienvenu, A. Galland, *et al.* 2011. A rule-based language for Web data management. *30th ACM Symposium on Principles of Database Systems Proceedings*. ACM Press. 283–292.
 31. Gelfond, M. 2008. *Answer sets. Handbook of knowledge representation*. Elsevier. 285–316.
 32. Motik, B., and R. Rosati. 2007. Closing semantic Web ontologies. University of Manchester Report. Available at: <http://www.cs.ox.ac.uk/people/boris.motik/pubs/mr06closing-report.pdf> (accessed November 21, 2013).
 33. Chen, P. P., B. Thalheim, and L. Y. Wong. 1999. Future directions of conceptual modeling. *Conceptual modeling*. Eds. P. P. Chen, J. Akoka, H. Kangassulu, and B. Thalheim. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 1565:287–301.
 34. Costal, D., C. Gómez, and G. Guizzardi. 2011. Formal semantics and ontological analysis for understanding subsetting, specialization and redefinition of associations in UML. *Conceptual modeling — ER 2011*. Eds. M. Jeusfeld, L. Delcambre, and T. W. Ling. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6998:189–203.
 35. Martinez, Y., C. Cachero, M. Matera, S. Abrahao, and S. Lujan. 2011. Impact of MDE approaches on the maintainability of Web applications: An experimental evaluation. *Conceptual modeling — ER 2011*. Eds. M. Jeusfeld, L. Delcambre, and T. W. Ling. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6998:233–246.
 36. Castro, L. F. Baião, and G. Guizzardi. 2011. A semantic oriented method for conceptual data modeling in OntoUML based on linguistic concepts. *Conceptual modeling — ER 2011*. Eds. M. Jeusfeld, L. Delcambre, and T. W. Ling. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6998:486–494.
 37. Fillottrani, P. R., E. Franconi, and S. Tessaris. 2011. The ICOM 3.0 intelligent conceptual modelling tool and methodology. *Semantic Web. IOS Press*. 1–14. Available at: http://www.semantic-web-journal.net/sites/default/files/swj105_1.pdf (accessed November 21, 2013).
 38. Maus, C., S. Rybacki, and A. M. Uhrmache. 2011. Rule-based multi-level modeling of cell biological systems. *BMC Syst. Biol.* 5:166. Available at: <http://www.biomedcentral.com/content/pdf/1752-0509-5-166.pdf> (accessed November 21, 2013).
 39. Calvanese, D., G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati. 2007. MASTRO-I: Efficient integration of relational data through DL ontologies. *2007 Workshop (International) on Description Logic (DL 2007) Proceedings. CEUR Workshop Proceedings*. 250.
 40. Cosentino, V., M. D. Del Fabro, and A. El Ghali. 2012. A model driven approach for bridging ILOG rule language and RIF. *RuleML 2012 Proceedings. CEUR Workshop Proceedings*. 874. Available at: <http://ceur-ws.org/Vol-874/paper9.pdf> (accessed November 21, 2013).
 41. Gonzalez-Moriyon, G., L. Polo, D. Berrueta, and C. Tejo-Alonso. 2012. Final steel industry public demonstrators. ONTORULE deliverable D5.5. Available at: <http://ontorule-project.eu/outcomes%3Ffunc=fileinfo&id=94.html> (accessed November 21, 2013).
 42. Loo, B. T., T. Condie, M. Garofalakis, D. E. Gay, J. M. Hellerstein, P. Maniatis, R. Ramakrishnan, T. Roscoe, and I. Stoica. 2006. Declarative networking: Language, execution and optimization. *ACM SIGMOD Conference Proceedings*. 97–108.
 43. Grumbach, S., and F. Wang. 2010. Netlog, a rule-based language for distributed programming. *Practical aspects of declarative languages*. Eds. M. Carro and R. Peña. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 5937:88–103.
 44. Alvaro, P., W. R. Marczak, N. Conway, J. M. Hellerstein, D. Maier, and R. Sears. 2010. Dedalus: Datalog in time and space. *Datalog reloaded*. Eds. O. de Moor, G. Gottlob, T. Furche, and A. Sellers. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6702: 262–281.
 45. Grossmann, G., R. Thiagarajan, M. Schrefl, and M. Stumptner. 2011. Conceptual modeling approaches for dynamic Web service composition. *The evolution of conceptual modeling*. Eds. R. Kaschek and L. M. L. Delcambre. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6520:180–204.
 46. Henderson-Sellers, B. 2011. Random thoughts on multi-level conceptual modelling. *The evolution of conceptual modeling*. Eds. R. Kaschek and L. M. L. Delcambre. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6520:93–116.
 47. Falbo, R. A., F. B. Ruy, and R. D. Moro. 2005. Using ontologies to add semantics to a software engineering en-

- vironment. *SEKE 2005 Proceedings*. Curran Associates. 151–156.
48. Berre, A., and M. Missikoff (Moderators). 2011. Panel: Modeling for the future Internet. *Conceptual modeling — ER 2011*. Eds. M. Jeusfeld, L. Delcambre, and T. W. Ling. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6998: 526–527.
49. Porto, F., and S. Spaccapietra. 2011. Data model for scientific models and hypotheses. *The evolution of conceptual modeling*. Eds. R. Kaschek and L. M. L. Delcambre. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 6520: 285–305.
50. Racunas, S. A., N. H. Shah, I. Albert, and N. V. Fedoroff. 2004. Hybrow: A prototype system for computer-aided hypothesis evaluation. *Bioinformatics* 20(1):257–264.
51. Sharpe, W., G. J. Alexander, and J. W. Bailey. 1998. *Investments*. Prentice Hall. 962 p.
52. Gabbay, D. M. 2012. What is negation as failure? *Logic programs, norms and action*. Eds. A. Artikis, R. Craven, N. K. Çiçekli, et al. Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag. 7360:52–78.
53. Briukhov, D. O., A. E. Vovchenko, V. N. Zakharov, O. P. Zhelenkova, L. A. Kalinichenko, D. O. Martynov, N. A. Skvortsov, and S. A. Stupnikov. 2008. Arkhitektura promezhutochnogo sloya predmetnykh posrednikov dlya resheniya zadach nad mnozhestvom integriruemyykh neodnorodnykh raspredelennykh informatsionnykh resursov v gibridnoy grid-infrastrukture virtual'nykh observatoriy [The middle ware architecture of the subject mediators for problem solving over a set of integrated heterogeneous distributed information resources in the hybrid grid-infrastructure of virtual observatories]. *Informatika i ee Primeneniya — Inform. Appl.* 2(1):2–34.
54. ATL Project. 2013. Available at: <http://www.eclipse.org/atl/> (accessed November 21, 2013).

Received November 25, 2013

Contributors

Kalinichenko Leonid A. (b. 1937) — Doctor of Science in physics and mathematics; Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow 119991, Russian Federation; leonidandk@gmail.com

Stupnikov Sergey A. (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; ssa@ipi.ac.ru

Vovchenko Alexey E. (b. 1984) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; itsnein@gmail.com

Kovalev Dmitry Yu. (b. 1988) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; programmer, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; dm.kovalev@gmail.com

КОНЦЕПТУАЛЬНЫЕ ДЕКЛАРАТИВНЫЕ СПЕЦИФИКАЦИИ И РЕШЕНИЕ ЗАДАЧ В ОБЛАСТЯХ С ИНТЕНСИВНЫМ ИСПОЛЬЗОВАНИЕМ ДАННЫХ

Л. Калиниченко¹, С. Ступников², А. Вовченко³, Д. Ковалев⁴

¹Институт проблем информатики Российской академии наук, Москва, Россия, leonidandk@gmail.com

²Институт проблем информатики Российской академии наук, Москва, Россия, ssa@ipi.ac.ru

³Институт проблем информатики Российской академии наук, Москва, Россия, itsnein@gmail.com

⁴Институт проблем информатики Российской академии наук, Москва, Россия, dm.kovalev@gmail.com

Аннотация: В сфере концептуального моделирования долгое время исследовались разнообразные нотации, предназначенные для определения семантики вычислений в терминах предметных областей. Подход «сущность—связь» и диаграммы UML позволяют определять семантику лишь неформально. Онтологические языки, основанные на дескриптивной логике, разрабатывались для формализации семантики данных. Однако сейчас общепризнано, что одной лишь семантики данных недостаточно — требуется еще и представление алгоритмов анализа данных для спецификации данных и поведения в одной парадигме. Более того, все усиливающееся разнообразие разноструктурированных моделей данных вызывает потребность в их унифицированной, интегрированной абстракции для получения спецификаций, неза-

висимых от реальных данных в предметных областях с интенсивным использованием данных. С целью преодоления названных недостатков предлагается новый подход к применению семантически различных языков на правилах (диалектов) для создания интероперабельных концептуальных спецификаций над различными системами на правилах. Подход основан на на технике преобразования логических программ, рекомендованной *Форматом обмена правилами* (RIF) W3C. Этот подход гармонично сочетается со спецификациями, предназначенными для определения семантических посредников на правилах, обеспечивающих интеграцию неоднородных баз данных. Определена инфраструктура, реализующая мультидиалектные концептуальные спецификации при помощи интероперабельных систем на правилах и систем поддержки посредников. Представлен подтверждающий предложенные концепции прототип инфраструктуры, основанный на системе поддержки посредников СИНТЕЗ и стандарте RIF. Подход к мультидиалектной концептуализации предметной области, делегированию правил, интероперабельности программ на правилах и посредников подробно рассмотрен и проиллюстрирован на реальном примере NP-полной задачи в финансовой области. Результаты исследования свидетельствуют о применимости подхода и инфраструктуры для концептуального, декларативного, независимого от ресурсов и повторно используемого анализа данных в различных предметных областях.

Ключевые слова: концептуальная спецификация; W3C RIF; логические языки на правилах; СИНТЕЗ; интеграция баз данных, посредники; RIF-BLD; RIF-CASPD; мультидиалектная инфраструктура; делегирование правил

DOI: 10.14357/19922264130412

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 11-07-00402-а) и Программы фундаментальных исследований Президиума Российской академии наук.

Литература

1. The fourth paradigm: Data-intensive scientific discovery / Eds. T. Hey, S. Tansley, K. Tolle. — Redmond: Microsoft Research, 2009. 252 p.
2. Challenges and opportunities with big data. A community white paper developed by leading researchers across the United States, 2012. Available at: <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf> (accessed November 21, 2013).
3. Kappel G., Wimmer M., Retschitzegger W., Schwinger W. Leveraging model-based tool integration by conceptual modeling techniques // The evolution of conceptual modeling. Eds. R. Kaschek and L. M. L. Delcambre. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6520. P. 254–284.
4. Kalinichenko L. A., Stupnikov S. A., Martynov D. O. SYNTHESIS: A language for canonical information modeling and mediator definition for problem solving in heterogeneous information resource environments. — M.: IPI RAN, 2007. 171 p.
5. Kalinichenko L. A., Briukhov D. O., Martynov D. O., Skvortsov N. A., Stupnikov S. A. Mediation framework for enterprise information system infrastructures // 9th Conference (International) on Enterprise Information Systems ICEIS 2007 Proceedings. — Funchal, 2007. Vol. Databases and information systems integration. P. 246–251.
6. Kalinichenko L. A. Methods and tools for equivalent data model mapping construction // Advances in database technology — EDBT'90. Eds. F. Buncilhon, C. Thanos, and D. Tsichritzis. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 1990. Vol. 416. P. 92–119.
7. Kalinichenko L. A., Stupnikov S. A. Synthesis of the canonical models for database integration preserving semantics of the value inventive data models // Advances in database and information systems / Eds. T. Morzy, T. Härder, R. Wrembel, *et al.* Lecture notes in computer science ser. Berlin, Heidelberg: Springer-Verlag, 2012. Vol. 7503. P. 223–239.
8. RIF overview. W3C working group note / Eds. H. Boley, M. Kifer. — 2nd ed., 2013. Available at: <http://www.w3.org/TR/rif-overview/> (accessed November 21, 2013).
9. Kalinichenko L., Stupnikov S., Vovchenko A., Kovalev D. Rule-based multi-dialect infrastructure for conceptual problem solving over heterogeneous distributed information resources. New trends in databases and information systems // Selected Papers of the 17th European Conference on Advances in Databases and Information Systems and Associated Satellite Events. Advances in Intelligent Systems and Computing, 2013. Vol. 241. P. 61–68.
10. Leone N., Pfeifer G., Faber W., Eiter T., Gottlob G., Perri S., Scarcello F. The DLV system for knowledge representation and reasoning // ACM Trans. Comput. Log., 2006. Vol. 7. No. 3. P. 499–562.
11. Kalinichenko L. A., Stupnikov S. A. Constructing of mappings of heterogeneous information models into the canonical models of integrated information systems // ADBIS 2008 Proceedings. — Pori: Tampere University of Technology, 2008. P. 106–122.
12. Abrial J.-R. The B-Book: Assigning programs to meanings. — Cambridge: Cambridge University Press, 1996. 816 p.
13. Скворцов Н. А. Отображение моделей данных NoSQL в объектные спецификации // Труды 14-й Всеросс. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL 2012. — Переславль-Залесский: Ун-т г. Переславля, 2012. Т. 934. С. 53–62.

14. Ступников С. А. Отображение графовой модели данных в каноническую объектно-фреймовую информационную модель при создании систем интеграции неоднородных информационных ресурсов // Труды 15-й Всеросс. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL 2013. — Ярославль: Ярославский гос. ун-т им. П. Г. Демидова, 2013. С. 193–202.
15. Скворцов Н. А. Отображение модели данных RDF в каноническую модель предметных посредников // Труды 15-й Всеросс. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL 2013. — Ярославль: Ярославский гос. ун-т им. П. Г. Демидова, 2013. С. 202–209.
16. Ступников С. А. 2013. Верифицируемое отображение модели данных, основанной на многомерных массивах, в объектную модель данных // Информатика и её применения, 2013. Т. 7. Вып. 3. С. 22–34.
17. Fagin R., Kolaitis P. G., Miller R. J., Popa L. Data exchange: Semantics and query answering // Theor. Comput. Sci., 2005. Vol. 336. P. 89–124.
18. Lenzerini M. Data integration: A theoretical perspective // ACM Symposium on Principles of Database Systems (PODS) Proceedings, 2002. P. 233–246.
19. Kalinichenko L. A., Martynov D. O., Stupnikov S. A. Query rewriting using views in a typed mediator environment // Advances in databases and information systems / Eds. G. Gottlob, A. Benczur, J. Demetrovics. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2004. Vol. 3255. P. 37–53.
20. RIF basic logic dialect. W3C recommendation / Eds. H. Boley, M. Kifer. — 2nd ed., 2013. Available at: <http://www.w3.org/TR/rif-bld/> (accessed November 21, 2013).
21. RIF production rule dialect. W3C recommendation / Eds. C. De Sainte Marie, G. Hallmark, A. Paschke. — 2nd ed., 2013. Available at: <http://www.w3.org/TR/rif-prd/> (accessed November 21, 2013).
22. RIF framework for logic dialects. W3C recommendation / Eds. H. Boley, M. Kifer. — 2nd ed., 2013. Available at: <http://www.w3.org/TR/rif-fld/> (accessed November 21, 2013).
23. RIF core logic programming dialect based on the well-founded semantics / Ed. M. Kifer. — 2010. Available at: <http://ruleml.org/rif/RIF-CLPWD.html> (accessed November 21, 2013).
24. Van Gelder A., Ross K. A., Schlipf J. S. The well-founded semantics for general logic programs // J. ACM, 1991. Vol. 8. No. 3. P. 620–650.
25. RIF core answer set programming dialect / Eds. S. Heymans, M. Kifer. — 2009. Available at: <http://ruleml.org/rif/RIF-CASPD.html> (accessed November 21, 2013).
26. Gelfond M., Lifschitz V. The stable model semantics for logic programming // Logic programming. 5th Conference and Symposium Proceedings, 1988. P. 1070–1080.
27. RIF RDF and OWL compatibility. W3C Recommendation / Eds. J. De Bruijn, C. Welty. — 2nd ed., 2013. Available at: <http://www.w3.org/TR/rif-rdf-owl/> (accessed November 21, 2013).
28. OWL 2 Web ontology language structural specification and functional-style syntax. W3C recommendation / B. Motik, P. F. Patel-Schneider, B. Parsia. — 2nd ed., 2012. Available at: <http://www.w3.org/TR/owl2-syntax/> (accessed November 21, 2013).
29. Shvaiko P., Euzenat J. A survey of schema-based matching approaches // J. Data Semantics, 2005. Vol. IV. P. 146–171.
30. Abiteboul S., Bienvenu M., Galland A., et al. A rule-based language for Web data management // 30th ACM Symposium on Principles of Database Systems Proceedings. — ACM Press, 2011. P. 283–292.
31. Gelfond M. Answer sets. Handbook of knowledge representation. — Elsevier, 2008. P. 285–316.
32. Motik B., Rosati R. Closing semantic Web ontologies. University of Manchester Report, 2007. Available at: <http://www.cs.ox.ac.uk/people/boris.motik/pubs/mr06closing-report.pdf> (accessed November 21, 2013).
33. Chen P. P., Thalheim B., Wong L. Y. Future directions of conceptual modeling // Conceptual modeling / Eds. P. P. Chen, J. Akoka, B. Thalheim. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 1999. Vol. 1565. P. 287–301.
34. Costal D., Gmbóxomez C., Guizzardi G. Formal semantics and ontological analysis for understanding subsetting, Specialization and redefinition of associations in UML // Conceptual modeling — ER 2011 / Eds. M. Juesfeld, L. Delcambre, T. W. Ling. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6998. P. 189–203.
35. Martinez Y., Cachero C., Matera M., Abrahao S., Lujan S. Impact of MDE approaches on the maintainability of Web applications: An experimental evaluation // Conceptual Modeling — ER 2011 / Eds. M. Juesfeld, L. Delcambre, T. W. Ling. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6998. P. 233–246.
36. Castro L., Baião F., Guizzardi G. A semantic oriented method for conceptual data modeling in OntoUML based on linguistic concepts // Conceptual modeling — ER 2011 / Eds. M. Juesfeld, L. Delcambre, T. W. Ling. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6998. P. 486–494.
37. Fillotrani P. R., Franconi E., Tessaris S. The ICOM 3.0 intelligent conceptual modelling tool and methodology. Semantic Web. — IOS Press, 2011. P. 1–14. Available at: http://www.semantic-web-journal.net/sites/default/files/swj105_1.pdf (accessed November 21, 2013).
38. Maus C., Rybacki S., Uhrmache A. M. Rule-based multi-level modeling of cell biological systems // BMC Syst. Biol., 2011. Vol. 5. P. 166. Available at: <http://www.biomedcentral.com/content/pdf/1752-0509-5-166.pdf> (accessed November 21, 2013).

39. *Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Poggi A., Rosati R.* MASTRO-I: Efficient integration of relational data through DL ontologies // 2007 Workshop (International) on Description Logic (DL 2007) Proceedings. CEUR Workshop Proceedings, 2007. P. 250.
40. *Cosentino V., Del Fabro M. D., El Ghali A.* A model driven approach for bridging ILOG rule language and RIF // RuleML 2012 Proceedings. CEUR Workshop Proceedings, 2012. P. 874. Available at: <http://ceur-ws.org/Vol-874/paper9.pdf> (accessed November 21, 2013).
41. *Gonzalez-Moriyon G., Polo L., Berrueta D., Tejo-Alonso C.* Final steel industry public demonstrators. ONTORULE deliverable D5.5. — 2012. Available at: <http://ontorule-project.eu/outcomes%3Ffunc=fileinfo&id=94.html> (accessed November 21, 2013).
42. *Loo B. T., Condie T., Garofalakis M., Gay D. E., Hellerstein J. M., Maniatis P., Ramakrishnan R., Roscoe T., Stoica I.* Declarative networking: Language, execution and optimization // ACM SIGMOD Conference Proceedings, 2006. P. 97–108.
43. *Grumbach S., Wang F.* Netlog, a rule-based language for distributed programming // Practical aspects of declarative languages / Eds. M. Carro, R. Peña. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2010. Vol. 5937. P. 88–103.
44. *Alvaro P., Marczak W. R., Conway N., Hellerstein J. M., Maier D., Sears R.* Dedalus: Datalog in time and space // Datalog reloaded / Eds. O. de Moor, G. Gottlob, T. Furche, and A. Sellers. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2010. Vol. 6702. P. 262–281.
45. *Grossmann G., Thiagarajan R., Schrefl M., Stumptner M.* Conceptual modeling approaches for dynamic Web service composition // The evolution of conceptual modeling / Eds. R. Kaschek, L. M. L. Delcambre. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6520. P. 180–204.
46. *Henderson-Sellers B.* Random thoughts on multi-level conceptual modelling // The evolution of conceptual modeling / Eds. R. Kaschek, L. M. L. Delcambre. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6520. P. 93–116.
47. *Falbo R. A., Ruy F. B., Moro R. D.* Using ontologies to add semantics to a software engineering environment // SEKE 2005 Proceedings. — Curran Associates, 2005. P. 151–156.
48. Berre A., Missikoff M. (Moderators). Panel: Modeling for the future Internet // Conceptual modeling — ER 2011 / Eds. M. Jeusfeld, L. Delcambre, T. W. Ling. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6998. P. 526–527.
49. *Porto F., Spaccapietra S.* Data model for scientific models and hypotheses // The evolution of conceptual modeling / Eds. R. Kaschek, L. M. L. Delcambre. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2011. Vol. 6520. P. 285–305.
50. *Racunas S. A., Shah N. H., Albert I., Fedoroff N. V.* 2004. Hybrow: A prototype system for computer-aided hypothesis evaluation // Bioinformatics, 2004. Vol. 20. No. 1. P. 257–264.
51. *Sharpe W., Alexander G. J., Bailey J. W.* Investments. — Prentice Hall, 1998. 962 p.
52. *Gabbay D. M.* What is negation as failure // Logic Programs, norms and action / Eds. A. Artikis, R. Craven, N. K. Çiçekli, et al. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer-Verlag, 2012. Vol. 7360. P. 52–78.
53. *Брюхов Д. О., Вовченко А. Е., Захаров В. Н., Желенкова О. П., Калиниченко Л. А., Мартынов Д. О., Скворцов Н. А., Ступников С. А.* Архитектура промежуточного слоя предметных посредников для решения задач над множеством интегрируемых неоднородных распределенных информационных ресурсов в гибридной грид-инфраструктуре виртуальных обсерваторий // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 2–34.
54. ATL Project. 2013. Available at: <http://www.eclipse.org/atl/> (accessed November 21, 2013).

Поступила в редакцию 25.11.13

PROBABILISTIC METHODS FOR SELF-CORRECTING HARDWARE DESIGN*

S. Dolev¹, S. Frenkel^{2,3}, and D. E. Tamir⁴

Abstract: This paper presents several ways for extending the scope of program self-correction methods, based on the “random self-reducibility” property, to hardware design. The concept can be utilized for both analog and digital hardware-design. The extension is based on sampling, polynomial-interpolation, and error-correcting codes. In particular, the authors suggest using the well-known reconstruction of real-numerical functions for correcting faults remaining in analog and digital hardware, e. g., arithmetic logic units (ALU), after manufacturing testing. The present approach can complement the state-of-the-art technique of program self-correction by uniformly testing samples of operations and verifying the results of these samples.

Keywords: self-correcting; real function computation; data analysis; interpolation

DOI: 10.14357/19922264130413

1 Introduction

The reliability of computation in the presence of errors is an important research topic. In particular, it is crucial in the scope of Digital Signal Processors (DSP) based classification tasks and for embedded devices of safety-critical systems. Robust methods for identifying incoming waveforms, referred to as Modulation Classification [1] are some examples of this type of classification problems. The objective in these computation tasks is to minimize the error probability.

Since it is very difficult to test and detect all of the possible manufacturing faults in the stage of fabrication of modern hardware, there is an increasing interest in methods for self-correction [2]. The methods presented in the literature (cf. [2]), however, require knowledge related to the logical structure of the target design as well as sophisticated models to analyze their reliability.

In addition, the task of verifying the correctness of information processing devices such as microprocessors is a very challenging task since, typically, an exhaustive verification is an exponential function of the device complexity.

Generally, self-correction is based on simple estimation of the error probability using sampling rather than proving correctness or exhaustive evaluation. Sampling can be used to efficiently identify the probability of a given

“black-box” device to correctly compute the results for uniformly selected inputs.

Originally, the scope of self-correction has involved a program that computes functions-over-finite-fields, overcoming computation errors on a small fraction (ϵ) of their input [3, 4]. In this paper, however, it is shown that the random reducibility-based self-correction approach, originally suggested to *amplify* the reliability of programs [3, 4], can be used in the scope of nonfinite fields for self-correcting hardware.

One of the main contributions of the current paper is the introduction of novel ways to extend the software-based self-correction paradigm to digital and analog hardware.

The proposed approach can complement state-of-the-art techniques by uniformly testing samples of operations and verifying the results of these samples. Hence, it enables tolerating a small percentage of incorrect results due to manufacturing defects, thereby facilitating the use of nonperfect hardware. In operation mode when an operation op_i has to be executed, a uniformly chosen set of operations op_j, op_k that can imply the result of the operation in hand are executed in order to maintain a verified result.

It is essential to note that the domain/range of computed functions cannot be restricted to finite fields and,

*Extended abstract of this work was presented at the 13th International Conference on Applied Stochastic Models and Data Analysis (ASMDA), 2009. This research is partially supported by the Russian Foundation for Basic Research (Grant RFBR 12-07-00109) and by the Rita Altura Trust Chair in Computer Sciences, Lynne and William Frankel Center for Computer Science, Israel Science Foundation (Grant No. 428/11), Cabarnit Cyber Security MAGNET Consortium, Grant from the Institute for Future Defense Technologies Research named for the Medvedi of the Technion, and the Israeli Internet Association.

¹Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel

²Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation

³Moscow Institute of Radio, Electronics, and Automation “MIREA,” Moscow 119454, Russian Federation

⁴Department of Computer Science, Texas State University, San Marcos, TX 78666, USA

by using some techniques [4, 5], can include the real numbers. In addition, sampling, error correction codes, polynomial interpolation, and segmentation are used to increase the efficiency of self-correction for any given function over the real numbers.

The paper, which is an extended version of paper [6], is organized as follows. Section 2 provides the problem definition and surveys related research. Section 3 presents a methodology for increasing the computation accuracy by polynomial interpolation with error correction. Section 4 considers possible ways to reconstruct real functions using interpolation and section 5 presents a synopsis of approaches to possible implementation of self-correcting based hardware. Conclusions and proposals for future work are included in section 6.

2 Problem Analysis and Related Work

The following aspects of self-correcting computations are considered in this section: (i) incorrect function on a small fraction of the inputs; and (ii) sampling-based self-correcting.

2.1 Incorrect function on a small fraction of the inputs

Consider a hardware computation device, such as an ALU, designed to compute a function $f(x)$ of input values from the domain X ; and assume that the device produces incorrect results $f^*(x) \neq f(x)$ for a small fraction of X . That is, $f^*(x) \neq f(x)$ for $x \in X_C \subset X$, such that $|X_C| \ll |X|$ where $|A|$ denotes the rank of a set A . This is depicted in Fig. 1.

Generally, the correctness of general-purpose and application-specific microprocessors is verified by manufacturer-testing at production time and/or self-checking procedures which are based on online detection.

The detection of all possible permanent faults through testing at the manufacturing cycle, however, is not feasible; and self-checking covers only a small fraction of erroneous bits. Hence, it requires specific knowledge about the logical structure of the target design.

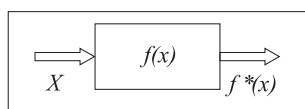


Figure 1 Incorrect results of the computation of the target function $f(x)$

The Floating Point divide instruction on the Pentium[®] processor is one of the well-known examples for this phenomenon. Despite more than 10 years of debugging and enhancements, the Pentium[®] processor Floating Point divide instructions have produced inaccurate results for a fraction of inputs [7].

Nevertheless, sampling can be used to efficiently identify the probability of a given device to correctly compute the results for inputs selected consistently according to a probability distribution such as uniform distribution. Indeed, sampling-based self-correction along with testing and self-checking is suggested in literature [3, 4].

2.2 Sampling-based self-correction

A function is *random self-reducible* of order k over a set D if its value at a given point can be efficiently reconstructed from its evaluation at random points [3, 4]. The reconstruction is possible if, and only if, there exists a function φ and a set of random functions $\sigma_1, \sigma_2, \dots, \sigma_n$ such that $f(x) = \varphi(x, f(\sigma_1(x, r)), \dots, f(\sigma_n(x, r)))$ for $x, r \in D$. This property allows reconstruction of the value of a function f using a finite number of elements taken from its domain without requiring any knowledge about the implementation of the device that implements the function, e.g., a hardware operational block or a program, which performs the computation. Note that polynomials of degree d over a finite field are random self-reducible using $d + 1$ random points [4].

Consider a function with no input/output domain restrictions. For example, these domains might include integer values, real numbers, vectors, etc. In order to use the *reliability amplification* technique utilizing the random self-reducible property [3, 8], it is necessary to provide a specific number of *batches* that yield sufficient probability for the majority of the batches to be correct; thereby, enabling using majority vote procedures for self-correction. In this context, the term *batch* denotes a *set of program input/output instances*. In other words, in the context of this paper, reliability amplification denotes the increase of correct computation probability due to computation reorganization, for example, using the fact that the functions are random self-reducible to reorganize the functions.

Let n be the number of batches and let p denote a fraction of the inputs for which the computations can be incorrect. The probability of correct computation can be calculated as the probability that the outputs obtained for more than $n/2$ of the batches are correct. This is given by:

$$\Pr \left(k \geq \left\lfloor \frac{n}{2} \right\rfloor + 1 \right) = 1 - \sum_{k=1}^L C_n^k p^k q^{n-k} \quad (1)$$

where k is the number of correct outputs; $p = 1 - q$; $L = \lfloor n/2 \rfloor + 1$ is the probability of correct computation for each batch; and C_n^k is the Binomial coefficient.

The reliability of the computations depends on the number of batches and on the choice of the reliability-parameter (or confidence level) r which is the probability of obtaining a majority of wrong results.

According to the Chernoff inequality, the required number of batches can be expressed as

$$n \geq \frac{1}{(p - 1/2)^2} \ln \left(\frac{1}{\sqrt{1 - r}} \right). \quad (2)$$

For example, if the function computed is a quadratic polynomial then $p = (1 - r)^3$ as each batch must include at least three input points (vectors). Equations (1) and (2) show that the use of majority-vote based choice among the results obtained from uniformly chosen batches can “amplify” the original reliability of devices if enough batches are used. Nevertheless, the minimum number of batches required for obtaining a correct computation results with a confidence level r might be very large, even when the device has a small probability of errors. For example, more than 10,000 batches are required in the case of quadratic polynomials for $\varepsilon = 0.2$ and $r = 0.05$, where ε is the small fraction the inputs and r is the reliability parameter. On the other hand, according to Eq. (1), the probability of correct computation with a reasonable number of batches is less than the original $\varepsilon = 0.2$. Hence, there is no amplification of the computation reliability, that is, using the Chernoff majority rule (1) will not lead to an increase in correct computations. Consequently, in this example, given r , it is impossible to provide correct computations on the basis of the majority vote rule, if the number of inputs is less than 10,000 batches, as the computations are erroneous on a fraction of inputs that is greater than ε .

A function over a group G is *linear* if it maps the group G to a group H so that $(x_1 \oplus x_2) = f(x_1) \otimes f(x_2)$ where \oplus and \otimes are the group operations. Integer multiplication and modular multiplication are some examples for such functions. From the point of view of the computation overhead, one benefit of the linearity is that given the values $f(x_1)$ and $f(x_2)$ and given that $x = x_1 \oplus x_2$, the function $f(x)$ can be computed as $f(x_1) \otimes f(x_2)$ which might be an easier computation task. An important aspect of the self-correction methods proposed in this paper is that the linearity properties of functions defined over finite fields can be utilized to increase the probability of success of a batch and, therefore, reduce the required number of batches. Error correcting codes can be used to obtain a better success rate for a batch result [8]. This approach, referred to as *batch self-corrector*, has been applied to the function $f(x) = x \bmod R$ over the positive integers domain [8]. In addition, it has been

used for self-testing, which is a part of the self-correction techniques. For example, Spielman suggested using the result of encoding functions defined over a finite field in order to increase the probability of correct computations of batches [9].

Assume that one is able to digitize (discretize) the input domain X for a set of integer or rational values [10], thereby transforming the given function to the domain/range of finite fields. Then, if the function is a polynomial, it becomes an integer function. Furthermore, if one is able to use the linearity properties to reduce the number of batches, then only the “improved” integer function has to be applied to the final real computation. This results in the following stages for computing $f(x)$:

- (1) digitize the original function to a finite field (cf. [10, 11]);
- (2) select a polynomial interpolation function over the finite field;
- (3) use error correction (e. g., the Berlekamp–Welch algorithm [11]) to correct a batch, leading to a reduction of the necessary number of batches according to the Chernoff-bound; and
- (4) reconstruct the digitized real number function from the discrete domain to the real domain using polynomial interpolation such as Taylor, Chebyshev, etc. Alternatively, this stage can be implemented by checking whether the original result of the device is close to the discrete value obtained.

3 Amplification by Polynomial Interpolation with Error Correction

Error correcting codes can be used to exploit linearity [10]. For example, the Reed–Solomon (RS) code of a polynomial over a finite field has linear properties [9]. Let (E, D) be an encoding-decoding pair for an error-correcting code of a code-word of length n with rate T (that is, a code that can correct T bits or symbols) for a polynomial function. For example, consider the Berlekamp–Welch algorithm, of RS codes polynomials computation in the presence of errors of interpolation over a finite field [11]. The polynomial P is unknown; and the only information about P is that it is of degree of l (say, $l = 2$). The polynomial E is unknown as well. Using the relationship defined above, one can produce a linear system whose solutions are the coefficients of P and E . This is shown in the following equations.

Let $Q(X) = aX^3 + bX^2 + cX + d = P(X)E(X)$ where a, b, c , and d are unknown coefficients. Substituting $P(X)$ by $R(X) = P(X)/Q(X)$, one obtains:

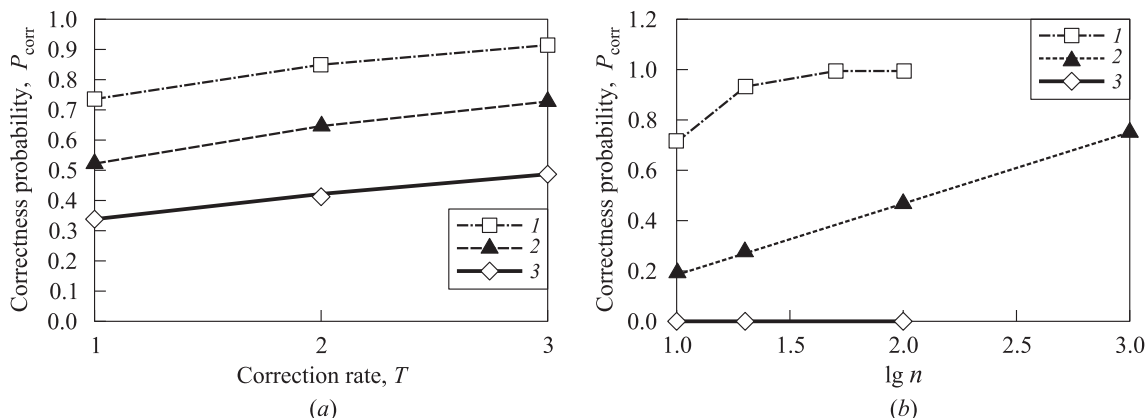


Figure 2 Correctness probability vs. correction rate (a) and vs. the logarithm of the number of batches (b) for different ALU error rates: 1 – 0.2; 2 – 0.3; and 3 – 0.4

$aX^3 + bX^2 + cX + d = R(X)E(X) = R(X)(X - e)$
 which can be rewritten as:

$$aX^3 + bX^2 + cX + d + R(X)e = R(X)X.$$

Now, one can substitute X by $\{0, 1, 2, 3, 4\}$ to obtain five linear equations in five unknowns. Solving this linear system for a, b, c, d , and e provides the polynomials $Q(X)$ and $E(X)$ which enable finding $P(X)$ by computing the quotient $Q(X)/E(X)$, and from P it is possible to recover the original (uncorrupted) values. In this case, the computation correctness probability can be defined as the probability that the number of incorrect symbols (from the specific finite field) is at most T . This probability is given by:

$$P_{\text{corr}} = \sum_{i=0}^T C_N^i \varepsilon^i (1 - \varepsilon)^{N-i} \quad (3)$$

where N is the degree of the polynomials, which is the number of points used for interpolation, referred to as the block size ($N = 2$ in the case of a quadratic interpolating polynomials), and ε corresponds to the probability of a symbol error. Obviously, the code distance is $T = (n - k)/2$, where k is the number of data symbols that has to be maintained. That is, for $T = 1$, a (5,3) code is obtained, $T = 2$ results in a (7,3) code, and $T = 3$ provides a (9,3) code, where the left value in a pair of numbers that describe a code is the code-word length, and the right value is the number of data symbols.

Figure 2a shows the batches correctness probability (P_{corr}) computed using Eq. (2) vs. the correction rate T (the x axis) for different device errors.

As seen in the figure, the probability of a correct result can be increased using the data correction encoding (e.g., the RS codes). Correspondingly, in some cases, due to this encoding, one can afford to work with

greater fraction of erroneously computed inputs. This means that amplification of the original probability (by using 3-symbols error correction) is possible even when the error probability of the device is approximately 0.3; moreover, such a correction is essential in the case that the error probability is around 0.4. Thus, if one deals with a discrete function (whose domain is a finite field) where RS-encoding can be used, it is possible to improve the probability of correct computation by repeating the computation using uniform random inputs (from this finite field), interpolating them, and choosing the result according to a majority vote rule.

Figure 2b presents an example of the computation of correctness probability vs. the logarithm (log base 10 is denoted as “lg”) of the number of batches for different ALU error rates with RS (7,3). This corresponds to RS encoding that corrects two symbols ($T = 2$), when the probability of a symbol error, which is equal in this instance to the erroneous fraction of inputs ε , is $P_s = 0.3$; and $P_{\text{corr}} = 0.647$. In this case, the amplification of the correct probability starts when 21 batches are used, a considerably lower number of batches than for the case of noncoded batches. In comparison, in the case of $\varepsilon = 0.3$, error correction is impossible for noncoded batches as $(1 - r)^3 < 1/2$ and is irrelevant for the majority-vote-based choice algorithm. One should take into account, however, the need for 4 extra points for each batch (7 instead of 3) to achieve this improvement.

Nevertheless, adding points to a batch can significantly reduce the required number of batches even when p is only slightly larger than $1/2$.

4 Real Function Reconstruction

The feasibility of implementation of the self-correcting algorithm proposed in this paper for an error correction

in analog hardware depends on existence of appropriate transformation of real signals input (numbers) to finite fields. As shown below, the state-of-the-art of analog-digital design allows finding proper solutions of this problem.

Note that the explicit reconstructing polynomials and rational functions over finite fields are presented by Sigal *et al.* in [3]. Sigal *et al.* use the fact that multiplication of any fixed element of finite fields by a random uniformly distributed element of the field gives a result that is uniformly distributed over the field. Therefore, in order to use the random self-reducibility-based approach to self-correction of real functions, one should consider discrete transformation of these functions to finite fields and commence with reconstruction of the functions.

In general, the reconstruction of real continuous function from digital data is governed by the *Nyquist sampling theorem* [8], which requires that a band-limited continuous function is sampled with a frequency equal to, or greater than, twice the maximum frequency of the signal. This digitization-reconstruction model, however, is not suitable in the context of self-correction, since the function is interpolated by algebraic polynomials. Moreover, the batches include randomly generated points. Hence, nonuniform discretization is required. This raises additional difficulties in the reconstruction [10]. In addition, the *quantization* of function values implies representation by a finite number of bits (say, n bits). Due to the finite precision representation of real numbers in computational devices; roundoff errors might occur during the calculations. The problem is finding the minimum accuracy necessary to ensure that the inverse quantization transformation that is a part of the digital-to-analog transformation can perform rounding and roundoff. This would make the function result equivalent to the rounding of the exact result (which could be obtained by the device) for all possible inputs. Since real-valued polynomial interpolations, say, Taylor polynomials, are defined over input variables given as real numbers, they cannot be used to express the finite bit-width limitations. Thus, one should coordinate the number of Taylor series terms and the number of bits in the aforementioned finite numbers representation. In order to resolve this problem, it is possible to use the technique presented in [5], where the coefficients of the series are expressed in terms of a finite number of bits referred to as fractional bits (FB). Several techniques for finding the necessary numbers of the Taylor series coefficients given a specific number of FBs are suggested in [5].

Another mean for increasing the accuracy is segmentation, which refers to dividing the input into subintervals, slices, or segments [12]. Generally, a set of coefficients of a low-degree polynomial can be used to evaluate

each segment, and the error probability is computed independently for each segment [12]. Note that the degree is an important parameter since a small degree enables correcting more errors using the Berlekamp–Welch algorithm. An evaluation of the obtained accuracy can be controlled by varying the number of segments and/or the polynomial degree. Using online segmentation requires predicting the interpolation error for each segment. For several differential functions, this error depends on the first $d + 1$ derivatives $f^{(d+1)}(x)$, where d is the degree of the interpolation polynomial, can be calculated during a preprocessing stage [13].

5 Synopsis

The computation of the value of $f(x)$ at a given point by evaluation at several random points can be implemented via interpolation of $f(x)$ using samples of x . For example, the input data of an ALU, which can be faulty with some known probability, can be considered as a set or series of batches, each of which is a series of k randomly generated arguments $r_{i,j}$ where $i = 1, \dots, m$ is the number of the ALU random inputs needed for interpolation; for example, $m = 2$ for the linear interpolation and $k = l + 1$ in the case of the interpolating polynomial of degree l . The variable $j = 1, \dots, n$ is the number of interpolations (number of batches of computations). The series quantity must provide a reliable choice of result of interpolation by majority in the sequence of results of interpolation obtained from the batches.

The model suggested in [3, 4] which uses uniform batches of random inputs is practical only in the case of relatively small error probability ε . Nevertheless, even in the relatively simple case of the quadratic polynomial, the batch correctness probability defined by the Chernoff-bound success probability of 0.512 for $\varepsilon = 0.2$ might imply a much smaller success probability for all polynomials of degrees $d > 2$.

Computation in a finite field is one possible way to increase the batches' correctness probability. This means operating with encoded data using coding methods such as RS codes and the Berlekamp–Welch algorithm [9]. This coding is an interpolation as it provides computation of polynomials in all the required points, using several points where the polynomial is known. In fact, the building of polynomials in the Berlekamp–Welch decoding algorithm is similar to Lagrange interpolation. Note that, in effect, different types of errors of computation (referred to as “ALU_errors”) are considered in the compared approaches. While the symbol (bit) error relates to a specific error rate T (say, $T = 3$), the error considered in self-correcting program theory (e. g., in [3]) is only a small fraction ε of erroneous computed inputs. This means that for the $(1 - \varepsilon)$ fraction of inputs

$\mu(f(x), f_C(x)) \leq \sigma$ where $f_C(x)$ is the function $f(x)$ computed by the ALU; $\mu(f(x), f_C(x))$ is the measure of distance between the exact value $f(x)$ and the computed function $f_C(x)$; and σ is the threshold error value. In general, σ corresponds to other erroneous quantities of bits.

As for DSP-based classification tasks, mentioned above as a prospective field of the self-correcting approach application, many current approaches use various polynomials to compute the classification characteristics, e. g., spline approximation in image processing. Therefore, in the event that the DSP computes the characteristics correctly on all but a small fraction ε of inputs, the algorithms mentioned above are suitable, and implementation of the computation schema, presented in sections 2–4, can essentially improve the classification reliability in comparison with the DSP characteristic ε .

Since the error correction can be interpreted as a “decoding of code-words,” one can borrow several ideas from Locally Decodable Codes (LDC) [14]. Locally decodable codes are the correcting codes where in order to retrieve the correct value of just one position of the input with high probability, it is sufficient to read a small number of positions of the corresponding possibly corrupted code-word. The locally decodable code can recover from a much higher error-rate [4]. One of the reasons for using LDC is that the previously used RS code consists of complete evaluations of polynomials of total degree up to d . In particular, there are LDCs which provide reduction of the error rate of the code with the number of queries which can be essentially higher than the polynomial degree d . Hence, the polynomial degree is not a limiting factor for the fraction of erroneous results reduction. In this context, the term *query* is a measure of complexity computed by the number of bits that need to be read from a corrupted code-word in order to recover a single bit of the encoded word [14]. It should be noted that LDCs are based on the classical Reed–Muller (RM) codes, which have rather simple and fast hardware implementation [15].

6 Concluding Remarks and Proposals for Future Research

In this paper, recent results in self-correcting computations have been presented. As the results show, in spite of essential reduction in the number of batches needed for suitable computation accuracy, this number might be rather significant. The complexity of the proposed approach depends on the number of batches as well as on the complexity of decoding the codes used for increasing correct computation probability for each batch. Feasible ways for improving the amplification have been pro-

posed and it has been demonstrated that these methods can minimize the number of batches of the computation function used to correct the computed value and provide a significant decrease in the error probability with the number of the batches used. Furthermore, a hardware implementation of this approach to self-correction can be derived from hardware implementation of the coding methods such as RS and RM [15, 16].

In the future, both the theory of random self-reducibility and new results in LDC will be explored for the problem of reconstruction of real numerical functions for correcting faults remaining in hardware after manufacturing testing. In addition, the authors plan to explore nonuniform sampling methods such as compressive sensing. Furthermore, a study of technical details of hardware implementation as well as DSP-based solutions will be performed.

References

1. Bilén, S., and A. Price. 2007. Modulation classification for radio interoperability via SDR. *SDR 07 Technical Conference and Product Exposition Proceedings*. <http://www.slideshare.net/kirill443/12-4-5647963> (accessed November 7, 2013).
2. Lala, P. 2000. *Self-checking and fault-tolerant digital design*. Morgan Kaufmann Publs. 400 p.
3. Sigal, A., R. Lipton, R. Rubinfeld, and M. Sudan. 1990. Reconstructing algebraic functions from mixed data. *33rd Annual Symposium on Foundations of Computer Science*. 503–512.
4. Gemmell, P., R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. 1991. Self-testing/correcting for polynomials and for approximate functions. *23rd Annual ACM Symposium on Theory of Computing Proceedings*. 32–34.
5. Lee, D. U., R. Cheung, W. Luk, and J. Villasenor. 2008. Hardware implementation trade-offs of polynomial approximations and interpolations. *IEEE Trans. Comput.* 57(5):686–701.
6. Dolev, Sh., and S. Frenkel. 2009. Extending the scope of self-correcting. *13th Conference (International) on Applied Stochastic Models and Data Analysis (ASMDA2009) Proceedings*. 458–462.
7. Nicely, T.R. Some results of computational research in prime numbers (Computational number theory). <http://www.trnicely.net/pentbug/pentbug.html> (accessed December 2010).
8. Rubinfeld, R. 1992. Batch checking with applications to linear functions. *Inform. Process. Lett.* 42:77–80.
9. Spielman, D. 1996. Highly fault-tolerant parallel computation. *37th IEEE Annual Symposium on Foundations of Computer Science Proceeding*. 154–163.
10. Oppenheim, A. V., R. W. Schaffer, and J. R. Buck. 1999. *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice-Hall. 871 p.

11. Berlekamp, E., and L. Welch. 1986. Error correction of algebraic block codes. U.S. Patent No. 4,633,470.
12. Tertinek, S., and C. Vogel. 2008. Reconstruction of nonuniformly sampled bandlimited signals using a differentiator-multiplier cascade. *IEEE Trans. Circuits Syst.* 55(8):2273–2286.
13. Pang, Y., and K. Radecka. 2008. Optimizing imprecise fixed-point arithmetic circuits specified by Taylor series through arithmetic transform. *Design Automation Conference DAC'08 Proceedings*. 397–402.
14. Yekhanin, S. 2011. Locally decodable codes. *Foundations Trends Theoretical Computer Sci.* 7(1):1–117.
15. Rahardja, S., and B.J. Falkowski. 2001. Efficient algorithm to calculate Reed–Muller expansions over GF(4). *IEE Proceedings — Circuits, Devices and Systems*. 148(6):289, 297.
16. Leroux, C., G. Le Mestre, C. Jego, P. Adde, and M. Jezequel. 2008. A 5-Gbps FPGA prototype of a $(31,29)^2$ Reed–Solomon turbo decoder. *5th Symposium (International) on Turbo Codes and Related Topics Proceedings*. 67–72.

Received October 23, 2013

Contributors

Dolev Shlomi (b. 1958) — professor, Doctor of Science in computer science, Dean of the Faculty of Natural Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; dolev@cs.bgu.ac.il

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, Russian Federation; associate professor, Moscow Institute of Radio, Electronics, and Automation (MIREA), Moscow 119454, Russian Federation; sfergei@mail.ru

Tamir Dan E. (b. 1955) — PhD-CS, associate professor in the Department of Computer Science, Texas State University, San Marcos, TX 78666, USA; dt19@txstate.edu

ВЕРОЯТНОСТНЫЙ ПОДХОД К САМОКОРРЕКТИРУЮЩИМ ВЫЧИСЛЕНИЯМ В ПРОЕКТИРОВАНИИ АППАРАТУРЫ

Ш. Долев¹, С. Френкель², Д. Е. Тамир³

¹Университет им. Бен-Гуриона в Негаве, Беэр-Шева, Израиль, dolev@cs.bgu.ac.il

²Институт проблем информатики Российской академии наук; Московский государственный технический университет радиотехники, электроники и автоматики (МГТУ МИРЭА), Москва, Россия, sfergei@mail.ru

³Университет Техаса, г. Сан-Маркос, США, dt19@txstate.edu

Аннотация: Описаны некоторые подходы к распространению метода самокоррекции программ, основанного на свойстве «случайной самосокращаемости» (random self-reducibility), на задачи проектирования аппаратной части вычислительных систем. Данная концепция может быть использована для проектирования как цифровой, так и аналоговой аппаратуры. Расширение метода основано на использовании случайных выборок, полиномиальной интерполяции и теории самокорректирующихся кодов. В частности, предлагается использовать известные методы реконструкции числовых функций для коррекции ошибок, вызываемых неисправностями, остающимися в аппаратуре после производственного контроля. Предлагаемый подход может дополнять известные методы тестирования цифровых и аналоговых приборов посредством использования равновероятной выборки операций и верификации результатов их выполнения, обеспечивая приемлемый уровень (небольшую долю) неправильных результатов.

Ключевые слова: самокоррекция; вычисление действительных функций; анализ данных; интерполяция

DOI: 10.14357/19922264130413

Расширенные тезисы данной статьи были представлены на 13-й Международной конференции по прикладным стохастическим моделям и анализу данных (ASMDA-2009). Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 12-07-00109), Фонда главы отделения информатики Риты Алтура, Центра вычислительной техники им. Линне и Уильяма Франкелей, Израильского научного фонда (грант № 428/11), Кабарнит кибербезопасности консорциума «Магнит», гранта Института перспективных оборонных технологий им. Медведи (Технион) и Израильской Интернет ассоциации.

Литература

1. *Bilén S., Price A.* Modulation classification for radio interoperability via SDR // SDR 07 Technical Conference and Product Exposition Proceedings, 2007. <http://www.slideshare.net/kirill443/12-4-5647963> (accessed November 7, 2013).
2. *Lala P.* Self-checking and fault-tolerant digital design. — Morgan Kaufmann Publ., 2000. 400 с.
3. *Sigal A., Lipton R., Rubinfeld R., Sudan M.* Reconstructing algebraic functions from mixed data // 33rd Annual Symposium on Foundations of Computer Science, 1990. P. 503–512.
4. *Gemmell P., Lipton R., Rubinfeld R., Sudan M., Wigderson A.* Self-testing/correcting for polynomials and for approximate functions // 23rd Annual ACM Symposium on Theory of Computing Proceedings, 1991. P. 32–34.
5. *Lee D. U., Cheung R., Luk W., Villasenor J.* Hardware implementation trade-offs of polynomial approximations and interpolations // IEEE Trans. Comput., 2008. Vol. 57. No. 5. P. 686–701.
6. *Dolev Sh., Frenkel S.* Extending the scope of self-correcting // 13th Conference (International) on Applied Stochastic Models and Data Analysis (ASMDA2009) Proceedings, 2009. С. 458–462.
7. *Nicely T. R.* Some results of computational research in prime numbers (Computational number theory). <http://www.trnicely.net/pentbug/pentbug.html> (last retrieved December 2010).
8. *Rubinfeld R.* Batch checking with applications to linear functions // Inform. Process. Lett., 1992. Vol. 42. P. 77–80.
9. *Spielman D.* Highly fault-tolerant parallel computation // 37th IEEE Annual Symposium on Foundations of Computer Science Proceeding, 1996. P. 154–163.
10. *Oppenheim A. V., Schaffer R. W., Buck J. R.* Discrete-time signal processing. — Upper Saddle River, NJ: Prentice-Hall, 1999. 871 p.
11. *Berlekamp E., Welch L.* Error correction of algebraic block codes. U.S. Patent No. 4,633,470, 1986.
12. *Tertinek S., Vogel C.* Reconstruction of nonuniformly sampled bandlimited signals using a differentiator-multiplier cascade // IEEE Trans. Circuits Syst., 2008. Vol. 55. No. 8. P. 2273–2286.
13. *Pang Y., Radecka K.* Optimizing imprecise fixed-point arithmetic circuits specified by Taylor series through arithmetic transform // Design Automation Conference DAC'08 Proceedings, 2008. P. 397–402.
14. *Yekhanin S.* Locally decodable codes // Foundations Trends Theoretical Computer Sci., 2011. Vol. 7. Iss. 1. P. 1–117.
15. *Rahardja S., Falkowski B. J.* Efficient algorithm to calculate Reed–Muller expansions over GF(4) // IEE Proceedings — Circuits, Devices and Systems, 2001. Vol. 148. No. 6. P. 289, 297.
16. *Leroux C., Le Mestre G., Jego C., Adde P., Jezequel M.* A 5-Gbps FPGA prototype of a $(31,29)^2$ Reed–Solomon turbo decoder // 5th Symposium (International) on Turbo Codes and Related Topics Proceedings, 2008. P. 67–72.

Поступила в редакцию 23.10.13

Абрамова Екатерина Максимовна (р. 1989) — студентка кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Архипов Олег Петрович (р. 1948) — кандидат технических наук, директор Орловского филиала ИПИ РАН

Вовченко Алексей Евгеньевич (р. 1984) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Ворфоломеева Ольга Викторовна (р. 1987) — аспирант математического факультета Челябинского государственного университета

Гайдамака Юлия Васильевна (р. 1971) — кандидат физико-математических наук, доцент Российского университета дружбы народов

Голованов Сергей Алексеевич (р. 1950) — доктор медицинских наук, заведующий лабораторией НИИ урологии

Григорьева Мария Евгеньевна (р. 1986) — биостатистик II компании Paraxel International, Москва

Долев Шломи (р. 1958) — профессор, доктор наук по информатике, декан факультета естественных наук Университета им. Бен-Гуриона в Негеве, Беэр-Шева, Израиль

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) — доктор технических наук, заведующий отделом ИПИ РАН

Зыкова Зоя Петровна (р. 1953) — кандидат физико-математических наук, заведующая лабораторией Орловского филиала ИПИ РАН

Калиниченко Леонид Андреевич (р. 1937) — доктор физико-математических наук, профессор факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; заведующий лабораторией

Композиционных методов и средств построения информационных систем ИПИ РАН

Ковалев Дмитрий Юрьевич (р. 1988) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; программист ИПИ РАН

Коновалов Михаил Григорьевич (р. 1950) — доктор 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

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник ИПИ РАН

Кузнецова Мария Георгиевна (р. 1991) — студентка кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Печинкин Александр Владимирович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник ИПИ РАН, профессор Российского университета дружбы народов

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН

Савченко Павел Анатольевич (р. 1967) — инженер-программист ИПИ РАН

Садовой Иван Андреевич (р. 1990) — студент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Самуйлов Андрей Константинович (р. 1988) — аспирант Российского университета дружбы народов

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой Российского университета дружбы народов

Сивков Андрей Владимирович (р. 1957) — доктор медицинских наук, заместитель директора НИИ урологии

Соколов Игорь Анатольевич (р. 1954) — академик (действительный член) Российской академии наук, доктор технических наук, директор ИПИ РАН

Сопин Эдуард Сергеевич (р. 1987) — аспирант Российского университета дружбы народов

Ступников Сергей Александрович (р. 1978) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Сучков Александр Павлович (р. 1954) — доктор технических наук, ведущий научный сотрудник ИПИ РАН

Тамир Дан Е. (р. 1955) — PhD в области информатики, доцент (associate professor) отделения информатики Университета Техаса, г. Сан-Маркос, США

Тырсин Александр Николаевич (р. 1961) — доктор технических наук, ведущий научный сотрудник Научно-инженерного центра «Надежность и ресурс

больших систем и машин» УрО РАН, г. Екатеринбург

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, старший научный сотрудник ИПИ РАН; доцент Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА)

Хавансков Валерий Александрович (р. 1950) — научный сотрудник ИПИ РАН

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора ИПИ РАН

Шубников Сергей Константинович (р. 1955) — старший научный сотрудник ИПИ РАН

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2013 г.

	Выпуск	Стр.
Абрамова Е. М. см. Захарова Т. В.		
Андреев А. М., Березкин Д. В., Козлов И. А., Симаков К. В. Подход к автоматизированному контролю работы системы извлечения данных с веб-сайтов.....	3	2
Архипов О. П., Зыкова З. П. Оптимизация функций Lab-контрастного градиционного преобразования.....	4	44
Беляев М. Г., Бурнаев Е. В. Аппроксимация многомерных зависимостей на основе разложения по словарю параметрических функций.....	3	114
Бенинг В. Е., Галиева Н. К., Королев В. Ю. Асимптотические разложения для функций распределения статистик, построенных по выборкам случайного объема.....	2	75
Бенинг В. Е., Галиева Н. К., Королев В. Ю. Об оценках функций концентрации регулярных статистик, построенных по выборкам случайного объема.....	1	116
Березкин Д. В. см. Андреев А. М.		
Бородина А. В., Морозов Е. В. Об оценивании эффективной пропускной способности системы с регенеративным входным процессом.....	2	26
Бурнаев Е. В. см. Беляев М. Г.		
Вовченко А. см. Калиниченко Л.		
Ворфоломеева О. В. см. Тырсин А. Н.		
Гайдамака Ю. В., Печинкин А. В., Разумчик Р. В., Самуйлов А. К., Самуйлов К. Е., Соколов И. А., Сопин Э. С., Шоргин С. Я. Распределение времени выхода из множества состояний перегрузки в системе $M M 1\langle L, H \rangle\langle H, R \rangle$ с гистерезисным управлением нагрузкой.....	4	20
Галиева Н. К. см. Бенинг В. Е.		
Галиева Н. К. см. Бенинг В. Е.		
Галина И. В. см. Шарнин М. М.		
Глушановский А. В., Калёнов Н. Е. Проблемы сетевого доступа к научным журналам..	3	56
Голованов С. А. см. Кривенко М. П.		
Голованов С. А. см. Кривенко М. П.		
Горшенин А. К. см. Королев В. Ю.		
Григорьева М. Е., Королев В. Ю. О сходимости распределений случайных сумм к скошенным экспоненциально-степенным законам.....	4	66
Григорьева М. Е., Королев В. Ю., Соколов И. А. Предельная теорема для геометрических сумм независимых неодинаково распределенных случайных величин и ее применение к прогнозированию вероятности катастроф в неоднородных потоках экстремальных событий.....	4	11
Грушо А. А., Грушо Н. А., Тимонина Е. Е. Статистические методы определения запретов вероятностных мер на дискретных пространствах.....	1	54
Грушо Н. А. см. Грушо А. А.		
Гуда С. А. Операции над представлениями кусочно-квазиаффинных функций в виде деревьев.....	1	58
Долев Ш., Френкель С., Тамир Д. Е. Вероятностный подход к самокорректирующимся вычислениям в проектировании аппаратуры.....	4	140
Закс Л. М., Королев В. Ю. Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм.....	1	105
Закс Л. М. см. Королев В. Ю.		
Закс Л. М. см. Королев В. Ю.		
Зализняк А. А. см. Луазо С.		

	Выпуск	Стр.
Захарова Т. В., Абрамова Е. М. О методе прогнозирования и классификации для цензурированных данных	4	105
Зацман И. М., Хавансков В. А., Шубников С. К. Метод извлечения библиографической информации из полнотекстовых описаний изобретений	4	52
Зацман И. М. см. Луазо С.		
Зацман И. М. см. Минин В. А.		
Зейфман А. И. см. Королев В. Ю.		
Зыкова З. П. см. Архипов О. П.		
Каганов В. Ю., Королёв А. К., Крылов М. Н., Машечкин И. В., Петровский М. И. Методы активной аутентификации на основе анализа динамики работы пользователей с клавиатурой	3	40
Калёнов Н. Е. см. Глушановский А. В.		
Калиниченко Л., Ступников С., Вовченко А., Ковалев Д. Концептуальные декларативные спецификации и решение задач в областях с интенсивным использованием данных	4	112
Капнин А. В. см. Кузнецов Л. А.		
Кириков И. А., Колесников А. В., Листопад С. В. Моделирование систем поддержки принятия решений синергетическим искусственным интеллектом	3	62
Клеменков П. А. Построение новостного рекомендательного сервиса реального времени с использованием NoSQL СУБД	3	14
Ковалев Д. см. Калиниченко Л.		
Ковалёв С. П. Семантика аспектно-ориентированного моделирования данных и процессов	3	70
Кожунова О. С. Когнитивная интероперабельность экспертного взаимодействия в задаче обработки русско-французских параллельных текстов: лингвокогнитивные аспекты	3	81
Козеренко Е. Б. Стратегии выравнивания параллельных текстов: семантические аспекты	1	82
Козеренко Е. Б. см. Шарнин М. М.		
Козлов И. А. см. Андреев А. М.		
Колесников А. В. см. Кириков И. А.		
Коновалов М. Г. Об одной задаче оптимального управления нагрузкой на сервер	4	34
Кореньков В. В., Нечаевский А. В., Трофимов В. В. Разработка имитационной модели сбора и обработки данных экспериментов на ускорительном комплексе НИКА	3	94
Королев В. Ю., Закс Л. М. Оценки скорости сходимости распределений некоторых случайных сумм к устойчивым законам	3	102
Королев В. Ю., Закс Л. М., Зейфман А. И. О сходимости случайных блужданий, порожденных обобщенными процессами Кокса, к процессам Леви	2	84
Королев В. Ю., Смелянский Р. Л., Смелянский Т. Р., Шалимов А. В. Статистическая проверка неисполняемости фрагментов кода последовательной программы	2	50
Королев В. Ю., Черток А. В., Корчагин А. Ю., Горшенин А. К. Вероятностно-статистическое моделирование информационных потоков в сложных финансовых системах на основе высокочастотных данных	1	12
Королев В. Ю. см. Бенинг В. Е.		
Королев В. Ю. см. Бенинг В. Е.		
Королев В. Ю. см. Григорьева М. Е.		
Королев В. Ю. см. Григорьева М. Е.		
Королев В. Ю. см. Закс Л. М.		
Королев В. Ю. см. Соколов И. А.		
Королёв А. К. см. Каганов В. Ю.		
Корчагин А. Ю. см. Королев В. Ю.		
Кривенко М. П., Голованов С. А., Савченко П. А., Сивков А. В., Сучков А. П. Информационно-аналитическая автоматизированная система «Мегалит» в оптимизации диагностики и лечения мочекаменной болезни	4	82

	Выпуск	Стр.
Кривенко М. П., Голованов С. А., Сивков А. В. Анализ однородности данных о химическом составе камней при уrolитиазе	4	94
Кружков М. Г. см. Минин В. А.		
Крылов М. Н. см. Каганов В. Ю.		
Кудрявцев А. А., Соколов И. А., Шоргин С. Я. Байесовская рекуррентная модель роста надежности: равномерное распределение параметров	2	55
Кузнецов И. П. см. Шарнин М. М.		
Кузнецов Л. А., Кузнецова В. Ф. Оценка семантической адекватности текстов информационным методом	1	94
Кузнецов Л. А., Кузнецова В. Ф., Капнин А. В. Универсальный метрический тезаурус русского языка	3	106
Кузнецова В. Ф. см. Кузнецов Л. А.		
Кузнецова В. Ф. см. Кузнецов Л. А.		
Кузнецова М. Г. см. Шестаков О. В.		
Купалов-Ярополк И. К., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Методы оценки эффективности и директивных сроков выполнения ресурсоемких вычислительных заданий	2	17
Листопад С. В. см. Кириков И. А.		
Луазо С., Сичинава Д. В., Зализняк А. А., Зацман И. М. Информационные технологии создания баз данных эквивалентных глагольных форм в русско-французском поливариантном параллельном корпусе	2	100
Лукашенко О. В., Морозов Е. В. О сходимости в пространстве L_p максимума процесса нагрузки для одного класса гауссовских систем обслуживания	1	36
Малашенко Ю. Е. см. Купалов-Ярополк И. К.		
Машечкин И. В. см. Каганов В. Ю.		
Милованова Т. А., Печинкин А. В. Стационарные характеристики системы обслуживания с инверсионным порядком обслуживания, вероятностным приоритетом и гистерезисной политикой	1	22
Минин В. А., Зацман И. М., Кружков М. Г., Норекиян Т. П. Методологические основы создания информационных систем для вычисления индикаторов тематических взаимосвязей науки и технологий	1	70
Морозов Е. В. см. Бородина А. В.		
Морозов Е. В. см. Лукашенко О. В.		
Морозова Ю. И. Построение семантических векторных пространств различных предметных областей	1	90
Морозова Ю. И. см. Шарнин М. М.		
Назарова И. А. см. Купалов-Ярополк И. К.		
Нечаевский А. В. см. Кореньков В. В.		
Норекиян Т. П. см. Минин В. А.		
Петровский М. И. см. Каганов В. Ю.		
Печинкин А. В. см. Гайдамака Ю. В.		
Печинкин А. В. см. Милованова Т. А.		
Разумчик Р. В. Стационарное распределение времени ожидания в системе обслуживания с отрицательными заявками, бункером для вытесненных заявок, различными интенсивностями обслуживания при дисциплине FIRST–FIFO–FIFO	2	34
Разумчик Р. В. см. Гайдамака Ю. В.		
Ронжин А. Ф. см. Купалов-Ярополк И. К.		
Рудой Г. И., Стрижов В. В. Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных	1	44
Савченко П. А. см. Кривенко М. П.		
Садовой И. А. см. Шестаков О. В.		
Самуйлов А. К. см. Гайдамака Ю. В.		
Самуйлов К. Е. см. Гайдамака Ю. В.		

	Выпуск	Стр.
Сивков А. В. см. Кривенко М. П.		
Сивков А. В. см. Кривенко М. П.		
Симаков К. В. см. Андреев А. М.		
Синицын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с разрывными характеристиками	1	3
Синицын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях	2	4
Сичинава Д. В. см. Луазо С.		
Смелянский Р. Л. см. Королев В. Ю.		
Смелянский Т. Р. см. Королев В. Ю.		
Соколов И. А., Королев В. Ю. Предисловие	2	3
Соколов И. А. см. Гайдамака Ю. В.		
Соколов И. А. см. Григорьева М. Е.		
Соколов И. А. см. Кудрявцев А. А.		
Сомин Н. В. см. Шарнин М. М.		
Сопин Э. С. см. Гайдамака Ю. В.		
Стрижов В. В. см. Рудой Г. И.		
Ступников С. А. Верифицируемое отображение модели данных, основанной на многомерных массивах, в объектную модель данных	3	22
Ступников С. А. см. Калиниченко Л.		
Сучков А. П. см. Кривенко М. П.		
Тамир Д. Е. см. Долев Ш.		
Тимонина Е. Е. см. Грушо А. А.		
Тимофеев В. С., Щеколдин В. Ю., Тимофеева А. Ю. Идентификация зависимостей признаков стохастической природы на основе регрессии Деминга	2	60
Тимофеева А. Ю. см. Тимофеев В. С.		
Трофимов В. В. см. Кореньков В. В.		
Тырсин А. Н., Ворфоломеева О. В. Исследование динамики многомерных стохастических систем на основе энтропийного моделирования	4	3
Френкель С. см. Долев Ш.		
Хавансков В. А. см. Зацман И. М.		
Хапланов А. Ю. Асимптотическая нормальность оценки параметров многомерной логистической регрессии	2	69
Черток А. В. см. Королев В. Ю.		
Шалимов А. В. см. Королев В. Ю.		
Шарнин М. М., Сомин Н. В., Кузнецов И. П., Морозова Ю. И., Галина И. В., Козеренко Е. Б. Статистические механизмы формирования ассоциативных портретов предметных областей на основе естественно-языковых текстов больших объемов для систем извлечения знаний	2	92
Шевцова И. Г. Об абсолютных константах в неравенстве Берри—Эссеена и его структурных и неравномерных уточнениях	1	124
Шестаков О. В. Центральная предельная теорема для функции обобщенной кросс-валидации при пороговой обработке вейвлет-коэффициентов	2	40
Шестаков О. В., Кузнецова М. Г., Садовой И. А. Обращение сферического преобразования Радона в классе дискретных случайных функций	4	75
Шкотин А. В. Исследование графа категорий английской версии Википедии	3	35
Шоргин С. Я. см. Гайдамака Ю. В.		
Шоргин С. Я. см. Кудрявцев А. А.		
Шубников С. К. см. Зацман И. М.		
Щеколдин В. Ю. см. Тимофеев В. С.		

2013 AUTHOR INDEX

	Issue	Page
Abramova E. M. see Zakharova T. V.		
Andreev A., Berezkin D., Kozlov I., and Simakov K. Unsupervised Approach to Web Wrapper Maintenance	3	2
Arkhipov O. P. and Zykova Z. P. Functions Optimization of Lab-Contrast Graded Transformation	4	44
Belyaev M. G. and Burnaev E. V. Approximation of a Multidimensional Dependency Based on Linear Expansion in a Dictionary of Parametric Functions.	3	114
Bening V. E., Galieva N. K., and Korolev V. Yu. Asymptotic Expansions for the Distribution Functions of Statistics Constructed from Samples with Random Sizes	2	75
Bening V. E., Galieva N. K., and Korolev V. Yu. On Bounds for the Concentration Functions of Regular Statistics Constructed from Samples with Random Sizes	1	116
Berezkin D. see Andreev A.		
Borodina A. V. and Morozov E. V. On Estimation of the Effective Bandwidths in a System with Regenerative Input	2	26
Burnaev E. V. see Belyaev M. G.		
Charnine M. M., Somin N. V., Kuznetsov I. P., Morozova Yu. I., Galina I. V., and Kozerenko E. B. Statistical Mechanisms of the Subject Domains Associative Portraits Formation on the Basis of Big Natural Language Texts for the Systems of Knowledge Extraction	2	92
Chertok A. V. see Korolev V. Yu.		
Dolev S., Frenkel S., and Tamir D. E. Probabilistic Methods for Self-Correcting Hardware Design	4	140
Frenkel S. see Dolev S.		
Gaidamaka Yu. V., Pechinkin A. V., Razumchik R. V., Samuylov A. K., Samouylov K. E., Sokolov I. A., Sopin E. S., and Shorgin S. Ya. The Distribution of the Return Time from the Set of Overload States to the Set of Normal Load States in a System $M M 1 \langle L, H \rangle \langle H, R \rangle$ with hysteretic load control	4	20
Galieva N. K. see Bening V. E.		
Galieva N. K. see Bening V. E.		
Galina I. V. see Charnine M. M.		
Glushanovskii A. V. and Kalenov N. E. Problems of the Online Access to Scientific Journals	3	56
Golovanov S. A. see Krivenko M. P.		
Golovanov S. A. see Krivenko M. P.		
Gorshenin A. K. see Korolev V. Yu.		
Grigor'eva M. E. and Korolev V. Yu. On Convergence of the Distributions of Random Sums to Skew Exponential Power Laws	4	66
Grigor'eva M. E., Korolev V. Yu., and Sokolov I. A. A Limit Theorem for Geometric Sums of Independent Nonidentically Distributed Random Variables and Its Application to the Prediction of the Probabilities of Catastrophes in Nonhomogeneous Flows of Extremal Events	4	11
Grusho A. A., Grusho N. A., and Timonina E. E. Statistical Techniques of Bans Determination of Probability Measures in Discrete Spaces	1	54
Grusho N. A. see Grusho A. A.		
Guda S. A. Operations on the Tree Representations of Piecewise Quasi-Affine Functions	1	58
Havanskov V. A. see Zatsman I. M.		
Kaganov V. Yu., Korolyov A. K., Krylov M. N., Mashechkin I. V., and Petrovskiy M. I. Active Authentication Methods Using Keystroke Dynamics	3	40
Kalenov N. E. see Glushanovskii A. V.		

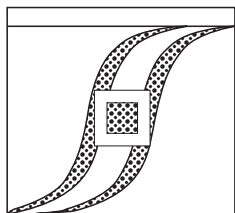
	Issue	Page
Kalinichenko L., Stupnikov S., Vovchenko A., and Kovalev D. Conceptual Declarative Problem Specification and Solving in Data Intensive Domains	4	112
Kapnin A. V. see Kuznetsov L. A.		
Khaplanov A. Yu. Asymptotic Normality of the Estimation of the Multivariate Logistic Regression	2	69
Kirikov I., Kolesnikov A., and Listopad S. Decision Support Systems Modeling with Synergetic Artificial Intelligence	3	62
Klemenkov P. A. Building Real-Time News Recommendation Service Using NoSQL DBMS	3	14
Kolesnikov A. see Kirikov I.		
Konovalov M. G. About One Task of Overload Control	4	34
Korchagin A. Yu. see Korolev V. Yu.		
Korenkov V. V., Nechaevskiy A. V., and Trofimov V. V. Data Acquisition Simulation for Nica Experiment	3	94
Korolev V. Yu., Chertok A. V., Korchagin A. Yu., and Gorshenin A. K. Probability and Statistical Modeling of Information Flows in Complex Financial Systems Based on High-Frequency Data	1	12
Korolev V. Yu., Smelyansky R. L., Smelyansky T. R., and Shalimov A. V. Statistical Testing for the Nonexecutability of Fragments of the Code of a Linear Program	2	50
Korolev V. Yu. and Zaks L. M. Estimates of the Rate of Convergence of the Distributions of Some Random Sums to Stable Laws	3	102
Korolev V. Yu., Zaks L. M., and Zeifman A. I. On Convergence of Random Walks Generated by Compound Cox Processes to Levy Processes	2	84
Korolev V. Yu. see Bening V. E.		
Korolev V. Yu. see Bening V. E.		
Korolev V. Yu. see Grigor'eva M. E.		
Korolev V. Yu. see Grigor'eva M. E.		
Korolev V. Yu. see Zaks L. M.		
Korolyov A. K. see Kaganov V. Yu.		
Kovalev D. see Kalinichenko L.		
Kovalyov S. P. Semantics of Aspect-Oriented Modeling of Data And Processes	3	70
Kozerenko E. B. Parallel Texts Alignment Strategies: The Semantic Aspects	1	82
Kozerenko E. B. see Charnine M. M.		
Kozhunova O. S. Cognitive Interoperability of Expert Collaboration in the Task of the Russian-French Parallel Texts Processing: Linguistic and Cognitive Aspects	3	81
Kozlov I. see Andreev A.		
Krivenko M. P., Golovanov S. A., and Sivkov A. V. Analysis of Data Homogeneity of the Chemical Compositions of Stones in Case of Urolithiasis	4	94
Krivenko M. P., Golovanov S. A., Savchenko P. A., Sivkov A. V., and Suchkov A. P. The Information-Analytical Computer System "MEGALITH" in Optimization of the Diagnosis and Treatment of Urolithiasis	4	82
Kruzhkov M. G. see Minin V. A.		
Krylov M. N. see Kaganov V. Yu.		
Kudriavtsev A. A., Sokolov I. A., and Shorgin S. Ya. Bayesian Recurrent Model of Reliability Growth: Uniform Distribution of Parameters	2	55
Kupalov-Yaropolk I. K., Malashenko Yu. E., Nazarova I. A., and Ronzhin A. F. Evaluation Methods for Efficiency and Directive Terms of Performance of Resource-Intensive Computing Tasks	2	17
Kuznetsov I. P. see Charnine M. M.		
Kuznetsov L. A. and Kuznetsova V. F. Information Method for Assessment Semantic Adequacy of Texts	1	94
Kuznetsov L. A., Kuznetsova V. F., and Kapnin A. V. Universal Metric Thesaurus of Russian Language	3	106
Kuznetsova M. G. see Shestakov O. V.		
Kuznetsova V. F. see Kuznetsov L. A.		

	Issue	Page
Kuznetsova V. F. see Kuznetsov L. A.		
Listopad S. see Kirikov I.		
Loiseau S., Sitchinava D. V., Zaluzniak A. A., and Zatsman I. M. Information Technologies for Creating the Database of Equivalent Verbal Forms in the Russian-French Multivariant Parallel Corpus.....	2	100
Lukashenko O. V. and Morozov E. V. On Convergence in the Space L_p of the Workload Maximum for a Class of Gaussian Queueing Systems.....	1	36
Malashenko Yu. E. see Kupalov-Yaropolk I. K.		
Mashechkin I. V. see Kaganov V. Yu.		
Milovanova T. A. and Pechinkin A. V. Stationary Characteristics of the Queueing System with LIFO Service, Probabilistic Priority, and Hysteric Policy.....	1	22
Minin V. A., Zatsman I. M., Kruzhkov M. G., and Norekyan T. P. Methodological Basis for the Creation of Information Systems for the Calculation of Indicators of Thematic Linkages Between Science and Technology.....	1	70
Morozov E. V. see Borodina A. V.		
Morozov E. V. see Lukashenko O. V.		
Morozova Yu. I. Semantic Vector Spaces for Different Knowledge Domains.....	1	90
Morozova Yu. I. see Charnine M. M.		
Nazarova I. A. see Kupalov-Yaropolk I. K.		
Nechaevskiy A. V. see Korenkov V. V.		
Norekyan T. P. see Minin V. A.		
Pechinkin A. V. see Gaidamaka Yu. V.		
Pechinkin A. V. see Milovanova T. A.		
Petrovskiy M. I. see Kaganov V. Yu.		
Razumchik R. V. Stationary Waiting Time Distribution in Queueing System with Negative Customers and Bunker for Ousted Customers Under FIRST–FIFO–FIFO Service Discipline	2	34
Razumchik R. V. see Gaidamaka Yu. V.		
Ronzhin A. F. see Kupalov-Yaropolk I. K.		
Rudoy G. I. and Strijov V. V. Algorithms for Inductive Generation of Superpositions for Approximation of Experimental Data.....	1	44
Sadovoy I. A. see Shestakov O. V.		
Samouylov K. E. see Gaidamaka Yu. V.		
Samuylov A. K. see Gaidamaka Yu. V.		
Savchenko P. A. see Krivenko M. P.		
Schekoldin V. Yu. see Timofeev V. S.		
Shalimov A. V. see Korolev V. Yu.		
Shestakov O. V. On the Rate of Convergence to the Normal Law of Risk Estimate for Wavelet Coefficients Thresholding When Using Robust Variance Estimates.....	2	40
Shestakov O. V., Kuznetsova M. G., and Sadovoy I. A. Inversion of Spherical Radon Transform in the Class of Discrete Random Functions.....	4	75
Shevtsova I. G. On the Absolute Constants in the Berry–Esseen Inequality and Its Structural and Nonuniform Improvements.....	1	124
Shkotin A. V. Study of the Wikipedia(En) Categories Graph.....	3	35
Shorgin S. Ya. see Gaidamaka Yu. V.		
Shorgin S. Ya. see Kudriavtsev A. A.		
Shubnikov S. K. see Zatsman I. M.		
Simakov K. see Andreev A.		
Sinitsyn I. N. Analytical Modeling of Invariant Measure Distributions in Stochastic Systems with Discontinuous Characteristics.....	1	3
Sinitsyn I. N. Parametrical Statistical and Analytical Modeling of Distributions in Nonlinear Stochastic Systems on Manifolds.....	2	4
Sitchinava D. V. see Loiseau S.		

	Issue	Page
Sivkov A. V. see Krivenko M. P.		
Sivkov A. V. see Krivenko M. P.		
Smelyansky R. L. see Korolev V. Yu.		
Smelyansky T. R. see Korolev V. Yu.		
Sokolov I. A. see Gaidamaka Yu. V.		
Sokolov I. A. see Grigor'eva M. E.		
Sokolov I. A. see Kudriavtsev A. A.		
Somin N. V. see Charnine M. M.		
Sopin E. S. see Gaidamaka Yu. V.		
Strijov V. V. see Rudoy G. I.		
Stupnikov S. A. A Verifiable Mapping of a Multidimensional Array Data Model into an Object Data Model	3	22
Stupnikov S. A. see Kalinichenko L.		
Suchkov A. P. see Krivenko M. P.		
Tamir D. E. see Dolev S.		
Timofeev V. S., Schekoldin V. Yu., and Timofeeva A. Yu. The Error-in-Variables Model Identification on the Basis of Deming's Approach	2	60
Timofeeva A. Yu. see Timofeev V. S.		
Timonina E. E. see Grusho A. A.		
Trofimov V. V. see Korenkov V. V.		
Tyrsin A. N. and Vorfolomeeva O. V. Study of the Dynamics of Multidimensional Stochastic Systems Based on Entropy Modeling	4	3
Vorfolomeeva O. V. see Tyrsin A. N.		
Vovchenko A. see Kalinichenko L.		
Zakharova T. V. and Abramova E. M. Prediction and Classification Method for Censored Data	4	105
Zaks L. M. and Korolev V. Yu. Variance-Generalized-Gamma-Distributions as Limit Laws for Random Sums	1	105
Zaks L. M. see Korolev V. Yu.		
Zaks L. M. see Korolev V. Yu.		
Zalizniak A. A. see Loiseau S.		
Zatsman I. M., Havanskov V. A., and Shubnikov S. K. Method of Bibliographic Information Extraction from Full-Text Descriptions of Inventions	4	52
Zatsman I. M. see Loiseau S.		
Zatsman I. M. see Minin V. A.		
Zeifman A. I. see Korolev V. Yu.		
Zykova Z. P. see Arkhipov O. P.		

ОБЪЯВЛЕНИЯ О КОНФЕРЕНЦИЯХ

<http://www.ipiran.ru/conference/stabil2014/>



XXXII Международный семинар по проблемам устойчивости стохастических моделей (XXXII International Seminar on Stability Problems for Stochastic Models)

16–24 июня 2014 г.

Норвежский университет науки и технологии
Тронхейм, Норвегия

XXXII Международный семинар по проблемам устойчивости стохастических моделей будет организован Норвежским университетом науки и технологии (Тронхейм) (НТНУ), МГУ им. М. В. Ломоносова и Институтом проблем информатики Российской академии наук (ИПИ РАН). Семинар будет проведен в НТНУ.

Проведение традиционного Международного семинара по проблемам устойчивости стохастических моделей имеет давнюю традицию, начинающуюся с 70-х гг. XX в. Основателем семинара был профессор В. М. Золотарев. Семинар проходил во многих странах; в XXI в. — в Венгрии, Болгарии, Испании, Латвии, Италии, Израиле, Румынии, Польше и России.

Главные темы семинара

- Предельные теоремы теории вероятностей
- Асимптотическая теория случайных процессов
- Устойчивые распределения и процессы
- Асимптотические методы математической статистики
- Теория риска
- Теория вероятностных метрик
- Характеризация вероятностных распределений
- Дискретные вероятностные модели
- Актуарная и финансовая математика
- Теория массового обслуживания и моделирование информационных систем

Международный программный и организационный комитет

В. М. Золотарев (Россия) — почетный председатель
В. Ю. Королев (Россия / МГУ им. М. В. Ломоносова, ИПИ РАН) — председатель
Н. Г. Ушаков (Норвегия) — зам. председателя
И. Г. Шевцова — (Россия / МГУ им. М. В. Ломоносова, ИПИ РАН) — генеральный секретарь
Ш. Баран (Венгрия)
В. Е. Бенинг (Россия / МГУ им. М. В. Ломоносова, ИПИ РАН)
А. В. Булинский (Россия / МГУ им. М. В. Ломоносова)
А. И. Зейфман (Россия / Вологодский ГУ, ИПИ РАН)
И. Мисевич (Польша)
Ю. С. Нефедова (Россия / МГУ им. М. В. Ломоносова, ИПИ РАН)

Э. Омей (Бельгия)
Д. Пап (Венгрия)
Ю. С. Хохлов (Россия / РУДН)
С. Я. Шоргин (Россия / ИПИ РАН)

Важные даты

1 декабря 2013 г. — начало регистрации

1 марта 2014 г. — крайний срок подачи тезисов

15 марта 2014 г. — извещение о включении доклада в программу конференции

Публикации

Тезисы докладов XXXII Международного семинара по проблемам устойчивости стохастических моделей будут опубликованы к началу семинара. Избранные труды семинара будут в дальнейшем опубликованы в журналах “Journal of Mathematical Sciences” (издательство “Springer Science+Business Media,” ISSN: 1072-3374, индексируется в системе Scopus), «Информатика и её применения» (издательство ТОРУС ПРЕСС, ISSN: 1992-2264) и «Системы и средства информатики» (издательство ТОРУС ПРЕСС, ISSN: 0869-6527).

<http://www.scs-europe.net/conf/ecms2014/index.html>



Специальная сессия
«Вероятностные и статистические методы
математического и имитационного
моделирования информационных систем
высокой производительности»

Май 2014 г.

Брешия, Италия

В рамках 28-й Европейской конференции по математическому и имитационному моделированию (28th European Conference on Modelling and Simulation — ECMS 2014), которая состоится в Брешии (Италия) с 27 по 30 мая 2014 г., будет проведена специальная сессия «Вероятностные и статистические методы математического и имитационного моделирования информационных систем высокой производительности» (Probability and Statistical Methods for Modelling and Simulation of High Performance Information Systems), организуемая с участием Института проблем информатики Российской академии наук (ИПИ РАН) и Российского университета дружбы народов (РУДН).

Сессия проводится во второй раз. В 2013 г. она была проведена в рамках конференции ECMS 2013 в Олесунне (Норвегия).

Сопредседатели сессии

А. И. Зейфман (Вологодский государственный университет, ИПИ РАН)
П. О. Абаев (РУДН)
Р. В. Разумчик (ИПИ РАН, РУДН)

Члены программного комитета

А. А. Грушо (ИПИ РАН, МГУ им. М. В. Ломоносова)
В. Ю. Королев (МГУ им. М. В. Ломоносова, ИПИ РАН)
А. В. Печинкин (ИПИ РАН, РУДН)
К. Е. Самуйлов (РУДН)
С. Я. Шоргин (ИПИ РАН)

Тематика специальной сессии

- Математические и имитационные модели систем массового обслуживания
- Моделирование и анализ производительности информационных и телекоммуникационных систем
- Моделирование и анализ информационных потоков
- Математическое и имитационное моделирование перегрузок и управления потоками
- Оценка эффективности, надежности и устойчивости информационных систем высокой производительности

Труды конференции публикуются в изданиях, индексируемых в системах **Web of Sciences** и **Scopus**.

Важные даты

13 февраля 2014 г. — представление полного доклада

14 марта 2014 г. — извещение о принятии доклада

9 апреля 2014 г. — представление окончательной версии доклада, регистрация и оплата взносов

СОБЫТИЯ

О МЕЖДУНАРОДНЫХ КОНФЕРЕНЦИЯХ ПО ОСНОВАМ ИНФОРМАЦИОННОЙ НАУКИ, ПРОВЕДЕННЫХ В 2013 ГОДУ

С 21 по 23 мая 2013 г. в Москве при поддержке Президиума РАН, Отделения математических наук и Отделения нанотехнологий и информационных технологий РАН, а также Московского гуманитарного университета и Института проблем информатики РАН состоялась **Пятая Международная конференция по фундаментальным основам информационной науки**. Это — традиционная конференция, проводящаяся под эгидой Международного общества информационных исследований (The International Society for Information Studies), центральный офис которого находится в Вене (Австрия). Конференция по данной тематике проведена в России впервые. Предыдущие конференции состоялись в Мадриде (1994), Вене (1997), Париже (2005) и Пекине (2010). Они оказали существенное влияние на развитие фундаментальных основ информационной науки в развитых странах и содействовали созданию в них исследовательских институтов и научных центров. На конференции в Москве были заслушаны 18 пленарных докладов ведущих ученых из Австрии, Испании, Китая, России и Франции, в которых были рассмотрены:

- философские и научно-методологические проблемы развития информационной науки как комплексного междисциплинарного научного направления, а также ее место в системе науки и образования;
- основные результаты исследований в области фундаментальных основ информатики и опыт их использования в науке, образовании и социально-экономическом развитии общества;
- перспективные направления дальнейшего развития фундаментальных основ информацион-

ной науки, а также использования ее концепций и методов в интересах развития науки, образования, решения глобальных проблем развития цивилизации;

- предложения по развитию международной кооперации ученых в области исследования фундаментальных основ информатики и их практическому использованию.

Основные материалы конференции представлены на сайте Московского гуманитарного университета, на базе которого проводилась эта конференция (см. <http://mosgu.ru/nauchnaya/conference/2013/FIS/>).

В период с 18 по 21 октября 2013 г. в древней столице Китая городе Сиань состоялась **Первая Международная конференция по философии информации (ISPI-2013)**. Ее инициаторами и основными организаторами стали Международный центр философии информации (в составе Сианьского транспортного университета) и Международное общество информационных исследований. В конференции участвовало более 80 ученых из различных стран мира, включая Россию. Тезисы отобранных Программным комитетом докладов опубликованы в материалах конференции, которые доступны по адресу http://is4is.unileon.es/index.php?option=com_content&view=article&id=81%3A2013-c1-start&catid=83%3Aconference&Itemid=59&lang=en.

В мае 2015 г. в Вене (Австрия) планируется провести **Шестую Международную конференцию по фундаментальным основам информационной науки**. Параллельно с ней там же будет проводиться **Вторая Международная конференция по философии информации (ISPI-2015)**.

К. К. Колин
д.т.н., проф., главный научный сотрудник
Института проблем информатики РАН

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Рекомендуется для составления списка “References” пользоваться размещенной на сайте <http://translit.ru/> (опция BGN) бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaistvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’—kompozit” [Experimental study of the strength of joints “steel—composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publ. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

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 (Сейфуль-Мулюков Рустем Бадриевич)

<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
 - mathematical methods for studying complex systems and processes;
 - information systems and networks;
 - information technologies; and
 - architecture and software of computational complexes and networks.
1. The Journal publishes original articles which have not been published before and are not intended for publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:
 - title;
 - author’s name and surname;
 - affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
 - data on authors according to the format: (see site)
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
 - abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;

- indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.

Important! Keywords must not be sentences.

9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.ru>, option BGN).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Article from the continuing publication (collection of works, proceedings):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy "stal'-kompozit" [Experimental study of the strength of joints "steel-composite"]. *Trudy MGTU "Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem"* [Bauman MSTU "Mathematical Modeling of Complex Technical Systems" Proceedings]. 593:125–130.

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publ. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author's name in the "Subject" field;
- an article and additional materials have to be attached using the "attach" function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. "Informatics and Applications" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

IPI RAN, Vavilova Str., 44, block 2, Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)

<http://www.ipiran.ru/english/journal.asp>