

Информатика и её применения

Том 13 Выпуск 2 Год 2019

СОДЕРЖАНИЕ

Доказательство унимодальности целевой функции в задаче порогового управления нагрузкой на сервер Я. М. Агаларов, М. Г. Коновалов	2
О развитии концепции условно-минимаксной нелинейной фильтрации: модифицированный фильтр и его анализ А. В. Босов, Г. Б. Миллер	7
Свойства вейвлет-оценок сигналов, регистрируемых в случайные моменты времени О. В. Шестаков	16
Архитектурные решения в задаче выявления мошенничества при анализе информационных потоков в цифровой экономике А. А. Грушо, М. И. Забежайло, Н. А. Грушо, Е. Е. Тимонина	22
О теоретико-автоматных моделях блокчейн-среды В. С. Анашин	29
Об одном методе решения систем квадратичных булевых уравнений, использующем локальные аффинности булевых функций О. А. Логачев, А. А. Сукаев, С. Н. Федоров	37
Применение методов поддержки принятия решений для многокритериальной задачи отбора многомасштабных композиций К. К. Абгарян, В. А. Осипова	47
Исследование проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением и случайной задержкой поставки П. В. Шнурков, Н. А. Вахтанов	54
Определение релевантности параметров нейросети А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов	62
Байесовские модели баланса факторов, имеющих априорные распределения Вейбулла и Накагами Е. Н. Арутюнов, А. А. Кудрявцев, А. И. Титова	71
Протокол гетерогенного мышления гибридной интеллектуальной многоагентной системы для решения проблемы восстановления распределительной электросети А. В. Колесников, С. В. Листопад	76
Сочетаемость логико-семантических отношений: количественные методы анализа О. Ю. Инькова, М. Г. Кружков	83
Моделирование процессов сетевого планирования портфеля проектов с неоднородными ресурсами в условиях нечеткой информации А. А. Зацаринный, В. В. Коротков, М. Г. Матвеев	92
О генезисе информационного общества: информатико-кибернетическое модельное представление С. Н. Гринченко	100
A Gaussian approximation of the distributed computing process О. V. Lukashenko, E. V. Morozov, and M. Pagano	109
Virtual experiments in data intensive research D. Y. Kovalev and E. A. Tarasov	117
Об авторах	126
Правила подготовки рукописей	128
Requirements for manuscripts	131

ДОКАЗАТЕЛЬСТВО УНИМОДАЛЬНОСТИ ЦЕЛЕВОЙ ФУНКЦИИ В ЗАДАЧЕ ПОРОГОВОГО УПРАВЛЕНИЯ НАГРУЗКОЙ НА СЕРВЕР*

Я. М. Агаларов¹, М. Г. Коновалов²

Аннотация: Рассматривается задача ограничения нагрузки в системе $M/M/N/\infty$ с помощью простой пороговой стратегии. Процесс обслуживания характеризуется наличием дедлайна для времени выполнения заданий. Другая особенность постановки задачи заключается в системе доходов и штрафов, которые получает система в зависимости от качества обслуживания. Качество управления оценивается в терминах предельного среднего дохода, а оптимальным считается значение порога, которое максимизирует эту величину. Нахождение оптимального порога существенно облегчается, когда целевая функция имеет единственный максимум. Результаты экспериментов свидетельствуют об унимодальности целевой функции для широкого класса входных потоков. Однако строгое доказательство этого факта отсутствует, и в статье этот пробел восполняется для пуассоновской нагрузки. При доказательстве используются результаты теории марковских цепей и теории массового обслуживания.

Ключевые слова: цепи Маркова; система $M/M/N/\infty$; ограничение нагрузки; пороговое управление; дедлайн

DOI: 10.14357/19922264190201

1 Введение

Эта заметка продолжает и дополняет работы [1, 2], в которых на простых моделях сервера с фиксированным числом мест обслуживания и потенциально бесконечной очередью изучались стратегии ограничения входной нагрузки. Основным результатом [1] заключался в том, что в многоканальной системе с дедлайном и экспоненциальным временем обслуживания оптимальная в смысле предельного среднего дохода стратегия ограничения нагрузки является простой пороговой стратегией и при этом утверждение справедливо для достаточно широкого класса входных потоков. Отсюда естественным образом возникает проблема отыскания оптимального значения порога, которая в [1] изучалась с помощью имитационного моделирования. Все результаты численных экспериментов показали, что целевая функция вогнута и имеет единственный максимум по значению порога. В [2] доказана унимодальность целевой функции для случая одноканальной системы без дедлайна. В данной статье приведено формальное доказательство аналогичного результата для многоканальной системы с пуассоновским входным потоком.

Стратегии порогового типа возникают во многих задачах управления и имеют большое значение, в частности, для решения проблемы контроля над

перегрузками в объектах, трактуемых как системы обслуживания. Привлекательность таких стратегий заключается в сочетании простоты реализации с удовлетворительным качеством, которое они обеспечивают даже в сравнении с более изощренными алгоритмами. Материал по тематике порогового управления весьма обширен и труднообозрим. В части, касающейся контроля над перегрузками, полезно опираться на фундаментальную обзорную статью [3]. Применительно к содержанию конкретной работы отметим статью [4].

В разд. 2 приводятся строгое описание постановки задачи и формулировка результата об унимодальности предельного среднего дохода в многоканальной системе с пороговым ограничением нагрузки, пуассоновским входом, дедлайном и системой доходов и штрафов за качество обслуживания. Раздел 3 содержит доказательство теоремы.

2 Описание модели и результат

Рассматривается сервер, имеющий $N < \infty$ обслуживающих мест (процессоров) для выполнения заданий, поступающих извне в виде рекуррентного потока. В каждый момент появления очередной заявки должно быть принято одно из двух возможных решений: задание может быть либо оставлено

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00692 и 19-07-00739).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, agglar@yandex.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, mkonovalov@ipiran.ru

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 неограниченной. Продвижение в очереди осуществляется по принципу «первый пришел – первый обслуживается». Время выполнения задания на любом процессоре является случайным и имеющим экспоненциальное распределение с параметром μ . Общее время пребывания задания в системе не должно превышать заданный предельный срок δ (так называемый дедлайн).

Процесс выполнения заданий сопровождается получением доходов и штрафов по следующей схеме. Если поступившее задание принято и если время его выполнения не превысило дедлайн, то система получает доход, равный $C_0 > 0$. Если поступившее задание принято, но время его выполнения превысило величину δ , то система платит штраф $C_1 > 0$. Наконец, если поступившая заявка отклонена, то штраф составляет $C_2 > 0$.

Обозначим через $\tau_n, n = 1, 2, \dots$, последовательные моменты поступления заявок, а через v_n – общее число заданий, находящихся на сервере в момент τ_n . Предположим, что управление поступающими заявками осуществляется согласно (простой) пороговой стратегии с порогом $h \geq N$. Таким образом, задание, поступившее в момент τ_n , принимается на обслуживание, если $v_n < h$, и отклоняется, если $v_n = h$.

Пусть $g_n^{(h)}$ – средний доход, связанный с заявкой, поступившей в момент τ_n . Определим предельный средний доход, усредненный по числу поступивших заявок, как

$$w(h) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=1}^T g_n^{(h)}.$$

Теорема. Если входной поток пуассоновский, то функция $w(h)$ имеет единственный максимум по h .

3 Доказательство теоремы

При сделанных предположениях относительно процесса обслуживания и рекуррентном входящем потоке последовательность v_n образует марковскую цепь с множеством состояний $\{0, \dots, h\}$ и переходными вероятностями

$$p_{ij}^{(h)} = \begin{cases} q_{ij} & \text{при } j \leq i+1 \leq N-1; \\ r_{i+1-j} & \text{при } N \leq j \leq i+1 \leq h; \\ s_{ij} & \text{при } j < N \leq i+1 \leq h; \\ 0 & \text{при } j > i+1. \end{cases}$$

Здесь

$$\begin{aligned} q_{ij} &= \int_0^\infty \binom{i+1}{j} [1 - e^{-\mu t}]^{i+1-j} e^{-\mu t j} dF(t); \\ r_{i+1-j} &= \int_0^\infty \frac{(N\mu t)^{i+1-j}}{(i+1-j)!} e^{-N\mu t} dF(t); \\ s_{ij} &= \int_0^\infty \binom{N}{j} e^{-\mu t j} \times \\ &\times \left[\int_0^t \frac{(N\mu y)^{i-N}}{(i-N)!} (e^{-\mu y} - e^{-\mu t})^{N-j} N\mu dy \right] dF(t), \end{aligned}$$

где $F(t)$ – функция распределения между последовательными поступлениями заявок.

Стандартные рассуждения показывают, что стационарные вероятности состояний $\{\pi_i^{(h)}, 0 \leq i \leq h\}$ рассматриваемой марковской цепи удовлетворяют следующей системе уравнений равновесия:

$$\begin{aligned} \pi_0^{(h)} &= \sum_{i=0}^h \pi_i^{(h)} p_{i0}^{(h)}; \\ \pi_j^{(h)} &= \sum_{i=j-1}^h \pi_i^{(h)} p_{ij}^{(h)}, \quad j = 1, \dots, h-1; \\ \pi_h^{(h)} &= \pi_{h-1}^{(h)} p_{h-1,h}^{(h)} + \pi_h^{(h)} p_{h-1,h}^{(h)}; \\ \sum_{i=0}^h \pi_i^{(h)} &= 1, \quad \pi_i^{(h)} \geq 0, \quad i = 0, \dots, h. \end{aligned}$$

Решив данную систему, получим для $j = 0, \dots, h-1$ следующие соотношения:

$$\left. \begin{aligned} \pi_h^{(h)} &= \left(1 + \sum_{i=0}^{h-1} D_i^{(h)} \right)^{-1}; \\ \pi_j^{(h)} &= D_j^{(h)} \pi_h^{(h)} = D_j^{(h)} \left(1 + \sum_{i=0}^{h-1} D_i^{(h)} \right)^{-1}. \end{aligned} \right\} (1)$$

Здесь

$$\begin{aligned} D_h^{(h)} &= 1; \quad D_{h-1}^{(h)} = \frac{1-r_0}{r_0}; \\ D_{j-1}^{(h)} &= \frac{1}{r_0} \left(D_j^{(h)} (1-r_1) - \right. \\ &\left. - \sum_{i=j+1}^{h-1} D_i^{(h)} r_{i+1-j} - r_{h-j} \right), \quad N \leq j \leq h-1; \\ D_{N-2}^{(h)} &= \frac{1}{q_{N-2,N-1}} \left(D_{N-1}^{(h)} (1-r_0) - \right. \\ &\left. - \sum_{i=N}^{h-1} D_i^{(h)} s_{i,N-1} - s_{h,N-1} \right); \end{aligned}$$

$$D_{j-1}^{(h)} = \frac{1}{q_{j-1,j}} \left(D_j^{(h)}(1 - q_{jj}) - \sum_{i=j+1}^{N-1} D_i^{(h)} q_{ij} - \sum_{i=N}^{h-1} D_i^{(h)} s_{ij} - s_{hj} \right), \quad 1 \leq j < N - 1.$$

Среднее значение дохода, получаемого системой в состоянии i , равно

$$R_i^{(h)} = \begin{cases} C_0 \Gamma_{\mu,1}(\delta) - C_1 [1 - \Gamma_{\mu,1}(\delta)], & \text{если } 0 \leq i < N; \\ C_0 \Gamma_{N\mu,i+1}(\delta) - C_1 [1 - \Gamma_{N\mu,i+1}(\delta)], & \text{если } N \leq i < h; \\ -C_2, & \text{если } i = h, \end{cases}$$

где $\Gamma_{a,b}(\cdot)$ — функция гамма-распределения с параметрами a и b .

Представим последнюю формулу в виде:

$$R_i^{(h)} = \begin{cases} (C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1, & \text{если } 0 \leq i < N; \\ (C_0 + C_1) \Gamma_{N\mu,i+1}(\delta) - C_1, & \text{если } N \leq i < h; \\ -C_2, & \text{если } i = h. \end{cases} \quad (2)$$

Отсюда при $0 \leq i \leq h - 2$ получим, что

$$R_{i+1}^{(h)} = \begin{cases} R_i^{(h)}, & \text{если } 0 \leq i < N; \\ R_i^{(h)} - (C_0 + C_1) \frac{(N\mu\delta)^{i+1}}{(i+1)!} e^{-N\mu\delta}, & \text{если } N \leq i \leq h - 2. \end{cases} \quad (3)$$

Как следует из (1), справедливы также формулы:

$$\pi_{j+1}^{(h+1)} = A_{h+1} \pi_j^{(h)}, \quad j = N - 1, \dots, h. \quad (4)$$

Здесь

$$A_{h+1} = \frac{1 - B_{N-1}^{(h+1)}}{1 - B_{N-2}^{(h)}} < 1,$$

где

$$B_{N-1}^{(h+1)} = \sum_{i=0}^{N-1} \pi_i^{(h+1)};$$

$$B_{N-2}^{(h)} = \begin{cases} 0, & \text{если } N = 1; \\ \sum_{i=0}^{N-2} \pi_i^{(h)}, & \text{если } N \geq 2. \end{cases}$$

Обозначив

$$\Delta_{i+1}^{(h+1)} = R_{i+1}^{(h+1)} - R_i^{(h+1)}, \quad i = 1, \dots, h,$$

и применив формулы (4), находим:

$$\begin{aligned} w(h) - w(h+1) &= \sum_{i=0}^h \pi_i^{(h)} R_i^{(h)} - \sum_{i=0}^{h+1} \pi_i^{(h+1)} R_i^{(h+1)} = \\ &= [(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] \sum_{i=0}^{N-2} \pi_i^{(h)} + \sum_{i=N-1}^h \pi_i^{(h)} R_i^{(h)} - \\ &\quad - [(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] \sum_{i=0}^{N-1} \pi_i^{(h+1)} + \\ &\quad + \sum_{i=N}^{h+1} \pi_i^{(h+1)} R_i^{(h+1)} = \\ &= [(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] (B_{N-2}^{(h)} - B_{N-1}^{(h+1)}) + \\ &\quad + \sum_{i=N-1}^h \pi_i^{(h)} R_i^{(h)} - \\ &\quad - A_{h+1} \left[\sum_{i=N-1}^{h-1} \pi_i^{(h)} (R_i^{(h+1)} + \Delta_{i+1}^{(h+1)}) + \pi_h^{(h)} R_{h+1}^{(h+1)} \right]. \end{aligned}$$

Из (2) следует, что $R_i^{(h)} = R_i^{(h+1)}$ при $i = 0, \dots, h - 1$ и $R_h^{(h)} = R_{h+1}^{(h+1)}$. Продолжая преобразование с помощью (3), получим

$$\begin{aligned} w(h) - w(h+1) &= \\ &= [(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] (Q_{N-2}^{(h)} - Q_{N-1}^{(h+1)}) + \\ &\quad + (1 - A_{h+1}) \sum_{i=N-1}^h \pi_i^{(h)} R_i^{(h)} - A_{h+1} \sum_{i=N-1}^{h-1} \pi_i^{(h)} \Delta_{i+1}^{(h+1)} = \\ &= -[(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] \frac{B_{N-1}^{(h+1)} - B_{N-2}^{(h)}}{1 - B_{N-2}^{(h)}} + \\ &\quad + \left[[(C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_1] B_{N-2}^{(h)} + \right. \\ &\quad \left. + \sum_{i=N-1}^h \pi_i^{(h)} R_i^{(h)} \right] \frac{B_{N-1}^{(h+1)} - B_{N-2}^{(h)}}{1 - B_{N-2}^{(h)}} - \\ &\quad - A_{h+1} \sum_{i=N-1}^{h-1} \pi_i^{(h)} \Delta_{i+1}^{(h+1)} = \\ &= (1 - A_{h+1}) \left[w(h) - (C_0 + C_1) \Gamma_{\mu,1}(\delta) + C_1 + \right. \\ &\quad \left. + (C_0 + C_1) \frac{A_{h+1}}{1 - A_{h+1}} \sum_{i=N-1}^{h-1} \pi_i^{(h)} \frac{(N\mu\delta)^{i+1}}{(i+1)!} e^{-N\mu\delta} \right] = \\ &= (1 - A_{h+1}) \left[w(h) - (C_0 + C_1) \Gamma_{\mu,1}(\delta) + C_1 + \right. \\ &\quad \left. + (C_0 + C_1) \frac{e^{-N\mu\delta}}{1 - A_{h+1}} \sum_{i=N}^h \pi_i^{(h+1)} \frac{(N\mu\delta)^i}{i!} \right]. \end{aligned}$$

Таким образом, рассматриваемое равенство приводится к виду:

$$w(h) - w(h+1) = (1 - A_{h+1}) [w(h) - G(h)], \quad (5)$$

где

$$G(h) = (C_0 + C_1) \Gamma_{\mu,1}(\delta) - C_2 - (C_0 + C_1) \frac{e^{-N\mu\delta}}{1 - A_{h+1}} \sum_{i=N}^h \pi_i^{(h+1)} \frac{(N\mu\delta)^i}{i!}.$$

Докажем, что функция $G(h)$ убывает с возрастанием переменной $h \geq N$. Для этого достаточно показать, что функция $\varphi(h) = (1/(1 - A_{h+1})) \sum_{i=N}^h \pi_i^{(h+1)} (N\mu\delta)^i / (i!) -$ возрастающая по $h \geq N$. Начиная с этого места, воспользуемся тем, что по условию доказываемого утверждения

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

Применим известные формулы для величин $\pi_i^{(h)}$, которые имеют место для систем массового обслуживания типа $M/M/N/h$ [5, с. 123]:

$$\pi_i^{(h)} = \begin{cases} \pi_0^{(h)} \frac{\rho^i}{i!}, & \text{если } 0 \leq i \leq N-1; \\ \pi_0^{(h)} \frac{\rho^N}{N!} \left(\frac{\rho}{N}\right)^{i-N}, & \text{если } N \leq i \leq h, \end{cases}$$

где

$$\rho = \frac{\lambda}{\mu}; \quad \pi_0^{(h)} = \left(\sum_{j=0}^{N-1} \frac{\rho^j}{j!} + \frac{\rho^N}{N!} \sum_{j=0}^{h-N} \left(\frac{\rho}{N}\right)^j \right)^{-1}.$$

Имеем:

$$\varphi(h) = \psi(h) \frac{\rho^N}{N!} \sum_{i=N}^h \left(\frac{\rho}{N}\right)^{i-N} \frac{(N\mu\delta)^i}{i!},$$

где $\psi(h) = \pi_0^{(h+1)} / (1 - A_{h+1})$. Поскольку сумма в последнем выражении — возрастающая функция по h , то достаточно показать, что функция $\psi(h)$ не зависит от h . Но этот факт следует из того, что

$$\begin{aligned} \psi(h+1) - \psi(h+2) &= \sum_{j=0}^{h+1-N} \left(\frac{\rho}{N}\right)^j \left/ \left(\left[1 - \left(\frac{\rho}{N}\right)^{h+2-N} \right] \sum_{j=0}^{N-1} \frac{\rho^j}{j!} + \left(\frac{\rho^N}{N!}\right) \sum_{j=0}^{h+1-N} \left(\frac{\rho}{N}\right)^j \right) \right. - \sum_{j=0}^{h+2-N} \left(\frac{\rho}{N}\right)^j \left/ \left(\left[1 - \left(\frac{\rho}{N}\right)^{h+3-N} \right] \sum_{j=0}^{N-1} \frac{\rho^j}{j!} + \frac{\rho^N}{N!} \sum_{j=0}^{h+2-N} \left(\frac{\rho}{N}\right)^j \right) \right. = 0. \end{aligned}$$

Следовательно, функция $\varphi(h)$ возрастает, а функция $G(h)$, соответственно, убывает по h , если $h \geq N$.

Далее доказательство практически идентично рассуждениям, проведенным в теореме 1 из [2], в которых фигурирует аналог функции $w(h)$, представленной в виде (5), где $G(h)$ — невозрастающая функция, $0 < A_{h+1} < 1$. Эти рассуждения приводят к утверждению теоремы об унимодальности $w(h)$.

4 Заключение

В работе приведено доказательство существования и единственности оптимального значения порога в задаче управления нагрузкой на сервер с несколькими параллельными местами обслуживания и бесконечной очередью. Модель предполагает пуассоновскую нагрузку, экспоненциальное время обслуживания, возможность регулировать число заданий в системе, а также доходы или штрафы в зависимости от качества обслуживания. Целевой функцией выступает предельный средний доход.

Результат статьи частично восполняет пробел в теории, где свойства целевой функции в задачах порогового управления изучены недостаточно. Анализ доказательства показывает, что большая часть рассуждений справедлива также для произвольного рекуррентного входного потока. Это дает основание продолжить работу над получением аналогичного, более общего результата для системы $G/M/N/\infty$.

Литература

1. Коновалов М. Г. Об одной задаче оптимального управления нагрузкой на сервер // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 34–43.
2. Агаларов Я. М., Шоргин В. С. Об одной задаче максимизации дохода системы массового обслуживания типа $G/M/1$ с пороговым управлением очередью // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 55–64.
3. Adams R. Active queue management: A survey // IEEE Commun. Surv. Tut., 2013. Vol. 15. No. 3. P. 1425–1476.
4. Жерновский Ю. В. Решение задач оптимального синтеза для некоторых марковских моделей обслуживания // Информационные процессы, 2010. Т. 10. № 3. С. 257–274.
5. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.

Поступила в редакцию 20.02.19

PROOF OF THE UNIMODALITY OF THE OBJECTIVE FUNCTION IN $M/M/N$ QUEUE WITH THRESHOLD-BASED CONGESTION CONTROL

Ya. M. Agalarov and M. G. Konovalov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The problem of limiting the load in the system $M/M/N/\infty$ is considered using a simple threshold strategy. In addition to the service time, each task is characterized by a deadline. Depending on the quality of service, the system receives either a fixed income or a penalty. The quality of control is determined by the marginal average income and the threshold value that maximizes this value is considered as optimal. Usually, it is much easier to find the optimal threshold if the objective function has a single maximum. The experimental results show the unimodality of the objective function for a wide class of arrival flows. However, there is no rigorous proof of this fact and in the paper, this gap is filled up for the Poisson arrivals. The proof is based on the results of the Markov chain theory and queueing theory.

Keywords: Markov chains; $M/M/N/\infty$ system; congestion control; threshold control; deadline

DOI: 10.14357/19922264190201

Acknowledgments

The reported study was partly funded by the Russian Foundation for Basic Research according to the research projects No. 18-07-00692 and No. 19-07-00739.

References

1. Konovalov, M. G. 2013. Ob odnoy zadache optimal'nogo upravleniya nagruzkoy na server [About one task of overload control]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):34–43.
2. Agalarov, Ya. M., and V. S. Shorgin. 2017. Ob odnoy zadache maksimizatsii dokhoda sistemy massovogo obsluzhivaniya tipa $G/M/1$ s porogovym upravleniem ochered'yu [About the problem of profit maximization in $G/M/1$ queueing system with threshold control of the queue]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):55–64.
3. Adams, R. 2013. Active queue management: A survey. *IEEE Commun. Surv. Tut.* 15(3):1425–1476.
4. Zhernovyy, Ju. V. 2010. Reshenie zadach optimal'nogo sinteza dlya nekotorykh markovskikh modeley obsluzhivaniya [Solution of optimum synthesis problem for some Markov models of service]. *Information Processes* 10(3):257–274.
5. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.

Received February 20, 2019

Contributors

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agglar@yandex.ru

Konovalov Mikhail G. (b. 1950) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkonovalov@ipiran.ru

О РАЗВИТИИ КОНЦЕПЦИИ УСЛОВНО-МИНИМАКСНОЙ НЕЛИНЕЙНОЙ ФИЛЬТРАЦИИ: МОДИФИЦИРОВАННЫЙ ФИЛЬТР И ЕГО АНАЛИЗ*

А. В. Босов¹, Г. Б. Миллер²

Аннотация: Представлен новый субоптимальный фильтр, основанный на методе условно-минимаксной нелинейной фильтрации (УМНФ), используемом в задаче фильтрации состояний нелинейных стохастических систем наблюдения с дискретным временем. Идея модификации УМНФ состоит в отказе от априорного расчета параметров условно-минимаксного фильтра, выполняемого в оригинальном УМНФ методом имитационного моделирования, в пользу процедуры аппроксимации параметров фильтра, совмещенной с оценением текущего состояния. Методом Монте-Карло аппроксимируются условные моментные характеристики, определяющие параметры модифицированного УМНФ, в отличие от безусловных, вычисляемых в оригинальном фильтре. Дано минимаксное обоснование предлагаемого варианта аппроксимации, аналогичное базовой концепции УМНФ. Обсуждаются простые модельные эксперименты, подтверждающие работоспособность представленного алгоритма оценивания, его сравнение с исходным фильтром.

Ключевые слова: нелинейная стохастическая система наблюдения с дискретным временем; условно-минимаксная нелинейная фильтрация; метод Монте-Карло

DOI: 10.14357/19922264190202

1 Введение

Фильтрация, т. е. оценивание состояний динамических систем по косвенным наблюдениям в режиме реального времени, — ключевая задача теории стохастических систем. Фундаментальные результаты в этой области [1, 2] значительно обогатили теорию, составили основу многих развиваемых и в настоящее время направлений. При этом в утилитарном плане ценность уравнений оптимальной фильтрации невелика: решать их в практически значимых задачах, в условиях реального времени не удается даже с учетом колоссальных возможностей современных вычислительных средств.

Для практики значимыми являются адаптированные к условиям проведения реальных расчетов вычислительные процедуры. Такие алгоритмы обосновываются чаще всего некоторыми приближенными соотношениями. Соответствующий огромный пласт прикладных результатов в целом определяется понятием «субоптимальная фильтрация».

Альтернативой эвристическому подходу, характерному для этого направления, выступает теория условно-оптимальной фильтрации В. С. Пугачева [3, 4] и ее развитие в отношении динамических

систем с дискретным временем — теория УМНФ А. Р. Панкова [5, 6]. В настоящее время эти подходы продолжают развиваться (см., например, [7–10]), хотя имеют менее широкую область применения по сравнению с субоптимальными фильтрами, чего не заслуживают хотя бы по причине своей теоретической обоснованности.

В данной работе методика условно-оптимальной фильтрации развивается для модели нелинейной динамической системы наблюдения с дискретным временем, предлагается новая модификация оригинальных фильтров, обоснование структуры фильтра представлено в минимаксной форме УМНФ.

2 Задачи нелинейной фильтрации. Условно-оптимальные и условно-минимаксные фильтры

Рассматриваемая дискретная стохастическая система наблюдения описывается следующими рекуррентными соотношениями ($t = 1, 2, \dots$):

* Работа выполнена при частичной поддержке РФФИ (проект 19-07-00187-А).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ABosov@frccsc.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, GMiller@frccsc.ru

$$\left. \begin{aligned} x_t &= \varphi_t^{(1)}(x_{t-1}) + \varphi_t^{(2)}(x_{t-1})w_t, \quad x_0 = \eta; \\ y_t &= \psi_t^{(1)}(x_t) + \psi_t^{(2)}(x_t)v_t, \end{aligned} \right\} \quad (1)$$

где состояние $x_t \in \mathbb{R}^p$ и наблюдения $y_t \in \mathbb{R}^q$ — гильбертовы случайные последовательности (наличие вторых моментов обеспечивается, например, линейными ограничениями на скорость роста функций $\varphi_t^{(1)}, \varphi_t^{(2)}, \psi_t^{(1)}$ и $\psi_t^{(2)}$); возмущения $\{w_t\}$ и ошибки наблюдения $\{v_t\}$ — независимые дискретные белые шумы второго порядка, гильбертовым и независимым является и вектор начальных условий η . Соответствующие моменты первого и второго порядка обозначаются $m_w(t), D_w(t), m_v(t), D_v(t), m_\eta$ и D_η .

Модель (1) по форме аналогична дифференциальным уравнениям диффузионного типа, что позволяет интерпретировать ее как результат дискретизации соответствующего стохастического дифференциального уравнения, а коэффициенты $\varphi_t^{(1)}, \varphi_t^{(2)}, \psi_t^{(1)}$ и $\psi_t^{(2)}$ — как аналоги коэффициентов сноса и диффузии. Соответственно, линейные ограничения на скорость роста этих функций на бесконечности и наличие вторых моментов у возмущений можно рассматривать как дискретный аналог условий Ито существования решения, обеспечивающих процессам x_t и y_t существование вторых моментов. В этих условиях корректной является задача фильтрации, т. е. вычисления оценки \hat{x}_t состояния x_t по наблюдениям $y_\tau, \tau = 1, \dots, t$, характеризующейся среднеквадратичным критерием качества $E\{||x_t - \hat{x}_t||^2\}, E\{\cdot\}$ — математическое ожидание. Обозначив через $\mathfrak{F}_{y_1^t}$ σ -алгебру, порожденную наблюдениями $y_\tau, \tau = 1, \dots, t$, заметим, что предложенные условия обеспечивают существование как оптимального решения $\hat{x}_t = E\{x_t | \mathfrak{F}_{y_1^t}\}$ — условного математического ожидания x_t относительно $\mathfrak{F}_{y_1^t}$, если \hat{x}_t выбирается из всего класса $\mathfrak{F}_{y_1^t}$ -измеримых функций, так и различных вариантов его аппроксимации, сужающих этот класс, в частности условно-оптимальных оценок. Это класс оценок типа «прогноз–коррекция», настраиваемых на конкретную модель (1) с помощью выбираемых структурных функций $\xi_t = \xi_t(x)$ (базовый прогноз) и $\zeta_t = \zeta_t(x, y)$ (базовая коррекция), которые для формирования оценки подвергаются линейному преобразованию [3, 4]. Именно: пусть имеется оценка \hat{x}_{t-1} состояния x_{t-1} . Условно-оптимальный прогноз \tilde{x}_t ищется в виде:

$$\tilde{x}_t = F_t \xi_t + f_t, \quad \xi_t = \xi_t(\hat{x}_{t-1}), \quad (2)$$

а оценка \hat{x}_t состояния x_t в виде:

$$\hat{x}_t = \tilde{x}_t + H_t \zeta_t + h_t, \quad \zeta_t = \zeta_t(\tilde{x}_t, y_t). \quad (3)$$

Типовым вариантом базового прогноза ξ_t выступает прогноз «в силу системы» $\xi_t(x) = \varphi_t^{(1)}(x) + \varphi_t^{(2)}(x)m_w(t)$, базовой коррекции ζ_t — невязка $\zeta_t(x, y) = y - \psi_t^{(1)}(x) - \psi_t^{(2)}(x)m_v(t)$, но разнообразие подходов здесь ограничено только теми же условиями существования вторых моментов, что и в исходной модели. Отметим, что типовой вариант коррекции требует наличия вторых моментов у наблюдений y_t , что и предполагалось выше.

Коэффициенты линейных преобразований в (2) и (3) определяются из условий минимума $E\{||x_t - F_t \xi_t + f_t||^2\}$ для прогноза и $E\{||x_t - \tilde{x}_t - H_t \zeta_t - h_t||^2\}$ для коррекции и имеют вид:

$$\left. \begin{aligned} F_t &= \text{cov}(x_t, \xi_t) \text{cov}^+(\xi_t, \xi_t); \\ f_t &= E\{x_t\} - F_t E\{\xi_t\}; \\ H_t &= \text{cov}(x_t - \tilde{x}_t, \zeta_t) \text{cov}^+(\zeta_t, \zeta_t); \\ h_t &= -H_t E\{\zeta_t\}. \end{aligned} \right\} \quad (4)$$

Здесь и далее через $\text{cov}(x, y)$ обозначена ковариационная матрица x и y , через $+$ — операция псевдообращения матрицы по Муру–Пенроузу. Псевдообращение используется по той причине, что решение задач оптимизации в отношении F_t и/или H_t может быть не единственным, так что выбор будет сделан в пользу решения с минимальной нормой. Условно-оптимальный прогноз \tilde{x}_t и оценка фильтрации \hat{x}_t при этом являются несмещенными и обладают гарантированным качеством оценивания:

$$\left. \begin{aligned} \tilde{K}_t &= \text{cov}(x_t - \tilde{x}_t, x_t - \tilde{x}_t) = \\ &= \text{cov}(x_t, x_t) - F_t \text{cov}(\xi_t, x_t); \\ \hat{K}_t &= \text{cov}(x_t - \hat{x}_t, x_t - \hat{x}_t) = \\ &= \tilde{K}_t - H_t \text{cov}(\zeta_t, x_t - \tilde{x}_t). \end{aligned} \right\} \quad (5)$$

Изложенная концепция условно-оптимальной фильтрации дополняется двумя существенными моментами, которые составляют суть концепции УМНФ: минимаксным обоснованием структуры фильтра и методом расчета коэффициентов линейных преобразований.

Минимаксное обоснование обеспечивает следующую вспомогательная задача. Пусть для случайного вектора $z = \text{col}(x, y), x \in \mathbb{R}^p, y \in \mathbb{R}^q$, с известным математическим ожиданием $m_z = \text{col}(m_x, m_y)$ и ковариацией

$$D_z = \begin{pmatrix} D_x & D_{xy} \\ D_{yx} & D_y \end{pmatrix}$$

относительно его закона распределения \mathcal{F}_z известно, что $\mathcal{F}_z \in \Phi(m_z, D_z)$ — классу всех вероятност-

ных распределений со средним m_z и ковариацией D_z . Требуется найти оценку \hat{x} вектора x по наблюдениям y . При этом допустимыми оценками считаются любые измеримые функции $\theta(y)$, критерий — среднеквадратичный при условии заданного множества неопределенности Φ , т.е. $\mathcal{J}(\theta, \mathcal{F}_z) = E \{ \|x - \theta(y)\|^2 \}$. Решение такой задачи определяет седловая точка $(\theta^*, \mathcal{F}_z^*)$ [5, 6], а именно: распределение \mathcal{F}_z^* — гауссовское с параметрами m_z и D_z и оценка $\hat{x} = \theta^*(y) = D_{xy}D_y^+y + (m_x - D_{xy}D_y^+m_y)$ — наилучшая линейная оценка x по y . При этом $\mathcal{J}(\theta^*, \mathcal{F}_z^*) = D_x - D_{xy}D_y^+D_{yx}$. Это решение и составляет минимаксное обоснование УМНФ, т.е. линейные преобразования базовых прогнозов и коррекции в (2) и (3). Именно: оценку фильтрации \hat{x}_t состояния x_t в соответствии с концепцией УМНФ следует определять как результат решения следующих минимаксных задач:

$$\left. \begin{aligned} \tilde{x}_t &= \tilde{\theta}_t(\xi_t), \\ \tilde{\theta}_t &= \arg \min_{\tilde{\theta}_t} \max_{\mathcal{F}_z} E \left\{ \|x_t - \tilde{\theta}_t(\xi_t)\|^2 \right\}, \\ & z = \text{col}(x_t, \xi_t); \\ \hat{x}_t &= \tilde{x}_t + \hat{\theta}_t(\zeta_t), \\ \hat{\theta}_t &= \arg \min_{\hat{\theta}_t} \max_{\mathcal{F}_z} E \left\{ \|x_t - \tilde{x}_t - \hat{\theta}_t(\zeta_t)\|^2 \right\}, \\ & z = \text{col}(x_t - \tilde{x}_t, \zeta_t), \end{aligned} \right\} (6)$$

и эти решения задаются соотношениями (2)–(4) и обладают свойствами (5). Кроме того, наилучшими в обеих задачах распределениями оказываются гауссовские, параметры которых определяются параметрами вектора $\text{col}(x_t, \xi_t)$ в части прогнозирования и вектора $\text{col}(x_t - \tilde{x}_t, \zeta_t)$ в части коррекции.

Вторым положением, составляющим неперемный атрибут УМНФ, является метод расчета F_t , f_t , H_t и h_t путем имитационного моделирования. Условно-минимаксная нелинейная фильтрация получается заменой в соотношениях (4) математических ожиданий и ковариаций их оценками по методу Монте-Карло, т.е. в практической реализации УМНФ операции $E\{\cdot\}$ и $\text{cov}(\cdot, \cdot)$ заменяются соответствующими статистическими оценками, рассчитанными априорно по смоделированному пучку траекторий. Обозначим соответствующие операторы усреднения через $\overline{E}\{\cdot\}$ и $\overline{\text{cov}}(\cdot, \cdot)$. При реализации УМНФ в соотношениях (4), например, вместо $E\{x_t\}$ используется

$$\overline{E}\{x_t\} = \frac{1}{N} \sum_{i=1}^N x_t^i,$$

вместо $\text{cov}(\zeta_t, x_t - \tilde{x}_t)$ —

$$\overline{\text{cov}}(\zeta_t, x_t - \tilde{x}_t) = \frac{1}{N-1} \sum_{i=1}^N \zeta_t^i (x_t^i - \tilde{x}_t^i)^T,$$

где N — число смоделированных траекторий, а x_t^i , \tilde{x}_t^i и ζ_t^i , $i = 1, \dots, N$, — реализации.

3 Основной результат — модифицированный условно-минимаксный нелинейный фильтр

Концепция УМНФ позволяет получать реализуемые и высокоточные решения широкого класса задач нелинейной фильтрации. Основываясь на опыте применения, можно отметить, что для алгоритма УМНФ не критичны такие известные практические проблемы, как рост размерности, неточность в задании начальных условий, вырожденность возмущений. Вообще, условно-оптимальный подход представляется уникальным направлением с колоссальным практическим потенциалом, не имеющим в настоящее время ни аналогов, ни конкурентов. Вместе с тем возможности для его дальнейшего совершенствования имеются, и одна из них представлена далее. Источником для предлагаемой модификации служит, как принято называть, калмановская структура УМНФ. Объяснение такому термину в задаче нелинейной фильтрации, не имеющей общих черт с линейно-гауссовской постановкой классического фильтра Калмана, можно представить, записав самый простой из субоптимальных фильтров, предназначенных для решения задач нелинейной фильтрации, — расширенный фильтр Калмана (РФК). Соотношения для РФК применительно к оцениванию в модели (1) таковы:

$$\left. \begin{aligned} \tilde{x}_t &= \xi_t, \quad \xi_t = \xi_t(\hat{x}_{t-1}), \\ \xi_t(x) &= \varphi_t^{(1)}(x) + \varphi_t^{(2)}(x)m_w(t); \\ \tilde{K}_t &= f_t \tilde{K}_{t-1} f_t^T + \varphi_t^{(2)} D_w(t) \varphi_t^{(2)T}, \\ f_t &= \frac{\partial \varphi_t^{(1)}(x)}{\partial x} \Big|_{x = \tilde{x}_t}, \quad \varphi_t^{(2)} = \varphi_t^{(2)}(\tilde{x}_t); \\ \hat{x}_t &= \tilde{x}_t + H_t \zeta_t, \\ \zeta_t &= y_t - \psi_t^{(1)}(\tilde{x}_t) - \psi_t^{(2)}(\tilde{x}_t) m_v(t), \\ H_t &= \tilde{K}_t h_t^T \left(h_t \tilde{K}_t h_t^T + \psi_t^{(2)} D_v(t) \psi_t^{(2)T} \right)^+, \\ h_t &= \frac{\partial \psi_t^{(1)}(x)}{\partial x} \Big|_{x = \tilde{x}_t}, \quad \psi_t^{(2)} = \psi_t^{(2)}(\tilde{x}_t), \\ \hat{K}_t &= \tilde{K}_t - H_t h_t \tilde{K}_t. \end{aligned} \right\} (7)$$

Обратим внимание, что в (7) использованы те же обозначения для прогноза \tilde{x}_t и оценки \hat{x}_t , типового прогноза «в силу системы» ξ_t и невязки ζ_t , повторно использованы обозначения f_t , h_t , H_t , \tilde{K}_t и \hat{K}_t , поскольку в структуре фильтров, в выражениях для характеристик точности оценок эти коэффициенты играют идентичные роли. Но в выражениях для УМНФ фигурируют точные значения соответствующих моментов, заменяемых при реализации результатами имитационного моделирования, в выражениях для РФК используются аппроксимации, полученные путем линейризации около прогноза \tilde{x}_t . И это последнее обстоятельство заслуживает определенного внимания.

Действительно, коэффициенты f_t , h_t и H_t в РФК рассчитываются вместе с траекторией конкретной оценки по мере поступления наблюдений в отличие от аналогичных коэффициентов УМНФ, вычисляемых априорно. Значит, можно предполагать, что в этих коэффициентах используются оценки, аппроксимирующие не априорные, а условные моменты и даже потраекторный критерий оценивания $E \left\{ \|x_t - \hat{x}_t\|^2 \mid \mathfrak{J}_{y_1^t} \right\}$. Конечно, РФК не минимизирует этот критерий, но пытается аппроксимировать некоторую оценку, имеющую смысл в отношении данного критерия. Другое дело, что для такой аппроксимации используется не слишком работоспособный вариант линейризации, но в идее использования в фильтре оценок условных моментов, вычисляемых вместе с оценками состояния на каждом шаге фильтрации, и видится вариант усовершенствования УМНФ.

Для формального описания алгоритма модифицированной УМНФ (МУМНФ) потребуется манипулирование условными моментами. Так, для случайного вектора y будем обозначать через \mathfrak{J}_y порожденную им σ -алгебру, через $E \{x \mid \mathfrak{J}_y\}$ — условное математическое ожидание x относительно \mathfrak{J}_y , через $\text{cov} \{x, x \mid \mathfrak{J}_y\} = E \left\{ (x - E \{x \mid \mathfrak{J}_y\})^T (x - E \{x \mid \mathfrak{J}_y\}) \mid \mathfrak{J}_y \right\}$ — условную ковариацию. Говоря далее об условных распределениях, будем под этим понимать сужение вероятностной меры, заданной на исходном вероятностном пространстве, на подходящую σ -алгебру \mathfrak{J}_y .

Пусть теперь $(\hat{x}_{t-1}, \hat{K}_{t-1})$ — оценка МУМНФ, вычисленная в момент $t - 1$. Пусть, кроме того, существует σ -алгебра $\tilde{\mathfrak{J}}_{y_1^{t-1}}$, являющаяся σ -подалгеброй $\mathfrak{J}_{y_1^{t-1}}$, такая что условное распределение x_{t-1} относительно $\tilde{\mathfrak{J}}_{y_1^{t-1}}$ обладает следующими свойствами:

$$E \{x_{t-1} \mid \tilde{\mathfrak{J}}_{y_1^{t-1}}\} = \hat{x}_{t-1};$$

$$\text{cov} \left(x_{t-1} - \hat{x}_{t-1}, x_{t-1} - \hat{x}_{t-1} \mid \tilde{\mathfrak{J}}_{y_1^{t-1}} \right) = \hat{K}_{t-1}.$$

Отметим, что форма распределения при этом не важна, а обосновывать выбор оценщика, как в оригинальном УМНФ, будет некоторая минимаксная задача.

Прогноз \tilde{x}_t будем искать в виде:

$$\tilde{x}_t = E \left\{ x_t \mid \tilde{\mathfrak{J}}_{y_1^{t-1}} \right\}. \quad (8)$$

Точность этого прогноза очевидным образом характеризуется выражением:

$$\begin{aligned} \tilde{K}_t &= \text{cov} \left(x_t - \tilde{x}_t, x_t - \tilde{x}_t \mid \tilde{\mathfrak{J}}_{y_1^{t-1}} \right) = \\ &= \text{cov} \left(x_t, x_t \mid \tilde{\mathfrak{J}}_{y_1^{t-1}} \right). \end{aligned} \quad (9)$$

Далее будем предполагать, что существует σ -алгебра $\hat{\mathfrak{J}}_{y_1^{t-1}}$, являющаяся σ -подалгеброй $\mathfrak{J}_{y_1^{t-1}}$, такая, что условное распределение x_t относительно $\hat{\mathfrak{J}}_{y_1^{t-1}}$ обладает следующими свойствами:

$$E \left\{ x_t - \tilde{x}_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right\} = 0;$$

$$\text{cov} \left(x_t - \tilde{x}_t, x_t - \tilde{x}_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right) = \tilde{K}_t,$$

где \tilde{x}_t определено в (8), а \tilde{K}_t — в (9). Для заданной корректирующей структурной функции $\zeta_t = \zeta_t(x, y)$ оценку \hat{x}_t будем искать в виде:

$$\hat{x}_t = \tilde{x}_t + H_t \zeta_t + h_t, \quad (10)$$

где

$$\begin{aligned} \zeta_t &= \zeta_t(\tilde{x}_t, y_t); \\ H_t &= \text{cov} \left(x_t - \tilde{x}_t, \zeta_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right) \text{cov}^+ \left(\zeta_t, \zeta_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right); \\ h_t &= -H_t E \left\{ \zeta_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right\}. \end{aligned}$$

Точность оценки МУМНФ определяется выражением

$$\begin{aligned} \hat{K}_t &= \text{cov} \left(x_t - \hat{x}_t, x_t - \hat{x}_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right) = \\ &= \tilde{K}_t - H_t \text{cov} \left(\zeta_t, x_t - \tilde{x}_t \mid \hat{\mathfrak{J}}_{y_1^{t-1}} \right). \end{aligned} \quad (11)$$

Приведенные соотношения, по сути, в точности повторяют выражения для оценки УМНФ в момент $t = 1$ с единственным изменением — вместо безусловных моментов фигурируют условные относительно некоторых σ -алгебр, ограничивающих объем доступной для оценивания информации. Обоснование всей процедуры оценивания дается следующими минимаксными задачами, вполне аналогичными задачам (6):

$$\left. \begin{aligned} \tilde{x}_t &= \tilde{\theta}_t, \\ \tilde{\theta}_t &= \arg \min_{\tilde{\theta}_t} \max_{\mathcal{F}_z} E \left\{ \|x_t - \tilde{\theta}_t\|^2 \middle| \tilde{\mathcal{J}}_{y_1^{t-1}} \right\}, \\ \hat{x}_t &= \tilde{x}_t + \hat{\theta}_t(\zeta_t), \\ \hat{\theta}_t &= \arg \min_{\hat{\theta}_t} \max_{\mathcal{F}_z} E \left\{ \|x_t - \tilde{x}_t - \hat{\theta}_t(\zeta_t)\|^2 \middle| \hat{\mathcal{J}}_{y_1^{t-1}} \right\}, \\ & z = \text{col}(x_t - \tilde{x}_t, \zeta_t). \end{aligned} \right\} (12)$$

Заметим, что решение (12) дает та же вспомогательная задача из разд. 2, поскольку отличие постановок, состоящее в замене безусловных моментов условными, никак не влияет на решение. Это верно и для первой задачи, вырожденной в том смысле, что у $\tilde{\theta}_t$ нет аргумента, и составляющей, таким образом, просто свойство условного математического ожидания. А главное, что этот же результат обосновывает переход от оценки в момент $t - 1$ к прогнозу и далее к коррекции на основе гауссовского распределения, являющегося наилучшим, т. е. минимаксно обоснованным (12).

Принципиальное же отличие МУМНФ состоит в отсутствии точной априорной информации о его точности. Вместо этого приходится использовать некие потраекторные характеристики \tilde{K}_t и \hat{K}_t , которые, будучи привязанными к условному распределению, не дают однозначного понимания качества оценки фильтрации в смысле исходного среднеквадратичного критерия.

В связи с предложенной модификацией следует коснуться двух вопросов.

Во-первых, это технически важный вопрос о σ -алгебрах $\tilde{\mathcal{J}}_{y_1^{t-1}}$ и $\hat{\mathcal{J}}_{y_1^{t-1}}$, т. е. о том, можно ли на очередном шаге оценивания ограничить объем доступной информации так, чтобы выполнялись указанные свойства для условных моментов. В принципе, в области субоптимальной фильтрации редко рассматриваются вопросы такого рода. Действительно, использовать только тот объем информации, который приведет к эффективной и качественной вычислительной процедуре, описав его некоторыми вербальными правилами, вполне допустимо, если это приводит к практическому результату. Но для МУМНФ важно еще и обоснование, формальное и основанное на конкретном математическом аппарате. Записывая условные моменты, нужно быть уверенным, что это можно сделать, что манипуляции выполняются со случайными векторами, имеющими конечные вторые моменты, что σ -алгебры, порождающие условные распределения, существуют. Для этого нужно быть уверенным, что для любого вектора $z = \text{col}(x, y)$ с $\mathcal{F}_z \in \Phi(m_z, D_z)$ можно так преобразовать y , что условное распределение x относительно преобра-

зованной $\tilde{\mathcal{J}}_y$ будет обеспечивать заданные значения условным моментам — математическому ожиданию и ковариации. Легко привести примеры, когда у такой задачи есть решения и когда их нет, т. е. все зависит от конкретной системы (1). Поэтому здесь ограничимся формальным предположением, что система и выбранные для МУМНФ структурные функции ξ_t и ζ_t обеспечивают существование таких σ -алгебр.

Второй вопрос — о практической реализации МУМНФ. Это существенный вопрос, так как вычислять аналитически моменты, фигурирующие в этих соотношениях, априорно нельзя. Равно как и рассчитывать их в процессе фильтрации также не удастся. Здесь проблема в огромной вычислительной трудоемкости, неминуемо сопровождающей любые попытки манипулирования условными вероятностями в динамических системах. Поэтому проводить вычисления нужных параметров придется приближенно. Следуя концепции УМНФ, используем для этого тот же метод Монте-Карло, но уже не априорно, а в процессе расчетов для каждой траектории системы-оценки.

Основания для реализации имитационного моделирования дает решение задач (12), отвечая на вопрос, какими в наилучшем случае будут условные распределения относительно σ -алгебр $\tilde{\mathcal{J}}_{y_1^{t-1}}$ и $\hat{\mathcal{J}}_{y_1^{t-1}}$. Действительно, поскольку во второй задаче наилучшее распределение $z = \text{col}(x_t - \tilde{x}_t, \zeta_t)$ гауссовское, а $\hat{\theta}_t(\zeta_t)$ — линейная, значит, и наилучшее распределение \hat{x}_t — гауссовское. Такова же ситуация и на предыдущем шаге, поэтому в (8) и (9) наилучшее распределение x_{t-1} из возможных условных распределений относительно $\tilde{\mathcal{J}}_{y_1^{t-1}}$ является гауссовским. Аналогично в первой задаче в (12) относительно $z = \text{col}(x_t, \xi_t)$ наилучшим распределением оказывается гауссовское, $\hat{\theta}_t(\xi_t)$ — линейна, поэтому в (10) и (11) наилучшее распределение $x_t - \tilde{x}_t$ из возможных условных распределений относительно $\hat{\mathcal{J}}_{y_1^{t-1}}$ будет гауссовским. Отметим еще, что при выполнении предположений о гауссовском распределении вектора $\text{col}(x_t - \tilde{x}_t, \zeta_t)$ соотношение (11) позволяет интерпретировать \hat{K}_t не только как условную ковариацию относительно $\tilde{\mathcal{J}}_{y_1^{t-1}}$, но и как условную ковариацию относительно $\tilde{\mathcal{J}}_{y_1^t}$, ссылаясь на теорему о нормальной корреляции [1], т. е. обосновать рекуррентный переход к следующему шагу оценивания. Эти соображения дают основание для реализации расчетов по формулам (8)–(11) методом Монте-Карло.

Пусть в момент t имеются \hat{x}_{t-1} — оценка МУМНФ состояния x_{t-1} по наблюдениям $y_\tau, \tau = 1, \dots, t - 1$, и \hat{K}_{t-1} — оценка условной ковариации ошибки. На шаге t выполняется моде-

лирование выборок: $\{x_{t-1}^i\}_{i=1}^N$ — из гауссовского распределения со средним \hat{x}_{t-1} и ковариацией \hat{K}_{t-1} и $\{w_t^i, v_t^i\}_{i=1}^N$ — дискретных белых шумов в соответствии с моделью (1) — и рассчитываются наборы $\{x_t^i, y_t^i\}_{i=1}^N$ по формулам (1). Прогноз МУМНФ задается в виде $\tilde{x}_t = \overline{E}\{x_t\}$, его точность $\tilde{K}_t = \overline{\text{cov}}(x_t - \tilde{x}_t, x_t - \tilde{x}_t)$. Отметим, что здесь не нужна функция базового прогноза ξ_t , коэффициент $F_t = 0$, поскольку не моделируется выборка оценок \hat{x}_{t-1} , и расчет выполняется так же, как и на первом шаге УМНФ.

На этапе коррекции формируются наборы $\{x_t^i - \tilde{x}_t, \zeta_t^i\}_{i=1}^N$ и вычисляются коэффициенты

$$H_t = \overline{\text{cov}}(x_t - \tilde{x}_t, \zeta_t) \overline{\text{cov}}^+(\zeta_t, \zeta_t);$$

$$h_t = -H_t \overline{E}\{\zeta_t\}, \quad \zeta_t^i = \zeta_t(\tilde{x}_t, y_t^i)$$

и оценка точности

$$\hat{K}_t = \tilde{K}_t - H_t \overline{\text{cov}}(\zeta_t, x_t - \tilde{x}_t).$$

Окончательная оценка фильтрации, как и в УМНФ, ищется в виде:

$$\hat{x}_t = \tilde{x}_t + H_t \zeta_t + h_t, \quad \zeta_t = \zeta_t(\tilde{x}_t, y_t).$$

Важно отметить, что размеры моделируемых выборок в УМНФ и МУМНФ должны принципиально отличаться, поскольку расчет коэффициентов УМНФ выполняется априорно и, таким образом, ничем не ограничен, а расчет коэффициентов МУМНФ выполняется в процессе фильтрации, т. е. предполагает реализуемость в режиме реального времени. По этой причине в экспериментах для УМНФ моделировались пучки объемом в 10^6 траекторий, для МУМНФ N полагается далее равным 25, 100 и 1000.

4 Анализ качества модифицированной условно-минимаксной нелинейной фильтрации

В рамках данного исследования авторами был проведен значительный объем вычислительных экспериментов с целью сравнения качества оценок УМНФ и МУМНФ. В целом концепция МУМНФ продемонстрировала свою работоспособность, но говорить о регулярном превосходстве МУМНФ нельзя. В подавляющем большинстве рассмотренных модельных примеров точности оценок обоих фильтров близки, совпадая до процентных пунктов. Незначительная часть примеров давала преимущество МУМНФ в несколько процентов, аналогично

наблюдались и другие примеры с небольшим преимуществом УМНФ. При этом практически всегда удавалось нивелировать это преимущество за счет роста параметра N , числа моделируемых траекторий МУМНФ, что на практике, конечно, неприемлемо. Найти пример ощутимого превосходства МУМНФ удалось только в задаче идентификации параметров динамической системы. Иллюстрация сказанному — следующие два модельных примера.

Первый эксперимент проведен для модели наблюдения следующего вида:

$$\left. \begin{aligned} x_t &= \frac{x_{t-1}}{1 + x_{t-1}^2} + w_t, \quad t = 1, \dots, T, \quad T = 50, \\ y_t &= x_t + x_t^3 + v_t. \end{aligned} \right\} \quad x_0 = \eta; \quad (13)$$

В этой модели две характерных черты. Во-первых, нелинейная регрессия в уравнении состояния, функция которой $x/(1 + x^2)$ хотя и существенно нелинейная, но обладает довольно инертным характером, дифференцируема и деградирует на бесконечности. Во-вторых, это наблюдения в форме кубического сенсора, также нелинейные, но весьма информативные и удобные для нелинейной фильтрации. Шумы w_t и v_t в (13) стандартные гауссовские, начальное условие η также гауссовское с $m_\eta = 0,1$ и $D_\eta = 1$. Структура обоих фильтров типовая: прогноз строится «в силу системы», коррекция — в форме невязки. Параметры УМНФ вычислялись априорно методом Монте-Карло по пучку из 10^5 траекторий, качество оценок анализировалось по второму пучку из 10^6 траекторий. Модифицированный условно-минимаксный нелинейный фильтр рассчитывался в трех вариантах: для N , равного 25, 100 и 1000 (такие же параметры использованы и в следующем примере). Таким образом, получены четыре оценки, качество которых проиллюстрировано на рис. 1.

Как видно, МУМНФ начинает демонстрировать приемлемое качество оценивания уже при $N = 100$, а при $N = 1000$ дает преимущество порядка 2%–3%.

Вторая задача — идентификация параметра θ динамической системы по косвенным линейным наблюдениям — представлена следующей моделью:

$$\left. \begin{aligned} x_t &= \theta x_{t-1} + w_t, \quad t = 1, \dots, T, \quad T = 50, \\ y_t &= x_t + 0,1v_t. \end{aligned} \right\} \quad x_0 = \eta; \quad (14)$$

Неизвестный коэффициент θ авторегрессии в уравнении динамики предполагается распределенным равномерно на отрезке $[-0,9; 0,9]$ и подлжит идентификации на каждой траектории.

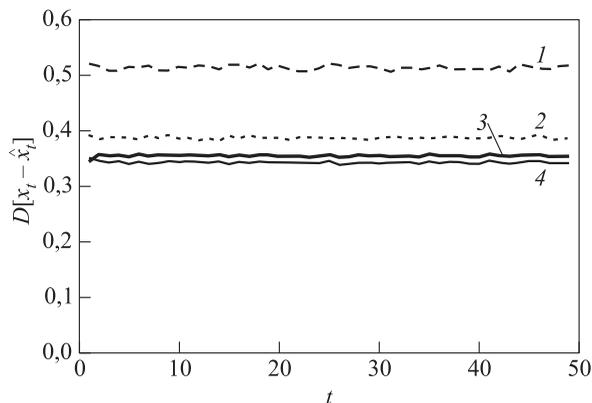


Рис. 1 Показатели качества оценок в модели (13): 1 — $D[x_T - \hat{x}_T] = 0,52$, МУМНФ ($N = 25$); 2 — $D[x_T - \hat{x}_T] = 0,39$, МУМНФ ($N = 100$); 3 — $D[x_T - \hat{x}_T] = 0,35$, УМНФ; 4 — $D[x_T - \hat{x}_T] = 0,34$, МУМНФ ($N = 1000$)

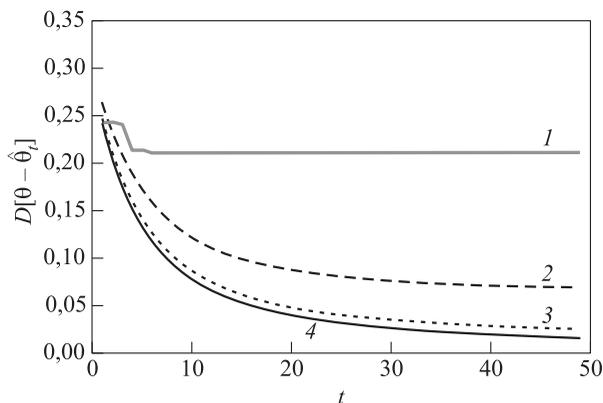


Рис. 2 Показатели качества идентификации в модели (14): 1 — $D[\theta - \hat{\theta}_T] = 0,21$, УМНФ; 2 — $D[\theta - \hat{\theta}_T] = 0,07$, МУМНФ ($N = 25$); 3 — $D[\theta - \hat{\theta}_T] = 0,03$, МУМНФ ($N = 100$); 4 — $D[\theta - \hat{\theta}_T] = 0,02$, МУМНФ ($N = 1000$)

С этой целью использованы те же фильтры, что и в предыдущем примере, примененные к расширенному вектору состояния $\text{col}(x_t, \theta)$. Результат целевой задачи идентификации иллюстрирует рис. 2.

В данном примере МУМНФ имеет существенное преимущество уже в том, что обеспечивает сходимость к значению θ на каждой траектории, в то время как УМНФ ограничивается оценкой с фиксированной, но не уменьшающейся дисперсией, лишь незначительно улучшающей качество $D_\theta = E\{\theta^2\} = 0,27$ тривиальной оценки $m_\theta = E\{\theta\} = 0$.

5 Выводы

Главным результатом работы представляется в очередной раз продемонстрированная высокая эффективность концепций условно-оптимальной и условно-минимаксной фильтрации, подтвержденная новой модификацией классического фильтра.

Полученный алгоритм оценивания показал не только свою работоспособность, но и однозначное преимущество в том случае, когда априорные расчеты параметров фильтра по какой-либо причине невозможны, например при отсутствии достоверных предположений о требуемом горизонте фильтрации или невозможности хранения большого объема априорно вычисленных параметров.

Кроме того, МУМНФ продемонстрировал определенное преимущество в задаче идентификации параметров стохастической динамической системы. Если к условиям фильтрации конкрет-

ная задача специальных требований (бесконечный горизонт, невозможность априорных расчетов, необходимость идентификации части параметров) не предъявляет, то метод УМНФ представляется лучшим вариантом практической нелинейной фильтрации.

Литература

1. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов: Нелинейная фильтрация и смежные вопросы. — М.: Наука, 1974. 696 с.
2. Kallianpur G. Stochastic filtering theory. — New York, NY, USA: Springer-Verlag, 1980. 318 p.
3. Пугачев В. С. Рекуррентное оценивание переменных и параметров в стохастических системах, описываемых разностными уравнениями // Докл. Акад. наук СССР, 1978. Т. 243. № 5. С. 1131–1133.
4. Пугачев В. С. Оценивание переменных и параметров в дискретных нелинейных системах // Автоматика и телемеханика, 1979. № 6. С. 63–79.
5. Панков А. Р. Рекуррентная условно-минимаксная фильтрация процессов в разностных нелинейных стохастических системах // Известия РАН. Теория и системы управления, 1992. № 3. С. 63–70.
6. Pankov A. R., Bosov A. V. Conditionally minimax algorithm for nonlinear system state estimation // IEEE T. Automat. Contr., 1994. Vol. 39. No. 8. P. 1617–1620.
7. Синицын И. Н., Корепанов Э. Р. Устойчивые линейные условно-оптимальные фильтры и экстраполяторы для стохастических систем с мультипликативны-

- ми шумами // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 70–75.
8. Сеницын И. Н., Корепанов Э. Р. Нормальные условно-оптимальные фильтры и экстраполяторы Пугачёва для стохастических систем, линейных относительно состояния // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 14–23.
9. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р. Модифицированные эллипсоидальные условно-оптимальные фильтры для нелинейных стохастических систем на многообразиях // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 101–111.
10. Борисов А. В., Босов А. В., Кибзун А. И., Миллер Г. Б., Семенихин К. В. Метод условно-оптимальной нелинейной фильтрации и современные подходы к оцениванию состояний нелинейных стохастических систем // Автоматика и телемеханика, 2018. № 1. С. 3–17.

Поступила в редакцию 27.12.18

ON THE CONDITIONALLY MINIMAX NONLINEAR FILTERING CONCEPT DEVELOPMENT: FILTER MODIFICATION AND ANALYSIS

A. V. Bosov and G. B. Miller

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The main result of the research is a new suboptimal filter developed from the conditionally minimax nonlinear filtering (CMNF) method for nonlinear stochastic systems in discrete time. The main idea of the proposed modification is to omit the time and resource consuming phase of *a priori* CMNF parameter calculation in favor of their online approximation together with the current state estimation. In the original CMNF filter, the simulation study is used in order to approximate dynamic system parameters’ unconditional expectation and covariances, while the modified version deals with the conditional moments which are also calculated by means of the Monte-Carlo method. The proposed filter modification is provided with the minimax justification, similar to the underlying CMNF concept. Simulation examples show the proposed algorithm effectiveness and performance gain in comparison with the original conditionally minimax nonlinear filter.

Keywords: nonlinear stochastic observation system in discrete time; conditionally minimax nonlinear filtering; Monte-Carlo simulation

DOI: 10.14357/19922264190202

Acknowledgments

This work was partially supported by the Russian Foundation for Basic Research (grant 19-07-00187-A).

References

1. Liptser, R., and A. N. Shiryaev. 2001. *Statistics of random processes*. Berlin – Heidelberg: Springer-Verlag. 427 p.
2. Kallianpur, G. 1980. *Stochastic filtering theory*. New York, NY: Springer-Verlag. 318 p.
3. Pugachev, V. S. 1978. Recursive estimation of variables and parameters in stochastic systems described by autoregression equations. *Sov. Math. Dokl.* 19:991–995.
4. Pugachev, V. S. 1979. Estimation of variables and parameters in discrete-time nonlinear systems. *Automat. Rem. Contr.* 40(4):39–50.
5. Pankov, A. R. 1993. Recurrent conditionally minimax filtering of processes in nonlinear difference stochastic systems. *J. Comput. Sys. Sc. Int.* 31(4):54–60.
6. Pankov, A. R., and A. V. Bosov. 1994. Conditionally minimax algorithm for nonlinear system state estimation. *IEEE T. Automat. Contr.* 39(8):1617–1620.
7. Sinitsyn, I. N., and E. R. Korepanov. 2015. Ustoychivye lineynye uslovno-optimal’nye fil’try i ekstrapolyatory dlya stokhasticheskikh sistem s mul’tiplikativnymi shumami [Stable linear conditionally optimal filters and extrapolators for stochastic systems with multiplicative noises]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):70–75.
8. Sinitsyn, I. N., and E. R. Korepanov. 2016. Normal’nye

- uslovno-optimal'nye fil'try i ekstrapolyatory Pugacheva dlya stokhasticheskikh sistem, lineynykh otnositel'no sostoyaniya [Normal Pugachev conditionally-optimal filters and extrapolators for state linear stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 10(2): 14–23.
9. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2017. Modifitsirovannye ellipsoidal'nye uslovno-optimal'nye fil'try dlya nelineynykh stokhasticheskikh sistem na mnogoobraznykh [Modified ellipsoidal conditionally optimal filters for nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):101–111.
10. Borisov, A. V., A. V. Bosov, A. I. Kibzun, G. B. Miller, and K. V. Semenikhin. 2018. The conditionally minimax nonlinear filtering method and modern approaches to state estimation in nonlinear stochastic systems. *Automat. Rem. Contr.* 79(1):1–11.

Received December 27, 2018

Contributors

Bosov Alexey V. (b. 1969) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AVBosov@ipiran.ru

Miller Gregory B. (b. 1980) — Candidate of Science (PhD) in physics and mathematics, PhD in informatics, scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; GMiller@ipiran.ru

СВОЙСТВА ВЕЙВЛЕТ-ОЦЕНОК СИГНАЛОВ, РЕГИСТРИРУЕМЫХ В СЛУЧАЙНЫЕ МОМЕНТЫ ВРЕМЕНИ*

О. В. Шестаков¹

Аннотация: Алгоритмы вейвлет-анализа в сочетании с процедурами пороговой обработки широко используются в задачах непараметрической регрессии при оценивании функции сигнала по зашумленным данным. Преимущества данных методов заключаются в их вычислительной эффективности и возможности адаптации к локальным особенностям оцениваемой функции. Анализ погрешностей методов пороговой обработки представляет собой важную практическую задачу, поскольку позволяет оценить качество как самих методов, так и используемого оборудования. Иногда природа данных такова, что регистрация наблюдений производится в случайные моменты времени. Если точки отсчета образуют вариационный ряд, построенный по выборке из равномерного распределения на отрезке регистрации данных, то использование обычных процедур пороговой обработки оказывается адекватным. В данной работе проведен анализ оценки среднеквадратичного риска пороговой обработки и показано, что при определенных условиях данная оценка оказывается сильно состоятельной и асимптотически нормальной.

Ключевые слова: вейвлеты; пороговая обработка; случайные отсчеты; оценка среднеквадратичного риска

DOI: 10.14357/19922264190203

1 Введение

Одним из популярных инструментов в задачах непараметрической регрессии, к которым относится подавление шума в наблюдаемом сигнале, является метод пороговой обработки. Наблюдаемые данные преобразуются в эмпирические вейвлет-коэффициенты, к которым применяется функция пороговой обработки, после чего осуществляется обратное преобразование. Наиболее распространены функции жесткой и мягкой пороговой обработки. Для модели отсчетов сигнала, заданных на равномерной неслучайной сетке и содержащих белый гауссовский шум, эти методы хорошо изучены. Предложены способы вычисления пороговых значений, обеспечивающих «почти» оптимальный порядок среднеквадратичного риска в различных классах функций сигнала [1–3]. Также изучены статистические свойства оценки среднеквадратичного риска. Показано, что при определенных условиях она оказывается сильно состоятельной и асимптотически нормальной [4–7].

В некоторых приложениях нет возможности (или она сильно затруднена) регистрировать отсчеты сигнала через равные промежутки времени [8]. Иногда природа сигнала такова, что регистрация его отсчетов производится в случайные моменты времени. В работе [9] показано, что если точки

отсчета образуют вариационный ряд, построенный по выборке из равномерного распределения на отрезке регистрации сигнала, то при использовании обычной пороговой обработки вейвлет-коэффициентов порядок среднеквадратичного риска остается с точностью до логарифмического множителя равным оптимальному порядку в классе функций, регулярных по Липшицу. В данной работе доказывается, что статистические свойства оценки риска также не меняются при переходе от фиксированной равномерной сетки отсчетов к случайной.

2 Модель непараметрической регрессии со случайными точками отсчета

Пусть функция сигнала $f(x)$ задана на отрезке $[0, 1]$ и равномерно регулярна по Липшицу с некоторым показателем $\gamma > 0$ и константой Липшица $L > 0$: $f \in \text{Lip}(\gamma, L)$. Предположим, что отсчеты $f(x)$ регистрируются в случайные моменты времени и содержат аддитивный белый гауссовский шум, т. е. рассмотрим следующую модель данных:

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, N \quad (N = 2^J),$$

где x_i независимы и равномерно распределены на $[0, 1]$, а ε_i независимы между собой и от x_i и име-

* Работа выполнена при финансовой поддержке Российского научного фонда (проект 18-11-00155).

¹ Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

ют нормальное распределение с нулевым средним и дисперсией σ^2 . Пусть $0 \leq x_{(1)} < \dots < x_{(N)} \leq 1$ — вариационный ряд, построенный по выборке x_i , $i = 1, \dots, N$. Тогда, перенумеровав Y_i и ε_i , получаем модель:

$$Y_i = f(x_{(i)}) + \varepsilon_i, \quad i = 1, \dots, N \quad (N = 2^J). \quad (1)$$

Наблюдения состоят из пар $(x_{(1)}, Y_1), \dots, (x_{(N)}, Y_N)$, в которых расстояния между отсчетами в общем случае не равны. При этом $E x_{(i)} = i/(N+1)$. Наряду с (1) рассмотрим выборку с равными расстояниями между отсчетами:

$$\left(\frac{1}{N+1}, Z_1 \right), \dots, \left(\frac{N}{N+1}, Z_N \right), \quad (2)$$

где

$$Z_i = f\left(\frac{i}{N+1}\right) + \varepsilon_i, \quad i = 1, \dots, N.$$

Для выборки (2) Донохо и Джонстон разработали методы пороговой обработки, эффективно подавляющие шум и обеспечивающие «почти» оптимальный порядок среднеквадратичного риска [1]. К (2) применяется дискретное ортогональное вейвлет-преобразование и получается набор эмпирических вейвлет-коэффициентов:

$$W_{j,k} = \mu_{j,k} + \xi_{j,k}, \quad j = 0, \dots, J-1, \quad k = 0, \dots, 2^j - 1,$$

где $\xi_{j,k}$ независимы и имеют такое же распределение, как и ε_i , а $\mu_{j,k}$ — коэффициенты дискретного вейвлет-преобразования выборки $f(1/(N+1)), \dots, f(N/(N+1))$.

Далее к $W_{j,k}$ применяется функция жесткой пороговой обработки $\rho_H(y, T) = x \mathbf{1}(|y| > T)$ или мягкой пороговой обработки $\rho_S(y, T) = \text{sgn}(x) (|y| - T)_+$ с некоторым порогом T и получаются оценки коэффициентов $\widehat{W}_{j,k}$. После этого осуществляется обратное вейвлет-преобразование. Идея пороговой обработки заключается в том, что вейвлет-преобразование обеспечивает «разреженное» представление функции полезного сигнала, т.е. сигнал представляется относительно небольшим числом больших по модулю коэффициентов. При этом характеристики шума остаются неизменными. Следовательно, при обнулении малых по модулю коэффициентов полезный сигнал будет затронут несильно и практически весь шум будет удален. Чтобы обеспечить «разреженное» представление функции, равномерно регулярной по Липшицу с показателем γ , вейвлет-функция, участвующая в дискретном вейвлет-преобразовании, должна удовлетворять определенным условиям. В частности, она должна иметь M непрерывных

производных ($M \geq \gamma$), M нулевых моментов и достаточно быстро убывать на бесконечности [10]. Далее предполагается, что все необходимые условия выполнены.

Если применить дискретное вейвлет-преобразование к выборке (1), то получится набор эмпирических вейвлет-коэффициентов

$$V_{j,k} = \nu_{j,k} + \xi_{j,k}, \quad j = 0, \dots, J-1, \quad k = 0, \dots, 2^j - 1.$$

Здесь $\nu_{j,k}$ — коэффициенты дискретного вейвлет-преобразования выборки $f(x_{(1)}), \dots, f(x_{(N)})$.

В общем случае $V_{j,k}$ не равны $W_{j,k}$, а $\nu_{j,k}$ не равны $\mu_{j,k}$. Однако к $V_{j,k}$ можно применить ту же процедуру, что и к коэффициентам $W_{j,k}$, и получить оценки $\widehat{V}_{j,k}$. В следующих разделах обсуждаются свойства полученных оценок.

3 Среднеквадратичный риск пороговой обработки

Среднеквадратичный риск пороговой обработки для выборки со случайными точками отсчетов определим как

$$R_\nu(f, T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} E \left(\widehat{V}_{j,k} - \mu_{j,k} \right)^2. \quad (3)$$

Определим также среднеквадратичный риск для выборки с равными расстояниями между отсчетами:

$$R_\mu(f, T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} E \left(\widehat{W}_{j,k} - \mu_{j,k} \right)^2.$$

Способ выбора порога является одной из основных проблем при пороговой обработке. При выборе так называемого универсального порога $T_U = \sigma \sqrt{2 \ln 2^J}$ в работе [9] доказано следующее утверждение о порядке $R_\nu(f, T)$.

Теорема 1. Пусть $f \in \text{Lip}(\gamma, L)$ на отрезке $[0, 1]$ с $\gamma \geq 1/2$. Для модели (1) при использовании вейвлет-функции, удовлетворяющей перечисленным выше условиям, и выборе порога T_U справедливо

$$R_\nu(f, T) \leq C \cdot 2^{J/(2\gamma+1)} J^{2\gamma/(2\gamma+1)},$$

где C — некоторая положительная константа.

Аналогичное утверждение справедливо и для $R_\mu(f, T)$ [2]. Таким образом, замена равноотстоящих точек отсчета на случайные не оказывает влияния на оценку порядка среднеквадратичного риска.

В [11] для класса $\text{Lip}(\gamma, L)$ предложено использовать порог

$$T_\gamma = \sigma \sqrt{\frac{4\gamma}{2\gamma + 1} \ln 2^J}.$$

При этом оценка порядка $R_\mu(f, T)$ останется такой же. Используя лемму 3 из [9] и повторяя рассуждения, приводимые в этой работе при доказательстве теоремы 1, несложно убедиться, что теорема 1 также останется верна при замене порога T_U на T_γ . Порог T_γ является более адекватным для класса $\text{Lip}(\gamma, L)$, чем порог T_U , поскольку при том же порядке средневквдратичного риска он меньше «сглаживает» полезный сигнал [11].

4 Оценка средневквдратичного риска

В выражении (3) присутствуют неизвестные величины «чистых» коэффициентов $\mu_{j,k}$, поэтому вычислить значение риска $R_\nu(f, T)$ на практике нельзя. Однако его можно оценить непосредственно по наблюдаемым данным. По аналогии с [10] в качестве оценки риска будем использовать величину

$$\widehat{R}_\nu(f, T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} F[V_{j,k}, T], \quad (4)$$

где

$$F[V_{j,k}, T] = \begin{cases} (V_{j,k}^2 - \sigma^2) \mathbf{1}(|V_{j,k}| \leq T) + \sigma^2 \mathbf{1}(|V_{j,k}| > T) \\ \text{в случае жесткой пороговой обработки;} \\ (V_{j,k}^2 - \sigma^2) \mathbf{1}(|V_{j,k}| \leq T) + \\ \quad + (\sigma^2 + T^2) \mathbf{1}(|V_{j,k}| > T) \\ \text{в случае мягкой пороговой обработки.} \end{cases}$$

В ситуации с неслучайной равномерной сеткой отсчетов при использовании мягкой пороговой обработки такая оценка риска является несмещенной. Если же используется жесткая пороговой обработка, то оценка содержит смещение, которое, однако, при достаточной гладкости функции сигнала не оказывает влияния на статистические свойства оценки.

Оценка риска (4) дает возможность получить представление о погрешности, с которой оценивается функция сигнала, используя только наблюдаемые данные. Докажем утверждение об асимптотической нормальности оценки (4), которое,

в частности, позволяет строить асимптотические доверительные интервалы для теоретического риска $R_\nu(f, T)$.

Теорема 2. Пусть $f \in \text{Lip}(\gamma, L)$ на отрезке $[0, 1]$ с $\gamma > 1/2$ и вейвлет-функция удовлетворяет перечисленным выше условиям. Тогда при жесткой и мягкой пороговой обработке при выборе порога T_γ

$$P\left(\frac{\widehat{R}_\nu(f, T_\gamma) - R_\nu(f, T_\gamma)}{\sigma^2 \sqrt{2^{J+1}}} < x\right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty,$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Доказательство. Докажем теорему для случая жесткой пороговой обработки. В случае мягкой пороговой обработки доказательство аналогично.

Наряду с $\widehat{R}_\nu(f, T_\gamma)$ рассмотрим

$$\widehat{R}_\mu(f, T_\gamma) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} F[W_{j,k}, T_\gamma]$$

и запишем разность $\widehat{R}_\nu(f, T_\gamma) - R_\nu(f, T_\gamma)$ в виде

$$\widehat{R}_\nu(f, T_\gamma) - R_\nu(f, T_\gamma) = \widehat{R}_\mu(f, T_\gamma) - R_\mu(f, T_\gamma) + \widetilde{R},$$

где

$$\widetilde{R} = \widehat{R}_\nu(f, T_\gamma) - \widehat{R}_\mu(f, T_\gamma) - (R_\nu(f, T_\gamma) - R_\mu(f, T_\gamma)).$$

В [6] показано, что

$$P\left(\frac{\widehat{R}_\mu(f, T_\gamma) - R_\mu(f, T_\gamma)}{\sigma^2 \sqrt{2^{J+1}}} < x\right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty.$$

Следовательно, для доказательства теоремы достаточно показать, что

$$\frac{\widetilde{R}}{\sqrt{2^J}} \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

Если $\gamma > 1/2$, то в силу аналога теоремы 1 для порога T_γ и аналогичного утверждения из работы [2]

$$\frac{R_\nu(f, T_\gamma) - R_\mu(f, T_\gamma)}{\sqrt{2^J}} \rightarrow 0 \text{ при } J \rightarrow \infty.$$

Далее, поскольку $f \in \text{Lip}(\gamma, L)$, то выполняется

$$|\mu_{j,k}| \leq \frac{A \cdot 2^{j/2}}{2^{j(\gamma+1/2)}}$$

для некоторой константы A , зависящей от L [10]. Пусть

$$j_0 \approx \frac{J}{2\gamma + 1} + \frac{1}{2\gamma + 1} \log_2 J.$$

Представим $\widehat{R}_\nu(f, T_\gamma) - \widehat{R}_\mu(f, T_\gamma)$ в виде

$$\widehat{R}_\nu(f, T_\gamma) - \widehat{R}_\mu(f, T_\gamma) = S_1 + S_2,$$

где

$$S_1 = \sum_{j=0}^{j_0-1} \sum_{k=0}^{2^j-1} (F[V_{j,k}, T_\gamma] - F[W_{j,k}, T_\gamma]);$$

$$S_2 = \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (F[V_{j,k}, T_\gamma] - F[W_{j,k}, T_\gamma]).$$

Поскольку как в случае жесткой, так и в случае мягкой пороговой обработки

$$\left. \begin{aligned} |F[V_{j,k}, T_\gamma]| &\leq T_\gamma^2 + \sigma^2; \\ |F[W_{j,k}, T_\gamma]| &\leq T_\gamma^2 + \sigma^2 \text{ п. в.}, \end{aligned} \right\} \quad (5)$$

то, если $\gamma > 1/2$,

$$\frac{S_1}{\sqrt{2^J}} \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

Далее

$$S_2 = \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (F[V_{j,k}, T_\gamma] - F[W_{j,k}, T_\gamma]) =$$

$$= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (V_{j,k}^2 - W_{j,k}^2) +$$

$$+ \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (W_{j,k}^2 - 2\sigma^2) \mathbf{1}(|V_{j,k}| \leq T_\gamma, |W_{j,k}| > T_\gamma) +$$

$$+ \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (2\sigma^2 - V_{j,k}^2) \mathbf{1}(|V_{j,k}| > T_\gamma, |W_{j,k}| \leq T_\gamma) +$$

$$+ \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (W_{j,k}^2 - V_{j,k}^2) \mathbf{1}(|V_{j,k}| > T_\gamma, |W_{j,k}| > T_\gamma). \quad (6)$$

Рассмотрим сумму $\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (V_{j,k}^2 - W_{j,k}^2)$:

$$\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (V_{j,k}^2 - W_{j,k}^2) = \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (\nu_{j,k}^2 - \mu_{j,k}^2) +$$

$$+ 2 \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \xi_{j,k} (\nu_{j,k} - \mu_{j,k}).$$

Учитывая результаты работы [9], можно показать, что условное распределение этой суммы при фиксированных x_i нормально с математическим ожиданием $\sum_{i=1}^N (f_*^2(x_{(i)}) - f_*^2(i/(N+1)))$ и дисперсией $4\sigma^2 \sum_{i=1}^N (f_*(x_{(i)}) - f_*(i/(N+1)))^2$, где

$f_*(x)$ — функция, вейвлет-разложение которой состоит только из коэффициентов при $j_0 \leq j \leq J-1$. Эта функция имеет ту же гладкость, что и $f(x)$ [10].

Учитывая, что $f \in \text{Lip}(\gamma, L)$, имеем:

$$\left| \mathbb{E}_x \sum_{i=1}^N \left(f_*^2(x_{(i)}) - f_*^2\left(\frac{i}{N+1}\right) \right) \right| =$$

$$= \left| N \left(\int_0^1 f_*^2(x) dx - \frac{1}{N} \sum_{i=1}^N f_*^2\left(\frac{i}{N+1}\right) \right) \right| \leq$$

$$\leq C_1 N^{1-\min(2,\gamma)} = C_1 2^{J(1-\min(2,\gamma))},$$

где C_1 — некоторая константа. Кроме того, в силу вида функции $f_*(x)$ для некоторой константы $C_2 > 0$

$$\frac{1}{N} \sum_{i=1}^N f_*^2\left(\frac{i}{N+1}\right) \leq$$

$$\leq C_2 N^{-2\gamma/(2\gamma+1)} = C_2 \cdot 2^{-(2\gamma/(2\gamma+1))J}.$$

Следовательно,

$$\mathbb{E}_x \sum_{i=1}^N f_*^2(x_{(i)}) \leq C_1 \cdot 2^{J(1-\min(2,\gamma))} +$$

$$+ C_2 \cdot 2^{(1-2\gamma/(2\gamma+1))J}.$$

Также в [9] доказана оценка

$$\mathbb{E}_x \sum_{i=1}^N \left(f_*(x_{(i)}) - f_*\left(\frac{i}{N+1}\right) \right)^2 \leq$$

$$\leq C_3 N^{1-\min(1,\gamma)} = C_3 2^{J(1-\min(1,\gamma))}, \quad (7)$$

где C_3 — некоторая константа. Следовательно, если $\gamma > 1/2$, то, применяя неравенство Маркова, получаем, что

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(f_*^2(x_{(i)}) - f_*^2\left(\frac{i}{N+1}\right) \right) \xrightarrow{P} 0;$$

$$\frac{1}{N} \sum_{i=1}^N \left(f_*(x_{(i)}) - f_*\left(\frac{i}{N+1}\right) \right)^2 \xrightarrow{P} 0$$

при $N \rightarrow \infty$. Таким образом,

$$\frac{\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (V_{j,k}^2 - W_{j,k}^2)}{\sqrt{2^J}} \xrightarrow{P} 0 \text{ при } J \rightarrow \infty.$$

В оставшихся суммах в (6) содержатся индикаторы, в которых либо $|V_{j,k}| > T_\gamma$, либо $|W_{j,k}| > T_\gamma$, причем для всех слагаемых

$$|\mu_{j,k}| \leq C_4 J^{-1/2}$$

с некоторой константой $C_4 > 0$. Повторяя рассуждения из работы [12] с использованием (7), можно показать, что эти суммы при делении на $\sqrt{2^J}$ также сходятся к нулю по вероятности. Теорема доказана.

Помимо асимптотической нормальности оценки (4) также обладает свойством сильной состоятельности.

Теорема 3. Пусть выполнены условия теоремы 2. Тогда при жесткой и мягкой пороговой обработке для любого $\lambda > 1/2$

$$\frac{\widehat{R}_\nu(f, T_\gamma) - R_\nu(f, T_\gamma)}{2^{\lambda J}} \rightarrow 0 \text{ п. в. при } J \rightarrow \infty.$$

Поскольку выполнено (5) и при фиксированных x_i слагаемые в (4) условно независимы, доказательство этой теоремы аналогично доказательству соответствующего утверждения из работы [7].

Замечание. Теоремы 2 и 3 останутся справедливыми, если вместо порога T_γ выбрать универсальный порог T_U .

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Donoho D., Johnstone I. M., Kerkyacharian G., Picard D. Wavelet shrinkage: Asymptopia? // *J. Roy. Stat. Soc. B*, 1995. Vol. 57. No. 2. P. 301–369.

3. Marron J. S., Adak S., Johnstone I. M., Neumann M. H., Patil P. Exact risk analysis of wavelet regression // *J. Comput. Graph. Stat.*, 1998. Vol. 7. P. 278–309.
4. Маркин А. В. Предельное распределение оценки риска при пороговой обработке вейвлет-коэффициентов // *Информатика и её применения*, 2009. Т. 3. Вып. 4. С. 57–63.
5. Шестак О. В. Асимптотическая нормальность оценки риска пороговой обработки вейвлет-коэффициентов при выборе адаптивного порога // *Докл. Акад. наук*, 2012. Т. 445. № 5. С. 513–515.
6. Шестак О. В. Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. — М.: Аргамак-Медиа, 2016. 200 с.
7. Shestakov O. V. On the strong consistency of the adaptive risk estimator for wavelet thresholding // *J. Math. Sci.*, 2016. Vol. 214. No. 1. P. 115–118.
8. Cai T., Brown L. Wavelet shrinkage for nonequispaced samples // *Ann. Stat.*, 1998. Vol. 26. No. 5. P. 1783–1799.
9. Cai T., Brown L. Wavelet estimation for samples with random uniform design // *Stat. Probabil. Lett.*, 1999. Vol. 42. P. 313–321.
10. Mallat S. *A wavelet tour of signal processing*. — New York, NY, USA: Academic Press, 1999. 857 p.
11. Jansen M. *Noise reduction by wavelet thresholding*. — Lecture notes in statistics ser. — New York, NY, USA: Springer Verlag, 2001. Vol. 161. 196 p.
12. Шестак О. В. Аппроксимация распределения оценки риска пороговой обработки вейвлет-коэффициентов нормальным распределением при использовании выборочной дисперсии // *Информатика и её применения*, 2010. Т. 4. Вып. 4. С. 73–81.

Поступила в редакцию 28.02.19

PROPERTIES OF WAVELET ESTIMATES OF SIGNALS RECORDED AT RANDOM TIME POINTS

O. V. Shestakov^{1,2}

¹Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Wavelet analysis algorithms in combination with threshold processing procedures are widely used in nonparametric regression problems when estimating the signal function from noisy data. The advantages of these methods are their computational efficiency and the ability to adapt to the local features of the function being estimated. The error analysis of threshold processing methods is an important practical task, since it allows assessing the quality of both the methods themselves and the equipment used. Sometimes, the nature of the data is such that observations are recorded at random times. If the sampling points form a variation series constructed from a sample of a uniform distribution over the data recording interval, then the use of conventional threshold processing procedures is adequate. In this paper, the author analyzes the estimate of the mean square risk of threshold processing and shows that under certain conditions, this estimate is strongly consistent and asymptotically normal.

Keywords: wavelets; threshold processing; random samples; mean square risk estimate

DOI: 10.14357/19922264190203

Acknowledgments

This research is supported by Russian Science Foundation (project No. 18-11-00155).

References

1. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455.
2. Donoho, D., I. M. Johnstone, G. Kerkyacharian, and D. Picard. 1995. Wavelet shrinkage: Asymptopia? *J. Roy. Stat. Soc. B* 57(2):301–369.
3. Marron, J. S., S. Adak, I. M. Johnstone, M. H. Neumann, and P. Patil. 1998. Exact risk analysis of wavelet regression. *J. Comput. Graph. Stat.* 7:278–309.
4. Markin, A. V. 2009. Predel'noe raspredelenie otsenki riska pri porogovoy obrabotke veyvlet-koeffitsientov [Limit distribution of risk estimate of wavelet coefficient thresholding]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):57–63.
5. Shestakov, O. V. 2012. Asymptotic normality of adaptive wavelet thresholding risk estimation. *Dokl. Math.* 86(1):556–558.
6. Shestakov, O. V. 2016. *Veroyatnostno-statisticheskie metody analiza i obrabotki signalov na osnove veyvlet-algoritmov* [Probabilistic-statistical methods of signal analysis and processing based on wavelet algorithms]. Moscow: Argamak-Media Publs. 200 p.
7. Shestakov, O. V. 2016. On the strong consistency of the adaptive risk estimator for wavelet thresholding. *J. Math. Sci.* 214(1):115–118.
8. Cai, T., and L. Brown. 1998. Wavelet shrinkage for nonequidspaced samples. *Ann. Stat.* 26(5):1783–1799.
9. Cai, T., and L. Brown. 1999. Wavelet estimation for samples with random uniform design. *Stat. Probabil. Lett.* 42:313–321.
10. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
11. Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. New York, NY: Springer Verlag. Vol. 161. 196 p.
12. Shestakov, O. V. 2010. Approksimatsiya raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov normal'nym raspredeleniem pri ispol'zovanii vyborochnoy dispersii [Normal approximation for distribution of risk estimate for wavelet coefficients thresholding when using sample variance]. *Informatika i ee Primeneniya — Inform. Appl.* 4(4):73–81.

Received February 28, 2019

Contributor

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; oshestakov@cs.msu.su

АРХИТЕКТУРНЫЕ РЕШЕНИЯ В ЗАДАЧЕ ВЫЯВЛЕНИЯ МОШЕННИЧЕСТВА ПРИ АНАЛИЗЕ ИНФОРМАЦИОННЫХ ПОТОКОВ В ЦИФРОВОЙ ЭКОНОМИКЕ*

А. А. Грушо¹, М. И. Забейло², Н. А. Грушо³, Е. Е. Тимонина⁴

Аннотация: Сформулирован подход к исследованию некоторых видов мошенничества в цифровой экономике с использованием причинно-следственных связей. Во всех видах рассматриваемых мошенничеств должно наблюдаться несоответствие между целями финансовых транзакций и реальной стоимостью достижения этих целей. Данные о транзакциях можно собирать, наблюдая информационные потоки, в которых отражаются эти транзакции. Архитектура сбора данных и их анализа может быть организована с помощью распределенных реестров с централизованным консенсусом, что позволяет создать аналог электронной бухгалтерской книги, фиксирующей финансово-экономическую деятельность субъектов цифровой экономики в регионе. Рассматриваемые методы выявления мошенничества основаны на противоречиях между действиями, описанными в транзакциях, и информацией, содержащейся в планах, стандартах, прецедентах и др. Рассмотрен метод, основанный на некоторой упрощенной схеме реализации абстрактного проекта. Для выявления противоречий необходимо проводить анализ от следствия к причине, т. е. искать аномалии в информации, описывающей порождение наблюдаемых следствий. Показано, как в реализации проекта можно выделять простые «необходимые условия» нарушения причинно-следственных связей, т. е. множество «необходимых условий», нарушение которых свидетельствует о наличии мошенничества. Это множество «необходимых условий» можно назвать метаданными для контроля проекта на выявление мошенничества.

Ключевые слова: цифровая экономика; информационные потоки; причинно-следственные связи; выявление мошеннических схем

DOI: 10.14357/19922264190204

1 Введение

В работе сформулирован подход к исследованию некоторых видов мошенничества в цифровой экономике с использованием причинно-следственных связей. Рассматриваются три вида мошенничества, а именно:

- (1) отмыв денег;
- (2) обман при выполнении договорных обязательств при реализации технических проектов (строительные проекты и др.);
- (3) незаконный вывод денег.

Названные виды мошенничества могут быть сведены к решению одного типа задач. Для отмывания денег источник должен заключать фиктивные контракты, в соответствии с которыми будут пе-

реводиться средства за заведомо ненужную работу и материалы.

Мошенничество, связанное с невыполнением договорных обязательств, связано со снижением качества услуг, качества и количества закупаемых материалов, выполнением работ с ненадлежащим качеством.

Вывод денег связан с переводом средств фирмам-однодневкам, которые заведомо не могут выполнить обязательства по контрактам, за которые им переводятся средства.

Таким образом, во всех трех видах рассматриваемых мошенничеств должно наблюдаться несоответствие между целями финансовых транзакций и реальной стоимостью достижения этих целей. Данные о транзакциях можно собирать, наблюдая информационные потоки, в которых отражаются эти транзакции.

* Работа частично поддержана РФФИ (проекты 18-29-03081 и 18-07-00274).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, info@itake.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

Однако для наблюдения таких информационных потоков необходимо создавать архитектуру телекоммуникационной системы, позволяющей перехватывать и собирать данные о всех транзакциях. Например, такая архитектура может быть организована с помощью распределенных реестров с централизованным консенсусом, т. е. все информационные потоки, сформированные в цифровой экономике и несущие информацию о транзакциях, проходят через некоторый центральный узел, запоминающий их в форме распределенного реестра. Такие реестры могут дублироваться в аналогичных центрах различных регионов, что позволяет создать аналог электронной бухгалтерской книги, фиксирующей финансово-экономическую деятельность субъектов цифровой экономики. Такой подход предложено реализовать на базе системы ситуационных центров, что отражено в работах [1, 2].

Собранная из информационных потоков информация о транзакциях, т. е. о контрактах, договорах, платежах, отчетах, закупленных материалах, характеристиках исполнителей работ и др., собирается в базе данных в указанном центре. Согласно теории интеллектуальных систем [3], эту базу данных можно называть базой фактов (БФ). Базу фактов можно представить как бинарную матрицу, строки которой описывают характеристики, входящие в транзакции, а столбцы нумеруются характеристиками. Строки матрицы будем называть объектами [4, 5].

Рассматриваемые в работе методы выявления мошенничества будут основаны на противоречиях между действиями, описанными в транзакциях, и информацией, содержащейся в планах, стандартах, прецедентах и др. Для нахождения противоречий в архитектуре центра предусмотрена другая база данных — база знаний (БЗ) [3, 6], которая устроена так же, как БФ.

Информация в БЗ собирается на основе положительного опыта или расчетов. Используя БЗ, можно выводить факты нарушения причинно-следственных связей. Нарушения причинно-следственных связей будем называть *аномалиями*.

Для упрощения дальнейшее изложение будет вестись в рамках поиска противоречий при выполнении некоторого абстрактного проекта. Выявление аномалий будет происходить на основе фактов из БФ с помощью знаний из БЗ методами искусственного интеллекта и интеллектуального анализа данных [6].

2 Модели

Наиболее сложная из рассмотренных выше задач — выявление противоречий, т. е. использование

БЗ для получения новых знаний и выявление аномалий из полученных фактов.

Все способы выявления противоречий основаны на определении причинно-следственных связей. При этом противоречия в параметрах транзакций по отношению к требуемым в БЗ составляют сущность аномалий.

Далее будет рассмотрен метод, основанный на некоторой упрощенной схеме реализации абстрактного проекта.

Каждый проект имеет цель: например, цель представляет собой построение некоторой системы. Воспользуемся структурным подходом, который позволяет строить проект на основе разбиения системы на подсистемы и определения взаимодействий подсистем [7]. При этом каждая подсистема также представима структурной моделью.

Как сама система, так и каждая ее подсистема имеют свой функционал и спецификацию, параметры настройки и домены параметров настройки. Кроме этих характеристик существует множество характеристик, связанных с «жизненным циклом» создания системы. Сюда входят работы, ресурсы, сроки выполнения работ по созданию подсистем и самой системы, стоимости компонентов и материалов, стоимости работ, схемы поставок, договорные обязательства и др. Все характеристики связаны между собой, поэтому можно говорить о стоимости и времени изготовления структурных компонентов системы.

Одной из важнейших характеристик является смета (система смет для подсистем). Смета сопоставляет каждому компоненту системы стоимость его изготовления и настройки.

Схема построения системы может быть представлена диаграммой, изображенной на рис. 1.



Рис. 1 Диаграмма достижения цели

Представленная на рис. 1 диаграмма позволяет описать основные классы возможных противоречий при достижении цели. Противоречия возникают, когда данные БФ не соответствуют требуемым характеристикам.

3 Потенциальные классы аномалий при достижении цели

Выделим четыре потенциальных класса противоречий, которые показывают, каким образом нужно искать эти противоречия.

Противоречие цели и проекта (рис. 2) возникает при отсутствии обоснования или в случае логического противоречия между возможностями проектируемого функционала и целью системы. Отметим, что в проект входят сроки, перечень работ, материалы, настройки, которые описываются соответствующими параметрами и допустимыми значениями этих параметров. Проект формируется на основе БЗ и расчетов, исходя из информации, полученной по аналогии с другими проектами и решениями, которые считаются апробированными.

Отметим, что цель порождает проект и в этом смысле является причиной проекта. Однако для анализа противоречий необходимо двигаться по штриховой стрелке диаграммы (см. рис. 2) от проекта к цели. В самом деле, любой компонент проекта направлен на теоретическое достижение цели. Цель — сложный объект, поэтому в проекте могут возникнуть характеристики, противоречащие хотя бы некоторым характеристикам цели. Это делает проект противоречивым, но вывод об этом может быть сделан только на уровне описания цели.

Противоречия между проектом и его реализацией, исключая настройки (рис. 3), могут возникать, например, при закупке исполнителем материалов более низкого качества по более низким ценам, при попытках достижения требуемых сроков работы за счет снижения качества выполнения работ, за счет нахождения «объективных» причин для увеличения сроков работы и, следовательно, увеличения цены реализации проекта.

Для выявления указанных противоречий необходимо двигаться по диаграмме (см. рис. 3) в обратную сторону в соответствии со штриховыми стрелками. Действительно, выявить противоречия между характеристиками закупленных материалов и требуемыми по проекту можно только при обращении к проекту и его спецификациям. Манипуляции со сроками работы также можно выявить только при обращении к соответствующим расчетам в проекте. Задержки в сроках работы, связанные с поставками материалов, можно определить только на предыдущем этапе диаграммы (см. рис. 3) в описании проекта.

Противоречия между реализацией проекта и его настройкой (рис. 4) возникает, когда не удается добиться требуемых значений параметров функционала, не удается обеспечить необходимый уровень

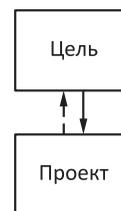


Рис. 2 Противоречия цели и проекта



Рис. 3 Противоречия проекта и его реализации (без настройки)

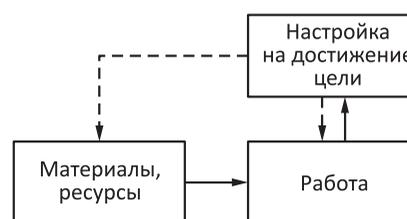


Рис. 4 Противоречия реализации проекта и его настройки

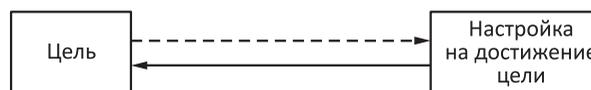


Рис. 5 Противоречия цели и достигнутой реализации проекта

качества реализации проекта. Для определения противоречия в настройках надо опять же двигаться по диаграмме (см. рис. 4) в обратную сторону по штриховым стрелкам, так как для выявления характеристик результатов работы, которые не дают возможности реализации определенного функционала, необходимо иметь информацию о результатах этой работы.

Противоречие между целью и достигнутой реализацией проекта (рис. 5) возникает, когда реализованная система не позволяет достичь цели. В этом случае опять противоречие нужно искать, двигаясь от цели к реальному достигнутому функционалу по штриховой стрелке (см. рис. 5).

Суммируя положения, изложенные в данном разделе, приходим к выводу, что для выявления противоречий необходимо проводить анализ от следствия к причине, т.е. искать аномалии в информации, описывающей порождение наблюдаемых следствий.

4 Связь противоречий и причин

Прежде чем построить связь между причинами и противоречиями, кратко опишем простейшую модель связи этих понятий. Причины и противоречия будут сформулированы для представления компонентов системы как объектов, обладающих наборами известных характеристик [4, 5].

Пусть $U = \{\alpha, \beta, \dots\}$ — совокупность характеристик (пространство характеристик). Согласно [5] объектом O называется любое подмножество характеристик $O \subseteq U$. Рассмотрим последовательность объектов, возможно в различных пространствах характеристик.

Определение 1. Объект P с числом характеристик, большим или равным 2, является *причиной* объекта (*свойства*) B в цепочке наблюдаемых объектов тогда и только тогда, когда выполнены следующие условия:

- (1) для каждого объекта C , если $P \subseteq C$, то $C \mapsto B$, где $C \mapsto B$ означает, что объект B присутствует в объекте, следующем за объектом C ;
- (2) объект P является минимальным объектом, удовлетворяющим условию 1, а именно: $\forall \alpha \in P$ объект $P \setminus \{\alpha\}$ не является причиной, т. е. $\exists C : \alpha \notin C, P \setminus \{\alpha\} \subseteq C$ и $C \not\mapsto B$, где $C \not\mapsto B$ означает, что B не может содержаться в объекте, следующем за объектом C .

Приведенное определение причины является упрощением причин, возникающих в реальном мире. Например, реальные причины могут возникать как совокупность характеристик из разных пространств. Одно следствие может порождаться разными причинами или возникать из внешних и ненаблюдаемых характеристик. Однако представленная далее формализация позволяет доступно изложить причинно-следственные истоки противоречий, которые инициируют в дальнейшем глубокое исследование рассматриваемых процессов.

Будем считать, что для любого интересующего нас свойства B существует причина. Тогда справедлива следующая теорема.

Теорема 1. Для любого свойства B существует единственная причина.

Доказательство. Доказательство будем вести от противного, т. е. предположим, что существуют две причины свойства B : P и P' , $P \neq P'$. Тогда существует $\alpha \in U$, которое удовлетворяет одному из двух условий:

- (а) $\alpha \in P, \alpha \notin P'$;
- (б) $\alpha \notin P, \alpha \in P'$.

Пусть выполняется условие (б). Тогда $P' \setminus \{\alpha\}$ не является причиной по условию 2 определения 1, т. е. $\exists C$ такое, что $\alpha \notin C, P' \setminus \{\alpha\} \subseteq C$ и $C \not\mapsto B$. Но если B произошло и P его причина, то $C \mapsto B$, что противоречит предположению. Теорема 1 доказана.

Лемма. Если P — причина появления свойства B , то объект B определяет существование свойства P в объекте, предшествующем B .

Доказательство. Из предположения, что у каждого свойства B есть причина, и условия, что P является причиной B , следует, что при появлении в данных свойства B объект C , предшествующий появлению B , содержит как часть объект P . Это следует из теоремы 1 и определения причины.

Докажем принцип «необходимого условия», который, несмотря на простоту доказательства, будет играть в дальнейшем существенную роль.

Теорема 2. Если P — причина появления свойства B и $A \subseteq P$, то объект B определяет наличие свойства A в объекте, предшествующем B .

Доказательство. Пусть в данных имеется объект B и $P \mapsto B$, тогда в силу существования и единственности причины B в данных должен существовать объект C , предшествующий B и содержащий причину P . Поскольку $A \subseteq P$ и B содержит причину P , то $B \mapsto A$. С учетом леммы теорема 2 доказана.

Пусть даны пространства U_1, U_2, \dots и имеется последовательность данных (процесс выполнения этапов проекта в соответствии с рис. 1) A, B, \dots , где каждый объект является подмножеством некоторого пространства $U_i, i = 1, \dots$. Тогда в объекте A присутствует причина P появления интересующего нас свойства C в объекте B . Пусть $P \subseteq A$, тогда по теореме 2 $\forall \alpha \in P: C \mapsto \{\alpha\}$, т. е. из появления C следует появление характеристики α в предшествующем объекте. Это необходимое условие того, что C удовлетворяет причинно-следственным связям развития процесса выполнения проекта. Если для C нет характеристики α , которую можно отнести к причине C , то можно считать, что нарушена причинно-следственная связь и C — аномальный объект.

Пример. Если объект C состоит в получении суммы a фирмой K , то согласно теореме 2 в предшествующем объекте должна существовать причина перевода суммы a на фирму K . Если эта причина в проекте отсутствует, то это можно считать признаком мошеннической схемы. Все проекты по предположению собираются из «кубиков», содержащихся в БЗ. Тогда можно сравнить цену объекта C , породившего получение суммы a , и сумму, присутствующую в смете проекта. Если разница

велика, то это либо ошибка проекта, либо признак мошеннической схемы.

5 Поиск противоречий на основе принципа «необходимых условий»

Как было показано в разд. 3, нахождение противоречий соответствуют движению от следствия к причине. Для каждого объекта в наблюдаемых данных выявление причин его появления является трудоемкой задачей. Кроме того, при реализации контроля соблюдения причинно-следственных связей на большом множестве участников экономической деятельности задача анализа причин становится трудоемкой. Поэтому процедуру контроля необходимо разбить на два этапа, где первый этап состоит в анализе простых «необходимых условий» проявления мошенничества, когда используется хотя бы одна известная характеристика причины. Второй этап (в режиме офлайн) состоит в выявлении причин, позволяющих провести анализ источников мошеннических схем.

Один из подходов к выбору «необходимых условий» состоит в построении множества подцелей исходной цели проекта (структурный метод построения проекта [7]). Каждая подцель описывается диаграммой на рис. 1, и реализации подцелей должны образовывать полный функционал цели. Это является необходимым, но не достаточным условием достижения цели, так как при таком подходе отсутствует компонент согласования всех подцелей в единую систему. Однако такой подход значительно упрощает анализ выполнения проекта на предмет поиска мошенничества. Если признаки мошенничества будут обнаружены в реализации хотя бы одной из подцелей, то это значит, что мошенничество присутствует в реализации всего проекта.

Аналогично в реализации каждого этапа в любой из подцелей можно выделять простые «необходимые условия» нарушения причинно-следственных связей.

Таким образом, получается множество «необходимых условий», нарушение которых свидетельствует о наличии мошенничества. Это множество «необходимых условий» можно назвать метаданными [8, 9] для контроля проекта на выявление мошенничества.

6 Заключение

В поиске противоречий необходимо от транзакций, соответствующих следствиям причинно-следственных связей, переходить к анализу причин наблюдаемых следствий. Это сложная задача, которая связана с описанием причин определенных свойств.

В работе представлена модель, позволяющая строить множество необходимых условий соответствия наблюдаемого следствия вызвавшей его причине. Этот подход делает поиск противоречий вполне вычислимой задачей, но не гарантирует успех.

Литература

1. Грушо А. А., Зацаринный А. А., Тимонина Е. Е. Блокчейны цифровой экономики на базе системы ситуационных центров и централизованного консенсуса // Радиолокация, навигация, связь: Мат-лы XXV Междунар. научн.-технич. конф. — Воронеж: Издательский дом ВГУ, 2019. Т. 6. С. 183–191.
2. Grusho A., Zatsarinny A., Timonina E. A system approach to information security in distributed ledgers on the situational centers platform. — Lecture notes in computer science ser. — Springer, 2019 (in press).
3. Финн В. К. Искусственный интеллект: Методология, применения, философия. — М.: Красанд, 2011. 448 с.
4. Анишаков О. М., Фабрикантова Е. Ф. ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания. — М.: Либроком, 2009. 432 с.
5. Poelmans J., Elzinga P., Viaene S., Dedene G. Formal concept analysis in knowledge discovery: A survey // Conceptual structures: From information to intelligence / Eds. M. Croitoru, S. Ferré, and D. Lukose. — Lecture notes in computer science ser. — Berlin—Heidelberg: Springer, 2010. Vol. 6208. P. 139–153.
6. Панкратова Е. С., Финн В. К. Автоматическое порождение гипотез в интеллектуальных системах. — М.: Либроком, 2009. 528 с.
7. Денисов А. А., Колесников Д. Н. Теория больших систем управления. — Л.: Энергоиздат, 1982. 488 с.
8. Грушо А. А., Грушо Н. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. Параметризация в прикладных задачах поиска эмпирических причин // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 62–66.
9. Грушо А. А., Грушо Н. А., Левыкин М. В., Тимонина Е. Е. Методы идентификации захвата хоста в распределенной информационно-вычислительной системе, защищенной с помощью метаданных // Информатика и её применения, 2018. Т. 12. Вып. 4. С. 41–45.

Поступила в редакцию 03.04.19

ARCHITECTURAL DECISIONS IN THE PROBLEM OF IDENTIFICATION OF FRAUD IN THE ANALYSIS OF INFORMATION FLOWS IN DIGITAL ECONOMY

A. A. Grusho, M. I. Zabezhailo, N. A. Grusho, and E. E. Timonina

Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: An approach to a research of some types of fraud in digital economy with the usage of relationships of cause and effect is formulated. In all types of the considered frauds, the discrepancy between the purposes of financial transactions and actual cost of achievement of these purposes has to be observed. Data on transactions can be collected by observing information flows in which these transactions are reflected. The architecture of data collection and their analysis can be organized by means of the distributed ledgers with the centralized consensus that allows creating an analog of the electronic account book fixing financial and economic activity of subjects of digital economy in the region. The methods of fraud identification considered are based on the contradictions between actions described in transactions and information, which is contained in plans, standards, precedents, etc. The method based on a simplified scheme of implementation of the abstract project is considered. For identification of contradictions, it is necessary to carry out the analysis from the effect to the cause, i. e., to look for anomalies in information describing the generation of the observed effects. It is shown how in implementation of the project it is possible to allocate simple “necessary conditions” of violation of cause and effect relationships, i. e., a set of “necessary conditions” violation of which demonstrates fraud existence. It is possible to call this set of “necessary conditions” by metadata for control of the project for fraud identification.

Keywords: digital economy; information flows; relationships of reason and effect; detection of fraudulent schemes

DOI: 10.14357/19922264190204

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (projects 18-29-03081 and 18-07-00274).

References

1. Grusho, A. A., A. A. Zatsarinny, and E. E. Timonina. 2019. Blokcheyny tsifrovoy ekonomiki na baze sistemy situatsionnykh tsevtrov i tsentralizovannogo konsensusa [Blockchains of digital economy on the basis of the system of the situational centres and the centralized consensus]. *25th Scientific and Technical Conference (International) “Radar-Location, Navigation, Communication” Proceedings*. Voronezh: VSU Pubs. 6:183–191.
2. Grusho, A., A. Zatsarinny, and E. Timonina. 2019 (in press). A system approach to information security in distributed ledgers on the situational centers platform. Lecture notes in computer science ser. Springer.
3. Finn, V. K. 2011. *Iskusstvennyy intellekt: Metodologiya, primeneniya, filosofiya* [Artificial intelligence: Methodology, applications, philosophy]. Moscow: KRASAND. 448 p.
4. Anshakov, O. M., and E. F. Fabrikantova. 2009. *DSM-metod avtomaticheskogo porozhdeniya gipotez: Logicheskoe i epistemologicheskoe osnovaniya* [JSM-method of automatic hypothesis generation: Logical and epistemological]. Moscow: KD LIBROKOM. 432 p.
5. Poelmans, J., P. Elzinga, S. Viaene, and G. Dedene. 2010. Formal concept analysis in knowledge discovery: A survey. *Conceptual structures: From information to intelligence*. Eds. M. Croitoru, S. Ferré, and D. Lukose. Lecture notes in computer science ser. Berlin–Heidelberg: Springer. 6208:139–153.
6. Pankratov, E. S., and V. K. Finn. 2009. *Avtomaticheskoe porozhdenie gipotez v intellektual'nykh sistemakh* [Automatic hypotheses generation in intelligent systems]. Moscow: KD LIBROKOM. 528 p.
7. Denisov, A. A., and D. N. Kolesnikov. 1982. *Teoriya bol'shikh sistem upravleniya* [Theory of big control systems]. Leningrad: Energoizdat. 488 p.
8. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2018. Parametrizatsiya v prikladnykh zadachakh poiska empiricheskikh prichin [Parametrization in applied problems of search of the empirical reasons]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):62–66.
9. Grusho, A. A., N. A. Grusho, M. V. Levykin, and E. E. Timonina. 2018. Metody identifikatsii zakhvata khosta v raspredelennoy informatsionno-vychislitel'noy sisteme, zashchishchennoy s pomoshch'yu metadannykh [Methods of identification of host capture in the distributed information system which is protected on the base of meta data]. *Informatika i ee Primeneniya — Inform. Appl.* 12(4):41–45.

Received April 3, 2019

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; grusho@yandex.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; m.zabezhailo@yandex.ru

Grusho Nikolai A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; info@itake.ru

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

О ТЕОРЕТИКО-АВТОМАТНЫХ МОДЕЛЯХ БЛОКЧЕЙН-СРЕДЫ*

В. С. Анашин¹

Аннотация: Рассматриваются методы анализа и моделирования блокчейн-среды, основанные на теоретико-автоматных моделях, в первую очередь на так называемых «автоматах с метками времени» (timed automata). Также предлагается новая версия автоматов с метками времени, позволяющая избежать некоторых неудобств моделирования с помощью классических автоматов с метками времени, а при моделировании блокчейн-среды на основе последних приходится использовать переменные разных типов, действительные и булевы, что вызывает ряд сложностей как теоретического, так и практического характера. Предлагаемый подход основан на применении 2-адического анализа, что дает возможность использовать переменные одного и того же типа, а именно булева.

Ключевые слова: блокчейн-среда; смарт-контракт; автомат с метками времени

DOI: 10.14357/19922264190205

1 Введение

При построении математических моделей блокчейн-среды теоретико-автоматная модель возникает естественным образом, поскольку функционирование блокчейн-среды — это детерминированный процесс, в ходе которого решение о включении или невключении блока в реестр зависит как от предыдущих блоков (булев тип данных), так и от времени (тип данных — действительные числа). Таким образом, если рассматривать текущее состояние реестра как предыдущее состояние, а состояние реестра сразу после включения в него очередного блока как последующее состояние, то процесс изменения содержимого реестра можно описать с помощью понятия «автомат с метками времени» (timed automaton), для которого далее в тексте статьи используется термин Т-автомат или кратко — ТА.

Основы математической теории ТА были заложены в работе [1]. В плане математических моделей блокчейн-среды на основе ТА представляет интерес работа [2], поскольку в ней моделируется функционирование смарт-контрактов в биткойн-среде (последняя представляет собой частный случай блокчейн-среды), что является одним из наиболее важных моментов для полноценного функционирования цифровой экономики на основе блокчейн-среды (а не только для реализации криптовалют). Отметим, что в этой работе для моделирования смарт-контрактов использован язык Uppaal, разработанный для описания и верификации моделей, основанных на ТА (см. описание Uppaal в [3]). На основе этой же модели описыва-

ется и биткойн-среда [4]. Подчеркнем, что с точки зрения математического описания и моделирования любой юридически правильно составленный контракт можно рассматривать как конечный автомат (см., например, [5]). Более того, известны методы извлечения описания смарт-контрактов как конечных автоматов из функционирующей во времени блокчейн-среды [6].

Сказанное означает, что функционирование смарт-контрактов в блокчейн-среде можно рассматривать как взаимодействие автоматов во времени, т. е. ТА можно рассматривать как релевантную модель описания такого взаимодействия.

Отметим, что моделирование функционирования смарт-контрактов в блокчейн-среде является одним из важных методов проверки стойкости против компрометации. Смарт-контракт использует входные данные от других смарт-контрактов, пользователей, а также о текущем времени и выдает выходные данные, которые используются другими пользователями и/или другими смарт-контрактами, а потому ошибочное функционирование одного смарт-контракта может привести к сбою в работе всех связанных с ним смарт-контрактов и узлов сети. Однако смарт-контракт может быть очень сложно устроен даже уже на уровне юридического документа, не говоря уже о его программной реализации, поэтому требуется тщательная проверка правильного функционирования смарт-контракта и в юридическом плане, и как компьютерной программы. Такую проверку достаточно сложно выполнить вручную, однако моделирование смарт-контракта позволяет поставить ряд машинных

* Работа выполнена при поддержке РФФИ (проект 18-20-03124).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, anashin@iisi.msu.ru

экспериментов для изучения поведения смарт-контракта как автомата при подаче на него тех или иных входных данных, т. е. смоделировать его поведение в самых разных условиях. Именно эти соображения послужили мотивацией работы [2], в которой используется теоретико-автоматная модель смарт-контракта на основе ТА, где время представляется действительными числами.

В данной работе предлагается принципиально иной подход к построению теоретико-автоматных моделей функционирования блокчейн-среды (в частности, функционирования смарт-контрактов в этой среде) как автомата с временными метками, в котором физическое время представляется не действительными, а 2-адическими числами. Такой подход представляется автору оправданным и плодотворным по нескольким причинам:

- итоговая модель представляет собой автомат в стандартном понимании этого термина; при этом задаваемое автоматом преобразование слов реализуется не в виде таблицы переходов состояний, а в виде программы без ветвления, представляющей собой последовательность стандартных команд процессора, а именно: арифметических и поразрядных логических операций;
- поскольку полученный автомат является «обычным» автоматом, описание его функционирования может быть сведено к изучению функции, заданной и принимающей значения в пространстве целых 2-адических чисел ввиду того, что каждая детерминированная (по С. В. Яблонскому) функция (т. е. функция, задаваемая автоматом) есть p -адическая функция, удовлетворяющая условию Липшица с константой 1 относительно подходящей p -адической метрики, и обратно: все такие функции являются детерминированными (см., например, [7]);
- сказанное дает возможность применять к изучению таких автоматов (а значит, и к изучению функционирования смарт-контрактов в блокчейн-среде) развитый аппарат p -адического анализа и, шире, алгебраической динамики в духе монографии [8];
- наконец, 2-адическое (и, шире, p -адическое) время является хоть и не общепринятой, но тем не менее довольно широко используемой и во многих случаях релевантной математической моделью физического времени, активно изучаемой уже почти три десятилетия в рамках p -адической математической физики (см. обзорную статью [9], а также соответствующие разделы и ссылки в [8]).

2 Автоматы и языки

Понятие «автомат» в русскоязычной литературе используется в различных смыслах, которым соответствуют английские термины *state machine*, *sequential machine*, *transducer* и т. д., поэтому во избежание недоразумений введем определения, используемые в данной работе. Везде далее под *алфавитом* понимается конечное непустое множество, содержащее хотя бы два элемента.

Определение 2.1. *Автомат-определитель* (далее — *d-автомат*) — это кортеж $\langle \mathcal{I}, \mathcal{S}, \mathcal{F}, S, s_0 \rangle$, где

- (1) \mathcal{I} есть *входной* алфавит;
- (2) \mathcal{S} есть непустое (необязательно конечное) множество, называемое множеством *состояний*;
- (3) \mathcal{F} есть конечное непустое подмножество множества \mathcal{S} , называемое множеством *принимающих состояний*;
- (4) $s_0 \in \mathcal{S}$ — *начальное* состояние;
- (5) $S: \mathcal{I} \times \mathcal{S} \rightarrow \mathcal{S}$ — *функция перехода*.

Автомат-определитель называется *конечным*, если конечно множество \mathcal{S} его состояний.

Множество всех конечных последовательностей $\mathbf{W}(\mathcal{I})$ над множеством \mathcal{I} называется множеством *слов*. Отметим, что $\mathbf{W}(\mathcal{I})$ не содержит пустого слова (слова нулевой длины) \emptyset ; полагаем $\mathbf{W}_0(\mathcal{I}) = \mathbf{W}(\mathcal{I}) \cup \{\emptyset\}$. Используем стандартное определение языка, распознаваемого *d-автоматом*, и регулярного языка (языка, распознаваемого конечным *d-автоматом*) (см., например, [10]). Отметим, что данному понятию конечного *d-автомата* в [10] соответствует понятие DFA — *deterministic finite automaton*.

Определение 2.2. *Автомат-преобразователь* (далее — *f-автомат*) есть кортеж $\langle \mathcal{I}, \mathcal{S}, \mathcal{O}, S, O, s_0 \rangle$, где $\mathcal{I}, \mathcal{S}, S$ и s_0 те же, что и в определении 2.1, \mathcal{O} — *выходной* алфавит, а $O: \mathcal{I} \times \mathcal{S} \rightarrow \mathcal{O}$ — *функция выхода*.

Вышеприведенное определение соответствует понятию автомата Мили, или, что то же самое, понятию 1-равномерного преобразователя (*1-uniform transducer*) из [10], с той лишь разницей, что множество состояний автомата в смысле определения 2.2 может быть и бесконечным; если же множество состояний конечно, то определение 2.2 превращается в стандартное определение автомата Мили (*Mealy sequential machine*).

Каждый преобразователь естественным образом задает отображение множества $\mathbf{W}(\mathcal{I})$ во множество $\mathbf{W}(\mathcal{O})$, а каждый *d-автомат* задает отображение

множества $\mathbf{W}(\mathcal{I})$ во множество $\mathcal{F} \cup R$, где $R \notin \mathcal{S}$, и в R отображаются те и только те слова, которые не принимаются d-автоматом.

3 Автоматы и время

T-автоматы, они же автоматы с метками времени, или *timed automata* в англоязычной литературе, введенные в [1], использовались для моделирования функционирования блокчейн-среды, поскольку разные узлы, функционирующие в блокчейн-среде, получают очередной блок, вообще говоря, в разные моменты времени. Целый ряд атак на блокчейн основан именно на факте «разновременности» получения очередных блоков пользователями.

Чтобы ввести понятие T-автомата, сначала понадобится определение (бесконечного) слова с метками времени (*timed word*), или, для краткости, t-слова.

Определение 3.1. (Бесконечное) слово с метками времени (t-слово) над алфавитом \mathcal{A} есть (бесконечная) последовательность пар $((a_i, \tau_i))_{i=0}^{\infty}$, где $a_i \in \mathcal{A}$, $\tau_i \in \mathbb{R}_{\geq 0}$, причем последовательность $(\tau_i)_{i=0}^{\infty}$ действительных неотрицательных чисел τ_i строго и неограниченно возрастает: $\tau_0 < \tau_1 < \tau_2 < \dots$ и $\lim_{i \rightarrow \infty} \tau_i = \infty$.

На содержательном уровне слово с метками времени $((a_i, \tau_i))_{i=0}^{\infty}$ интерпретируется как последовательность символов $(a_i)_{i=0}^{\infty}$ алфавита \mathcal{A} , поданных в автомат в моменты времени τ_0, τ_1, \dots соответственно.

Ниже, следуя [1], дадим определение T-автоматов. В этом определении будет использоваться понятие *таблицы переходов состояний с метками времени*, а также понятие *временных ограничений* на переходы.

Определение 3.2. Пусть T есть (счетное) множество переменных, называемых далее *временными переменными*, пусть $t_j \in T$. *Временным ограничением* называется любая булева комбинация предикатов вида $t_j \leq q$ и $q \leq t_j$, где $q \in \mathbb{Q}_{\geq 0}$ — неотрицательные рациональные константы, а символ \leq интерпретируются обычным образом как бинарное отношение «меньше либо равно». Пусть даны: (входной) алфавит \mathcal{I} , конечное множество \mathcal{S} состояний, конечное множество \mathcal{C} , называемое множеством *таймеров*, и множество $\Phi(\mathcal{C})$ временных ограничений от временных переменных $t_1, \dots, t_{|\mathcal{C}|}$. Тогда *таблица переходов состояний с метками времени* (далее — ТПТ) есть некоторое множество кортежей вида (s, a, s', G, φ) , где $s, s' \in \mathcal{S}$, $\varphi \in \Phi(\mathcal{C})$, $G \in 2^{\mathcal{C}}$ — подмножество множества таймеров (возможно, и пустое).

Определение 3.3. *Детерминированный автомат с метками времени* (далее — T-автомат) есть кортеж $\langle \mathcal{I}, \mathcal{S}, \mathcal{E}, \mathcal{C}, \Phi(\mathcal{C}), s_0 \rangle$, где $\mathcal{I}, \mathcal{S}, s_0$ — те же, что и в определении 2.1, \mathcal{S} — конечное множество, \mathcal{C} и $\Phi(\mathcal{C})$ — из определения 3.2, а \mathcal{E} — ТПТ в смысле определения 3.2, причем

- (1) в начальном состоянии s_0 текущие значения всех временных переменных равны 0;
- (2) для любых $a \in \mathcal{I}$, $s \in \mathcal{S}$ и любой пары элементов ТПТ \mathcal{E} вида $(s, a, *, *, \varphi_1)$, $(s, a, *, *, \varphi_2)$ временные ограничения φ_1 и φ_2 являются взаимно исключающими, т. е. $\varphi_1 \wedge \varphi_2$ тождественно ложно.

Замечание. Если в определении 3.3 допустить наличие нескольких начальных состояний и опустить условие (2), то получим определение недетерминированного T-автомата.

4 d- и f-автоматы вместо T-автоматов в моделях блокчейн-среды

В этом разделе обсудим возможность сведения T-автоматной модели блокчейн-среды, используемой в литературе (например, в работах [2, 4]), и покажем, что описанные T-автоматные модели блокчейн-среды могут быть «без потери точности» сведены к существенно более простым моделям, а именно: к «обычным» f-автоматам в смысле определения 2.2. Отметим сразу, что «физической основой» для такой «аппроксимации» функционирования блокчейн-среды с помощью более простой f-автоматной модели, чем с помощью используемой в литературе более сложной T-автоматной, служит следующее ограничение: *в реальной жизни время, разделяющее два следующих один за другим события, не может быть произвольно малым*, оно всегда ограничено снизу некоторой величиной. Например, в квантовой физике предполагается, что события, разделенные планковским временем, т. е. промежутком примерно в $5,4 \cdot 10^{-44}$ с, происходят одновременно, и, более того, в настоящее время минимальный интервал времени, доступный физическому измерению, составляет примерно 10^{-20} с. Отсюда следует, что любой временной интервал *кратен* некоторому минимальному временному интервалу (в предельном случае — планковскому времени) и, таким образом, с точностью до множителя, равного длине этого минимального интервала, является *натуральным числом*.

С другой стороны, T-автоматные модели позволяют рассматривать поведение моделируемой системы в «предельных» ситуациях; например, когда

временной промежутков между соседними «событиями» стремится к 0 (см. [1, example 3.22]). Рассмотрение «предельного» поведения часто оказывается очень полезным для описания не только качественных, но нередко и количественных характеристик моделируемой системы. Возможность «перехода к пределу» в Т-автоматных моделях основана на том, что множество $\mathbb{Q}_{\geq 0}$ всюду плотно в множестве $\mathbb{R}_{\geq 0}$ относительно обычной действительной метрики.

Таким образом, для моделирования блокчейн-среды хотелось бы построить аналог ТА, в которых время, с одной стороны, было бы «дискретным», т. е. в качестве «исходных» меток времени выступали бы натуральные, а не рациональные (как в определении ТА) числа, но, тем не менее, чтобы сохранялась и возможность «предельного перехода» как в целях получения описания поведения всей системы во времени на (хотя бы) качественном уровне, так и получения оценок «точности» модели. Поскольку речь идет о предельном переходе, то необходимо задать некоторую метрику на множестве всех натуральных чисел, относительно которой такой предельный переход был бы возможен и относительно которой натуральные числа образовывали бы всюду плотное подмножество подобно тому, как множество $\mathbb{Q}_{\geq 0}$ является всюду плотным подмножеством в $\mathbb{R}_{\geq 0}$ относительно действительной метрики. Такие метрики существуют: это p -адические метрики.

Будем рассматривать t -слова с метками времени из $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ (а не из $\mathbb{R}_{\geq 0}$, см. определение 3.1) над конечным алфавитом \mathcal{A} , содержащим хотя бы два символа. Без ограничения общности можно считать, что если $p \geq 2$ — это мощность алфавита \mathcal{A} , то символами алфавита \mathcal{A} являются числа $0, 1, \dots, p - 1$. Заметим, что такие t -слова представляют собой частный случай так называемых *слов с данными* (data word), введенных в [11], а именно: когда множество данных \mathcal{D} совпадает с \mathbb{N}_0 . Введенные в определении 3.1 слова с метками времени также представляют собой слова с данными для случая, когда данные лежат в $\mathbb{R}_{\geq 0}$. На основе понятия слов с данными естественным образом вводится понятие *языка с данными* (data language), а также *автомата с данными* (data automaton, далее — D-автомат).

Определение 4.1. Автомат \mathcal{A} с данными \mathcal{D} есть кортеж $\mathcal{A} = \langle \mathcal{I}, \mathcal{S}, \mathcal{F}, \mathcal{T}, k, \sim, s_0 \rangle$, где $\mathcal{I}, \mathcal{S}, \mathcal{F}, s_0$ — те же, что и в определении 2.1;

- k есть натуральное число (называемое числом регистров данных);
- \sim есть отношение эквивалентности конечного индекса, определенное на \mathcal{D}^k ;

— $\mathcal{T} \subseteq \mathcal{S} \times \mathcal{D}^k / \sim \times \mathcal{I} \times \mathcal{D}^k / \sim \times \mathcal{S}$ есть конечное множество переходов;

— \mathcal{U} есть множество модификаций состояний регистров $\text{upd}: \mathcal{D}^k \rightarrow \mathcal{D}^k$, удовлетворяющие следующим ограничениям:

- для любого кортежа

$$(s, g, a) \in \mathcal{S} \times \mathcal{D}^k / \sim \times \mathcal{I}$$

существует (единственная) модификация регистров $\text{upd} \in \mathcal{U}$ такая, что если $(s, g, a, \text{upd}', g', s') \in \mathcal{T}$ для некоторого $\text{upd}' \in \mathcal{U}$, то $\text{upd}' = \text{upd}$;

- если $(s, g, a, \text{upd}, g', s), (s, g, a, \text{upd}, g', s') \in \mathcal{T}$, то $s' = s$.

Как показано в [11, теорема 13], для любого детерминированного Т-автомата из определения 3.3, имеющего n таймеров (см. определение 3.2), существует D-автомат с $2n + 2$ регистрами, распознающий в точности тот же самый язык. Далее построим f-автоматы (см. определение 2.2), аппроксимирующие с любой наперед заданной точностью (в некотором точно определенном ниже смысле) данный детерминированный Т-автомат и тем самым сведем задачу моделирования функционирования блокчейн-среды (в первую очередь, моделирования смарт-контрактов) Т-автоматами к моделированию «обычными» автоматами с двоичными входами и двоичными выходами. Для этого сначала понадобится ввести новый частный тип D-автоматов, а именно: *автоматы с p -адическим временем*.

Зафиксируем некоторое простое число p (в контексте данной статьи наиболее важным является случай $p = 2$) и рассмотрим в качестве меток данных (в словах с данными) *целые p -адические числа*, т. е. элементы *пространства \mathbb{Z}_p целых p -адических чисел*. С введением в теорию p -адических чисел и p -адический анализ можно ознакомиться, например, по вводным главам в [8]. Здесь же введем лишь самые необходимые понятия из p -адического анализа, и притом на неформальном уровне.

Множество \mathbb{Z}_p можно рассматривать как множество $\mathbf{W}^\infty(\mathcal{A})$ всех бесконечных (в одну сторону, в данной статье — влево) слов над алфавитом $\mathcal{A} = \{0, 1, \dots, p - 1\}$, символы которого можно считать элементами кольца $\mathbb{Z}/p\mathbb{Z}$ вычетов по модулю p , т. е. элементами поля из p элементов. На множестве \mathbb{Z}_p можно задать операции сложения и умножения с помощью стандартных «школьных» алгоритмов сложения и умножения «в столбик» чисел, представленных в системе счисления с основанием p . Если $p = 2$, то бесконечные бинарные строчки можно мыслить себе как представления чисел

в обобщенном *обратном двоичном коде* (см., например, [12, с. 213]). Использование обобщенного обратного двоичного кода дает возможность записывать в регистр бесконечной длины как все целые неотрицательные числа (им соответствуют строчки с конечным числом единиц), так и все целые отрицательные числа (им соответствуют строчки с конечным числом нулей), а также все рациональные числа, представимые в виде простых несократимых дробей с нечетными знаменателями (им соответствуют периодические с какого-то момента строчки).

Множество \mathbb{Z}_p является полным компактным метрическим пространством относительно p -адической метрики d_p , которая задается следующим образом: $d_p(\mathbf{a}, \mathbf{b}) = 1/p^i$ тогда и только тогда, когда $\mathbf{a} = \dots a_{i+1} a_i c_{i-1} \dots c_0$, $\mathbf{b} = \dots b_{i+1} b_i c_{i-1} \dots c_0$ и $a_i \neq b_i$ (по определению $d_p(\mathbf{a}, \mathbf{b}) = 0$, если такого i не существует, т.е. если бесконечные слова \mathbf{a} и \mathbf{b} совпадают). Абсолютная величина $\|\mathbf{a}\|_p$ вводится стандартным образом как расстояние до числа 0 (этому числу соответствует бесконечная строчка из одних только нулей): $\|\mathbf{a}\|_p = d_p(\mathbf{a}, 0)$.

Можно ввести понятия «приведения по модулю p^n » и «сравнения по модулю p^n » для целых p -адических чисел, а именно: приведение по модулю p^n бесконечного слова в алфавите $\mathcal{A} = \{0, 1, \dots, p-1\}$ означает всего лишь переход к конечному начальному отрезку длины n этого бесконечного слова, т.е. $\text{mod } p^n: \mathbf{W}^\infty(\mathcal{A}) \rightarrow \mathbf{W}^n(\mathcal{A})$, где $\mathbf{W}^n(\mathcal{A})$ есть множество всех слов длины n над алфавитом \mathcal{A} . Отметим, что элементы множества $\mathbf{W}^n(\mathcal{A})$ естественным образом отождествляются с числами $0, 1, \dots, p^n - 1$, представленными в системе счисления с основанием p , а эти числа, в свою очередь, отождествляются с элементами кольца $\mathbb{Z}/p^n\mathbb{Z}$ вычетов по модулю p^n . Более того, оказывается, что любое отображение $f_{\mathfrak{A}}: \mathbf{W}^\infty(\mathcal{A}) \rightarrow \mathbf{W}^\infty(\mathcal{A})$, задаваемое автоматом-преобразователем \mathfrak{A} , входной и выходной алфавиты которого суть \mathcal{A} , т.е. $\mathcal{I} = \mathcal{O} = \mathcal{A} = \{0, 1, \dots, p-1\}$ (см. определение 2.2) есть функция, определенная на \mathbb{Z}_p и принимающая значения в \mathbb{Z}_p , которая удовлетворяет p -адическому условию Липшица с константой 1 (и, следовательно, является непрерывной относительно метрики d_p функцией): $\|f_{\mathfrak{A}}(\mathbf{a}) - f_{\mathfrak{A}}(\mathbf{b})\|_p \leq \|\mathbf{a} - \mathbf{b}\|_p$ для любых $\mathbf{a}, \mathbf{b} \in \mathbb{Z}_p$. Верно и обратное: любое отображение из \mathbb{Z}_p в \mathbb{Z}_p , удовлетворяющее p -адическому условию Липшица с константой 1, задается некоторым автоматом-преобразователем (не обязательно конечным), входной и выходной алфавиты которого суть $\{0, 1, \dots, p-1\}$ (см., например, [7]).

Отметим, что каждый из двух типов автоматов: автоматы-определители, т.е. d -автоматы из опре-

деления 2.1, и автоматы-преобразователи, т.е. f -автоматы из определения 2.2, — может быть сведен один к другому (см. подробнее [10, теорема 4.3.2]). Таким образом, в случае $p = 2$ задачи о d -автоматах и распознаваемых ими языках могут быть сведены к задачам о функциях, удовлетворяющих 2-адическому условию Липшица с константой 1. Такие функции называются в литературе также функциями треугольного вида, двоичными совместимыми функциями, T-функциями.

В контексте данной статьи весьма важным является тот факт, что T-функции допускают особую простую реализацию в виде компьютерной программы, а именно: *компьютерная реализация детерминированной функции автомата, входной и выходной алфавиты которого состоят из двух символов, не требует реализации его таблицы переходов состояний, а может быть записана (в том числе и для автоматов с бесконечным числом состояний) в виде T-функции, которая, в свою очередь, представляет собой программу без ветвлений, состоящую из последовательности стандартных компьютерных команд, таких как арифметические команды (сложение и умножение натуральных чисел) и поразрядные логические команды OR, AND, XOR, NOT, а также производных от них команд, таких как сдвиг в сторону старших разрядов, маскирование и ряда других, таких как деление на нечетные числа, возведение нечетных чисел в степень и др.* [8].

Сказанное остается в силе и для автоматов, входной и выходной алфавиты которых состоят из соответственно 2^n и 2^m символов, поскольку такие автоматы можно рассматривать как автоматы, имеющие n двоичных входов и m двоичных выходов, а значит, как многомерные T-функции, т.е. как непрерывные относительно 2-адической метрики отображения из \mathbb{Z}_2^n в \mathbb{Z}_2^m , удовлетворяющие многомерному 2-адическому условию Липшица с константой 1. Для T-функций имеется хорошо развитая математическая теория, основанная на 2-адическом анализе и имеющая многочисленные (в первую очередь — криптографические) приложения (см. [8]).

Дадим теперь формальное определение автомата с 2-адическим временем.

Определение 4.2. D-автомат из определения 4.1 назовем *автоматом с 2-адическим временем* (\mathbb{Z}_2 -автоматом), если $\mathcal{I} = \{0, 1\}$ и $\mathcal{D} = \mathbb{Z}_2$.

Разумеется, похожим образом можно сформулировать и понятие \mathbb{Z}_2 -автомата с входным алфавитом из 2^r символов, т.е. \mathbb{Z}_2 -автомата с r двоичными входами. Язык, распознаваемый \mathbb{Z}_2 -автоматом, определяется обычным образом на основе определения 4.2.

Говоря неформально, любой T-автомат (см. определение 3.3) можно рассматривать как «автомат с двумя входами»: временным и алфавитным, где на каждом такте работы подается на алфавитный вход очередной символ входного слова, а на временной вход — действительное число, служащее меткой времени этого входного символа. С этой точки зрения \mathbb{Z}_2 -автомат тоже имеет два входа; при этом на алфавитный вход подается символ входного алфавита, т. е. 0 или 1, а на временной вход — метка времени, т. е. целое 2-адическое число.

Все t-слова могут быть равномерно приближены словами с 2-адическими метками времени (далее — \mathbb{Z}_2 -словами) в следующем смысле. Вначале все символы входного алфавита T-автомата пронумеруем и запишем в виде двоичных представлений соответствующих чисел. Таким образом, можно считать, что на алфавитный вход автомата всегда подается r бинарных последовательностей, где r — число двоичных разрядов, необходимых для записи всех символов входного алфавита.

Далее зафиксируем любое действительное $\varepsilon > 0$ и выберем рациональные числа $z_i(\mathbf{w})$, представимые в виде простых несократимых дробей с нечетными знаменателями (все эти рациональные числа лежат в \mathbb{Z}_2) так, чтобы $|\tau_i(\mathbf{w}) - z_i(\mathbf{w})| < \varepsilon$, где $\tau_i(\mathbf{w})$ есть i -я метка времени в t-слове \mathbf{w} . Такой выбор всегда можно сделать, например, следующим образом. Представим

$$\tau_i(\mathbf{w}) = \lfloor \tau_i(\mathbf{w}) \rfloor + (\tau_i(\mathbf{w}) - \lfloor \tau_i(\mathbf{w}) \rfloor),$$

где $\lfloor \tau_i(\mathbf{w}) \rfloor$ есть целая (с недостатком) часть числа $\tau_i(\mathbf{w})$. Выберем $h \in \mathbb{N}$ таким, чтобы $1/3^h < \varepsilon$, запишем дробную часть $(\tau_i(\mathbf{w}) - \lfloor \tau_i(\mathbf{w}) \rfloor)$ числа $\tau_i(\mathbf{w})$ в троичной системе счисления с точностью до h троичных разрядов после запятой. Тогда эта дробная часть есть число вида $c/3^h$, где $c \in \{0, 1, \dots, 3^h - 1\}$, и, следовательно, является целым 2-адическим числом. Прибавляя к полученному таким образом числу целую (с недостатком) часть $\lfloor \tau_i(\mathbf{w}) \rfloor$ числа $\tau_i(\mathbf{w})$, получаем целое 2-адическое число $z_i(\mathbf{w})$. В этом смысле каждое t-слово $\mathbf{w} = ((a_i, \tau_i))_{i=0}^{\infty}$ приближается с точностью не хуже чем ε словом $((a_i, z_i(\mathbf{w})))_{i=0}^{\infty}$, которое является входным \mathbb{Z}_2 -словом для \mathbb{Z}_2 -автомата с r алфавитными входами, причем алфавит каждого алфавитного входа бинарный.

Далее все \mathbb{Z}_2 -слова могут быть равномерно приближены \mathbb{Z}_2 -словами с метками времени из \mathbb{N}_0 (и даже из $\mathbb{Z}/2^h\mathbb{Z}$) с любой наперед заданной 2-адической точностью $1/2^h$. Действительно, для этого достаточно каждую из 2-адических меток времени в каждом \mathbb{Z}_2 -слове привести по модулю 2^h . Таким образом, на основе вышеописанной процедуры

«аппроксимации» t-автомата \mathbb{Z}_2 -автоматом можно построить последовательность \mathbb{Z}_2 -автоматов \mathfrak{A}_h с метками времени из $\mathbb{Z}/2^h\mathbb{Z}$, $h = 1, 2, 3, \dots$, аппроксимирующих в вышеуказанном смысле исходный t-автомат.

Используя описанную выше процедуру построения f-автомата на основе данного d-автомата, можно любому D-автомату сопоставить *детерминированную функцию с метками времени*, считая, например, что i -й символ выходного слова имеет ту же метку времени, что и i -й символ соответствующего ему входного слова. Таким образом на основе данного \mathbb{Z}_2 -автомата из определения 4.2 можно построить детерминированную функцию с метками времени из \mathbb{Z}_2 , полагая i -й выходной символ равным 1, если автомат находится в принимающем состоянии (т. е. в состоянии из множества \mathcal{F}), и 0 в противном случае.

Наконец, этим способом каждому из построенных выше аппроксимирующих автоматов \mathfrak{A}_h можно сопоставить детерминированную функцию с метками времени из $\mathbb{Z}/2^h\mathbb{Z}$. Итак, для данного T-автомата построена последовательность аппроксимирующих его (в вышеописанном смысле) T-функций, т. е. «обычных» f-автоматов в смысле определения 2.2, имеющих $r + h$ двоичных входов и $h + 1$ двоичный выход. Таким образом, доказано следующее

Предложение. *Каждый T-автомат аппроксимируется с любой наперед заданной точностью (в описанном выше смысле) некоторым f-автоматом над двухсимвольным алфавитом.*

5 Выводы

В данной работе показано, что для моделирования функционирования блокчейн-среды (в частности, моделирования работы смарт-контрактов), даже несмотря на то что эта среда функционирует в реальном физическом времени, нет необходимости прибегать к сложным (и достаточно ресурсоемким) моделям, основанным на концепции автоматов с метками времени, представляющими собой действительные числа, а достаточно ограничиться (без потери точности) моделированием этой среды с помощью детерминированных функций над 2-символьным алфавитом (известных также под названием T-функций), т. е. с помощью «обычных» автоматов с бинарным входным/выходным алфавитом. Эти функции могут быть реализованы в виде программ без ветвления, выполненных как последовательности стандартных команд любого процессора, что позволяет надеяться на относительную простоту их программной реализации и высокое быстродействие соответствующих программ.

Литература

1. Alur R., Dill D. The theory of timed automata // Real-time: Theory and practice / Eds. J. W. de Bakker, C. Huizing, W. P. de Roever, G. Rozenberg. — Lecture notes in computer science ser. — Springer, 1992. Vol. 600. P. 45–73.
2. Andrychowicz M., Dziembowski S., Malinowski D., Mazurek L. Modelling bitcoin contracts by timed automata // Formal modelling and analysis of timed systems / Eds. A. Legay, M. Bozga. — Lecture notes in computer science ser. — Springer, 2014. Vol. 8711: P. 7–22.
3. David A., Larsen K.G., Legay A., Mikučionis M., Poulsen D.B. Uppaal SMC tutorial // Int. J. Softw. Tools Te., 2015. Vol. 17. P. 397–415.
4. Chaudhary K., Fehnker A., van de Pol J., Stoelinga M. Modeling and verification of the bitcoin protocol // Electronic Proc. Theor. Comput. Sci., 2015. Vol. 196. P. 46–60.
5. Flood M.D., Goodenough O.R. Contract as automaton: The computational representation of financial agreements // SSRN Electronic J., 2015. doi: 10.2139/ssrn.2538224.
6. Guth F., Wüstholtz V., Christakis M., Müller P. Specification mining for smart contracts with automatic abstraction tuning // arXiv.org, 2018. arXiv:1807.07822 [cs.SE]. 12 p.
7. Anashin V. The non-Archimedean theory of discrete systems // Math. Comput. Sci., 2012. Vol. 6. P. 375–393.
8. Anashin V., Khrennikov A. Applied algebraic dynamics. — Gruyter expositions in mathematics ser. — Berlin – New York: Walter de Gruyter GmbH & Co, 2009. Vol. 49. 533 p.
9. Dragovich B., Khrennikov A. Yu., Kozyrev S.V., Volovich I.V. On p -adic mathematical physics // p -Adic Numbers Ultrametric Analysis Appl., 2009. Vol. 1. P. 1–17.
10. Allouche J.-P., Shallit J. Automatic sequences. Theory, applications, generalizations. — Cambridge: Cambridge University Press, 2003. 583 p.
11. Bouyer P., Petit A., Thérien D. An algebraic approach to data languages and timed languages // Inform. Comput., 2003. Vol. 182. P. 137–162.
12. Knuth D. The art of computer programming. Vol. 2: Seminumerical algorithms. — 3rd ed. — Addison-Wesley, 1997. 791 p.

Поступила в редакцию 09.02.19

ON AUTOMATA MODELS OF BLOCKCHAIN

V. S. Anashin

Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: The author considers automata models of blockchain, mostly based on timed automata. The author suggests a new version of timed automata that avoids some inconveniences that occur in modeling by using standard timed automata where time is represented by real numbers. In the latter case, one should use variables of two types, Boolean and real; when applied to blockchain modeling, this fact causes some difficulties both in obtaining theoretical estimates and in program implementation. The present approach is based on 2-adic analysis since in that case, both time and digital variables are of one type only; namely, Boolean.

Keywords: blockchain; smart contract; timed automaton

DOI: 10.14357/19922264190205

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (grant 18-29-03124).

References

1. Alur, R., and D. Dill. 1992. The theory of timed automata. *Real-time: Theory and practice*. Eds. J. W. de Bakker, C. Huizing, W. P. de Roever, and G. Rozenberg. Lecture notes in computer science ser. Springer. 600:45–73.
2. Andrychowicz, M., S. Dziembowski, D. Malinowski, and L. Mazurek. 2014. Modelling bitcoin contracts by timed automata. *Formal modelling and analysis of timed systems*. Eds. A. Legay and M. Bozga. Lecture notes in computer science ser. Springer. 8711:7–22.
3. David, A., K.G. Larsen, A. Legay, M. Mikučionis, and D. B. Poulsen. 2015. Uppaal SMC tutorial. *Int. J. Softw. Tools Te.* 17:397–415.
4. Chaudhary, K., A. Fehnker, J. van de Pol, and M. Stoelinga. 2015. Modeling and verification of the bitcoin protocol. *Electronic Proc. Theor. Comput. Sci.* 196:46–60.
5. Flood, M.D., and O.R. Goodenough. 2015. Contract as automaton: The computational representation of financial agreements. *SSRN Electronic J.* doi: 10.2139/ssrn.2538224.

6. Guth, F., V. Wüsthold, M. Christakis, and P. Müller. 2018. Specification mining for smart contracts with automatic abstraction tuning. arXiv:1807.07822 [cs.Se]. 12 p. Available at: <https://arxiv.org/abs/1807.07822> (accessed January 17, 2019).
7. Anashin, V. 2012. The non-Archimedean theory of discrete systems. *Math. Comput. Sci.* 6:375–393.
8. Anashin, V., and A. Khrennikov. 2009. *Applied algebraic dynamics*. Gruyter expositions in mathematics ser. Berlin – New York: Walter de Gruyter GmbH & Co. Vol. 49. 533 p.
9. Dragovich, B., A.Yu. Khrennikov, S.V. Kozyrev, and I.V. Volovich. 2009. On p -adic mathematical physics. *p-Adic Numbers Ultrametric Analysis Appl.* 1:1–17.
10. Allouche, J.-P., and J. Shallit. 2003. *Automatic sequences. Theory, applications, generalizations*. Cambridge: Cambridge University Press. 583 p.
11. Bouyer, P., A. Petit, and D. Thérien. 2003. An algebraic approach to data languages and timed languages. *Inform. Comput.* 182:137–162.
12. Knuth, D. 1997. *The art of computer programming. Vol. 2: Seminumerical algorithms*. 3rd ed. Addison-Wesley. 791 p.

Received February 9, 2019

Contributor

Anashin Vladimir S. (b. 1951) — Doctor of Science in physics and mathematics, professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; anashin@iisi.msu.ru

ОБ ОДНОМ МЕТОДЕ РЕШЕНИЯ СИСТЕМ КВАДРАТИЧНЫХ БУЛЕВЫХ УРАВНЕНИЙ, ИСПОЛЬЗУЮЩЕМ ЛОКАЛЬНЫЕ АФФИННОСТИ БУЛЕВЫХ ФУНКЦИЙ*

О. А. Логачев¹, А. А. Сукаев², С. Н. Федоров³

Аннотация: Как известно, вычислительная задача решения систем нелинейных уравнений над полем из двух элементов является NP-трудной. Этим обстоятельством обуславливается стремление исследователей разрабатывать алгоритмы ее решения, минимизирующие необходимые вычислительные ресурсы для тех или иных классов систем уравнений. В статье предлагается метод решения систем квадратичных булевых уравнений, использующий представление функций их аффинными нормальными формами, т.е. в некотором смысле аппроксимацию квадратичных функций кусочно-аффинными. Для каждого уравнения на основе такого представления строится набор систем небольшого числа линейных уравнений, а затем ищется совместная комбинация этих линейных систем для различных исходных уравнений. Исходная нелинейная задача, таким образом, сводится, по большому счету, к проверке совместности серии линейных систем от того же числа переменных. Метод может быть эффективно распараллелен и, несмотря на большую трудоемкость в худшем случае, допускает ряд эвристик, уменьшающих время его выполнения.

Ключевые слова: булева функция; система квадратичных булевых уравнений; разбиение векторного пространства; плоскость; локальная аффинность; аффинная нормальная форма; алгебраический криптоанализ

DOI: 10.14357/19922264190206

1 Введение

Фундаментом сложившегося за последние два десятилетия направления исследований, которое принято называть алгебраическим криптоанализом, служит идея, состоящая в том, чтобы описать используемые в анализируемой криптосистеме преобразования системой алгебраических уравнений (с некой секретной информацией в качестве неизвестных переменных) и затем решить эту систему. Сама эта идея далеко не нова, ее высказывал еще Клод Шеннон в 1940-х гг. [1], но более или менее активный и постоянный интерес к такой постановке вопроса наблюдается с середины 1990-х гг.

Для первой части этого подхода, состоящей в построении системы уравнений, вряд ли можно дать универсальные рекомендации. Все зависит от криптосистемы и от того, каким методом предполагается решать эту систему уравнений. При этом нередко структура исследуемой криптосистемы сама подсказывает, как составить соответствующую систему уравнений. Данная статья не касается это-

го вопроса. Следует только отметить, что, как правило, строится переопределенная система уравнений, левые части которых представляют собой полиномы над полем \mathbb{F}_2 или \mathbb{F}_{2^k} .

Что касается второй части подхода, проблема решения системы уравнений — это весьма важная задача для математики вообще, и поэтому довольно интенсивно изучалась. Однако для криптоанализа, в отличие от многих других областей математики, крайне важно не принципиальное наличие решения, а конструктивный метод его поиска, а также эффективность (невысокая трудоемкость) соответствующих алгоритмов, возможно, в ущерб универсальности и теоретической обоснованности метода.

Здесь можно выделить несколько основных классов (практических) методов решения систем полиномиальных уравнений: использование базисов Грёбнера [2, разд. 12.2], применение программных систем поиска выполняющего набора булевой формулы (SAT-solvers) [3], линеаризационные методы [2, разд. 12.3]. Основная идея последних — применение «линейных» методов к не-

* Работа выполнена при частичной поддержке РФФИ (проект 18-29-03124 мк).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ollog@inbox.ru

² Московский государственный университет имени М. В. Ломоносова, asukaev@gmail.com

³ Московский государственный университет имени М. В. Ломоносова, s.n.fedorov@yandex.ru

линейным системам, т. е. построение систем линейных уравнений, решение которых дает возможность найти решение исходной системы.

Скорее всего, именно к этому классу следует отнести описанный в настоящей работе метод решения системы m квадратичных уравнений от n переменных над полем \mathbb{F}_2 (см. систему (1) ниже), хотя он существенно, если не сказать — кардинально, отличается от ставшего уже классическим метода линеаризации.

Предлагается рассматривать вместо каждой функции f_i ($1 \leq i \leq m$) из левой части системы (1) ее аффинную нормальную форму — выражение, предполагающее, что векторное пространство \mathbb{F}_2^n разбито на плоскости (смежные классы по аффинным подпространствам) таким образом, что на каждой из этих плоскостей f_i совпадает с некоторой аффинной (линейной) функцией. Вопросу о построении таких выражений для функций системы посвящена работа [4], где представлен один из возможных способов их нахождения — вместе с оценками трудоемкости, поэтому дальнейшее изложение не касается этого этапа метода. Заметим, что решение системы уравнений лежит в пересечении некоторого набора плоскостей (π_1, \dots, π_m) , где π_i , $1 \leq i \leq m$, берется из указанного разбиения пространства для функции f_i . Идея метода состоит в том, чтобы, отбрасывая заведомо неподходящие комбинации плоскостей, попытаться угадать набор (π_1, \dots, π_m) . На этом наборе все уравнения системы можно заменить их аффинными «аппроксимациями», и тогда останется только решить линейную систему, которая, подчеркнем это, в стандартном варианте метода состоит не более чем из $m(\lfloor n/2 \rfloor + 1)$ уравнений от n неизвестных, т. е. в отличие от других известных линеаризационных методов, здесь число переменных не увеличивается. Еще одним преимуществом является то, что метод вполне естественным образом допускает распараллеливание. Кроме того, таким способом решение находится во всех случаях, хотя это может потребовать большого объема вычислений: в худшем случае трудоемкость, конечно, экспоненциальна (см. разд. 5). Чтобы преодолеть это препятствие, необходимо использовать специальные приемы. Подробно основная идея подхода описана в разд. 3, а соображения по его усовершенствованию приведены в разд. 4. Необходимо отметить, что подобный предлагаемому метод для систем уравнений, описывающих преобразования в фильтрующих генераторах, был описан в работе [5].

Данная работа ограничивается рассмотрением многочленов не более чем второй степени, поскольку существует алгоритм (см., например, [2, § 11.4.2]), который для произвольной системы по-

линомиальных уравнений фиксированной степени над полем из двух элементов за полиномиальное (от числа переменных) время строит такую систему *квадратичных* уравнений над тем же полем, что из ее решений однозначно восстанавливаются все решения исходной системы. Следует, однако, заметить, что если не накладывать ограничений на степень полиномов системы, то верхняя оценка трудоемкости упомянутого алгоритма становится уже экспоненциальной [2, § 11.4.6].

Отметим еще некоторые ограничения на рассматриваемую систему. Они не являются необходимыми для применения метода, однако несколько упрощают изложение, оставаясь при этом и довольно естественными в рамках криптоанализа. Предполагается, что число уравнений больше числа неизвестных, что, как правило, справедливо для систем уравнений, получаемых при анализе криптографических преобразований. Кроме того, рассматриваются лишь те системы уравнений, которые заведомо имеют решение и притом только одно. Это условие объясняется тем, что алгебраический криптоанализ имеет дело не с произвольными уравнениями, а лишь с описывающими преобразования некоторых объектов, например секретного ключа или открытого текста, в тех или иных криптосистемах; таким образом, решением системы уравнений является этот самый объект (ключ или открытый текст). Отметим здесь кстати, что в отличие от линеаризационных методов, где нужна система с достаточно большим количеством уравнений, чтобы компенсировать рост числа переменных, в предлагаемом методе чем меньше уравнений в исходной системе, тем лучше, — лишь бы они имели только одно общее решение.

2 Необходимые понятия и обозначения

В данной работе V_n — n -мерное векторное пространство над полем из двух элементов \mathbb{F}_2 . Знак \oplus будет использоваться для записи суммы по модулю 2 булевых переменных и операций сложения в \mathbb{F}_2 и покомпонентного сложения в V_n .

Далее под булевой функцией часто будет подразумеваться ее запись в виде полинома

$$f(x) = \bigoplus_{\varepsilon \in \{0,1\}^n} a_\varepsilon x^\varepsilon,$$

где $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$; $a_\varepsilon \in \mathbb{F}_2$; $x^\varepsilon = x_1^{\varepsilon_1} \cdots x_n^{\varepsilon_n}$ (считаем $x_i^0 = 1$ и $x_i^1 = x_i$).

В настоящей работе рассматриваются системы уравнений

$$\left. \begin{aligned} f_1(x_1, \dots, x_n) &= 0; \\ f_2(x_1, \dots, x_n) &= 0; \\ &\dots \\ f_m(x_1, \dots, x_n) &= 0 \end{aligned} \right\} \quad (1)$$

с квадратичными булевыми функциями $f_i, 1 \leq i \leq m$, и $m > n$. Предполагается, что все рассматриваемые системы квадратичных уравнений имеют единственное решение.

Плоскость π в V_n — это множество вида $v + L$, где v и L — соответственно вектор и подпространство пространства V_n . Сужение булевой функции f на плоскость π будем обозначать через $f|_\pi$. Таким образом, $f|_\pi: \pi \rightarrow \mathbb{F}_2$ и $f|_\pi(u) = f(u)$ для всех $u \in \pi$.

Локальной аффинностью булевой функции f называется такая плоскость π , что $f|_\pi$ можно продолжить до аффинной функции, т.е. существует аффинная функция l со свойством $f|_\pi = l|_\pi$.

Для произвольного разбиения $\Pi = \{\pi_1, \dots, \pi_\lambda\}$ пространства V_n на локальные аффинности булевой функции f от n переменных аффинной нормальной формой функции f называется выражение вида:

$$f(x) = \bigoplus_{j=1}^{\lambda} \chi_{\pi_j}(x) l_j(x), \quad (2)$$

где для каждого $j, 1 \leq j \leq \lambda$, функция l_j аффинна и $f|_{\pi_j}(x) = l_j|_{\pi_j}(x)$, а χ_{π_j} — характеристическая функция (индикатор) множества π_j . Функции l_j из этого выражения будем называть локальными аффинными аппроксимациями функции f . Длиной аффинной нормальной формы называется число плоскостей в разбиении Π , обозначим ее через $\lambda(\Pi)$.

Характеристическая функция плоскости в пространстве V_n имеет вполне определенный вид. Любая плоскость π , как уже отмечалось, может быть задана как множество решений системы d линейных уравнений (для некоторого d):

$$\left. \begin{aligned} h_1(x_1, \dots, x_n) &= 0; \\ h_2(x_1, \dots, x_n) &= 0; \\ &\dots \\ h_d(x_1, \dots, x_n) &= 0, \end{aligned} \right\} \quad (3)$$

где все $h_i(x)$ аффинные. Поскольку вектор x принадлежит плоскости π тогда и только тогда, когда все $h_i, 1 \leq i \leq d$, обращаются в нуль на нем, характеристическая функция π выражается следующим образом:

$$\chi_\pi(x) = \prod_{i=1}^d (h_i(x) \oplus 1).$$

Если система линейных уравнений задана в матричной форме — $xH \oplus (b_1, \dots, b_d) = 0, b_i = h_i(0)$, то выражение будет иметь вид:

$$\chi_\pi(x) = \prod_{i=1}^d (xH_i \oplus b_i \oplus 1),$$

где H_i — столбцы матрицы H .

Как видно из определения, аффинная нормальная форма представляет собой в некотором смысле кусочно-аффинную аппроксимацию булевой функции. На каждой локальной аффинности π_j из разбиения Π все, кроме одного, слагаемые в выражении (2) обращаются в нуль, и функция принимает вид $f(x) = \chi_{\pi_j}(x) l_j(x) = l_j(x)$ для всех $x \in \pi_j$.

Возможность заменить на плоскости π_j квадратичное уравнение $f(x) = 0$ линейным уравнением $l_j(x) = 0$ вместе с дописанной к нему системой (3) будет использоваться при решении систем полиномиальных уравнений.

3 Метод решения системы, использующий локальные аффинности

Пусть дана система квадратичных уравнений (1). Для применения описываемого далее метода следует получить аффинные нормальные формы для всех функций в левых частях уравнений системы. В работе [4] представлен один из возможных способов, кажущийся наиболее естественным в квадратичном случае.

3.1 Построение «локальных» систем линейных уравнений

Пусть для всех $f_i, 1 \leq i \leq m$, определены некоторые аффинные нормальные формы:

$$f_i(x) = \bigoplus_{j=1}^{\lambda(i)} \chi_{\pi_{i,j}}(x) l_{i,j}(x).$$

Здесь предполагается, что каждое разбиение пространства на локальные аффинности некоторым образом упорядочено:

$$\Pi_i = \{\pi_{i,1}, \dots, \pi_{i,\lambda(i)}\}, \lambda(i) = \lambda(\Pi_i)$$

(как именно лучше их упорядочить для дальнейшего использования — предмет отдельного исследования, несколько слов об этом будет сказано ниже). Исходя из этих аффинных нормальных форм можно для каждой пары i, j записать эквивалентную

уравнению $f_i = 0$ на π_{ij} систему линейных уравнений:

$$\left. \begin{aligned} l_{ij}(x) &= 0; \\ h_{ij}^1(x) &= 0; \\ &\dots \\ h_{ij}^{d(i,j)}(x) &= 0, \end{aligned} \right\} \quad (4)$$

в которой первое уравнение выражает равенство $f_i = 0$ через аффинную аппроксимацию $l_{ij}(x)$ функции $f_i(x)$ на плоскости π_{ij} , а остальные $d(i, j)$ уравнений задают эту плоскость (хотя это не принципиально, можно для удобства считать, что $d(i, j) = n - \dim \pi_{ij}$, т.е. что система содержит минимальное число уравнений).

Перепишем систему (4) в виде матричного уравнения

$$xA_{ij} = b_{ij} = \left(l_{ij}(0), h_{ij}^1(0), \dots, h_{ij}^{d(i,j)}(0) \right),$$

компоненты $(n \times (d(i, j) + 1))$ -матрицы A_{ij} представляют собой коэффициенты при переменных в полиномах Жегалкина аффинных функций l_{ij} (в первом столбце) и h_{ij}^k , а компоненты вектора b_{ij} — постоянные слагаемые в этих полиномах.

Обозначим также через H_{ij} матрицу системы уравнений, определяющей плоскость π_{ij} (т.е. подматрицу, полученную удалением первого столбца матрицы A_{ij}), а через b'_{ij} — вектор b_{ij} без первой компоненты. Таким образом, плоскость π_{ij} задается матричным уравнением:

$$xH_{ij} = b'_{ij}. \quad (5)$$

3.2 Опробование локальных аффинностей

Пусть для функций f_i системы (1) зафиксированы разбиения Π_i пространства V_n и соответствующие аффинные нормальные формы. Как уже говорилось, каждое Π_i некоторым образом упорядочено. Этот порядок естественным образом индуцирует лексикографический порядок на множестве $\Pi_1 \times \dots \times \Pi_m$ комбинаций локальных аффинностей функций системы.

В соответствии с этим порядком будем последовательно проверять комбинации плоскостей π_{i,j_i} , $1 \leq i \leq m$, и отбраковывать те из них, которые не могут содержать общего корня полиномов f_i , т.е. решения исходной системы. Основным инструментом такого отсеивания служит теорема Кронекера–Капелли.

Рассмотрим набор индексов

$$J = (j_1, \dots, j_m) \in \prod_{i=1}^m \{1, \dots, \lambda(i)\}.$$

Он определяет комбинацию плоскостей из $\Pi_1 \times \dots \times \Pi_m$, поэтому для краткости назовем его индексом комбинации. Первый раз нижеописанные шаги выполняются для набора $(1, \dots, 1)$.

Введем вспомогательные обозначения для $(n \times \sum_{i=1}^k d(i, j_i))$ -матрицы, составленной из матриц линейных систем вида (4):

$$A_k^J = (A_{1,j_1} | A_{2,j_2} | \dots | A_{k,j_k}),$$

и для соответствующего вектора правых частей уравнений в таких системах:

$$b_k^J = (b_{1,j_1} | b_{2,j_2} | \dots | b_{k,j_k}).$$

Шаг 1. Для плоскости $\pi_{1,j_1} \in \Pi_1$ проверим равенство

$$\text{rank } A_1^J = \text{rank } A_{1,j_1} = \text{rank} \left(\frac{A_{1,j_1}}{b_{1,j_1}} \right),$$

гарантирующее совместность системы вида (4) для локальной аффинности π_{1,j_1} .

Шаг s, $1 < s \leq m$. Пусть уже имеется такой набор локальных аффинностей $(\pi_{1,j_1}, \dots, \pi_{s-1,j_{s-1}}) \in \Pi_1 \times \dots \times \Pi_{s-1}$, что система $xA_{s-1}^J = b_{s-1}^J$ совместна. Возьмем плоскость $\pi_{s,j_s} \in \Pi_s$ и проверим равенство

$$\text{rank } A_s^J = \text{rank} \left(\frac{A_{1,j_1} | A_{2,j_2} | \dots | A_{s,j_s}}{b_{1,j_1} | b_{2,j_2} | \dots | b_{s,j_s}} \right). \quad (6)$$

Справедливость равенства (6) означает, что система $xA_s^J = b_s^J$ совместна и, значит, $\bigcap_{i=1}^s \pi_{i,j_i}$ содержит решения подсистемы $\{f_i = 0 \mid 1 \leq i \leq s\}$, т.е. отказываться от данной комбинации еще рано.

После k -го шага, $1 \leq k < m$, в случае успешной проверки условия переходим к следующему шагу $k + 1$.

Если на шаге, скажем, с номером t соответствующее равенство для локальной аффинности π_{t,j_t} , $j_t < \lambda(t)$, не выполняется, значит, выбранная комбинация плоскостей не содержит решения системы (1). В этом случае процедура повторяется для комбинации с индексом $(j_1, \dots, j_{t-1}, j_t + 1, 1, \dots, 1)$ начиная с шага 1. Таким образом, вся «ветвь» комбинаций, имеющих индексы с началом (j_1, \dots, j_t, \dots) , отбрасывается за бесперспективностью. Если же то же происходит при $j_t = \lambda(t)$, то переходить следует к комбинации с индексом $(j_1, \dots, j_{t-2}, j_{t-1} + 1, 1, \dots, 1)$.

В том случае, когда удалось пройти все m шагов, получаем линейную систему из $(m + \sum_{i=1}^m d(i, j_i))$ уравнений с n неизвестными. Остается решить ее

одним из вариантов метода Гаусса. Следующая теорема утверждает, что это даст решение исходной системы (1).

Теорема 1. Пусть система (1) квадратичных булевых уравнений имеет единственное решение. Тогда существует построенная согласно вышеописанной процедуре совместная система линейных булевых уравнений вида

$$xA_m^J = b_m^J, \quad (7)$$

она единственна и ее решение является в точности решением системы (1).

Доказательство. Пусть $a \in V_n$ — единственное решение системы (1), т.е. $f_i(a) = 0$ для всех i . Значит, для любого i существует единственный индекс j_i , такой что $a \in \pi_{i,j_i}$. Обозначим $J_a = (j_1, \dots, j_m)$. Тогда a является решением системы $xA_m^{J_a} = b_m^{J_a}$, так как a удовлетворяет всем уравнениям плоскостей π_{i,j_i} , $1 \leq i \leq m$, и обнуляет аффинную аппроксимацию l_{i,j_i} функции f_i : $l_{i,j_i}(a) = f_i(a) = 0$ (см. систему (4)).

Обратно, пусть a' — решение произвольной системы $xA_m^{J'} = b_m^{J'}$ для $J' = (j'_1, \dots, j'_m)$. Тогда для всех i имеем $a' \in \pi_{i,j'_i}$ и $l_{i,j'_i}(a') = 0$. Поэтому $f_i(a') = f_i|_{\pi_{i,j'_i}}(a') = 0$ для всех i . Следовательно, $a' = a$.

Таким образом, в любом случае рано или поздно будет построена полноразмерная совместная система (7), решение которой гарантированно является искомым в задаче.

Замечание 1. Если аффинные нормальные формы были получены способом, описанным в работе [4], то

$$d(i, j_i) = \frac{1}{2} \text{rank } \tilde{Q}_{f_i} \leq \left\lfloor \frac{n}{2} \right\rfloor,$$

где \tilde{Q}_{f_i} — матрица, определяющая билинейную форму $f_i(u \oplus v) \oplus f_i(u) \oplus f_i(v) \oplus f_i(0)$, ассоциированную с квадратичной функцией f_i . Таким образом, число уравнений в системе, построенной для любой комбинации локальных аффинностей, не превосходит $m(1 + \lfloor n/2 \rfloor)$.

Замечание 2. Очевидно, что если исключить условие единственности решения, то комбинация локальных аффинностей, содержащих решение, будет уже не обязательно одна. Соответственно, будет и некоторое число совместных линейных систем вида (7).

Такая ситуация может возникнуть из-за введения фиктивных переменных при построении исходной системы уравнений. Например, в случае (блоковых) криптосистем, использующих композицию нескольких криптографических преобразований, для упрощения процедуры и снижения

степени уравнений переменными назначаются результаты промежуточных вычислений. (Подчеркнем, что алгоритм сведения полиномиальной системы к квадратичной [2, § 11.4.2] новых решений не добавляет.)

Поскольку здесь речь идет об уже готовой системе, а для нее все ее решения равнозначны, в рамках метода нельзя заранее отсеять «ложные» решения. Поэтому на предварительном этапе следует строить (нелинейную) систему так, чтобы она имела как можно меньше фиктивных решений, а впоследствии при получении каждого решения — проверять, соответствует ли оно искомым секретным данным.

4 Замечания по оптимизации метода

Прежде всего надо заметить, что существенное влияние на работу метода оказывает выбор аффинных нормальных форм и соответствующих разбиений пространства. Здесь достаточно обратить внимание на то, что снижение на единицу размерности локальных аффинностей для одной функции может увеличить число опробуемых комбинаций плоскостей в два раза (подробнее см. в [4]). Поэтому вопрос построения аффинных нормальных форм должен быть одним из основных при исследовании возможности оптимизировать метод.

Рассмотрим некоторые из приемов, способных повысить эффективность метода.

Предварительное отсеивание непересекающихся плоскостей. Некоторые из комбинаций локальных аффинностей, которые, напомним, задаются матричными уравнениями (5), можно отбросить до вычисления рангов соответствующих матриц следующим способом. На шаге с номером s , $2 \leq s \leq m$, можно добавить проверку условия:

$$\begin{aligned} \text{rank} (H_{1,j_1} | H_{2,j_2} | \dots | H_{s,j_s}) &= \\ &= \text{rank} \left(\begin{array}{c|c|c|c} H_{1,j_1} & H_{2,j_2} & \dots & H_{s,j_s} \\ \hline b'_{1,j_1} & b'_{2,j_2} & \dots & b'_{s,j_s} \end{array} \right). \end{aligned}$$

Благодаря этому равенству из рассмотрения исключаются не пересекающиеся в совокупности локальные аффинности. Такую проверку, например для всех пар плоскостей, потом для всех троек и т.д., можно провести и до начала выполнения первого шага. Сравнение соответствующих рангов на каждом шаге имеет то преимущество, что требует меньше дополнительных вычислений, так как H_{i,j_i} является подматрицей A_{i,j_i} (то же верно для соответствующих расширенных матриц).

Уменьшение числа уравнений. Допустим, ранг матрицы будет вычисляться приведением матрицы к ступенчатому виду с помощью преобразования столбцов методом Гаусса. Для каждой комбинации плоскостей с индексом J последовательно вычисляются ранги матриц A_s^J и

$$\hat{A}_s^J = \left(\frac{A_s^J}{b_s^J} \right),$$

$s \geq 1$, при этом как A_{s+1}^J , так и \hat{A}_s^J являются подматрицами матрицы \hat{A}_{s+1}^J . Так что, приведя к ступенчатому виду \hat{A}_s^J и получив из него $\text{rank } \hat{A}_s^J$ и $\text{rank } A_s^J$, имеет смысл на шаге $(s + 1)$ оставить теперь уже подматрицу \hat{A}_s^J в полученном виде — с фиксированными линейно независимыми столбцами, тем самым сокращая до минимума число уравнений строящейся системы и количество операций, необходимых для приведения к ступенчатому виду матрицы \hat{A}_{s+1}^J . Также можно заранее привести все системы вида (4) к ступенчатому виду, что должно несколько ускорить вычисления.

Досрочное решение линейной системы. Возможно, уже на некотором шаге t , $t < m$, линейных уравнений будет достаточно для попытки найти решение, а именно: если $\text{rank } A_t^J$ равен или достаточно близок к n , линейная система будет иметь единственное решение или небольшое число решений соответственно. Достижение полного ранга на промежуточных шагах метода крайне вероятно, если учесть, что уравнений в линейной системе $m + \sum_{i=1}^m d(i, j_i)$, причем $m > n$.

Полученные таким образом решения необходимо проверить подстановкой в систему (1). Если эти решения не подходят, а такое возможно, поскольку на данном этапе не учтены уравнения $f_i = 0$ для $t < i \leq n$, то следует перейти к другой комбинации локальных аффинностей — с индексом $(j_1, \dots, j_{t-1}, j_t + 1, 1, \dots, 1)$.

«Прорежение» системы линейных уравнений. Поскольку искомая линейная система (7) сильно переопределенная, возникает идея брать не все уравнения из системы (4), а лишь некоторые из них. Тогда в уравнениях вида (6) нужно будет рассматривать вместо указанных там матриц их подматрицы, полученные удалением некоторых столбцов, что должно снизить трудоемкость метода.

При выборе уравнений из (4) следует стремиться к максимизации информативности составленной из них линейной системы. Возможно, в системах (4) целесообразнее брать первые уравнения ($l_{ij} = 0$) и еще по одному какому-нибудь уравнению. В таком случае для некоторой комбинации локальных аффинностей получим $2m > 2n$ совместных линейных уравнений (если, конечно, они удовлетворяют

всем равенствам типа (6) для $1 < s \leq m$). Если при этом ранг матрицы полученной линейной системы существенно меньше n , то можно взять не по два, а по три уравнения из систем (4) и повторить процедуру для той же комбинации плоскостей. Заметим, что если на некотором шаге s равенство типа (6) для усеченных матриц не выполняется, то и система $x A_s^J = b_s^J$ будет несовместной, так что данную комбинацию плоскостей можно сразу исключить из рассмотрения.

Упорядочение поиска. Порядок выполняемых действий играет существенную роль в методе. Поскольку перебор большого числа комбинаций локальных аффинностей с вычислением соответствующих рангов может оказаться затратнее, чем перебор всех значений переменных, крайне важно внимательно отнестись к выбору последовательности, в которой будут опробоваться локальные аффинности. Для этого следует определить «правильный» порядок уравнений в системе (1) и затем упорядочить тем или иным образом разбиения Π_i пространства на локальные аффинности для каждой функции f_i . Как именно лучше это сделать — тема отдельного исследования. Например, если локальные аффинности функции имеют разную размерность, кажется целесообразным первыми опробовать плоскости большей размерности, поскольку вероятность содержать решение у них при прочих равных больше.

5 О трудоемкости метода

Рассмотрим вычислительную сложность задач, связанных с решением систем m квадратичных булевых уравнений от n переменных над полем \mathbb{F}_2 (об этом также см., в частности, [6]). Предположим, что m ограничено некоторым полиномом от n , так что $n < m < \text{poly}(n)$. Тогда можно говорить о полиномиальности (алгоритмов) от числа переменных, а не от длины записи задачи (считаем, что полиномы в левых частях уравнений системы кодируются в виде двоичных строк одним из естественных способов).

Хорошо известно, что в общем случае задача распознавания совместности системы такого типа NP-полна (см., например, [7] или [8, с. 321]). Задача поиска (произвольного) решения системы полна в классе FNP — функциональном (т. е. определенном для задач поиска) аналоге класса NP. Это можно доказать таким же способом, как и в случае распознавательных задач, — сведением к ней задачи поиска выполняющего набора 3-КНФ, например. При этом задача поиска решения системы является и NP-трудной — в том смысле, что к ней,

очевидно, сводится ее (NP-трудный) распознавательный вариант. Таким образом, если $P \neq NP$, то полиномиальных алгоритмов решения систем квадратичных уравнений не существует.

Еще менее определена ситуация с вариантом задачи, который интересен в криптоанализе: найти (единственное) решение заведомо совместной системы уравнений. Современное состояние исследований не дает оснований полагать, что эта задача существенно проще общего случая.

Однако на практике приходится решать не произвольные (совместные) системы уравнений, а лишь обладающие теми или иными свойствами. Эти частные случаи определяют подзадачи общей задачи, для которых уже могут найтись полиномиальные алгоритмы решения.

Замечание 3. Функции $\chi_\pi(x)$ и слагаемые в аффинной нормальной форме (2) являются мультиаффинными функциями. Теоретико-сложностные вопросы, связанные, в частности, с решением систем мультиаффинных уравнений, а также оценки числа таких функций рассматриваются в [6, 9].

Что касается собственно оценок трудоемкости предлагаемого метода, прежде всего оговоримся: для простоты считаем, что метод Гаусса требует выполнения $O(nm^2)$ операций (для системы m уравнений от n неизвестных). Положим также, что все локальные аффинности каждой функции f_i имеют одинаковую коразмерность r_i и, более того, что аффинные нормальные формы функций построены так, как описано в [4], — тогда $r_i = (1/2)\text{rank}(Q_{f_i} \oplus Q_{f_i}^T) \leq n/2$, $1 \leq i \leq m$, где Q_{f_i} — треугольная матрица коэффициентов квадратичной части полинома f_i .

Для начала приведем данные об алгоритме сведения произвольной полиномиальной системы к квадратичной, о котором говорилось во введении. В обозначениях книги [2] в строящейся квадратичной системе будет $(m + D - 2z)$ уравнений и $(n + D - 2z)$ переменных, где z — число различных мономов степени не ниже третьей в исходной системе; D — сумма их степеней. Если Δ — максимальная из степеней уравнений системы, то, считая этот параметр фиксированным, получаем $D = O(n^\Delta)$ [2, § 11.4.6]. Поскольку алгоритм заключается только в добавлении новых переменных и уравнений, то его трудоемкость имеет такой же порядок.

Далее, трудоемкость построения аффинных нормальных форм для функций f_i оценена в [4]: приведенный там алгоритм требует выполнения порядка $(r_i 2^{r_i} + n^3)$ операций. Значит, для всех функций имеем оценку $O(mn^3 + \sum_{i=1}^m r_i 2^{r_i})$. При этом представляется, что для специфических клас-

сов систем уравнений эта оценка становится полиномиальной от n (см. рассуждения в [4]).

Когда выбрана комбинация локальных аффинностей, на последовательное вычисление рангов (с учетом приемов из разд. 4) нужно не более $O(n \sum_{i=1}^m r_i^2)$ (а значит, и $O(mn^3)$) операций. Всего комбинаций $\prod_{i=1}^m 2^{r_i} \leq 2^{mn/2}$.

Таким образом, в худшем случае, когда для всех f_i матрицы $\tilde{Q}_{f_i} = Q_{f_i} \oplus Q_{f_i}^T$ имеют максимальный ранг и нужная система (7) строится последней, верхняя оценка трудоемкости метода имеет вид $O(mn^3 \cdot 2^{mn/2})$. Однако есть основания ожидать, что в среднем благодаря возможным усовершенствованиям (эвристики и т. п.) метод сможет работать не хуже полного перебора, а на определенных имеющих практическое значение классах систем будет гораздо более эффективным.

Некоторые аргументы в пользу эффективности метода в среднем содержатся в следующих примерах. Рассмотрим потоковый шифр LILI-128 [10] и его фильтрующий генератор, т. е. регистр сдвига с линейными обратными связями и фильтрующей булевой функцией f_d степени 6 от 10 переменных. Полином Жегалкина функции f_d приведен в [5, 11]. Задача нахождения ключа по известной паре открытого текста $a \in \mathbb{F}_2^N$ и шифртекста $c \in \mathbb{F}_2^N$ сводится к решению системы уравнений $\{f_d(xL^t) \oplus a_t = c_t\}_{t=0}^{N-1}$, где L — $(n \times n)$ -матрица линейного преобразования, реализуемого регистром сдвига, зависит от коэффициентов полинома обратных связей. Таким образом, получается линейная рекуррентная последовательность $\{x_i\}$, как только вычислен ее начальный отрезок (x_0, \dots, x_{n-1}) .

При использовании предлагаемого метода для решения данной системы при разных длинах n регистра и двух вариантах G'_n и G''_n полинома обратных связей получены следующие значения для доли несовместных систем, соответствующих частичным комбинациям длин 2 и 3 (т. е. комбинациям систем вида (4) для $i = 1, 2$ и $i = 1, 2, 3$ соответственно). Рассмотрим примитивные полиномы над \mathbb{F}_2 :

$$\begin{aligned} G'_{10}(y) &= y^{10} \oplus y^3 \oplus 1; \\ G''_{10}(y) &= y^{10} \oplus y^8 \oplus y^7 \oplus y^6 \oplus y^5 \oplus y^2 \oplus 1; \\ G'_{20}(y) &= y^{20} \oplus y^3 \oplus 1; \\ G''_{20}(y) &= y^{20} \oplus y^9 \oplus y^5 \oplus y^3 \oplus 1; \\ G'_{30}(y) &= y^{30} \oplus y^{23} \oplus y^2 \oplus y \oplus 1; \\ G''_{30}(y) &= y^{30} \oplus y^{25} \oplus y^{24} \oplus y^{23} \oplus y^{19} \oplus y^{18} \oplus y^{16} \oplus \\ &\quad \oplus y^{14} \oplus y^{11} \oplus y^8 \oplus y^6 \oplus y^4 \oplus y^3 \oplus y \oplus 1. \end{aligned}$$

Доля $\varepsilon_t(G)$ несовместных систем вида $xA_t^J = b_t^J$ для $t = 2, 3$ и $G = G'_n, G''_n$ ($n = 10, 20, 30$), принимает следующие значения, полученные экспе-

риментальным путем (при случайных начальных заполнениях регистра).

$$\begin{aligned} \varepsilon_2 (G'_{10}) &= 0,8290; & \varepsilon_3 (G'_{10}) &= 0,9966; \\ \varepsilon_2 (G'_{20}) &= 0,2870; & \varepsilon_3 (G'_{20}) &= 0,7461; \\ \varepsilon_2 (G'_{30}) &= 0,1322; & \varepsilon_3 (G'_{30}) &= 0,4031; \\ \varepsilon_2 (G''_{10}) &= 0,8271; & \varepsilon_3 (G''_{10}) &= 0,9966; \\ \varepsilon_2 (G''_{20}) &= 0,2577; & \varepsilon_3 (G''_{20}) &= 0,6628; \\ \varepsilon_2 (G''_{30}) &= 0,1324; & \varepsilon_3 (G''_{30}) &= 0,3948. \end{aligned}$$

Аналогичные вычисления были проведены для фильтрующего генератора, построенного на основе регистра сдвига (длины $n = 10, 20, 30$), с линейными обратными связями, заданными теми же полиномами, и с квадратичной фильтрующей функцией $f(x) = x_1x_2 \oplus x_3x_4 \oplus x_5x_6 \oplus x_7x_8 \oplus x_9x_{10}$.

Результаты показывают, что для квадратичной функции f предлагаемый метод (с теми же параметрами) более эффективен, нежели для функции f_d степени 6:

$$\begin{aligned} \varepsilon_2 (G'_{10}) &= 0,8197; & \varepsilon_3 (G'_{10}) &= 0,9964; \\ \varepsilon_2 (G'_{20}) &= 0,3967; & \varepsilon_3 (G'_{20}) &= 0,8742; \\ \varepsilon_2 (G'_{30}) &= 0,3275; & \varepsilon_3 (G'_{30}) &= 0,7428; \\ \varepsilon_2 (G''_{10}) &= 0,8195; & \varepsilon_3 (G''_{10}) &= 0,9964; \\ \varepsilon_2 (G''_{20}) &= 0,3329; & \varepsilon_3 (G''_{20}) &= 0,7742; \\ \varepsilon_2 (G''_{30}) &= 0,3186; & \varepsilon_3 (G''_{30}) &= 0,7371. \end{aligned}$$

Таким образом, можно предположить, что в типичном случае большая часть комбинаций будет отсекается уже после нескольких первых шагов алгоритма.

Для более точной оценки трудоемкости метода следует использовать оценки (математическое ожидание) таких параметров, как доля ε_t несовместных систем для частичных комбинаций длины t (т.е. среди систем $xA_t^J = b_t^J$ по всем J) и минимальное m_0 , такое что совместность системы для частичной комбинации длины m_0 уже гарантирует на шаге m_0 решения исходной системы уравнений, а именно: обозначим через $S_t, 2 \leq t \leq m$, множество совместных линейных систем для частичных комбинаций длины t , а через m_0 — минимальное целое число, такое что $|S_{m_0}| = 1$. Тогда в предположении единственности решения исходной системы, пренебрегая возможным случаем несовместности системы для комбинации длины 1, имеем:

$$\left. \begin{aligned} |S_2| &= 2^{r_1} \cdot 2^{r_2}; \\ &\dots \\ |S_{t+1}| &= |S_t| \cdot 2^{r_{t+1}}(1 - \varepsilon_{t+1}); \\ &\dots \\ |S_{m_0}| &= |S_{m_0-1}| \cdot 2^{r_{m_0}}(1 - \varepsilon_{m_0}) = 1. \end{aligned} \right\} (8)$$

Соотношения (8) соответствуют последовательному перебору и проверке на совместность систем линейных уравнений, доставляемых частичными комбинациями длин $t = 1, 2, \dots$

Обозначим через $\gamma_t, t \leq m$, трудоемкость алгоритма Гаусса для систем линейных уравнений, описываемых частичными комбинациями, соответствующими t первым уравнениям. Тогда для этапа опробования локальных аффинностей — самого трудоемкого — потребуется двоичных операций примерно в количестве $\gamma_2 |S_2| + \gamma_3 |S_3| + \dots + \gamma_{m_0} |S_{m_0}|$, где величины $|S_t|, 2 \leq t \leq m$, выражены соотношениями (8).

Отметим, что, учитывая вид систем линейных уравнений, получаемых способом из работы [4] (все, кроме одного, уравнения каждой «локальной» системы содержат единственный непостоянный моном), можно рассчитывать на более быстрое выполнение метода Гаусса, чем в общем случае.

Заключение

Предлагаемый метод по своей сути принадлежит к классу линеаризационных методов алгебраического криптоанализа. При этом он обладает рядом потенциальных преимуществ, связанных с тем, что вместо снижения степени уравнений посредством увеличения количества переменных предлагается рассматривать множество линейных систем от тех же переменных и с умеренно увеличенным числом уравнений. Очевидно, что надежность метода равна 1.

Основная проблема состоит в том, чтобы правильно определить стратегию перебора линейных систем, которая позволит наиболее быстро прийти к системе, дающей искомое решение. Кроме того, важной задачей является изучение способов оптимизации вычислений и разработка эвристик для повышения эффективности метода, в частности, использующих свойства анализируемых криптографических преобразований и, соответственно, систем полиномиальных уравнений. Также, поскольку для общей задачи решения систем квадратичных булевых уравнений предположительно нет полиномиальных алгоритмов, как это обычно и делается для криптоаналитических методов, следует определить классы систем полиномиальных уравнений, для которых метод работает наиболее быстро.

Литература

1. Шеннон К. Теория связи в секретных системах // Работы по теории информации и кибернетике / Пер.

- с англ. — М.: ИЛ, 1963. С. 333–369. (*Shannon C.* Communication theory of secrecy systems // *Bell Syst. Tech. J.*, 1949. Vol. 28. Iss. 4. P. 656–715.)
2. *Bard G. V.* Algebraic cryptanalysis. — Springer, 2009. 389 p.
 3. *Bard G., Courtois N., Jefferson C.* Efficient methods for conversion and solution of sparse systems of low-degree multivariate polynomials over GF(2) via SAT-solvers // *Cryptology ePrint Archive*. Report 2007/024. <http://eprint.iacr.org/2007/024.pdf>.
 4. *Логачев О. А., Сукаев А. А., Федоров С. Н.* Полиномиальные алгоритмы вычисления локальных аффинностей квадратичных булевых функций // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 67–74.
 5. *Логачев О. А.* Об использовании аффинных нормальных форм булевых функций для определения ключей фильтрующих генераторов // Мат-лы IV Междунар. научн. конф. по проблемам безопасности и противодействия терроризму. — М.: МЦНМО, 2009. Т. 2. С. 101–109.
 6. *Горшков С. П.* Применение теории NP-полных задач для оценки сложности решения систем булевых уравнений // Обзорение прикладной и промышленной математики, 1995. Т. 2. Вып. 3. С. 325–398.
 7. *Fraenkel A. S., Yesha Y.* Complexity of solving algebraic equations // *Inform. Process. Lett.*, 1980. Vol. 10. No. 4-5. P. 178–179.
 8. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи / Пер. с англ. — М.: Мир, 1982. 416 с. (*Garey M. R., Johnson D. S.* Computers and intractability: A guide to the theory of NP-completeness. — San Francisco, CA, USA: W. H. Freeman and Co., 1979. 348 p.)
 9. *Горшков С. П., Тарасов А. В.* Сложность решения систем булевых уравнений. — М.: Курск, 2017. 192 с.
 10. *Simpson L. R., Dawson E., Golić J. Dj., Millan W. L.* LILI keystream generator // *Selected areas in cryptography / Eds. D. R. Stinson, S. Tavares.* — Lecture notes in computer science ser. — Springer, 2001. Vol. 1212. P. 248–261.
 11. *Logachev O. A., Yashchenko V. V., Denisenko M. P.* Local affinity of Boolean mappings // *Boolean Functions in Cryptology and Information Security: Proc. NATO Advanced Study Institute.* — IOS Press, 2008. P. 148–172.

Поступила в редакцию 01.04.19

ON LOCAL AFFINITY BASED METHOD OF SOLVING SYSTEMS OF QUADRATIC BOOLEAN EQUATIONS

O. A. Logachev¹, A. A. Sukayev², and S. N. Fedorov²

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²M. V. Lomonosov Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, Russian Federation

Abstract: Solving nonlinear systems of Boolean equations is NP-hard. Nevertheless, certain classes of such systems can be solved by efficient algorithms. There are theoretical and applied reasons for studying these classes and designing corresponding efficient algorithms. The paper proposes an approach to solving the systems of quadratic equations over two-element field. The method makes use of the quadratic functions’ representation by their affine normal forms, i. e., in a sense, of their piecewise affine approximation. So, the initial nonlinear instance comes to a number of linear equations systems of the same variables. The paper also discusses possible ways to reduce the complexity of the proposed method.

Keywords: Boolean function; system of quadratic Boolean equations; vector space partition; flat; local affinity; affine normal form; algebraic cryptanalysis

DOI: 10.14357/19922264190206

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 18-29-03124 mk).

References

1. Shannon, C. 1949. Communication theory of secrecy systems. *Bell Syst. Tech. J.* 28 (4):656–715.
2. Bard, G. V. 2009. *Algebraic cryptanalysis*. Springer. 389 p.
3. Bard, G., N. Courtois, and C. Jefferson. 2007. Efficient methods for conversion and solution of sparse systems of low-degree multivariate polynomials over GF(2) via SAT-solvers. *Cryptology ePrint Archive*. Report 2007/024. Available at: <http://eprint.iacr.org/2007/024.pdf> (accessed August 30, 2018).
4. Logachev, O. A., A. A. Sukayev, and S. N. Fedorov. 2019. Polinomial’nye algoritmy vychisleniya lokal’nykh

- affinnostey kvadraticnykh bulevykh funktsiy [Polynomial algorithms for constructing local affinities of quadratic Boolean functions]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):67–74.
5. Logachev, O. A. 2009. Ob ispol'zovanii affinykh normal'nykh form bulevykh funktsiy dlya opredeleniya klyuchey fil'truyushchikh generatorov [On using affine normal forms of Boolean functions for identifying filter generator keys]. *4th Conference (International) on Security and Counteraction to Terrorism Proceedings*. Moscow. 2:101–109.
 6. Gorshkov, S. P. 1995. Primenenie teorii NP-polnykh zadach dlya otsenki slozhnosti resheniya sistem bulevykh uravneniy [Application of the NP-complete problem theory to assessment of complexity of solving the systems of Boolean equations]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Applied and Industrial Mathematics Review] 2(3):325–398.
 7. Fraenkel, A. S., and Y. Yesha. 1980. Complexity of solving algebraic equations. *Inform. Process. Lett.* 10(4-5):178–179.
 8. Garey, M. R., and D. S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W. H. Freeman and Co. 348 p.
 9. Gorshkov, S. P., and A. V. Tarasov. 2017. *Slozhnost' resheniya sistem bulevykh uravneniy* [Complexity of solving the systems of Boolean equations]. Moscow: Kurs. 192 p.
 10. Simpson, L. R., E. Dawson, J. Dj. Golić, and W. L. Millan. 2001. LILI keystream generator. *Selected areas in cryptography*. Eds. D. R. Stinson and S. Tavares. Lecture notes in computer science ser. Springer. 2012:248–261.
 11. Logachev, O. A., V. V. Yashchenko, and M. P. Denisenko. 2008. Local affinity of Boolean mappings. *Boolean Functions in Cryptology and Information Security: Proc. NATO Advanced Study Institute*. IOS Press. 148–172.

Received April 1, 2019

Contributors

Logachev Oleg A. (b. 1950) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; ollog@inbox.ru

Sukayev Al'bert A. (b. 1994) — student, Information Security Institute, M. V. Lomonosov University, 1 Michurinskiy Prosp., Moscow 119192, Russian Federation; asukaev@gmail.com

Fedorov Sergey N. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Information Security Institute, M. V. Lomonosov Moscow State University, 1 Michurinskiy Prosp., Moscow 119192, Russian Federation; s.n.feodorov@yandex.ru

ПРИМЕНЕНИЕ МЕТОДОВ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ МНОГОКРИТЕРИАЛЬНОЙ ЗАДАЧИ ОТБОРА МНОГОМАСШТАБНЫХ КОМПОЗИЦИЙ

К. К. Абгарян¹, В. А. Осипова²

Аннотация: Рассмотрены вопросы использования методов поддержки принятия решений для задачи отбора многомасштабных композиций (МК) — вычислительных аналогов многомасштабных физико-математических моделей, созданных для анализа различных гетерогенных процессов, связанных с формированием новых композиционных материалов с заранее заданными свойствами. При решении конкретных задач могут быть построены разные многомасштабные модели и соответствующие им МК. Возникает вопрос о сравнении этих моделей, об оценке их эффективности для данной конкретной задачи. В работе на этапе предсказательного моделирования предлагается методика сравнения многомасштабных моделей с помощью оценки и отбора соответствующих МК с использованием методов поддержки принятия решений при многих критериях качества. Для иллюстрации возможности выбора наилучшей альтернативы при наличии дополнительной информации о критериях оценки МК рассмотрен модельный пример, связанный с исследованием электронных и структурных свойств тонких пленок InN (GaN) на кремниевых подложках.

Ключевые слова: многомасштабное моделирование; теория принятия решений; критерии качества; альтернатива; методы поддержки принятия решений; многокритериальность; функция ценности

DOI: 10.14357/19922264190207

Использование методов многомасштабного моделирования — как последовательных, так и параллельных — открывает широкие возможности для изучения процессов и явлений, при анализе которых необходимо учитывать широкий спектр пространственных и/или временных масштабов. К таким явлениям, в частности, относятся сложные гетерогенные процессы, протекающие в многокомпонентных системах, связанных с формированием новых композиционных материалов с заранее заданными свойствами.

В работе [1] представлены основные положения разработанной информационной технологии построения многомасштабных моделей с использованием таких новых понятий, как «базовая модель-композиция» и «многомасштабная композиция». Для создания на их основе программных систем применялся модельно-ориентированный подход, который был развит в работах Ю. И. Бродского [2]. Особенностью изложенной в работе [1] технологии является применение информационных структур, названных базовыми композициями (БК), объединяющих данные и методы их обработки. Эти математические объекты ставятся в соответствие базовым математическим моделям, которые ис-

пользуются для решения различных многомасштабных задач. Для описания БК задействован теоретико-множественный аппарат, который позволяет передать вычислительную сущность исходных математических моделей. Базовые композиции служат композиционными элементами (объектами), из которых согласно представленной в работе [1] технологии строятся МК — информационные аналоги многомасштабных моделей, при помощи которых передается содержание многомасштабных вычислительных процессов и явлений. Далее на базе МК строятся сложные иерархические программные системы, применяемые для решения различных прикладных задач, в том числе связанных с созданием новых композиционных материалов.

При решении конкретных задач могут быть построены разные многомасштабные модели, и, естественно, возникает вопрос о сравнении этих моделей, об оценке их эффективности. В данной работе на этапе предсказательного моделирования предлагается методика сравнения многомасштабных моделей с помощью оценки и отбора соответствующих МК. Оценка и отбор проводятся с использованием методов поддержки принятия решений при многих критериях качества.

¹Вычислительный центр имени А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский авиационный институт (национальный исследовательский университет), kristal83@mail.ru

²Московский авиационный институт (национальный исследовательский университет), victoria.a.osipova@gmail.com

Приведем описание метода построения МК [1].

Базовую модель-композицию можно представить как объединение основных множеств разного структурного типа: VX_{ij} , MA_{ij} , E_{ij} , $\{MA_{ij}^k\}_{k=1}^p$, $\{E_{ij}^k\}_{k=1}^p$. Здесь i — номер масштабного уровня, $i = \overline{1, L}$, где L — число рассматриваемых уровней; j — номер базовой модели-композиции на текущем масштабном уровне, $j = \overline{1, N_i}$, где N_i — число моделей на i -м уровне; k — номер элементарного процесса БК. Опишем основные множества:

$VX_{ij} = \{V_{ij}, X_{ij}\}$ — множество данных, включающее:

V_{ij} — множество входных данных (внешние характеристики модели);

X_{ij} — множество выходных данных (фазовых переменных и данных — свойств модели);

MA_{ij} — множество методов обработки данных (модели и алгоритмы);

E_{ij} — множество событий, отнесенных к описанию выполняемых в рамках БК элементарных процессов;

$\{MA_{ij}^k\}_{k=1}^p$ — множество реализаций моделей и алгоритмов в зависимости от элементарного процесса p ;

$\{E_{ij}^k\}_{k=1}^p$ — множество реализаций событий по элементарным процессам.

Множество методов обработки данных опишем подробнее:

$$MA_{ij} = \{M_{ij}, A_{ij}\} = \{s_{ij}, f_{ij}, a_{ij}, a_{i, \dots, i^*, j}\}.$$

Множество моделей M_{ij} , входящих в множество MA_{ij} , состоит из статических (s_{ij}) и динамических (f_{ij}) методов обработки. Алгоритмические модели (алгоритмы) a_{ij} , $i = \overline{0, L}$, $j = \overline{1, N_i}$, могут быть специализированными, т. е. используемыми только в данной конкретной модели с определенного масштабного уровня, или универсальными, применяемыми в различных моделях с разных масштабных уровней $a_{i, \dots, i^*, j}$.

Определение 1. Под базовой моделью-композицией MC_i^j будем понимать однопараметрическое семейство основных множеств, задействованных в общем вычислительном процессе, разного структурного типа, включая данные (входные и выходные) и методы их обработки:

$$MC_i^j = \left\langle \left\{ VX_{ij}, MA_{ij}, E_{ij}, \left\{ MA_{ij}^k \right\}_{k=1}^p, \left\{ E_{ij}^k \right\}_{k=1}^p \right\} \right\rangle.$$

Здесь

$$\begin{aligned} VX_{ij} &= \{V_{ij}, X_{ij}\}; \\ MA_{ij} &= \{M_{ij}, A_{ij}\}; \\ \left\{ MA_{ij}^k \right\}_{k=1}^p &= \{MA_{ij}^1, MA_{ij}^2, \dots, MA_{ij}^p\}; \\ \left\{ E_{ij}^k \right\}_{k=1}^p &= \{E_{ij}^1, E_{ij}^2, \dots, E_{ij}^p\}. \end{aligned}$$

Параметром семейства основных множеств служит число элементарных процессов в базовой модели-композиции p . Индексы i и j позволяют идентифицировать MC_i^j на пространственном уровне i по ее номеру j .

Структура модели-композиции представлена в [1]. Такое представление полностью описывает структуру базовой модели-композиции и задает шаблон, который будет заполняться конкретными данными при создании реальных экземпляров БК для решения практических задач математического моделирования.

Далее приведем описание МК, позволяющее представить информацию, из каких именно моделей-композиций, с каких масштабных уровней она состоит, сколько и каких процессов задействовано в ее работе, каким образом происходит обмен данными между моделями-композициями с разных уровней.

Определение 2. Под МК будем понимать однопараметрическое семейство, полученное из экземпляров БК с разных масштабных уровней за счет объединения в общем вычислительном процессе их основных множеств разного структурного типа, включая данные (входные и выходные) и методы их обработки.

Многомасштабную композицию будем обозначать через $MC_{i, i^*, \dots, i^{***}}^{j, j^*, \dots, j^{***}}$. Здесь i, i^*, \dots, i^{***} — номера масштабных уровней, задействованных в данной МК; j, j^*, \dots, j^{***} — номера БК на конкретном масштабном уровне. В определенном смысле $MC_{i, i^*, \dots, i^{***}}^{j, j^*, \dots, j^{***}}$ схожа с БК, так как представляет собой объединение основных множеств разных структурных типов, связанных общим вычислительным процессом. Ее структуру можно представить набором таблиц, соответствующих экземплярам входящих в нее БК, расположенных в определенном порядке, соответствующем иерархии масштабов, задействованных в ней.

Пусть на i -м масштабном уровне имеется экземпляр MC_i^j и на i^* -м масштабном уровне — экземпляр $MC_{i^*}^{j^*}$. Здесь j и j^* — номера базовых моделей-композиций на масштабных уровнях i и i^* соответственно. Составим многомасштабную композицию MC_{i, i^*}^{j, j^*} из двух экземпляров БК MC_i^j и $MC_{i^*}^{j^*}$. Основными множествами, как и в случае создания композиции, будут:

$$\begin{aligned}
 & V_{ij} \cup V_{i^*j^*}; X_{ij} \cup X_{i^*j^*}; \\
 & MA_{ij} \cup MA_{i^*j^*}; E_{ij} \cup E_{i^*j^*}; \\
 & \left\{ MA_{ij}^k \right\}_{k=1}^p \cup \left\{ MA_{i^*j^*}^k \right\}_{k=1}^{p^*}; \\
 & \left\{ E_{ij}^k \right\}_{k=1}^p \cup \left\{ E_{i^*j^*}^k \right\}_{k=1}^{p^*}.
 \end{aligned}$$

Здесь p и p^* обозначают число процессов в БК MC_i^j и $MC_{i^*}^{j^*}$ соответственно.

Многомасштабную композицию можно описать следующим образом:

$$\begin{aligned}
 MK_{i,i^*}^{j,j^*} = & \left\langle \left\{ V_{ij} \cup V_{i^*j^*}, X_{ij} \cup X_{i^*j^*}, \right. \right. \\
 & MA_{ij} \cup MA_{i^*j^*}, E_{ij} \cup E_{i^*j^*}, \\
 & \left. \left\{ MA_{ij}^k \right\}_{k=1}^p \cup \left\{ MA_{i^*j^*}^k \right\}_{k=1}^{p^*}, \right. \\
 & \left. \left. \left\{ E_{ij}^k \right\}_{k=1}^p \cup \left\{ E_{i^*j^*}^k \right\}_{k=1}^{p^*} \right\} \right\rangle.
 \end{aligned}$$

Число процессов в МК равно сумме $p + p^*$.

Связующими элементами между вычислительными моделями с разных масштабных уровней, входящими в МК, служат глобальные параметры, которые играют основную роль при передаче информации между масштабными уровнями.

При составлении многомасштабной композиции $MK_i^j = MK_{i^*,i^{**}}^{j^*,j^{**}}$ из $MC_{i^*}^{j^*}$ и $MC_{i^{**}}^{j^{**}}$ под глобальными параметрами

$$\check{v} \in VX_{ij} = \{V_{i^*j^*} \cup V_{i^{**}j^{**}}, X_{i^*j^*} \cup X_{i^{**}j^{**}}\}$$

будем понимать элементы (параметры), относящиеся к множеству $X_{i^*j^*} \cap V_{i^{**}j^{**}}$, образованному в результате пересечения двух множеств выходных данных $X_{i^*j^*}$ (с нижнего масштабного уровня) и входных данных $V_{i^{**}j^{**}}$ с верхнего масштабного уровня:

$$\begin{aligned}
 X_{i^*j^*} \cap V_{i^{**}j^{**}} = \\
 = \{ \check{v} : (\check{v} \in X_{i^*j^*}) \cap (\check{v} \in V_{i^{**}j^{**}}), \check{v} \in VX_{ij} \}.
 \end{aligned}$$

Кроме того, при построении МК используются базовые модели-композиции специального вида, обозначенные DB_i , i — номер масштабного уровня, $i = \overline{1, L}$, где L — число рассматриваемых уровней. Они требуются для хранения и передачи дополнительной информации, необходимой для работы БК соответствующего уровня.

Описанная технология многомасштабного моделирования может быть применена для решения различных задач структурной оптимизации, в частности для задач материаловедения при моделировании свойств полупроводниковых наносистем [3–5].

При компьютерной реализации разработанной концепции многомасштабного моделирования требуется построение различных информационных

моделей — МК. В данной работе на этапе предсказательного моделирования предлагается методика сравнения многомасштабных моделей с помощью оценки и отбора соответствующих МК, основанная на методах принятия решений в многокритериальных ситуациях.

В соответствии с принятой терминологией каждую МК, позволяющую решить поставленную задачу моделирования, назовем альтернативой. Можно проследить связь между требуемыми свойствами исследуемого объекта и оценкой задействованных в ходе моделирования МК, используемых для анализа этих свойств.

Как показал анализ задачи моделирования, на первом этапе для оценки МК можно выбрать, например, следующие 5 критериев:

- (1) вычислительная точность: критерий K_1 с числовой шкалой $E_1 \subseteq R$, где R — множество действительных чисел;
- (2) число арифметических операций: критерий K_2 с числовой шкалой $E_2 \subseteq R$;
- (3) универсальность (возможность использования для скрининга): критерий K_3 со шкалой E_3 , состоящей из двух значений;
- (4) соотношение собственного программного обеспечения и пакетов прикладных программ: критерий K_4 с числовой шкалой $E_4 \subseteq R$;
- (5) число задействованных масштабных уровней: критерий K_5 с числовой шкалой $E_5 \subseteq R$.

Этот набор критериев не является полным, однако удовлетворяет основным требованиям, предъявляемым к перечню критериев, и основан на информации о смысле поставленной задачи [6]. Действительно, набор критериев $\{K_1, \dots, K_5\}$ соответствует существу поставленной задачи, минимален (в том смысле, что различные критерии характеризуют различные свойства исходов), критерии измеримы и операциональны (каждый критерий имеет однозначный и ясный смысл, характеризует определенное свойство исходов).

Таким образом, каждая МК, позволяющая решить задачу моделирования, характеризуется пятимерным вектором $x = (x_1, x_2, x_3, x_4, x_5)$, где x_i — значение, приписываемое данной альтернативе по критерию K_i , $i = 1, \dots, 5$. Используя известные процедуры [7], можно привести критерии к однородным шкалам.

Для выбора наилучших альтернатив можно сравнивать их по векторному отношению доминирования и рассмотреть парето-оптимальные варианты. Однако дополнительная информация

о критериях и свойствах решений может сузить множество парето-оптимальных вариантов. Эта информация предоставляется экспертами, работающими в данной предметной области, в коалиции с аналитиком, специалистом в области принятия решений. Затем в зависимости от конкретного вида полученной информации применяется математически обоснованный метод выбора наилучших альтернатив в многокритериальной ситуации [6, 7].

Распространенный подход к решению задачи оценки и выбора наилучшей альтернативы, одновременно учитывающий дополнительную информацию о критериях и свойствах альтернатив, содержится в построении числовой функции ценности.

Пусть S — множество альтернатив и считается, что при заданных критериях K_1, \dots, K_n каждая альтернатива $s \in S$ имеет векторную оценку $x = (x_1, \dots, x_n)$, где $x_i = K_i(s)$ — значение альтернативы по критерию $K_i, i = 1, \dots, n$.

Рассмотрим множество $X = \{x | x = (x_1, \dots, x_n), x_i = K_i(s), i = 1, \dots, n; s \in S\}$ — множество векторных оценок альтернатив из множества S . Пусть $a, b \in S$ и $x = (x_1, \dots, x_n)$ — векторная оценка альтернативы a ; $y = (y_1, \dots, y_n)$ — векторная оценка альтернативы b .

Числовая функция ценности $f: X \rightarrow R$ обладает следующим свойством: для любых двух альтернатив a и b : $f(x_1, \dots, x_n) \geq f(y_1, \dots, y_n) \Leftrightarrow$ альтернатива a не менее предпочтительна, чем b . Если известна функция ценности, то поиск оптимального варианта сводится к задаче нахождения аргумента максимума функции ценности на множестве X :

$$x^* = \arg \max f(x), \quad x \in X.$$

При использовании эвристических методов построения функции ценности используется метод обобщенного критерия, заключающийся в сведении многокритериальной задачи к однокритериальной, набор критериев «сворачивается» в числовую функцию, которая и будет служить функцией ценности.

Обычно функцию ценности строят в аддитивном виде как сумму функций ценности по каждому критерию с некоторыми весовыми коэффициентами $\alpha_1, \dots, \alpha_n$:

$$f(x_1, \dots, x_n) = \sum_{i=1}^n \alpha_i f_i(x_i),$$

где f_i — функция ценности критерия K_i .

Как известно, построение функции ценности в аддитивном виде правомерно только в случае взаимной независимости критериев [8]. Проверку взаимной независимости рассматриваемых критериев

K_1, \dots, K_5 можно провести с учетом результатов Леонтьева—Гормана: если любая пара критериев $\{K_i, K_j\}$, где $i, j = 1, \dots, 5$, не зависит по предпочтению от остальных двух критериев, то все критерии K_1, \dots, K_5 взаимно независимы по предпочтению. Построение функции ценности проводится известными методами, описанными, в частности, в работе [6].

Для решения конкретной практической задачи с использованием многомасштабной модели рассматривается определенный набор МК. Весовые коэффициенты $\alpha_i, i = 1, \dots, 5$, и конкретный вид функции ценности определяются содержательным смыслом задачи. Для их нахождения проводится поэтапная процедура, опирающаяся в числе прочего на интерактивный процесс диалога между специалистом в данной прикладной области и аналитиком, формирующим процедуру опроса [6, 9]. Полученные оценки позволяют сравнить по предпочтительности МК и отобрать для дальнейших исследований оптимальный набор МК.

Другие возможные подходы к решению задачи многокритериального выбора МК, в том числе не требующие нахождения весовых коэффициентов, рассмотрены в [7, 9, 10].

В качестве иллюстрации возможности выбора наилучшей альтернативы при наличии дополнительной информации о критериях оценки МК рассмотрим модельный пример.

Для исследования электронных и структурных свойств тонких пленок InN (GaN) на кремниевых подложках [1] можно построить условно 5 вариантов МК, которые реализуются с применением вычислительных средств ЦКП ФИЦ ИУ РАН. Обозначим их как $\text{МК}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\text{МК}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\text{МК}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$.

В работе [1] представлена структура многомасштабной композиции $\text{МК}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$. Указаны экземпляры БК и последовательность их использования в вычислительном процессе [1].

Рассматриваемые МК характеризуются следующими особенностями:

- $\text{МК}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ — задействованы 3 масштабных уровня, один пакет (VASP), МК может применяться для скрининга (универсальна), число арифметических операций (условно) 10 млн, точность расчетов 0,1;
- $\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ — задействованы 3 масштабных уровня, два пакета (VASP и Material Studio), МК может применяться для скрининга (универсальна), число арифметических операций (условно) 15 млн, точность расчетов 0,15;

Таблица 1 Оценки МК по критериям

№ критерия	$\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\text{МК}}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\text{МК}}}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\widetilde{\text{МК}}}}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\widetilde{\widetilde{\text{МК}}}}}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$
1	до 0,1	до 0,15	до 0,01	до 0,015	до 0,01
2	10 млн	15 млн	30 млн	25 млн	30 млн
3	Да	Да	Да	Нет	Нет
4	75% собственного программного обеспечения	50%	15%	15%	75%
5	3 уровня	3 уровня	4 уровня	3 уровня	4 уровня

Таблица 2 Оценки МК по десятибалльной шкале

№ критерия	$\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\text{МК}}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\text{МК}}}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\widetilde{\text{МК}}}}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$	$\widetilde{\widetilde{\widetilde{\widetilde{\widetilde{\text{МК}}}}}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$
1	2	1	10	9	10
2	10	8	2	4	2
3	10	10	10	1	1
4	7	4	1	1	7
5	6	6	8	6	8

- $\widetilde{\text{МК}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ — задействованы 4 масштабных уровня, три пакета (VASP, Material Studio, SRIM), МК может применяться для скрининга (универсальна), число арифметических операций (условно) 30 млн, точность расчетов 0,01;
- $\widetilde{\widetilde{\text{МК}}}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ — задействованы 3 масштабных уровня, три пакета (VASP, Material Studio и SRIM), не универсальна, число арифметических операций (условно) 30 млн, точность расчетов 0,015;
- $\widetilde{\widetilde{\widetilde{\text{МК}}}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ — задействованы 4 масштабных уровня, один пакет (VASP), не универсальна, число арифметических операций (условно) 20 млн, точность расчетов 0,01.

Пусть на этапе предсказательного моделирования следует сравнить эти пять возможных МК, оценки которых по критериям K_1, \dots, K_5 приведены в табл. 1, и выбрать лучшие с точки зрения лица, принимающего решение, альтернативы.

Из содержательного смысла задачи следует, что значения по критериям K_1 и K_2 следует минимизировать, а по критериям K_3, K_4 и K_5 — максимизировать.

В табл. 2 представлены оценки по критериям K_1, \dots, K_5 этих пяти альтернатив (т.е. пяти рассматриваемых МК), полученные после линейного преобразования шкал, сохраняющего упорядочение по предпочтению для каждого критерия. При

этом лучшему значению соответствует более высокая оценка.

По векторному отношению доминирования альтернатива $s_1 = (2, 10, 10, 7, 6)$ предпочтительней альтернативы $s_2 = (1, 8, 10, 4, 6)$.

Альтернативы $s_3 = (10, 2, 10, 1, 8)$, $s_4 = (9, 4, 1, 1, 6)$ и $s_5 = (10, 2, 1, 7, 8)$ не сравнимы между собой и не сравнимы с альтернативой s_1 .

Следовательно, без дополнительной информации о предпочтениях лица, принимающего решение, не может выбрать лучший вариант из четырех МК, именно из $\widetilde{\text{МК}}_{0,1,3}^{(A_{\alpha_1}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\widetilde{\text{МК}}}_{0,1,3,4}^{(A_{\alpha_1}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, $\widetilde{\widetilde{\widetilde{\text{МК}}}}_{0,1,3}^{(A_{\alpha_1}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$ и $\widetilde{\widetilde{\widetilde{\widetilde{\text{МК}}}}}_{0,1,3,4}^{(A_{\alpha_1}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$.

Считаем, что в этом модельном примере оценки в табл. 2 отражают также и результат приведения значений критериев к однородным шкалам и обосновано применение линейной свертки

$$f(x_1, \dots, x_5) = \sum_{i=1}^5 \alpha_i x_i.$$

Пусть в результате взаимодействия эксперта, аналитика и лица, принимающего решение, для конкретной задачи весовые коэффициенты критериев составят соответственно $\alpha_1 = 1$, $\alpha_2 = 0,5$, $\alpha_3 = 1$, $\alpha_4 = 2$ и $\alpha_5 = 0,5$, т.е. в этой задаче существенно, чтобы превалировало собственное программное обеспечение, была достаточная точность вычислений и выполнялось условие универсальности. Тогда лучшей альтернативой будет s_1 , т.е.

$\text{МК}_{0,1,3}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, затем в порядке убывания предпочтительности: s_5, s_3, s_2 и s_4 , несмотря на то что эта МК не предпочтительней их в точности вычислений.

Если же, например, весовые коэффициенты критериев составят соответственно $\alpha_1 = 1, \alpha_2 = 0,5, \alpha_3 = 2, \alpha_4 = 1$ и $\alpha_5 = 1,5$, т.е. в задаче наиболее существенно, чтобы выполнялось условие универсальности и было большее число масштабных уровней, то лучшей альтернативой будет s_3 , т.е.

$\widetilde{\text{МК}}_{0,1,3,4}^{(A_{\alpha_i}^i A_{\alpha_j}^j / A_{\alpha_k}^k)}$, затем в порядке убывания предпочтительности: s_1, s_2, s_5 и s_4 .

На следующих этапах решения задачи моделирования для рассматриваемой практической задачи можно попытаться расширить набор критериев для оценки МК, сохраняя все естественные требования к набору критериев. При этом значения критерия оценки МК могут быть результатом оценки частных критериев БК. Такой подход приведет к более точным оценкам критериев МК и, следовательно, к более обоснованному результату выбора.

В связи с тем, что разработанная технология дает возможность формализовать процесс отбора наилучших в определенном смысле МК и позволяет решить поставленную задачу с учетом предпочтений экспертов и лиц, принимающих решение, можно говорить о том, что данный подход может быть применен при создании человеко-машинных систем автоматизированного проектирования. В дальнейшем при расширении набора критериев можно использовать и последовательно наращивать в автоматическом режиме структуры данных, получаемых при помощи отобранных МК.

Помимо всего прочего предложенный выше подход может применяться в так называемых обратных задачах выбора, в которых заранее указаны желательные значения частных критериев, описывающих МК, или диапазоны их изменения.

Литература

1. *Абгарян К. К.* Информационная технология построения многомасштабных моделей в задачах вычислительного материаловедения // Системы высокой доступности, 2018. Т. 15. № 2. С. 9–15.
2. *Бродский Ю. И.* Модельный синтез и модельно-ориентированное программирование. — М.: ВЦ РАН, 2013. 142 с.
3. *Abgaryan K. K., Mutigullin I. V., Reviznikov D. L.* Theoretical investigation of 2DEG concentration and mobility in the AlGaIn/GaN heterostructures with various Al concentrations // Phys. Status Solidi C, 2015. Vol. 12. Iss. 12. P. 1376–1382.
4. *Абгарян К. К.* Задачи оптимизации наноразмерных полупроводниковых гетероструктур // Известия вузов. Материалы электронной техники, 2016. Т. 19. № 2. С. 108–114.
5. *Абгарян К. К., Ревизников Д. Л.* Численное моделирование распределения носителей заряда в наноразмерных полупроводниковых гетероструктурах с учетом поляризационных эффектов // ЖВМ и МФ, 2016. Т. 56. № 1. С. 155–166.
6. *Кини Р. Л., Райфа Х.* Принятие решений при многих критериях предпочтения и замещения / Пер. с англ. — М.: Радио и связь, 1981. 559 с. (*Keeney R., Raiffa H.* Decisions with multiple objectives: Preferences and value tradeoffs. — New York, NY, USA: Wiley, 1975. 592 p.)
7. *Подиновский В. В.* Введение в теорию важности критериев. — М.: Физматлит, 2007. 64 с.
8. *Фишберн П.* Теория полезности для принятия решений / Пер. с англ. — М.: Наука, 1978. (*Fishburn P. C.* Utility theory for decision making. — New York, NY, USA: Wiley, 1970. 224 p.)
9. *Осипова В. А., Алексеев Н. С.* Математические методы поддержки принятия решений. — М.: ИНФРА-М, 2019. 133 с.
10. *Петровский А. Б.* Теория принятия решений. — М.: Академия, 2009. 391 с.

Поступила в редакцию 15.11.18

APPLICATION OF DECISION SUPPORT METHODS FOR THE MULTICRITERIAL SELECTION OF MULTISCALE COMPOSITIONS

K. K. Abgaryan^{1,2} and V. A. Osipova²

¹Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44/2 Vavilov Str., Moscow, 119333, Russian Federation

²Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation

Abstract: The article discusses the use of decision-making support methods for the task of selecting multiscale compositions (MC) — computational analogues of multiscale physical and mathematical models created for analyzing

various heterogeneous processes associated with the formation of new composite materials with predetermined properties. When solving specific problems, different multiscale models and their corresponding MC can be constructed. The question arises of comparing these models and assessing their “effectiveness” for specific problem. On the stage of predictive modeling, the authors propose a methodology for comparison of multiscale models through evaluation and selection of appropriate MC using methods of decision-making support under multiple criteria. As an illustration of the possibility of choosing the best alternative in the presence of additional information on evaluation criteria of MC, a model example associated with the study of electronic and structural properties of thin films InN (GaN) on silicon substrates is considered.

Keywords: multiscale modeling; decision theory; quality criteria; alternative; decision support methods; multiple criteria; value function

DOI: 10.14357/19922264190207

References

1. Abgaryan, K. K. 2018. Informatsionnaya tekhnologiya postroeniya mnogomasshtabnykh modeley v zadachakh vychislitel'nogo materialovedeniya [Information technology is the construction of multiscale models in problems of computational materials science]. *Sistemy vysokoy dostupnosti* [High Availability Systems] 15(2):9–15.
2. Brodskij, Yu. I. 2013. *Model'nyy sintez i model'no-orientirovannoe programirovanie* [Model synthesis and model-oriented programming]. Moscow: CC RAS. 142 p.
3. Abgaryan, K. K., I. V. Mutigullin, and D. I. Reviznikov. 2015. Theoretical investigation of 2DEG concentration and mobility in the AlGaIn/GaN heterostructures with various Al concentrations. *Phys. Status Solidi C* 12(12):1376–1382.
4. Abgaryan, K. K. 2016. Zadachi optimizatsii nanorazmernykh poluprovodnikovyykh geterostruktur [Optimization problems of nanosized semiconductor heterostructures]. *Proceedings of Higher Schools, Materials of Electronics Engineering* 19(2):108–114.
5. Abgaryan, K. K., and D. L. Reviznikov. 2016. Numerical simulation of the distribution of charge carrier in nanosized semiconductor heterostructures with account for polarization effects. *Comp. Math. Math. Phys.* 56(1):161–172.
6. Keeney, R., and H. Raiffa. 1975. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York, NY: Wiley, 1975. 592 p.
7. Podinovskiy, V. V. 2007. *Vvedenie v teoriyu vazhnosti kriteriev* [Introduction to criteria importance theory]. Moscow: Fizmatlit. 64 p.
8. Fishburn, P. C. 1970. *Utility theory for decision making*. New York, NY: Wiley. 224 p.
9. Osipova, V. A., and N. S. Alekseev. 2019. *Matematicheskie metody podderzhki prinyatiya resheniy* [Mathematical decision support methods]. Moscow: INFRA-M. 133 p.
10. Petrovskiy, A. B. 2009. *Teoriya prinyatiya resheniy* [Decision making theory]. Moscow: Akademiya. 391 p.

Received November 15, 2018

Contributors

Abgaryan Karine K. (b. 1963) — Doctor of Science in physics and mathematics, Head of Department, Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; Head of Department, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation; kristal83@mail.ru
Osipova Victoria A. (b. 1963) — Candidate of Science (PhD) in physics and mathematics, professor, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation; victoria.a.osipova@gmail.com

ИССЛЕДОВАНИЕ ПРОБЛЕМЫ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ЗАПАСОМ ДИСКРЕТНОГО ПРОДУКТА В СТОХАСТИЧЕСКОЙ МОДЕЛИ РЕГЕНЕРАЦИИ С НЕПРЕРЫВНО ПРОИСХОДЯЩИМ ПОТРЕБЛЕНИЕМ И СЛУЧАЙНОЙ ЗАДЕРЖКОЙ ПОСТАВКИ

П. В. Шнурков¹, Н. А. Вахтанов²

Аннотация: Рассматривается проблема оптимального управления запасом дискретного продукта в схеме регенерации с пуассоновским потоком требований потребителей. В исследуемой системе допускается отложенный спрос, объем которого ограничен заданной величиной. В качестве параметра управления рассматривается уровень запаса, при достижении которого необходимо делать заказ на пополнение. Показателем эффективности управления служит средняя удельная прибыль, полученная на одном периоде регенерации. Задача оптимального управления решается на основе утверждения об экстремуме дробно-линейного интегрального функционала на множестве дискретных вероятностных распределений.

Ключевые слова: управление запасом дискретного продукта; управляемый регенерирующий процесс; экстремальная задача для дробно-линейного интегрального функционала

DOI: 10.14357/19922264190208

1 Введение

Стохастические модели регенерации, используемые для исследования задач управления запасом, рассматривались в целом ряде публикаций. В частности, в работах [1, 2] были рассмотрены различные варианты регенерационных моделей для систем управления запасом непрерывного продукта. В работе [3] была исследована специальная версия модели регенерации, описывающая систему управления запасом непрерывного продукта, в которой непосредственное пополнение запаса происходит не мгновенно, а в течение определенного периода времени, называемого периодом реального пополнения. Рассмотренная в [3] версия стохастической модели регенерации идейно связана с классической детерминированной моделью управления запасом, изложенной, например, в [4]. Во всех упомянутых работах исследуются задачи управления запасом непрерывного продукта. В моделях такого вида множество значений основного случайного процесса, описывающего объем запаса в системе, представляет собой некоторое подмножество множества вещественных чисел. Примерами непрерывных продуктов, находящихся в реальных системах, могут служить вода, нефть и нефтепродукты, зерно и др. Однако существует ряд важных продуктов, объем которых измеряется в дискретных величинах. К таковым относятся многие потребительские товары, прежде всего товары бытового

техники, продовольственные товары и ряд других. Таким образом, рассмотрение математических моделей управления запасом дискретного продукта является актуальной проблемой прикладной математики.

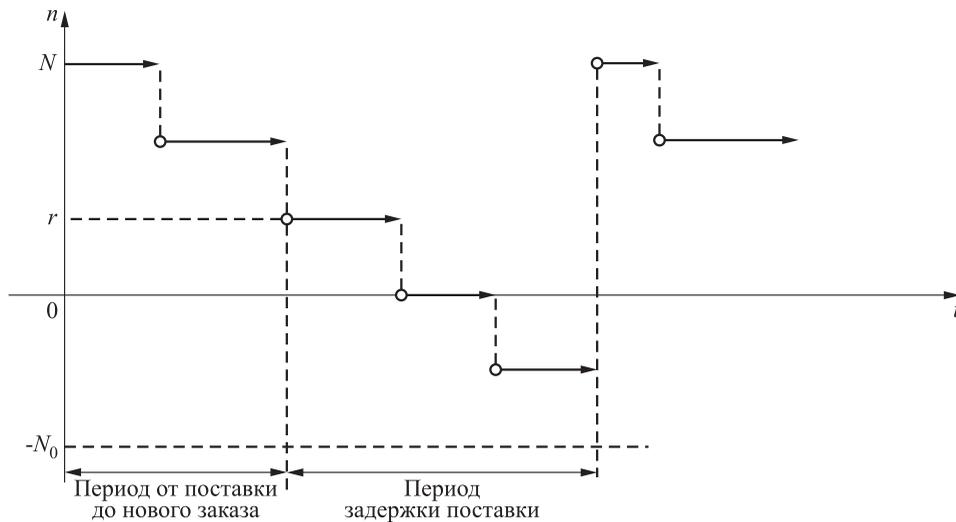
В зарубежной научной литературе по теории управления запасами стохастические модели регенерации известны достаточно давно. Модели, в которых моменты регенерации совпадают с моментами пополнения запаса до заданного фиксированного уровня, а показатель эффективности управления представляет собой средние удельные затраты на периоде регенерации, упоминаются в обзорах результатов, приведенных в [4, 5]. Однако среди научных исследований, выполненных в последние десятилетия, работы по управлению запасом дискретного продукта в схеме регенерации не встречаются.

2 Общее описание функционирования рассматриваемой системы управления дискретным запасом

Будем исследовать некоторую торговую систему (склад), предназначенную для временного хра-

¹ Национальный исследовательский университет «Высшая школа экономики», pshnurkov@hse.ru

² Национальный исследовательский университет «Высшая школа экономики», Vakhtanov1997@mail.ru



Эволюция объема запаса продукта в рассматриваемой системе

нения и поставки потребителю определенного однотипного товара, объем которого измеряется в дискретных целочисленных величинах. Предполагается, что в начальный момент времени система полностью заполнена и содержит N единиц продукта. Потребление товара осуществляется в случайные моменты поступления требований (покупателей). Моменты поступления требований образуют простейший (пуассоновский) поток с известным параметром $\lambda > 0$. В момент поступления очередного требования происходит потребление одной единицы продукта. Такое потребление осуществляется мгновенно.

В рассматриваемой системе периодически происходит пополнение запаса, которое осуществляется следующим образом. Заказ очередной партии товара происходит в момент времени, когда количество товара на складе достигает уровня r , где $r \leq N$ — некоторая целочисленная величина, которая в дальнейшем будет играть роль решения (управления). Заметим, что параметр r имеет вероятностную природу. Формальное описание процедуры выбора этого параметра приведено в разд. 3. Период времени от момента заказа товара до прибытия новой партии будем называть периодом задержки поставки или просто периодом задержки. Длительность этого периода является случайной величиной, имеющей заданное распределение $H_r(x)$, которое, вообще говоря, может зависеть от значения параметра управления r . Непосредственное пополнение запаса происходит мгновенно, в момент окончания периода задержки.

В период задержки поставки потребление товара продолжается согласно описанным правилам. После того как реальный запас товара будет израсходован, этот запас будет описываться отрица-

тельными величинами, что соответствует объему неудовлетворенного спроса или дефицита. При этом каждое поступающее требование принимается на учет и удовлетворяется при следующем пополнении. Такие требования образуют так называемый отложенный спрос. Объем отложенного спроса в данной модели не может превышать заданной величины N_0 . Если до момента очередного пополнения объем отложенного спроса достигает максимально допустимого уровня N_0 , то все поступающие после этого требования теряются. Предполагается, что пополнение запаса организовано так, что отложенный спрос полностью удовлетворяется, а запас товара пополняется до исходного уровня N . Дальнейшее функционирование системы происходит независимо от прошлого и в соответствии с описанными выше правилами. На рисунке приведена иллюстрация возможной эволюции объема запаса товара в системе. Заметим при этом, что параметр управления r может принимать любое целочисленное значение в пределах $-N_0 \leq r \leq N$.

3 Формальное построение математической модели, описывающей функционирование рассматриваемой системы

Будем предполагать, что все вводимые в дальнейшем стохастические объекты определены на некотором вероятностном пространстве (Ω, \mathcal{A}, P) , которое представляет собой формальную конструкцию, описывающую реализуемый случайный экс-

перимент с реальной системой. Прежде всего введем случайный процесс $\xi(t) = \xi(\omega, t)$, $\omega \in \Omega$, $t \in T = [0, \infty)$, который будет служить математической моделью функционирования рассматриваемой системы. Предположим, что значение этого процесса в произвольный момент времени t представляет собой объем запаса в системе, причём учитывается как реальный запас, так и дефицит, т. е. величина отложенного спроса. Таким образом, множество значений процесса $\xi(t)$ конечно: $X = \{-N_0, -N_0 + 1, \dots, -1, 0, 1, \dots, N - 1, N\}$.

Моментами изменения состояний процесса $\xi(t)$ являются моменты потребления (поступления требований от потребителей), если величина отложенного спроса не превышает значение N_0 , а также моменты непосредственного пополнения запаса. Для определенности будем предполагать, что траектории процесса непрерывны справа.

Обозначим через $t_n = t_n(\omega)$, $\omega \in \Omega$, $n = 0, 1, 2, \dots$, $t_0 = 0$, случайные моменты пополнения запаса. Согласно предположениям о поведении данной системы, описанным в предыдущем разделе, $\xi(t_n) = N$, $n = 0, 1, 2, \dots$. Кроме того, после каждого момента пополнения эволюция системы продолжается независимо от прошлого и по тем же закономерностям. Отсюда следует, что процесс $\xi(t)$ является регенерирующим [6], а моменты t_n , $n = 0, 1, 2, \dots$, представляют собой моменты его регенерации. Периоды регенерации данного процесса складываются из двух независимых частей:

$$\Delta_n = t_{n+1} - t_n = \Delta_n^{(0)} + \Delta_n^{(1)}, \quad (1)$$

где $\Delta_n^{(0)}$ — случайное время от момента очередного пополнения до момента заказа следующей поставки; $\Delta_n^{(1)}$ — случайная длительность задержки поставки (время от момента заказа до следующего пополнения запаса).

В соответствии с принятыми предположениями при выполнении дополнительного условия, что параметр управления принимает фиксированное значение r , случайная величина $\Delta_n^{(0)}$ имеет распределение Эрланга порядка $N - r$, а случайная величина $\Delta_n^{(1)}$ — заданное распределение $H_r(x)$.

Теперь опишем процедуру управления случайным процессом $\xi(t)$.

В качестве параметра управления будем рассматривать тот уровень запаса r , при котором происходит заказ очередной партии товара. Решение об управлении принимается в начальный момент каждого периода регенерации, после очередного пополнения запаса, т. е. в моменты времени $t_n + 0$, $n = 0, 1, 2, \dots$. В результате такого решения определяется уровень r , при котором будет осуществляться очередной заказ товара. Таким образом, множе-

ство допустимых значений параметра управления U оказывается конечным:

$$r \in U = \{N, N - 1, \dots, 0, 1, \dots, -N_0\}.$$

Заддим на множестве U совокупность всевозможных дискретных вероятностных распределений Γ_d . Каждое дискретное вероятностное распределение, принадлежащее этой совокупности, представляет собой вектор

$$\alpha = (\alpha_N, \alpha_{N-1}, \dots, \alpha_0, \alpha_{-1}, \dots, \alpha_{-N_0}),$$

удовлетворяющий условиям:

$$\alpha_i \geq 0, \quad i \in U, \quad \sum_{i \in U} \alpha_i = 1.$$

Любое фиксированное распределение α задает вероятностную стратегию управления в следующем смысле: в каждый момент принятия решения значение управления выбирается в соответствии с этим распределением. Иначе говоря, значение параметра управления r выбирается с вероятностью α_r , $r \in U$. Отметим, что стратегия управления одинакова для всех периодов регенерации. Задача оптимизации заключается в нахождении стратегии управления $\alpha^* \in \Gamma_d$, которая доставляет экстремум некоторому стационарному стоимостному показателю эффективности.

Указанный показатель зависит от исходных стоимостных характеристик модели. Перечислим эти характеристики, которые предполагаются известными:

- доход от реализации единицы продукции c_0 ;
- расходы на хранение единицы продукции за единицу времени c_1 ;
- расходы на приобретение единицы продукции c_2 ;
- расходы, связанные с дефицитом единицы продукции за единицу времени, c_3 ;
- расходы, связанные с потерей i клиентов, $c_4^{(i)}$, $i = 1, 2, \dots$.

Введенные стоимостные характеристики определяют все основные виды доходов и затрат, возникающих в моделях управления запасами.

4 Постановка задачи оптимального управления

Прежде всего введем некоторый аддитивный стоимостный функционал для рассматриваемой

модели. Пусть $\gamma(t) = \gamma(\omega, t)$, $\omega \in \Omega$, $t \in T = [0, \infty)$, — случайная прибыль, полученная в результате функционирования описанной выше стохастической системы на интервале времени $[0, t]$, $\gamma(0) = \gamma_0$ — заданная величина. Процесс зависит от траектории основного процесса $\xi(t)$, и его значения определяются также исходными стоимостными характеристиками $c_0, c_1, c_2, c_3, c_4^{(i)}$, $i = 1, 2, \dots$. Полное описание траекторий процесса $\gamma(t)$ сложно и не является необходимым для решения рассматриваемой задачи оптимального управления. Аналогичные стоимостные аддитивные функционалы известны в научной литературе. Общая схема их построения для управляемых марковских и полумарковских случайных процессов изложена в классических работах [7, 8]. В настоящей работе потребуются исследование вероятностных характеристик, связанных с приращениями процесса $\gamma(t)$ на периодах регенерации основного процесса $\xi(t)$.

Обозначим через $\gamma_n = \gamma(t_n)$, $n = 0, 1, 2, \dots$, значения процесса $\gamma(t)$ в моменты регенерации, полагая $\gamma(t_n + 0) = \gamma(t_n)$, $n = 0, 1, 2, \dots$. Пусть $\Delta\gamma_n = \gamma_{n+1} - \gamma_n$ есть приращение прибыли на периоде регенерации $(t_n, t_{n+1}]$, $n = 0, 1, 2, \dots$. Обозначим далее через $E_\alpha(\Delta\gamma_n)$ математическое ожидание приращения прибыли на периоде регенерации при условии, что стратегия управления процессом $\xi(t)$ определяется распределением $\alpha \in \Gamma_d$. Аналогично величина $E_\alpha(\Delta t_n) = E_\alpha(t_{n+1} - t_n)$ представляет собой математическое ожидание длительности периода регенерации при стратегии управления $\alpha \in \Gamma_d$. Введем также обозначение $E_\alpha\gamma(t)$ для математического ожидания всей прибыли, полученной на периоде времени $[0, t]$ при стратегии управления $\alpha \in \Gamma_d$.

В предположении, что решение на периоде регенерации фиксировано, т. е. параметр управления принимает значение $r \in U$, будем обозначать соответствующие условные математические ожидания через $E_r(\Delta\gamma_n)$ и $E_r(\Delta t_n)$, $n = 0, 1, 2, \dots$.

Изменение аддитивного функционала $\gamma(t)$ на отдельном периоде регенерации $(t_n, t_{n+1}]$ складывается из двух составляющих:

$$\Delta\gamma_n = \Delta\gamma_n^{(0)} + \gamma_n^{(1)}, \quad (2)$$

где $\Delta\gamma_n^{(0)} = \gamma(t_{n+1} - 0) - \gamma(t_n)$ — приращение функционала $\gamma(t)$ на открытом интервале (t_n, t_{n+1}) , в настоящей модели это приращение зависит от траектории процесса $\xi(t)$ на указанном интервале и определяется параметрами $c_0, c_1, c_2, c_3, c_4^{(i)}$, $i = 1, 2, \dots$; $\gamma_n^{(1)} = \gamma(t_{n+1}) - \gamma(t_{n+1} - 0)$ — приращение функционала $\gamma(t)$ в момент времени t_{n+1} , т. е. его мгновенное изменение в указанный момент времени. В рассматриваемой модели величина $\gamma_n^{(1)}$ пред-

ставляет собой затраты, связанные с очередным пополнением запаса. Отсюда следует, что данная величина отрицательна, а ее конкретное значение зависит от объема пополнения запаса в момент t_{n+1} и определяется равенством

$$\gamma_n^{(1)} = -c_2 [\xi(t_{n+1}) - \xi(t_{n+1} - 0)], \quad n = 0, 1, 2, \dots$$

С учетом отмеченных особенностей модели значение процесса $\gamma(t)$ в произвольный момент времени $t > 0$ также складывается из двух составляющих:

$$\gamma(t) = \gamma^{(0)}(t) + \gamma^{(1)}(t), \quad (3)$$

где $\gamma^{(0)}(t)$ — накопленная прибыль на интервале $[0, t]$, определяемая изменениями процесса $\xi(t)$ на периодах регенерации, исключая сами моменты регенерации, входящие в интервал $[0, t]$; $\gamma^{(1)}(t)$ — накопленные затраты на интервале $[0, t]$, связанные с пополнениями запаса в моменты регенерации процесса $\xi(t)$, входящие в интервал $[0, t]$.

Традиционно, начиная с первых фундаментальных работ по теории управления марковскими и полумарковскими случайными процессами [7, 8], в качестве показателя эффективности управления рассматривалась величина

$$I_\alpha = \lim_{t \rightarrow \infty} \frac{E_\alpha\gamma(t)}{t}, \quad (4)$$

которая по содержанию представляет собой среднюю удельную прибыль в рассматриваемой стохастической модели, определяемую при стратегии управления α . Однако в работах [7, 8] исследовались аддитивные стоимостные функционалы вида $\gamma^{(0)}(t)$, в которых не учитывались составляющие, возникающие в отдельные случайные моменты времени. Для того чтобы использовать показатель эффективности управления вида (4), необходимо доказать эргодическую теорему о существовании и явном представлении указанного предела для аддитивного функционала вида (3), учитывающего составляющую $\gamma^{(1)}(t)$. Докажем соответствующий результат для регенерирующего процесса $\xi(t)$ и определенного на нем стоимостного аддитивного функционала $\gamma(t)$.

Теорема. Пусть $\xi(t)$ — управляемый регенерирующий процесс, стратегия управления которым определяется вероятностным распределением $\alpha \in \Gamma_d$, $\gamma(t)$ — аддитивный стоимостный функционал, связанный с процессом $\xi(t)$, общая структура которого определяется соотношением (3). Предположим, что для любой допустимой стратегии управления выполняется условие $E_\alpha[\Delta t_n] > 0$. Тогда существует предел

$$I_\alpha = \lim_{t \rightarrow \infty} \frac{E_\alpha\gamma(t)}{t} = \frac{E_\alpha[\Delta\gamma_n]}{E_\alpha[\Delta t_n]}, \quad (5)$$

где $\Delta\gamma_n$ — полное изменение функционала $\gamma(t)$ на периоде регенерации $(t_n, t_{n+1}]$, определяемое соотношением (2).

Доказательство. Следуя классической теории восстановления [9], введем $\nu(t) = \sup\{n : t_n \leq t\}$. Процесс $\nu(t)$ обычно называют считающим для процесса восстановления, образованного моментами регенерации $\{t_n, n = 0, 1, 2, \dots\}$. Пусть $H(t) = E\nu(t)$ — функция восстановления процесса $\{t_n, n = 0, 1, 2, \dots\}$. Для фиксированного значения t случайная величина $\nu(t)$ совпадает с числом восстановлений на интервале $[0, t]$ (момент $t = 0$ не считается моментом восстановления).

По свойству аддитивности процесса $\gamma(t)$ можно утверждать, что

$$\gamma^{(1)}(t) = \sum_{n=0}^{\nu(t)-1} \gamma_n^{(1)} = \sum_{n=1}^{\nu(t)} \gamma_{n-1}^{(1)}. \quad (6)$$

Заметим, что из свойств построенного в данной работе регенерирующего процесса $\xi(t)$ и связанного с ним процесса $\gamma(t)$ следует, что для произвольного момента времени $t > 0$ и любого фиксированного значения $n = 1, 2, \dots$ случайное событие $(\nu(t) \leq n)$ не зависит от системы событий, порожденных случайными величинами $\{\gamma_{n+1}^{(1)}, \gamma_{n+2}^{(1)}, \dots\}$. Такое свойство в теории вероятностей называется независимостью от будущего [10]. Действительно, случайная величина $\nu(t)$ и событие $(\nu(t) \leq n)$ зависят от случайных моментов времени $t_1, t_2, \dots, t_n, t_{n+1}$ и соответствующих периодов регенерации $[0, t_1], (t_1, t_2], \dots, (t_n, t_{n+1}]$. В то же время случайные величины $\{\gamma_{n+1}^{(1)}, \gamma_{n+2}^{(1)}, \dots\}$ связаны с периодами регенерации $(t_{n+1}, t_{n+2}], (t_{n+2}, t_{n+3}], \dots$

При выполнении свойства независимости от будущего справедливо тождество Вальда [10, гл. 4, § 4]. Воспользовавшись этим результатом, получаем из (6):

$$E_\alpha \gamma^{(1)}(t) = E_\alpha \nu(t) E_\alpha \gamma_n^{(1)} = H(t) E_\alpha \gamma_n^{(1)}. \quad (7)$$

Из соотношения (3) с учетом (7) получаем:

$$\begin{aligned} I_\alpha &= \lim_{t \rightarrow \infty} \frac{E_\alpha \gamma(t)}{t} = \\ &= \lim_{t \rightarrow \infty} \frac{E_\alpha \gamma^{(0)}(t)}{t} + \lim_{t \rightarrow \infty} \frac{E_\alpha \gamma^{(1)}(t)}{t} = \\ &= \lim_{t \rightarrow \infty} \frac{E_\alpha \gamma^{(0)}(t)}{t} + E_\alpha \gamma_n^{(1)} \lim_{t \rightarrow \infty} \frac{H(t)}{t}. \end{aligned} \quad (8)$$

Первое слагаемое в правой части равенства (8) может быть определено на основании эргодической теоремы для аддитивного функционала. Классические формы такой теоремы для полумарковских

моделей приведены в [7, 8]. В данном случае для регенерирующего процесса $\xi(t)$ эргодическая теорема приводит к следующему результату:

$$\lim_{t \rightarrow \infty} \frac{E_\alpha \gamma^{(0)}(t)}{t} = \frac{E_\alpha [\Delta\gamma_n^{(0)}]}{E_\alpha [\Delta t_n]}. \quad (9)$$

Второе слагаемое в правой части равенства (8) определяется на основании элементарной теоремы восстановления [9]:

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{E_\alpha [\Delta t_n]}. \quad (10)$$

Заметим, что возможность применения указанной теоремы обеспечивается выполнением условия $E_\alpha [\Delta t_n] > 0$.

Тогда из (8) с учетом (9) и (10) получаем представление для средней удельной прибыли:

$$\begin{aligned} I_\alpha &= \lim_{t \rightarrow \infty} \frac{E_\alpha \gamma(t)}{t} = \frac{E_\alpha [\Delta\gamma_n^{(0)}] + E_\alpha \gamma_n^{(1)}}{E_\alpha [\Delta t_n]} = \\ &= \frac{E_\alpha [\Delta\gamma_n^{(0)} + \gamma_n^{(1)}]}{E_\alpha [\Delta t_n]}. \end{aligned} \quad (11)$$

Случайная величина $\Delta\gamma_n^{(0)} + \gamma_n^{(1)} = \Delta\gamma_n$ представляет собой полное изменение аддитивного стоимостного функционала на периоде регенерации $(t_n, t_{n+1}]$. Полученное соотношение (11) совпадает с (5). Теорема доказана.

Замечание 1. Утверждение теоремы остается справедливым для более общего варианта, в котором множество допустимых решений (управлений) U имеет произвольную структуру, а множество стратегий управления регенерирующим процессом $\xi(t)$ совпадает с множеством вероятностных распределений, заданных на U .

Замечание 2. Выполнение условия $E_\alpha [\Delta t_n] > 0$ можно обеспечить следующим образом. Из соотношения (1) следует, что

$$E_\alpha [\Delta t_n] = E_\alpha [\Delta_n^{(0)}] + E_\alpha [\Delta_n^{(1)}].$$

Здесь

$$\begin{aligned} E_\alpha [\Delta_n^{(0)}] &= \sum_{r \in U} \frac{N-r}{\lambda} \alpha_r; \\ E_\alpha [\Delta_n^{(1)}] &= \sum_{r \in U} \tau_r \alpha_r, \end{aligned}$$

где

$$\tau_r = E_r [\Delta_n^{(1)}] = \int_0^\infty x dH_r(x).$$

Очевидно, что $E_\alpha [\Delta_n^{(0)}] = 0$ только для распределения вида $(\alpha_N = 1, \alpha_r = 0, r \in U, r \neq N)$. Отсюда следует, что если $\tau_N > 0$, то $E_\alpha [\Delta t_n] > 0$ для указанного распределения вероятностей $(\alpha_N = 1, \alpha_r = 0, r \in U, r \neq N)$, а также и для любой другой стратегии принятия решений $\alpha \in \Gamma_d$. Если же потребовать выполнения более сильного естественного условия $\tau_r > 0, r \in U$ (средняя длительность задержки поставки не может быть равна нулю ни при каком допустимом решении), то будет иметь место оценка

$$E_\alpha [\Delta t_n] \geq E_\alpha [\Delta_n^{(1)}] > 0.$$

Из утверждения теоремы следует, что для нахождения явного представления показателя эффективности I_α необходимо определить аналитические представления математических ожиданий $E_\alpha [\Delta \gamma_n]$ и $E_\alpha [\Delta t_n]$.

Зафиксируем параметр управления $r \in U$ и обозначим:

$$A(r) = E_r [\Delta \gamma_n^{(0)} + \gamma_n^{(1)}] = E_r [\Delta \gamma_n]; \quad (12)$$

$$B(r) = E_r [\Delta t_n]. \quad (13)$$

По свойству математического ожидания справедливы следующие формулы:

$$E_\alpha [\Delta \gamma_n] = \sum_{r \in U} E_r [\Delta \gamma_n] \alpha_r = \sum_{r \in U} A(r) \alpha_r; \quad (14)$$

$$E_\alpha [\Delta t_n] = \sum_{r \in U} E_r [\Delta t_n] \alpha_r = \sum_{r \in U} B(r) \alpha_r. \quad (15)$$

Из (11) с учетом (14) и (15) получаем:

$$I_\alpha = \frac{E_\alpha [\Delta \gamma_n]}{E_\alpha [\Delta t_n]} = \frac{\sum_{r \in U} A(r) \alpha_r}{\sum_{r \in U} B(r) \alpha_r}. \quad (16)$$

Задача оптимального управления в рассматриваемой модели может быть сформулирована в виде следующей экстремальной проблемы:

$$I_\alpha = \frac{\sum_{r \in U} A(r) \alpha_r}{\sum_{r \in U} B(r) \alpha_r} \rightarrow \max, \alpha \in \Gamma_d. \quad (17)$$

Функционал (16) по форме представляет собой так называемый дробно-линейный интегральный функционал, заданный на множестве дискретных вероятностных распределений Γ_d . Изложим общий подход к решению поставленной задачи.

5 О решении задачи оптимального управления в рассматриваемой стохастической модели

Теория решения задачи безусловного экстремума для общего дробно-линейного интегрального функционала, заданного на множестве наборов вероятностных мер, изложена в работе [11]. Для решения поставленной экстремальной задачи (17) необходимо исследовать на экстремум специальный вариант дробно-линейного интегрального функционала, в котором вероятностные меры являются дискретными. Полное описание соответствующей экстремальной задачи и формулировка утверждения о ее решении не могут быть приведены в рамках настоящей статьи. В связи с этим упомянутые результаты включены в приложение [12]. Отметим также, что утверждение о безусловном экстремуме дробно-линейного интегрального функционала, заданного на множестве наборов дискретных вероятностных распределений, приведено в работе [13].

Основываясь на указанном теоретическом утверждении, сформулируем главный результат, связанный с решением экстремальной задачи (17). Если основная функция дробно-линейного дискретного интегрального функционала (16), определяемая формулой $C(r) = A(r)/B(r)$, где функции $A(r)$ и $B(r)$ определены соотношениями (12) и (13), достигает глобального максимума в некоторой фиксированной точке $r^* \in U = \{N, N-1, \dots, 0, -1, \dots, -N_0\}$, то решение исходной задачи (17) существует и достигается на вырожденном распределении α^* , сосредоточенном в точке r^* . Заметим, что в рассматриваемой задаче множество допустимых решений U конечно и глобальный максимум основной функции $C(r)$ достигается. Таким образом, для полного решения поставленной задачи необходимо найти явные аналитические выражения для функций $A(r)$ и $B(r)$. После этого точка глобального максимума функции $C(r)$ может быть определена численным методом.

6 Заключение

В настоящей статье изложена первая часть исследования задачи управления запасом дискретного продукта в схеме регенерации. В ней разработана новая стохастическая модель управления запасом дискретного продукта, которая представляет собой регенерирующий процесс с конечным множеством состояний. Сформулирована задача оптимального

управления, которая по своей математической форме представляет собой экстремальную задачу для функционала, заданного на множестве дискретных вероятностных распределений, каждое из которых определяет стратегию управления. Поскольку целевой функционал по своей структуре является дробно-линейным интегральным функционалом, для нахождения решения используется теорема о безусловном экстремуме функционалов данного вида. На основании этой теоремы установлено, что оптимальная стратегия управления в рассматриваемой задаче существует, является детерминированной и определяется точкой максимума некоторой функции, которая по своему идейному содержанию представляет собой среднюю удельную прибыль, зависящую от параметра управления.

Вторая часть исследования задачи управления запасом посвящена нахождению явного аналитического представления для указанной функции. Эта часть будет представлена авторами в следующей статье.

Литература

1. Шнурков П. В., Мельников Р. В. Оптимальное управление запасом непрерывного продукта в модели регенерации // Обозрение прикладной и промышленной математики, 2006. Т. 13. Вып. 3. С. 434–452.
2. Шнурков П. В., Мельников Р. В. Исследование проблемы управления запасом непрерывного продукта при детерминированной задержке поставки // Автоматика и телемеханика, 2008. № 10. С. 93–113.
3. Шнурков П. В., Пименова Е. Ю. Оптимальное управление запасом непрерывного продукта в схеме регенерации с детерминированной задержкой поставки и периодом реального пополнения // Системы и средства информатики, 2017. Т. 27. № 4. С. 80–94.
4. Porteus E. L. Foundations of stochastic inventory theory. — Stanford, CA, USA: Stanford Business Book, 2002. 299 p.
5. Simchi-Levi D., Chen X., Bramel J. The logic of logistics: Theory, algorithms, and applications for logistics management. — New York, NY, USA: Springer-Verlag, 2014. 447 p.
6. Рыков В. В., Козырев Д. В. Основы теории массового обслуживания. — М.: Инфра-М, 2016. 224 с.
7. Джевелл В. Управляемые полумарковские процессы // Кибернетический сборник. Новая серия. — М.: Мир, 1967. Вып. 4. С. 97–134.
8. Майн Х., Осаки С. Марковские процессы принятия решений / Пер. с англ. — М.: Наука, 1977. 176 с. (Mine H., Osaki S. Markovian decision processes. — New York, NY, USA: Elsevier, 1970. 142 p.)
9. Климов Г. П. Теория массового обслуживания. — М.: Изд-во Московского ун-та, 2011. 312 с.
10. Боровков А. А. Теория вероятностей. — М.: Либроком, 2009. 656 с.
11. Шнурков П. В. О решении задачи безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. Акад. наук. Сер. Математика, 2016. Т. 470. № 4. С. 387–392.
12. Шнурков П. В., Вахтанов Н. А. Приложение к статье «Исследование проблемы оптимального управления запасом дискретного продукта в стохастической модели регенерации с непрерывно происходящим потреблением и случайной задержкой поставки», 2019. 16 с. <http://www.ipiran.ru/publications/Приложение.pdf>.
13. Shnourkoff P. V., Novikov D. A. Analysis of the problem of intervention control in the economy on the basis of solving the problem of tuning // arXiv.org, 2018. arXiv:1811.10993 [q-fin.GN]. 15 p.

Поступила в редакцию 19.02.19

RESEARCH OF THE OPTIMAL CONTROL PROBLEM OF INVENTORY OF A DISCRETE PRODUCT IN THE STOCHASTIC REGENERATION MODEL WITH CONTINUOUSLY OCCURRING CONSUMPTION AND RANDOM DELIVERY DELAY

P. V. Shnurkov and N. A. Vakhtanov

National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

Abstract: The paper considers the optimal control problem of inventory of a discrete product in a regeneration scheme with a Poisson flow of customer requirements. In the system, deferred demand is allowed, the volume of which is limited by a given value. The control parameter is the level of the stock, at which achievement it is necessary to make an order for replenishment. The indicator of management effectiveness is the average specific profit received in one regeneration period. The optimal control problem is solved on the basis of the statement about the extremum of a fractional-linear integral functional on the set of discrete probability distributions.

Keywords: inventory management of a discrete product; controlled regenerative process; extremal problem for a fractional-linear integral functional

DOI: 10.14357/19922264190208

References

1. Shnurkov, P. V., and R. V. Mel'nikov. 2006. Optimal'noe upravlenie zapasom nepreryvnogo produkta v modeli regeneratsii [Optimal control of a continuous product inventory in the regeneration model]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Review of Applied and Industrial Mathematics] 13(3):434–452.
2. Shnurkov, P. V., and R. V. Mel'nikov. 2008. Analysis of the problem of continuous-product inventory control under deterministic lead time. *Automat. Rem. Contr.* 69(10):1734–1751.
3. Shnurkov, P. V., and E. Yu. Pimenova. 2017. Optimal'noe upravlenie zapasom nepreryvnogo produkta v skheme regeneratsii s determinirovannoy zaderzhkoy postavki i periodom real'nogo popolneniya [Optimal inventory control of continuous product in regeneration theory with determinate delay of the delivery and the period of real replenishment]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):80–94.
4. Porteus, E. L. 2002. *Foundations of stochastic inventory theory*. Stanford, CA: Stanford Business Book. 299 p.
5. Simchi-Levi, D., X. Chen, and J. Bramel. 2014. *The logic of logistics: Theory, algorithms, and applications for logistics management*. New York, NY: Springer-Verlag. 447 p.
6. Rykov, V. V., and D. V. Kozhev. 2016. *Osnovy teorii massovogo obsluzhivaniya* [Fundamentals of queuing theory]. Moscow: Infra-M. 224 p.
7. Jewell, W. S. 1963. Markov-renewal programming. *Oper. Res.* 11:938–971.
8. Mine, H., and S. Osaki. 1970. *Markovian decision processes*. New York, NY: Elsevier. 142 p.
9. Klimov, G. P. 2011. *Teoriya massovogo obsluzhivaniya* [Queuing theory]. Moscow: MSU Pubs. 312 p.
10. Borovkov, A. A. 2009. *Teoriya veroyatnostey* [Probability theory]. Moscow: Librokom. 656 p.
11. Shnurkov, P. V. 2016. Solution of the unconditional extremum problem for a linear-fractional integral functional on a set of probability measures. *Dokl. Math.* 94(2):550–554.
12. Shnurkov, P. V., and N. A. Vakhtanov. 2019. Prilozhenie k stat'ye "Issledovanie problemy optimal'nogo upravleniya zapasom diskretnogo produkta v stokhasticheskoy modeli regeneratsii s nepreryvno proiskhodyashchim potrebleniem i sluchaynoy zaderzhkoy postavki" [Appendix to article "Research of the optimal control problem of inventory of a discrete product in stochastic regeneration model with continuously occurring consumption and random delivery delay"]. 16 p. Available at: <http://www.ipiran.ru/publications/Приложение.pdf> (accessed May 6, 2019).
13. Shnourkoff, P. V., and D. A. Novikov. 2018. Analysis of the problem of intervention control in the economy on the basis of solving the problem of tuning. arxiv.org. arXiv:1811.10993 [q-fin.GN]. 15 p.

Received February 19, 2019

Contributors

Shnurkov Peter V. (b. 1953) — Candidate of Science (PhD) in physics and mathematics, associate professor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; pshnurkov@hse.ru

Vakhtanov Nikita A. (b. 1997) — Master student, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; Vakhtanov1997@mail.ru

ОПРЕДЕЛЕНИЕ РЕЛЕВАНТНОСТИ ПАРАМЕТРОВ НЕЙРОСЕТИ*

А. В. Грабовой¹, О. Ю. Бахтеев², В. В. Стрижов³

Аннотация: Работа посвящена оптимизации структуры нейронной сети. Предполагается, что число параметров нейросети можно существенно снизить без значимой потери качества и значимого повышения дисперсии функции ошибки. Предлагается метод прореживания параметров нейронной сети, основанный на автоматическом определении релевантности параметров. Для определения релевантности параметров предлагается проанализировать ковариационную матрицу апостериорного распределения параметров и удалить из нейросети мультикоррелирующие параметры. Для определения мультикорреляции используется метод Белсли. Для анализа качества представленного алгоритма проводятся эксперименты на выборке Boston Housing, а также на синтетических данных.

Ключевые слова: нейронные сети; оптимизация гиперпараметров; метод Белсли; релевантность параметров; прореживание нейронной сети

DOI: 10.14357/19922264190209

1 Введение

Решается задача выбора оптимальной структуры нейронной сети. В силу высокой вычислительной сложности время оптимизации нейронных сетей может занимать до нескольких дней [1]. Поэтому построение и выбор оптимальной структуры нейронной сети также является вычислительно сложной процедурой, которая значимо влияет на итоговое качество модели. Использование избыточно сложных моделей с избыточным числом неинформативных параметров служит препятствием для использования глубоких сетей на мобильных устройствах в режиме реального времени.

Существует ряд подходов к построению оптимальной сети. В работах [2, 3] предлагается использовать модель градиентного спуска для оптимизации сети. В [4] используются байесовские методы [5] оптимизации параметров нейронных сетей. Еще один метод поиска оптимальной структуры заключается в прореживании избыточно сложной модели [6–8]. В работе [6] предлагается удалять наименее релевантные параметры на основе значений первой и второй производных функции ошибки.

Данная работа посвящена прореживанию структуры сети. Предлагается удалять наименее релевантные параметры модели. Под релевант-

ностью [6] подразумевается то, насколько сильно параметр влияет на функцию ошибки. Малая релевантность указывает на то, что удаление этого параметра не влечет значимого изменения функции ошибки. Метод предлагает построение исходной избыточно сложной нейросети с большим числом избыточных параметров. Для определения релевантности параметров предлагается оптимизировать параметры и гиперпараметры в единой процедуре. Для удаления параметров предлагается использовать метод Белсли [9].

Проверка и анализ метода проводятся на выборке Boston Housing [10], Wine [11] и синтетических данных. Результат сравнивается с моделью, полученной при помощи базовых алгоритмов.

2 Постановка задачи

Задана выборка

$$\mathcal{D} = \{\mathbf{x}_i, y_i\}, \quad i = 1, \dots, N,$$

где $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \{1, \dots, Y\}$, Y — число классов. Рассмотрим модель $f(\mathbf{x}, \mathbf{w}) : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \{1, \dots, Y\}$, где $\mathbf{w} \in \mathbb{R}^n$ — пространство параметров модели,

$$f(\mathbf{x}, \mathbf{w}) = \text{softmax}(f_1(f_2(\dots(f_l(\mathbf{x}, \mathbf{w}))))),$$

где $f_k(\mathbf{x}, \mathbf{w}) = \tanh(\mathbf{w}^T \mathbf{x})$, $k \in \{1, \dots, l\}$; l — число слоев нейронной сети. Параметр w_j модели f

* Работа выполнена при поддержке РФФИ (проект 19-07-0875) и Правительства РФ (соглашение 05.Y09.21.0018). Настоящая статья содержит результаты проекта «Статистические методы машинного обучения», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ им. М. В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 11.12.2018 № 13/1251/2018.

¹ Московский физико-технический институт, grabovoy.av@phystech.edu

² Московский физико-технический институт, bakhteev@phystech.edu

³ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский физико-технический институт, strijov@ccas.ru

называется активным, если $w_j \neq 0$. Множество индексов активных параметров обозначим $\mathcal{A} \subset \mathcal{J} = \{1, \dots, n\}$. Задано пространство параметров модели

$$\mathbb{W}_{\mathcal{A}} = \{\mathbf{w} \in \mathbb{R}^n \mid w_j \neq 0, j \in \mathcal{A}\}.$$

Для модели f с множеством индексов активных параметров \mathcal{A} и соответствующего ей вектора параметров $\mathbf{w} \in \mathbb{W}_{\mathcal{A}}$ определим логарифмическую функцию правдоподобия выборки:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}, \mathcal{A}, \mathbf{w}) = \log p(\mathcal{D}|\mathcal{A}, \mathbf{w}), \quad (1)$$

где $p(\mathcal{D}|\mathcal{A}, \mathbf{w})$ — апостериорная вероятность выборки \mathcal{D} при заданных \mathbf{w}, \mathcal{A} . Оптимальные значения \mathbf{w}, \mathcal{A} находятся из минимизации $-\mathcal{L}_{\mathcal{A}}(\mathcal{D}, \mathcal{A})$ — логарифма правдоподобия модели:

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\mathcal{D}, \mathcal{A}) &= \log p(\mathcal{D}|\mathcal{A}) = \\ &= \log \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} p(\mathcal{D}|\mathbf{w})p(\mathbf{w}|\mathcal{A}) d\mathbf{w}, \quad (2) \end{aligned}$$

где $p(\mathbf{w}|\mathcal{A})$ — априорная вероятность вектора параметров в пространстве $\mathbb{W}_{\mathcal{J}}$.

Так как вычисление интеграла (2) является вычислительно сложной задачей, рассмотрим вариационный подход [12] для решения этой задачи. Пусть задано распределение:

$$q(\mathbf{w}) \sim \mathcal{N}(\mathbf{m}, \mathbf{A}_{ps}^{-1}).$$

Здесь $\mathbf{m}, \mathbf{A}_{ps}^{-1}$ — вектор средних и матрица ковариации, аппроксимирующее неизвестное апостериорное распределение $p(\mathbf{w}|\mathcal{D}, \mathcal{A})$, полученное при априорном предположении

$$p(\mathbf{w}|\mathcal{A}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{pr}^{-1}),$$

где $\boldsymbol{\mu}, \mathbf{A}_{pr}^{-1}$ — вектор средних и матрица ковариации априорного распределения.

Приблизим интеграл (2) методом, предложенным в [12]:

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\mathcal{D}, \mathcal{A}) &= \log p(\mathcal{D}|\mathcal{A}) = \\ &= \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} q(\mathbf{w}) \log \frac{p(\mathcal{D}, \mathbf{w}|\mathcal{A})}{q(\mathbf{w})} d\mathbf{w} - \\ &- \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} q(\mathbf{w}) \log \frac{p(\mathbf{w}|\mathcal{D}, \mathcal{A})}{q(\mathbf{w})} d\mathbf{w} \approx \\ &\approx \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} q(\mathbf{w}) \log \frac{p(\mathcal{D}, \mathbf{w}|\mathcal{A})}{q(\mathbf{w})} d\mathbf{w} = \end{aligned}$$

$$\begin{aligned} &= \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} q(\mathbf{w}) \log \frac{p(\mathbf{w}|\mathcal{A})}{q(\mathbf{w})} d\mathbf{w} + \\ &+ \int_{\mathbf{w} \in \mathbb{W}_{\mathcal{J}}} q(\mathbf{w}) \log p(\mathcal{D}|\mathcal{A}, \mathbf{w}) d\mathbf{w} = \\ &= \mathcal{L}_{\mathbf{w}}(\mathcal{D}, \mathcal{A}, \mathbf{w}) + \mathcal{L}_E(\mathcal{D}, \mathcal{A}). \quad (3) \end{aligned}$$

Первое слагаемое формулы (3) — это сложность модели. Оно определяется расстоянием Кульбака–Лейблера:

$$\mathcal{L}_{\mathbf{w}}(\mathcal{D}, \mathcal{A}, \mathbf{w}) = -D_{\text{KL}}(q(\mathbf{w})||p(\mathbf{w}|\mathcal{A})).$$

Второе слагаемое формулы (3) представляет собой математическое ожидание правдоподобия выборки $\mathcal{L}_{\mathcal{D}}(\mathcal{D}, \mathcal{A}, \mathbf{w})$. В данной работе оно является функцией ошибки:

$$\mathcal{L}_E(\mathcal{D}, \mathcal{A}) = \mathbb{E}_{\mathbf{w} \sim q} \mathcal{L}_{\mathcal{D}}(\mathcal{D}, \mathcal{A}, \mathbf{w}).$$

Требуется найти параметры, доставляющие минимум суммарному функционалу потерь $\mathcal{L}_{\mathcal{A}}(\mathcal{D}, \mathcal{A}, \mathbf{w})$ из (3):

$$\begin{aligned} \hat{\mathbf{w}} &= \arg \min_{\mathcal{A} \subset \mathcal{J}, \mathbf{w} \in \mathbb{W}_{\mathcal{A}}} -\mathcal{L}_{\mathcal{A}}(\mathcal{D}, \mathcal{A}, \mathbf{w}) = \\ &= \arg \min_{\mathcal{A} \subset \mathcal{J}, \mathbf{w} \in \mathbb{W}_{\mathcal{A}}} D_{\text{KL}}(q(\mathbf{w})||p(\mathbf{w}|\mathcal{A})) - \mathcal{L}_{\mathcal{D}}(\mathcal{D}, \mathcal{A}, \mathbf{w}). \quad (4) \end{aligned}$$

3 Базовые методы прореживания нейросети

3.1 Случайное удаление

Метод случайного удаления заключается в том, что случайным образом удаляется некоторый параметр w_{ξ} из множества активных параметров сети. Индекс параметра ξ из равномерного распределения — случайная величина, предположительно доставляющая оптимум в (4):

$$\xi \sim \mathcal{U}(\mathcal{A}).$$

3.2 Оптимальное прореживание

Метод оптимального прореживания [6] использует вторую производную целевой функции (1) по параметрам для выявления нерелевантных параметров. Рассмотрим функцию потерь \mathcal{L} (1), разложенную в ряд Тейлора в некоторой окрестности вектора параметров \mathbf{w} :

$$\delta \mathcal{L} = \sum_{j \in \mathcal{A}} g_j \delta w_j + \frac{1}{2} \sum_{i, j \in \mathcal{A}} h_{ij} \delta w_i \delta w_j + O(\|\delta \mathbf{w}\|^3), \quad (5)$$

где δw_j — компоненты вектора $\delta \mathbf{w}$; g_j — компоненты вектора градиента $\nabla \mathcal{L}$; h_{ij} — компоненты гессиана \mathbf{H} :

$$g_j = \frac{\partial \mathcal{L}}{\partial w_j}, \quad h_{ij} = \frac{\partial^2 \mathcal{L}}{\partial w_i \partial w_j}.$$

Задача является вычислительно сложной в силу высокой размерности матрицы \mathbf{H} . Введем предположение [6] о том, что удаление нескольких параметров приводит к такому же изменению функции потерь \mathcal{L} , как и суммарное изменение при индивидуальном удалении:

$$\delta \mathcal{L} = \sum_{j \in \mathcal{A}} \delta \mathcal{L}_j,$$

где \mathcal{A} — множество активных параметров; $\delta \mathcal{L}_j$ — изменение функции потерь при удалении одного параметра w_j .

В силу данного предположения будем рассматривать только диагональные элементы матрицы \mathbf{H} . После введенного предположения выражение (5) принимает вид:

$$\delta \mathcal{L} = \frac{1}{2} \sum_{j \in \mathcal{A}} h_{jj} \delta w_j^2.$$

Получаем следующую задачу оптимизации:

$$\xi = \arg \min_{j \in \mathcal{A}} h_{jj} \frac{w_j^2}{2},$$

где ξ — индекс наименее релевантного, удаляемого параметра, предположительно доставляющего оптимум в (4).

3.3 Удаление неинформативных параметров с помощью вариационного вывода

Для удаления параметров в работе [7] предлагается удалить параметры, которые имеют максимальное отношение плотности $p(\mathbf{w}|\mathcal{A})$ априорной вероятности в нуле к плотности априорной вероятности в математическом ожидании μ_j параметра w_j .

Для гауссовского распределения с диагональной матрицей ковариации получаем:

$$p_j(\mathbf{w}|\mathcal{A})(w) = \frac{1}{\sqrt{2\sigma_j^2}} \exp\left(-\frac{(w - \mu_j)^2}{2\sigma_j^2}\right),$$

где w — значение носителя распределенного параметра. Разделим плотность вероятности в нуле к плотности в математическом ожидании

$$\frac{p_j(\mathbf{w}|\mathcal{A})(0)}{p_j(\mathbf{w}|\mathcal{A})(\mu_j)} = \exp\left(-\frac{\mu_j^2}{2\sigma_j^2}\right)$$

и поставим следующую задачу оптимизации:

$$\xi = \arg \min_{j \in \mathcal{A}} \left| \frac{\mu_j}{\sigma_j} \right|,$$

где ξ — индекс наименее релевантного, удаляемого параметра.

4 Предлагаемый метод определения релевантности параметров нейросети

Предлагается метод, основанный на модификации метода Белсли. Пусть \mathbf{w} — вектор параметров, доставляющий минимум функционалу потерь $\mathcal{L}_{\mathcal{A}}$ из (3) на множестве $\mathbb{W}_{\mathcal{A}}$, а \mathbf{A}_{ps} — соответствующая ему ковариационная матрица.

Выполним сингулярное разложение матрицы

$$\mathbf{A}_{ps} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T.$$

Индекс обусловленности η_j определим как отношение максимального элемента к j -му элементу матрицы $\mathbf{\Lambda}$. Для нахождения мультикоррелирующих признаков требуется найти индекс ξ вида

$$\xi = \arg \max_{j \in \mathcal{A}} \eta_j.$$

Дисперсионный долевым коэффициент q_{ij} определим как вклад j -го признака в дисперсию i -го элемента вектора параметра \mathbf{w} :

$$q_{ij} = \frac{u_{ij}^2 / \lambda_{jj}}{\sum_{j=1}^n u_{ij}^2 / \lambda_{jj}}.$$

Большие значения дисперсионных долей указывают на наличие зависимости между параметрами. Находим долевым коэффициенты, которые вносят максимальный вклад в дисперсию параметра w_{ξ} :

$$\zeta = \arg \max_{j \in \mathcal{A}} q_{\xi j}.$$

Параметр с индексом ζ определим как наименее релевантный параметр нейросети.

Проиллюстрируем принцип работы метода Белсли на примере. Рассмотрим данные, порожденные следующим образом:

$$\mathbf{w} = \begin{bmatrix} \sin(x) \\ \cos(x) \\ 2 + \cos(x) \\ 2 + \sin(x) \\ \cos(x) + \sin(x) \\ x \end{bmatrix},$$

с матрицей ковариации, представленной на рис. 1, а, где $x \in [0, 0; 0, 02; \dots; 20, 0]$.

В табл. 1 приведены индексы обусловленности и соответствующие им дисперсионные доли, ко-

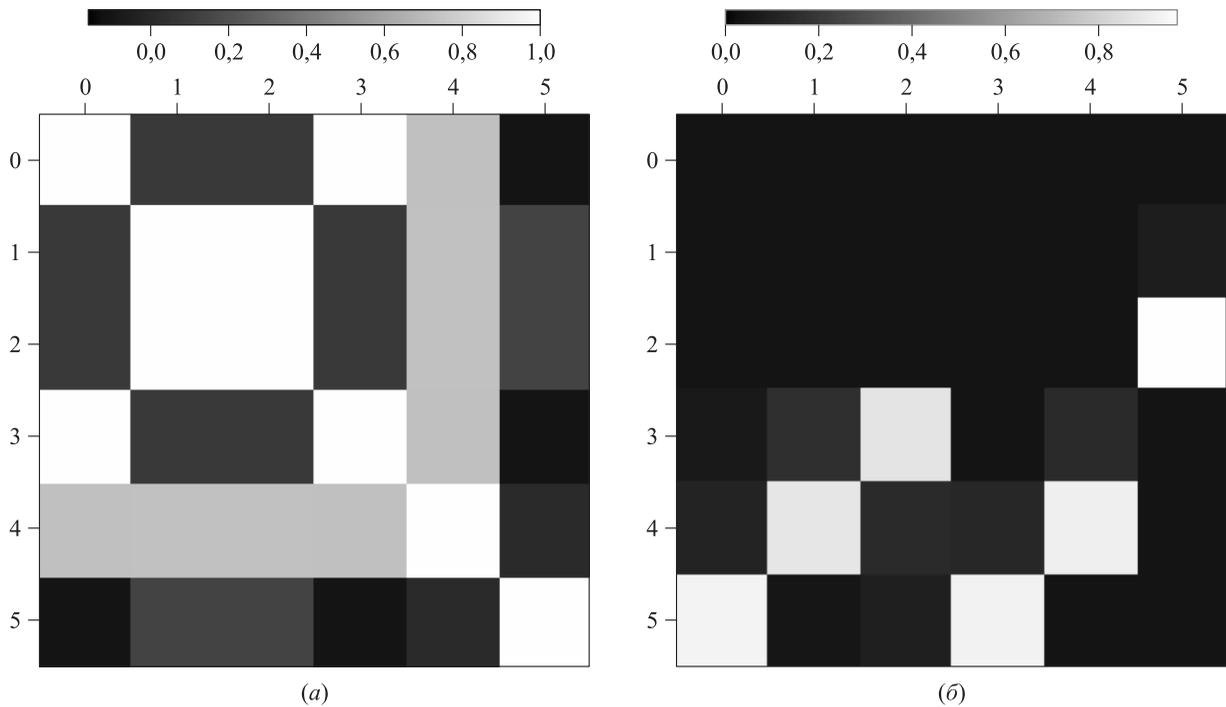


Рис. 1 Иллюстрация метода Белсли: (а) матрица ковариации; (б) дисперсионные доли

Таблица 1 Иллюстрация метода Белсли

η	q_1	q_2	q_3	q_4	q_5	q_6
1,0	$2 \cdot 10^{-17}$	$4 \cdot 10^{-17}$	$1 \cdot 10^{-16}$	$2 \cdot 10^{-17}$	$6 \cdot 10^{-17}$	$3 \cdot 10^{-4}$
1,5	$5 \cdot 10^{-17}$	$9 \cdot 10^{-17}$	$2 \cdot 10^{-16}$	$5 \cdot 10^{-17}$	$3 \cdot 10^{-20}$	$3 \cdot 10^{-2}$
3,3	$9 \cdot 10^{-18}$	$1 \cdot 10^{-17}$	$2 \cdot 10^{-17}$	$9 \cdot 10^{-18}$	$2 \cdot 10^{-19}$	$9 \cdot 10^{-1}$
$2 \cdot 10^{15}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$8 \cdot 10^{-1}$	$2 \cdot 10^{-3}$	$9 \cdot 10^{-2}$	$1 \cdot 10^{17}$
$8 \cdot 10^{15}$	$6 \cdot 10^{-2}$	$8 \cdot 10^{-1}$	$9 \cdot 10^{-2}$	$8 \cdot 10^{-2}$	$9 \cdot 10^{-1}$	$2 \cdot 10^{17}$
$1 \cdot 10^{16}$	$9 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$9 \cdot 10^{-1}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-21}$

торые также изображены на рис. 1, б. Согласно этим данным максимальный индекс обусловленности $\eta_6 = 1,2 \cdot 10^{16}$. Ему соответствуют максимальные дисперсионные доли признаков с индексами 1 и 4, которые, как видно из построения выборки, коррелируют между собой.

5 Вычислительный эксперимент

Для анализа свойств предложенного алгоритма и сравнения его с существующими был проведен вычислительный эксперимент, в котором параметры нейросети удалялись методами, описанными в подразд. 3.1–3.3, и методом Белсли.

В качестве данных использовались три выборки, представленные в табл. 2. Выборки Wine [11] и Boston Housing [10] — это реальные данные. Синтетические данные сгенерированы таким образом, чтобы параметры сети были мультикоррелируемыми. Генерация данных состояла из двух этапов.

Таблица 2 Описание выборок

Выборка	Тип задачи	Размер выборки	Число признаков
Wine	Классификация	178	13
Boston Housing	Регрессия	506	13
Synthetic data	Регрессия	10 000	100

На первом этапе генерировался вектор параметров $\mathbf{w}_{\text{synthetic}}$:

$$\mathbf{w}_{\text{synthetic}} \sim \mathcal{N}(\mathbf{m}_{\text{synthetic}}, \mathbf{A}_{\text{synthetic}}),$$

где

$$\mathbf{m}_{\text{synthetic}} = \begin{bmatrix} 1,0 \\ 0,0025 \\ \vdots \\ 0,0025 \end{bmatrix};$$

$$\mathbf{A}_{\text{synthetic}} = \begin{bmatrix} 1,0 & 10^{-3} & \dots & 10^{-3} & 10^{-3} \\ 10^{-3} & 1,0 & \dots & 0,95 & 0,95 \\ \dots & \dots & \dots & \dots & \dots \\ 10^{-3} & 0,95 & \dots & 0,95 & 1,0 \end{bmatrix}.$$

На втором этапе генерировалась выборка $\mathcal{D}_{\text{synthetic}}$:

$$\mathcal{D}_{\text{synthetic}} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \sim \mathcal{N}(\mathbf{1}, \mathbf{I}), y_i = x_{i0}, i = 1, \dots, 10\,000\}.$$

В приведенном выше векторе параметров $\mathbf{w}_{\text{synthetic}}$ для выборки $\mathcal{D}_{\text{synthetic}}$ наиболее релевантным является первый параметр, а все остальные параметры — нерелевантные. Матрица ковариации была выбрана таким образом, чтобы все нерелевантные параметры были зависимы и метод Белсли был максимально эффективен.

Для алгоритмов тренировочная и тестовая выборки составили 80% и 20% соответственно. Критерием качества прореживания служит доля параметров нейросети, удаление которых не влечет значимой потери качества прогноза. Также критерием качества служит устойчивость нейросети к зашумленности данных.

Качеством прогноза R_{cl} модели для задачи классификации выступает точность прогноза модели:

$$R_{\text{cl}} = \frac{\sum_{(\mathbf{x}, y) \in \mathcal{D}} [f(\mathbf{x}, \mathbf{w}) = y]}{|\mathcal{D}|}.$$

Качеством прогноза R_{rg} модели для задачи регрессии является среднее квадратичное отклонение результата модели от точного:

$$R_{\text{rg}} = \frac{\sum_{(\mathbf{x}, y) \in \mathcal{D}} (f(\mathbf{x}, \mathbf{w}) - y)^2}{|\mathcal{D}|}.$$

Выборка Wine. Рассмотрим нейронную сеть с 13 нейронами на входе, 13 нейронами в скрытом слое и 3 нейронами на выходе.

На рис. 2, а показано, как меняется точность прогноза R_{cl} при удалении параметров указанными методами. Из графика видно, что метод оптимального прореживания, вариационный метод и метод Белсли позволяют удалить $\approx 80\%$ параметров и качество всех этих методов падает при удалении $\approx 90\%$ параметров нейросети.

На рис. 3 представлены поверхности изменения уровня шума ответов нейросети при изменении доли удаленных параметров и уровня шума входных данных для разных методов прореживания. Из графиков видно, что при удалении параметров нейросети методом Белсли шум меньше, чем при удалении параметров другими методами. На это указывает то, что поверхность, которая соответствует методу Белсли, ниже других поверхностей.

Выборка Boston Housing. Рассмотрим нейронную сеть с 13 нейронами на входе, 39 нейронами в скрытом слое и одним нейроном на выходе.

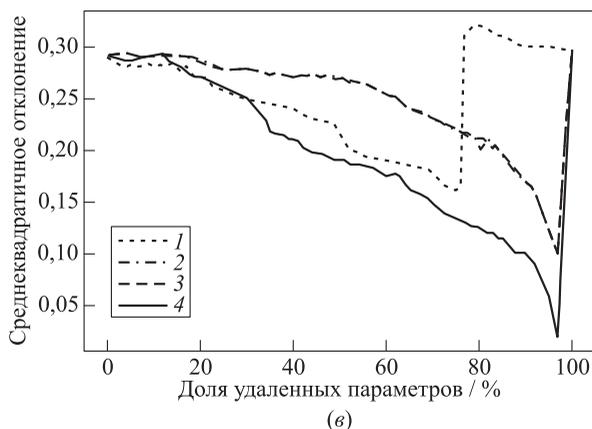
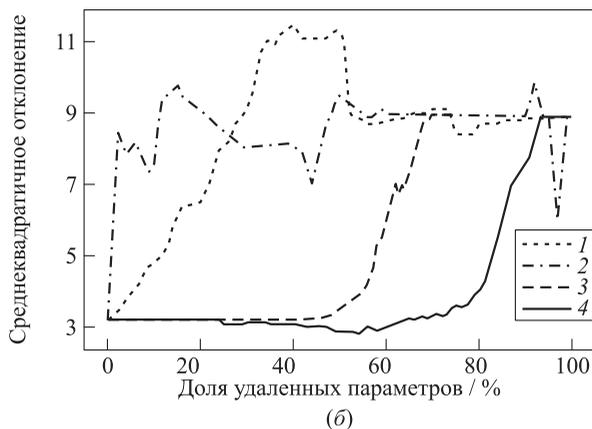
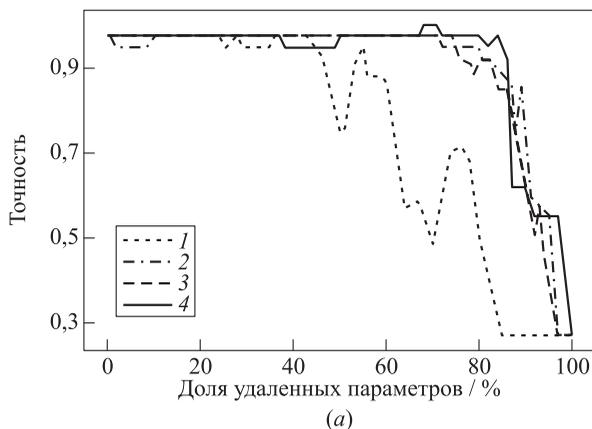
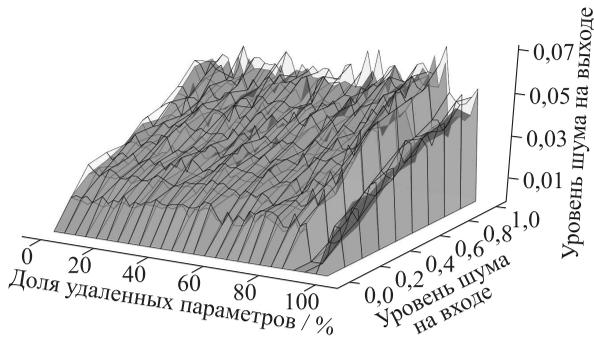
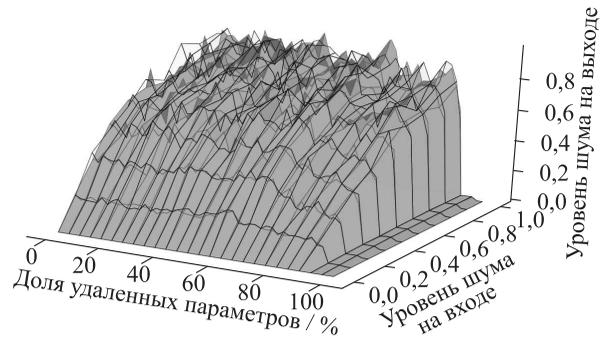


Рис. 2 Качество прогноза при удалении параметров на выборках Wine (а), Boston (б) и синтетической (в): 1 — произвольное удаление; 2 — оптимальное прореживание; 3 — вариационный метод; 4 — метод Белсли

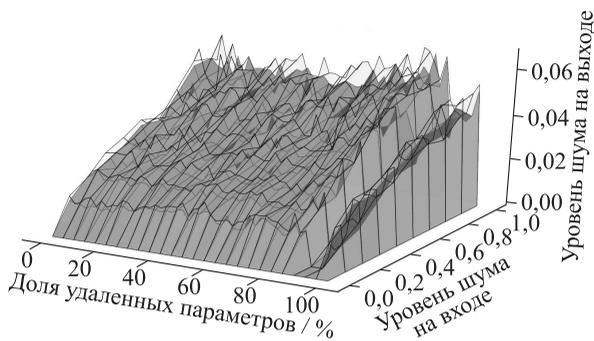
На рис. 2, б показано, как меняется среднее квадратичное отклонение прогноза R_{rg} от точного ответа при удалении параметров указанными методами. График показывает, что метод Белсли является более эффективным, чем другие методы, так как позволяет удалить больше параметров нейросети без потери качества.



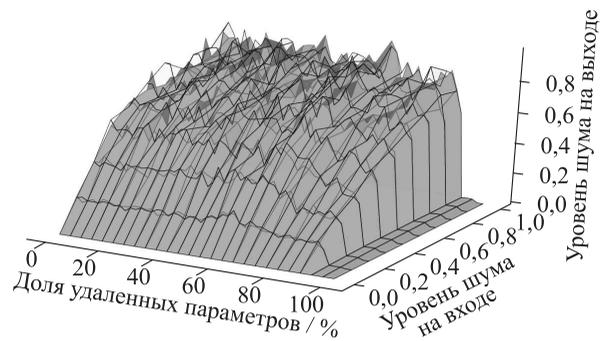
(a)



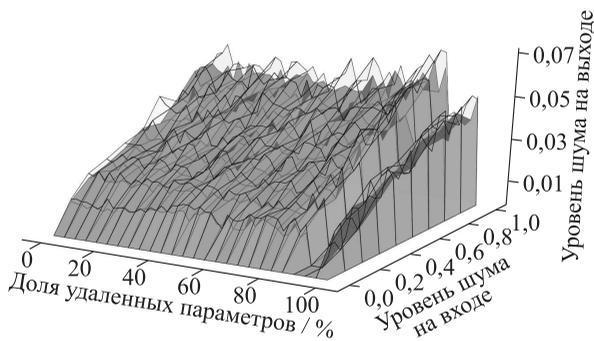
(a)



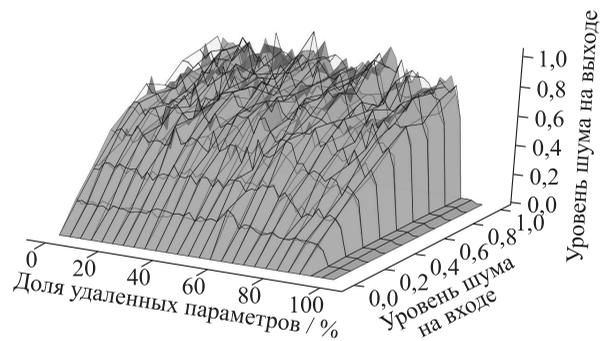
(б)



(б)



(в)



(в)

Рис. 3 Влияние шума в начальных данных на шум выхода нейросети на выборке Wine: (a) произвольное удаление параметров; (б) оптимальное прореживание; (в) вариационный метод. Серый цвет — метод Белсли

Рис. 4 Влияние шума в начальных данных на шум выхода нейросети на выборке Boston: (a) произвольное удаление параметров; (б) оптимальное прореживание; (в) вариационный метод. Серый цвет — метод Белсли

На рис. 4 представлены поверхности изменения уровня шума ответов нейросети при изменении доли удаленных параметров и уровня шума входных данных для разных методов прореживания. График показывает, что уровень шума всех методов одинаковый, так как поверхности всех методов находятся на одном уровне.

Синтетические данные. Рассмотрим нейронную сеть со 100 нейронами на входе и одним нейроном на выходе.

На рис. 2, в показано, как меняется среднеквадратичное отклонение прогноза от $R_{тг}$ точного ответа при удалении параметров указанными методами. График показывает, что удаление параметров методом Белсли является более эффективным, чем другие методы прореживания, так как качество прогноза нейросети улучшается при удалении шумовых параметров.

На рис. 5 представлены поверхности изменения уровня шума ответов нейросети при изменении

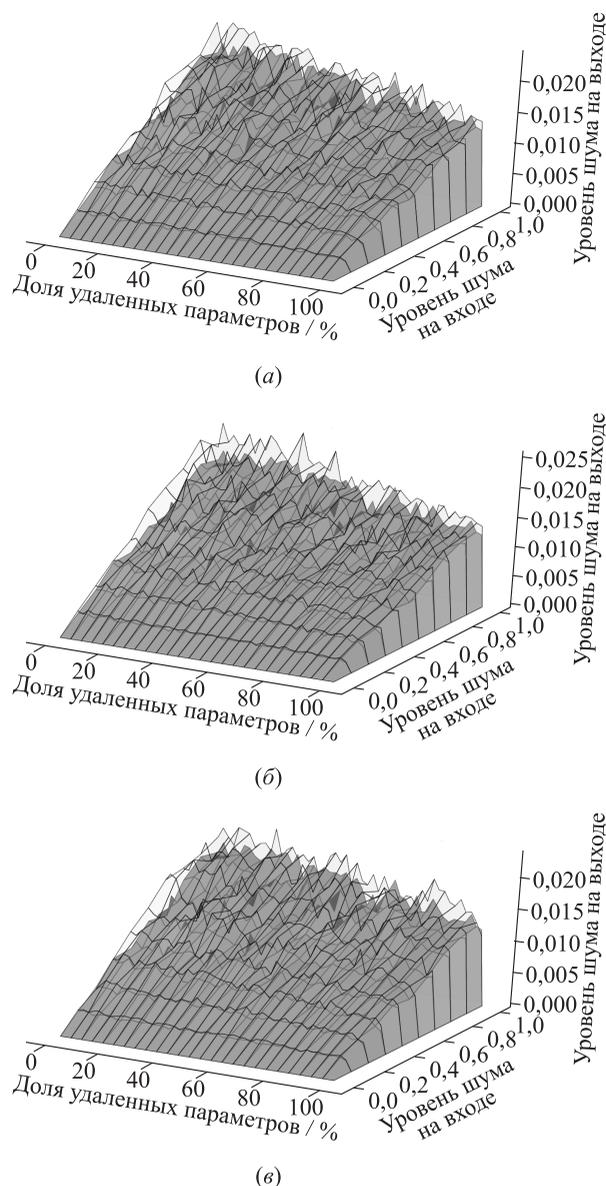


Рис. 5 Влияние шума в начальных данных на шум выхода нейросети на синтетической выборке: (а) произвольное удаление параметров; (б) оптимальное прореживание; (в) вариационный метод. Серый цвет — метод Белсли

доли удаленных параметров и уровня шума входных данных для разных методов прореживания. Из графиков видно, что при удалении параметров нейросети методом Белсли шум меньше, чем при удалении параметров другими методами, так как поверхность, которая соответствует методу Белсли, ниже других поверхностей.

6 Заключение

В работе рассматривалась задача прореживания моделей нейросетей. Рассматривались метод оптимального прореживания и метод, основанный на вариационном подходе. Был предложен алгоритм прореживания, основанный на методе Белсли, для удаления зависимых параметров модели. В ходе эксперимента было показано, что нейросети, прореженные методом Белсли, более устойчивы к шуму на входных данных. Качество прогноза нейросетей после прореживания методом Белсли не хуже качества прогноза нейросетей, прореженных другими методами.

Литература

1. *Sutskever I., Vinyals O., Le Q.* Sequence to sequence learning with neural networks // *Adv. Neur. Inf.*, 2014. Vol. 2. P. 3104–3112.
2. *Maclaurin D., Duvenaud D., Adams R.* Gradient-based hyperparameter optimization through reversible learning // *32th Conference (International) on Machine Learning Proceedings.* — Lille, 2015. Vol. 37. P. 2113–2122.
3. *Luketina J., Berglund M., Raiko T., Greff K.* Scalable gradient-based tuning of continuous regularization hyperparameters // *33th Conference (International) on Machine Learning Proceedings.* — New York, NY, USA, 2016. Vol. 48. P. 2952–2960.
4. *Molchanov D., Ashukha A., Vetrov D.* Variational dropout sparsifies deep neural networks // *34th Conference (International) on Machine Learning Proceedings.* — Sydney, 2017. Vol. 70. P. 2498–2507.
5. *Neal A., Radford M.* Bayesian learning for neural network. — Toronto, ON, Canada, 1995. Ph.D. Thesis. 195 p.
6. *LeCun Y., Denker J., Solla S.* Optimal brain damage // *Adv. Neur. Inf.*, 1989. Vol. 2. P. 598–605.
7. *Graves A.* Practical variational inference for neural networks // *Adv. Neur. Inf.*, 2011. Vol. 24. P. 2348–2356.
8. *Louizos C., Ullrich K., Welling M.* Bayesian compression for deep learning // *Adv. Neur. Inf.*, 2017. Vol. 30. P. 3288–3298.
9. *Neychev R., Katrutsa A., Strijov V.* Robust selection of multicollinear features in forecasting // *Factory Laboratory*, 2016. Vol. 82. No. 2. P. 68–74.
10. *Harrison D., Rubinfeld D.* Hedonic prices and the demand for clean air // *J. Environ. Econ. Manag.*, 1978. Vol. 5. P. 81–102. <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>.
11. *Aeberhard S.* Wine Data Set, 1991. <http://archive.ics.uci.edu/ml/datasets/Wine>.
12. *Bishop C.* Pattern recognition and machine learning. — Berlin: Springer, 2006. 758 p.

Поступила в редакцию 31.10.18

ESTIMATION OF THE RELEVANCE OF THE NEURAL NETWORK PARAMETERS

A. V. Grabovoy¹, O. Yu. Bakhteev¹, and V. V. Strijov^{1,2}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

²A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper investigates a method for optimizing the structure of a neural network. It is assumed that the number of neural network parameters can be reduced without significant loss of quality and without significant increase in the variance of the loss function. The paper proposes a method for automatic estimation of the relevance of parameters to prune a neural network. This method analyzes the covariance matrix of the posteriori distribution of the model parameters and removes the least relevant and multicorrelate parameters. It uses the Belsly method to search for multicorrelation in the neural network. The proposed method was tested on the Boston Housing data set, the Wine data set, and synthetic data.

Keywords: neural network; hyperparameters optimization; Belsly method; relevance of parameters; neural network pruning

DOI: 10.14357/19922264190209

Acknowledgments

This research was supported by the Russian Foundation for Basic Research, project 19-07-0875, and by Government of the Russian Federation, agreement 05.Y09.21.0018. This paper contains results of the project “Statistical methods of machine learning,” which is carried out within the framework of the Program “Center of Big Data Storage and Analysis” of the National Technology Initiative Competence Center supported by the Ministry of Science and Higher Education of the Russian Federation according to the agreement between M. V. Lomonosov Moscow State University and the Foundation of Project Support of the National Technology Initiative from 11.12.2018, No. 13/1251/2018.

References

1. Sutskever, I., O. Vinyals, and Q. Le. 2014. Sequence to sequence learning with neural networks. *Adv. Neur. Inf.* 2:3104–3112.
2. Maclaurin, D., D. Duvenaud, and R. Adams. 2015. Gradient-based hyperparameter optimization through reversible learning. *32th Conference (International) on Machine Learning Proceedings*. Lille. 37:2113–2122.
3. Luketina, J., M. Berglund, T. Raiko, and K. Greff. 2016. Scalable gradient-based tuning of continuous regularization hyperparameters. *33th Conference (International) on Machine Learning Proceedings*. New York, NY. 48:2952–2960.
4. Molchanov, D., A. Ashukha, and D. Vetrov. 2017. Variational dropout sparsifies deep neural networks. *34th Conference (International) on Machine Learning Proceedings*. Sydney. 70:2498–2507.
5. Neal, A., and M. Radford. 1995. Bayesian learning for neural networks. Toronto, ON: University of Toronto. Ph.D. Thesis. 195 p.
6. LeCun, Y., J. Denker, and S. Solla. 1989. Optimal brain damage. *Adv. Neur. Inf.* 2:598–605.
7. Graves, A. 2011. Practical variational inference for neural networks. *Adv. Neur. Inf.* 24:2348–2356.
8. Louizos, C., K. Ullrich, and M. Welling. 2017. Bayesian compression for deep learning. *Adv. Neur. Inf.* 30:3288–3298.
9. Neychev, R., A. Katrutsa, and V. Strijov. 2016. Robust selection of multicollinear features in forecasting. *Factory Laboratory* 82(3):68–74.
10. Harrison, D., and D. Rubinfeld. 1978. Hedonic prices and the demand for clean air. *J. Environ. Econ. Manag.* 5:81–102. Available at: <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html> (accessed June 4, 2019).
11. Aeberhard, S. 1991. Wine Data Set. Available at: <http://archive.ics.uci.edu/ml/datasets/Wine> (accessed June 4, 2019).
12. Bishop, C. 2006. *Pattern recognition and machine learning*. Berlin: Springer. 758 p.

Received October 31, 2018

Contributors

Grabovoy Andrey V. (b. 1997) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; grabovoy.av@phystech.edu

Bakhteev Oleg Yu. (b. 1991) — graduate student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; bakhteev@phystech.edu

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; strijov@ccas.ru

БАЙЕСОВСКИЕ МОДЕЛИ БАЛАНСА ФАКТОРОВ, ИМЕЮЩИХ АПРИОРНЫЕ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА И НАКАГАМИ*

Е. Н. Арутюнов¹, А. А. Кудрявцев², А. И. Титова³

Аннотация: Рассматриваются байесовские модели баланса, в рамках которых факторы, оказывающие влияние на состояние системы, условно разделяются на позитивные, т. е. способствующие функционированию, и негативные, т. е. препятствующие функционированию. В качестве показателя эффективности работы системы рассматривается отношение негативного фактора к позитивному — индекс баланса. Исследование проводится в предположении о зависимости факторов от условий внешней среды и невозможности определения точных значений факторов в каждый момент времени в силу внешних причин: несовершенства измерительного оборудования, нехватки материальных и временных ресурсов и т. п. Также предполагается, что законы изменения факторов априори известны и остаются постоянными. Данные предположения обуславливают применение байесовского метода, который заключается в рандомизации исходных параметров и, как следствие, индекса баланса, при этом предполагается, что априорные распределения факторов известны. Статья продолжает ряд исследований авторов по применению байесовских методов в задачах массового обслуживания и надежности. В работе приводятся полученные вероятностные характеристики индекса баланса факторов в случае априорных распределений Вейбулла и Накагами. Результаты представлены в терминах гамма-экспоненциальной функции.

Ключевые слова: байесовский подход; модели баланса; смешанные распределения; распределение Вейбулла; распределение Накагами; гамма-экспоненциальная функция

DOI: 10.14357/19922264190210

1 Введение

В условиях постоянного усложнения процессов, лежащих в основе большинства сфер человеческой деятельности, классические методы анализа эффективности теряют свою актуальность. Все чаще в качестве инструментов исследования различных систем используются показатели, рейтинги и индексы, что позволяет значительно экономить необходимые для проведения исследований ресурсы. При построении математических моделей в задачах исследования эффективности естественно разделять факторы, влияющие на систему, на способствующие функционированию целевого объекта (р-факторы) и препятствующие функционированию (п-факторы). В данных условиях для исследования системы наиболее наглядным показателем эффективности ее работы представляется отношение п-фактора к р-фактору, которое будет стремиться к единице при приближении системы к состоянию баланса и существенно отличаться от единицы в случае непродуктивной деятельности.

Назовем индексом баланса системы отношение $\rho = \lambda/\mu$, где λ и μ суть п- и р-факторы соответственно. В случае исследования сложных модифицируемых информационных систем примером индекса баланса может служить коэффициент загрузки системы, определяемый как отношение интенсивности входящего потока к интенсивности обслуживания, или ожидаемое время безотказной работы, представимое в виде отношения среднего времени безотказной работы к среднему времени восстановления системы.

В силу стохастичности среды функционирования любой современной системы значения факторов, влияющих на систему, меняются с течением времени, что создает предпосылки для рассмотрения факторов и индексов, зависящих от них, как случайных величин. Вместе с тем законы, которым подчиняются изменения факторов, можно считать неизменными в рамках конкретной модели, так как глобальные изменения окружающей среды происходят довольно редко. Эти предположения обосновывают применимость байесовского

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00577).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, enaroleon@mail.ru

² Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, nubigena@mail.ru

³ Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, onkelskroot@gmail.com

метода, в рамках которого происходит рандомизация параметров при помощи известных априорных распределений [1].

В статье в качестве априорных распределений факторов рассматриваются распределение Вейбулла и m -распределение Накагами. Ниже приводятся вероятностные характеристики индекса баланса, представляющего собой масштабную смесь двух распределений гамма-класса.

2 Основные результаты

Обозначим через $m(q, \theta)$ m -распределение Накагами [2] с плотностью

$$m_{q,\theta}(x) = \frac{2q^q x^{2q-1}}{\theta^q \Gamma(q)} \exp\left\{-\frac{qx^2}{\theta}\right\},$$

$$x > 0, q > 0, \theta > 0,$$

а через $W(p, \alpha)$ — распределение Вейбулла с плотностью

$$w_{p,\alpha}(x) = \frac{px^{p-1} e^{-(x/\alpha)^p}}{\alpha^p}, \quad x > 0, p > 0, \alpha > 0.$$

Рассмотрим гамма-экспоненциальную функцию [3]

$$Ge_{\alpha,\beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta),$$

$$x \in \mathbb{R}, 0 \leq \alpha < 1, \beta > 0.$$

Хорошо известен следующий результат.

Лемма 1. Для случайных величин ξ и η , имеющих m -распределение Накагами $m(q, \theta)$ и распределение Вейбулла $W(p, \alpha)$ соответственно, для $z \in \mathbb{R}$ выполняются соотношения:

$$E\xi^z = \left(\frac{\theta}{q}\right)^z \frac{\Gamma(z/2 + q)}{\Gamma(q)}, \quad z > -2q;$$

$$E\eta^z = \alpha^z \Gamma\left(1 + \frac{z}{p}\right), \quad z > -p.$$

Для дальнейших вычислений потребуется следующее утверждение.

Лемма 2. Для некоторых $q > 0, p > 0, a > 0$ и $b > 0$ справедливо

$$\int_0^{\infty} y^{2q+p-1} e^{-(y/a)^p - (y/b)^2} dy =$$

$$= \begin{cases} \frac{b^{2q+p}}{2} Ge_{p/2, (2q+p)/2} \left(-\left(\frac{b}{a}\right)^p\right), & p < 2; \\ \frac{a^{2q+p}}{p} Ge_{2/p, (2q+p)/p} \left(-\left(\frac{a}{b}\right)^2\right), & p > 2; \\ \frac{q\Gamma(q)}{2(a^{-2} + b^{-2})^{q+1}}, & p = 2. \end{cases}$$

Доказательство. Рассмотрим случай $p < 2$. Используя теорему Лебега о предельном переходе, получаем:

$$\int_0^{\infty} y^{2q+p-1} e^{-(y/a)^p - (y/b)^2} dy =$$

$$= \frac{a^{2q+p}}{p} \int_0^{\infty} t^{2q/p} \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} e^{-(a/b)^2 t^{2/p}} dt =$$

$$= \frac{b^{2q+p}}{2} \sum_{k=0}^{\infty} \frac{(-1)^k (b/a)^{pk}}{k!} \int_0^{\infty} z^{(pk+p+2q)/2-1} e^{-z} dz =$$

$$= \frac{b^{2q+p}}{2} \sum_{k=0}^{\infty} \frac{(-1)^k (b/a)^{pk}}{k!} \Gamma\left(\frac{pk+p+2q}{2}\right).$$

Случай $p > 2$ рассматривается аналогично. Случай $p = 2$ напрямую следует из определения гамма-функции. Лемма доказана.

Леммы 1 и 2 дают возможность находить вероятностные характеристики индекса баланса системы.

Теорема 1. Пусть негативный фактор λ имеет m -распределение Накагами $m(q, \theta)$, а позитивный фактор μ имеет распределение Вейбулла $W(p, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты индекса баланса ρ имеют вид:

$$f_{\rho}(x) =$$

$$= \begin{cases} \frac{2q^q \alpha^{2q} x^{2q-1}}{\theta^q \Gamma(q)} Ge_{2/p, 2q/p+1} \left(-\frac{q\alpha^2 x^2}{\theta}\right), & p > 2; \\ \frac{p\theta^{p/2} x^{-p-1}}{\alpha^p q^{p/2} \Gamma(q)} Ge_{p/2, q+p/2} \left(-\frac{\theta^{p/2}}{q^{p/2} \alpha^p x^p}\right), & p < 2; \end{cases} \quad (1)$$

$$F_{\rho}(x) = \begin{cases} \frac{2q^q \alpha^{2q} x^{2q}}{p\theta^q \Gamma(q)} Ge_{2/p, 2q/p} \left(-\frac{q\alpha^2 x^2}{\theta}\right), & p > 2; \\ \frac{1}{\Gamma(q)} \int_{(\sqrt{q/\theta} \alpha x)^{-p}}^{\infty} Ge_{p/2, q+p/2}(-z) dz, & p < 2; \end{cases}$$

$$E\rho^z = \frac{\theta^{z/2} \Gamma(q + z/2) \Gamma(1 - z/p)}{q^{z/2} \alpha^z \Gamma(q)}, \quad z < p.$$

При $p = 2$ индекс баланса ρ имеет распределение Дагума [4] с параметрами $(2, \sqrt{\theta/q}/\alpha, q)$.

Доказательство. Поскольку

$$f_{\rho}(x) = \int_0^{\infty} \frac{2q^q x^{2q-1} y^{2q+p-1}}{\alpha^p \theta^q \Gamma(q)} e^{-qx^2 y^2 / \theta - (y/\alpha)^p} dy,$$

соотношение (1) следует из леммы 2.

Для функции распределения ρ при $p > 2$ справедливо для $x > 0$:

$$\begin{aligned} F_{\rho}(x) &= \int_0^x \frac{2q^q \alpha^{2q} u^{2q-1}}{\theta^q \Gamma(q)} \text{Ge}_{2/p, 2q/p+1} \left(-\frac{q\alpha^2 u^2}{\theta} \right) du = \\ &= \frac{2q^q \alpha^{2q}}{\theta^q \Gamma(q)} \sum_{k=0}^{\infty} \frac{(-q\alpha^2)^k}{\theta^k k!} \Gamma \left(\frac{2k + 2q + p}{p} \right) \times \\ &\quad \times \int_0^x u^{2q+2k-1} du = \\ &= \frac{2q^q (\alpha x)^{2q}}{p \theta^q \Gamma(q)} \sum_{k=0}^{\infty} \frac{(-q\alpha^2 x^2)^k}{\theta^k k!} \Gamma \left(\frac{2k + 2q}{p} \right). \end{aligned}$$

В случае $p < 2$ имеем:

$$\begin{aligned} F_{\rho}(x) &= \int_0^x \frac{p\theta^{p/2} u^{-p-1}}{\alpha^p q^{p/2} \Gamma(q)} \text{Ge}_{p/2, q+p/2} \left(-\frac{\theta^{p/2}}{q^{p/2} \alpha^p u^p} \right) du = \\ &= \int_0^x \sum_{k=0}^{\infty} \frac{p\theta^{p/2} \left(-(\sqrt{q/\theta} \alpha u)^{-p} \right)^k}{\alpha^p q^{p/2} u^{p+1} \Gamma(q) k!} \times \\ &\quad \times \Gamma \left(\frac{pk + 2q + p}{2} \right) du = \\ &= \frac{1}{\Gamma(q)} \int_{(\sqrt{q/\theta} \alpha x)^{-p}}^{\infty} \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \Gamma \left(\frac{pk + 2q + p}{2} \right) dz. \end{aligned}$$

Для нахождения моментов ρ достаточно воспользоваться независимостью случайных величин λ и μ и леммой 1. Теорема доказана.

Рассмотрим симметричный случай априорных распределений факторов.

Теорема 2. Пусть негативный фактор λ имеет распределение Вейбулла $W(q, \theta)$, а позитивный фактор μ имеет t -распределение Накагами $m(p, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты индекса баланса ρ имеют вид:

$$\begin{aligned} f_{\rho}(x) &= \begin{cases} \frac{qx^{q-1}(\alpha/p)^{q/2}}{\theta^q \Gamma(p)} \text{Ge}_{q/2, q/2+p} \left(-\frac{\alpha^{q/2} x^q}{p^{q/2} \theta^q} \right), & q < 2; \\ \frac{2p^p \theta^{2p}}{\alpha^p x^{2p+1} \Gamma(p)} \text{Ge}_{2/q, 2p/q+1} \left(-\frac{p\theta^2}{\alpha x^2} \right), & q > 2; \end{cases} \\ F_{\rho}(x) &= \begin{cases} 1 - \frac{1}{\Gamma(p)} \text{Ge}_{q/2, p} \left(-\frac{\alpha^{q/2} x^q}{p^{q/2} \theta^q} \right), & q < 2; \\ \frac{2}{\Gamma(p)} \int_{\theta \sqrt{p/\alpha}/x}^{\infty} z^{2p-1} \text{Ge}_{q/2, q/2+p}(-z^2) dz, & q > 2; \end{cases} \\ E\rho^z &= \frac{p^{z/2} \theta^z \Gamma(1+z/q) \Gamma(p-z/2)}{\alpha^{z/2} \Gamma(p)}, \quad z < 2p. \end{aligned}$$

При $q = 2$ распределение индекса баланса ρ совпадает с распределением Бурра [5] с параметрами $(2, p, \theta \sqrt{p/\alpha})$.

Доказательство. Аналогично предыдущей теореме для получения выражения для плотности ρ при всех $q > 0$ достаточно воспользоваться леммой 2.

Найдем функцию распределения в случае $q < 2$. Имеем:

$$\begin{aligned} F_{\rho}(x) &= \frac{q(\alpha/p)^{q/2}}{\theta^q \Gamma(p)} \times \\ &\times \int_0^x u^{q-1} \text{Ge}_{q/2, q/2+p} \left(-\frac{\alpha^{q/2} u^q}{p^{q/2} \theta^q} \right) du = \frac{q(\alpha/p)^{q/2}}{\theta^q \Gamma(p)} \times \\ &\times \int_0^x u^{q-1} \sum_{k=0}^{\infty} \frac{(-\alpha^{q/2} u^q)^k}{p^{qk/2} \theta^{qk} k!} \Gamma \left(\frac{qk + q + 2p}{2} \right) du = \\ &= -\frac{1}{\Gamma(p)} \sum_{k=0}^{\infty} \frac{(-\alpha^{q/2} x^q)^{k+1}}{(p^{q/2} \theta^q)^{k+1} (k+1)!} \times \\ &\quad \times \Gamma \left(\frac{q(k+1) + 2p}{2} \right) = \\ &= 1 - \frac{1}{\Gamma(p)} \sum_{m=0}^{\infty} \frac{(-\alpha^{q/2} x^q)^m}{(p^{q/2} \theta^q)^m m!} \Gamma \left(\frac{qm + 2p}{2} \right). \end{aligned}$$

Для $q > 2$ функция распределения ρ имеет вид:

$$\begin{aligned} F_{\rho}(x) &= \int_0^x \frac{2p^p \theta^{2p}}{\alpha^p u^{2p+1} \Gamma(p)} \text{Ge}_{2/q, 2p/q+1} \left(-\frac{p\theta^2}{\alpha u^2} \right) du = \end{aligned}$$

$$\begin{aligned}
&= \int_0^x \frac{2p^p \theta^{2p}}{\alpha^p u^{2p+1} \Gamma(p)} \sum_{k=0}^{\infty} \frac{(-p\theta^2)^k}{(\alpha u^2)^k k!} \Gamma\left(\frac{2k+2p+q}{q}\right) du = \\
&= \frac{2}{p\Gamma(p)} \int_{\theta\sqrt{p/\alpha/x}}^{\infty} z^{2p-1} \sum_{k=0}^{\infty} \frac{(-z^2)^k}{k!} \Gamma\left(\frac{2k+2p+q}{q}\right) dz.
\end{aligned}$$

Соотношение для моментов следует из леммы 1 и независимости случайных величин λ и μ . Теорема доказана.

Следствие. Утверждение теоремы 1 справедливо для частных случаев распределения Накагами, таких как распределение максимума процесса броуновского движения ($q = 1/2$, $\theta = \theta$), распределение Рэлея ($q = 1$, $\theta = \sqrt{\theta}$), хи-распределение ($q = q/2$, $\theta = q$), распределение Максвелла–Больцмана ($q = 3/2$, $\theta = 3\theta$). Утверждение теоремы 2 также справедливо для перечисленных распределений с соответствующей заменой параметров.

3 Заключение

Распределения Вейбулла и Накагами вместе с их частными случаями служат адекватными мате-

матическими моделями многих реальных процессов и явлений из всевозможных областей знания (радиотехники, физики, экономики, управления и пр.). По этой причине полученные результаты могут найти широкое применение при анализе эффективности разнообразных реальных систем.

Литература

1. *Кудрявцев А. А.* Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
2. *Nakagami M.* The m-distribution, a general formula of intensity of rapid fading // Statistical Methods in Radio Wave Propagation: Symposium Proceedings / Ed. W. C. Homan. — New York, NY, USA: Pergamon Press, 1960. P. 3–36.
3. *Кудрявцев А. А., Титова А. И.* Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
4. *Dagum C.* A new model of personal income-distribution-specification and estimation // Econ. Appl., 1977. Vol. 30. No. 3. P. 413–437.
5. *Burr I. W.* Cumulative frequency functions // Ann. Math. Stat., 1942. Vol. 13. No. 2. P. 215–232.

Поступила в редакцию 17.03.19

BAYESIAN MODELS OF FACTORS BALANCE WITH *A PRIORI* WEIBULL AND NAKAGAMI DISTRIBUTIONS

E. N. Arutyunov¹, A. A. Kudryavtsev², and A. I. Titova²

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: Bayesian balance models are considered. Within this approach, it is assumed that the parameters affecting a system can be divided into positive, which support system functioning, and negative, which interfere with the functioning. Thus, the ratio of negative to positive factors — balance index — is considered as an indication of system’s functioning effectiveness. The study is carried out assuming that the factors depend on the environment state and their exact value cannot be obtained due to external reasons, e. g., equipment faults, lack of resources, etc. It is also assumed that the principles of factors’ changes are known *a priori* and remain invariable. Considering these assumptions, it is natural to use the Bayesian method, which implies randomization of the initial parameters supposing that their *a priori* distributions are known. As a result, the balance index becomes a random variable as well. This paper continues a series of studies by the authors devoted to the application of Bayesian methods in the problems of queuing and reliability. In this work, the obtained probability characteristics of the factor balance index in the case of *a priori* Weibull and Nakagami distributions are presented. The results are formulated using gamma-exponential function.

Keywords: Bayesian approach; balance models; mixed distributions; Weibull distribution; Nakagami distribution; gamma-exponential function

DOI: 10.14357/19922264190210

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

References

1. Kudryavtsev, A.A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
2. Nakagami, M. 1960. The m-distribution, a general formula of intensity of rapid fading. *Statistical Methods in Radio Wave Propagation: Symposium Proceedings*. Ed. W.C. Hoffman. New York, NY: Pergamon Press. 3–36.
3. Kudryavtsev, A.A., and A.I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
4. Dagum, C. 1977. A new model of personal income-distribution-specification and estimation. *Econ. Appl.* 30(3):413–437.
5. Burr, I.W. 1942. Cumulative frequency functions. *Ann. Math. Stat.* 13(2):215–232.

Received March 17, 2019

Contributors

Arutyunov Evgeny N. (b. 1952) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; enapoleon@mail.ru

Kudryavtsev Alexey A. (b. 1978) — Candidate of Sciences (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

Titova Anastasiia I. (b. 1995) — student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; onkelskroot@gmail.com

ПРОТОКОЛ ГЕТЕРОГЕННОГО МЫШЛЕНИЯ ГИБРИДНОЙ ИНТЕЛЛЕКТУАЛЬНОЙ МНОГОАГЕНТНОЙ СИСТЕМЫ ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ ВОССТАНОВЛЕНИЯ РАСПРЕДЕЛИТЕЛЬНОЙ ЭЛЕКТРОСЕТИ*

А. В. Колесников¹, С. В. Листопад²

Аннотация: Для проблемы восстановления электроснабжения в региональной распределительной электросети после масштабных аварий характерны высокая комбинаторная сложность, неоднородность, недоопределенность, неточность и нечеткость. Применение механизма коллективного решения проблем для преодоления перечисленных НЕ-факторов невозможно из-за временных ограничений. Для решения проблемы предлагается новый класс интеллектуальных систем, моделирующих коллективное принятие решений под руководством фасилитатора — гибридные интеллектуальные многоагентные системы гетерогенного мышления (ГИМСГМ). В отличие от традиционных гибридных интеллектуальных систем, интегрирующих модели знаний экспертов, они дополнительно моделируют взаимодействие в групповых процессах и эффектах, адаптируясь к динамичным проблемным ситуациям послеаварийного восстановления региональной распределительной электросети. В работе рассматривается один из компонентов таких систем — протокол организации коллективного гетерогенного мышления агентов.

Ключевые слова: гетерогенное мышление; гибридная интеллектуальная многоагентная система; проблема восстановления распределительной электросети

DOI: 10.14357/19922264190211

1 Введение

При авариях в энергосистемах, происходящих несмотря на проводимые предупредительно-профилактические мероприятия, принимаются меры, снижающие их интенсивность и продолжительность. Сразу после масштабной аварии повышение частоты и развал энергосистемы тормозятся автоматически изменением и отключением нагрузки управляемым разделением. В течение последующих часов выполняются восстановительные мероприятия, в ходе которых операторы распределительной сети и электростанций вручную поддерживают баланс нагрузки и генерации. Длительность выработки решений и действий по их реализации намного превышает допустимые ограничения по оборудованию, и возникает опасность, что управление электростанциями и энергосистемой не может обеспечить необходимую координацию [1].

Одна из важнейших задач оперативно-диспетчерского персонала — составление плана обеспечения безопасности, сохранности оборудования и быстрого восстановления электроснабжения потребителей, не противоречащего требованиям энергосистемы [2].

При разработке плана оперируют большим разнообразием ресурсов, их свойств и отношений, а также учитывают НЕ-факторы А. С. Нариньяни [3]: недоопределенность места аварии на момент планирования; неточность величины мощности, потребляемой клиентами и генерируемой источниками распределенной генерации; нечеткость времени восстановительных операций; некорректность работы датчиков; неполнота модели электросети. В силу временных ограничений организовать всестороннее коллективное обсуждение послеаварийной проблемной ситуации невозможно.

Для информационной подготовки решений («предрешений», по П. К. Анохину) по восстановлению электроснабжения в распределительной электросети [4] предлагается разработать новый класс интеллектуальных систем — ГИМСГМ, комбинирующих гибридный подход А. В. Колесникова [5], аппарат многоагентных систем В. Б. Тарасова [6] и методики гетерогенного мышления [7–9]. Такие системы интегрируют знания и взаимодействие экспертов различных специальностей, учитывают несколько критериев оптимальности и множество ограничений в условиях динамиче-

* Работа выполнена при поддержке РФФИ (проект 18-07-00448а).

¹ Балтийский федеральный университет им. И. Канта, avkolesnikov@yandex.ru

² Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

ских сред и дефицита времени на принятие решения.

2 Формализованная модель гибридной интеллектуальной многоагентной системы гетерогенного мышления

Формально ГИМСГМ определяется следующим образом [10]:

$$\begin{aligned} \text{himsht} &= \langle AG^*, \text{env}, \text{INT}^*, \text{ORG}, \{\text{ht}\} \rangle; \\ \text{act}_{\text{himsht}} &= \\ &= \left(\bigcup_{ag \in AG^*} \text{act}_{ag} \right) \cup \text{act}_{\text{dmsa}} \cup \text{act}_{\text{htmc}} \cup \text{act}_{\text{col}}; \\ \text{act}_{ag} &= (\text{MET}_{ag}, \text{IT}_{ag}), ag \in AG^*, \left| \bigcup_{ag \in AG^*} \text{IT}_{ag} \right| \geq 2, \end{aligned}$$

где $AG^* = \{ag_1, \dots, ag_n, ag^{dm}, ag^{fc}\}$ — множество агентов, включающее агентов-экспертов (АЭ) $ag_i, i \in \mathbb{N}, 1 \leq i \leq n$, агента, принимающего решения (АПР), ag^{dm} и агента-фасилитатора (АФ) ag^{fc} , n — число АЭ; env — концептуальная модель внешней среды ГИМСГМ; $\text{INT}^* = \{\text{prot}_{gm}, \text{lang}, \text{ont}, \text{dmscl}\}$ — элементы структурирования взаимодействий агентов: prot_{gm} — протокол взаимодействия агентов, позволяющий организовать их коллективное гетерогенное мышление, lang — язык передачи сообщений, ont — модель предметной области, dmscl — классификатор ситуаций коллективного решения проблемы, идентифицирующий стадии этого процесса; ORG — множество архитектур ГИМСГМ; $\{\text{ht}\}$ — множество концептуальных моделей макроуровневых процессов в ГИМСГМ: ht — модель процесса коллективного решения проблем

методами гетерогенного мышления — ромб группового принятия решений С. Кейнера, К. Толди, С. Фиск, Д. Бергера (рис. 1) [9]; $\text{act}_{\text{himsht}}$ — функция ГИМСГМ в целом; act_{ag} — функция АЭ из множества AG^* ; act_{dmsa} — функция «анализ ситуации коллективного решения проблемы» АФ, обеспечивающая идентификацию стадии процесса гетерогенного мышления ГИМСГМ на основе предлагаемых АЭ частных решений, напряженности конфликта между АЭ и предшествующей стадии процесса решения проблемы; act_{htmc} — функция «выбор метода гетерогенного мышления» АФ, которая реализуется с использованием нечеткой базы знаний об эффективности методов гетерогенного мышления в зависимости от характеристик проблемы, стадии процесса ее решения и текущей ситуации решения в ГИМСГМ; act_{col} — коллективная функция ГИМСГМ, конструируемая динамически; met_{ag} — метод решения задачи; it_{ag} — интеллектуальная технология, в рамках которой реализован метод met_{ag} .

Согласно модели, представленной на рис. 1, процесс решения проблемы ГИМСГМ трехстадийный: (1) дивергентное мышление; (2) бурление; (3) конвергентное мышление. На стадии дивергентного мышления АЭ генерируют множество вариантов решения проблемы, а АФ стимулирует их выработку соответствующими методами [11]. В случае если даже с применением методов дивергентного мышления противоречий не возникает, т.е. задача имеет очевидное решение, процесс завершается. В противном случае агенты ГИМСГМ конфликтуют по поводу знаний, убеждений, мнений, т.е. участвуют в своего рода когнитивных конфликтах [11, 12]. Конфликт — отличительная черта стадии бурления, позволяющая АФ предпринять меры по сближению точек зрения агентов. На стадии конвергентного мышления агенты совместно переформулируют, дорабатывают решения, пока не

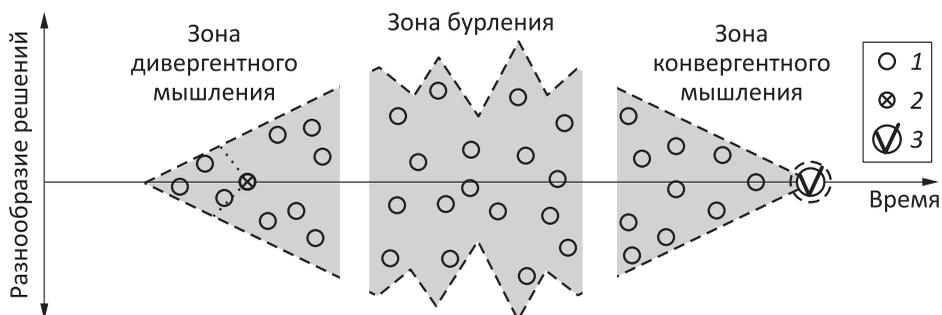


Рис. 1 Модель коллективного решения проблем методами гетерогенного мышления — ромб группового принятия решений С. Кейнера, К. Толди, С. Фиск, Д. Бергера: 1 — альтернатива; 2 — досрочное несогласованное решение; 3 — согласованное решение

получат коллективное решение, релевантное разнообразию моделей экспертов ГИМСГМ.

Функциональная структура ГИМСГМ для решения проблемы восстановления электроснабжения в региональной распределительной электросети после масштабных аварий представлена в [12]. Рассмотрим протокол взаимодействия агентов ГИМСГМ в процессе коллективного гетерогенного мышления.

3 Протокол гетерогенного мышления гибридной интеллектуальной многоагентной системы

Основная цель протоколирования агентов гибридной интеллектуальной многоагентной системы — инкапсуляция разрешенных взаимодействий. Протокол определяет схемы (распределенный алгоритм) обмена информацией, знаниями, координации агентов при решении поставленных задач [13]. С одной стороны, протокол служит для объединения агентов через концептуальный интерфейс и организации их совместной работы, с другой — определяет четкие границы компонентов системы [14]. При описании протокола должны быть специфицированы: роли агентов; типы сообщений между парами ролей; семантика каждого типа сообщения декларативно; любые дополнительные ограничения на сообщения, такие как порядок их следования или правила передачи информации из одного сообщения в другое. Такая спецификация протокола однозначно определяет, удовлетворяет ли конкретная реализация взаимодействия агентов указанному протоколу и совместим ли конкретный агент с ГИМСГМ [14].

К настоящему времени разработано множество протоколов многоагентных систем как общего назначения, так и для решения конкретных задач. Выделим наиболее известные классы протоколов [13, 15]:

- (1) на основе контрактной сети [16, 17] для автоматического планирования взаимодействия агентов и минимизации затрат посредством метафоры переговоров агентов на рыночных торгах;
- (2) на основе теории речевых актов [18], когда переговоры строятся с использованием небольшого числа примитивов, посредством обмена которыми агенты обсуждают некоторую тему, обновляют свои базы знаний, обмениваются

«мнениями» и приходят к общему решению [13];

- (3) переговорные [19, 20], предлагающие механизмы разрешения конфликтов для повышения суммарной полезности, достигаемой агентами;
- (4) на основе доски объявлений, когда выделяется общая область памяти для взаимодействия агентов.

Предлагаемый протокол организации коллективного гетерогенного мышления агентов предназначен для моделирования работы малого коллектива экспертов за круглым столом, где нет необходимости сбивать цену за услуги или подбирать экспертов, поэтому он основывается на теории речевых актов. Схема работы гибридной интеллектуальной многоагентной системы по протоколу гетерогенного мышления представлена на рис. 2.

Как показано на рис. 2, стандартный протокол речевых актов [18] расширен следующими типами сообщений: `request-ch-tt`, `commit-ch-tt`, `request-start-ps`, `request-stop-ps`, `request-task`, `report-decision`, используемых для организации взаимодействия АФ, АПР и АЭ. Взаимодействуют АЭ между собой и с другими агентами по протоколу речевых актов [18].

По предлагаемому протоколу процесс коллективного гетерогенного мышления начинается с отправки агентом-фасилитатором АПР и АЭ сообщения типа `request-ch-tt`, в теле которого указывается метод гетерогенного мышления, применяемый на данном этапе работы ГИМСГМ. Агент-фасилитатор приостанавливает свою работу в ожидании ответов-подтверждений от АЭ и АПР. Получив от АФ сообщение `request-ch-tt`, АПР и АЭ выбирают соответствующий алгоритм и переходят в режим ожидания сигнала на начало решения проблемы в соответствии с данным алгоритмом. Для подтверждения получения сообщения и готовности к работе по установленному алгоритму АПР и АЭ отправляют АФ ответ-подтверждение `commit-ch-tt`. Дождавшись подтверждений от АПР и всех АЭ, АФ отправляет АПР сообщение `request-start-ps`, сигнализирующее о том, что все агенты перешли на соответствующий метод гетерогенного мышления, а система готова к дальнейшей работе, и ожидает решений от АЭ.

Получив сообщение `request-start-ps`, АПР формирует и рассылает задачи АЭ с использованием сообщения типа `request-task`, в теле которого также описываются исходные данные задачи, после чего он приступает к сбору решений, поступающих от АЭ, и работает с ними в соответствии с установленным ранее алгоритмом гетерогенного мышления.

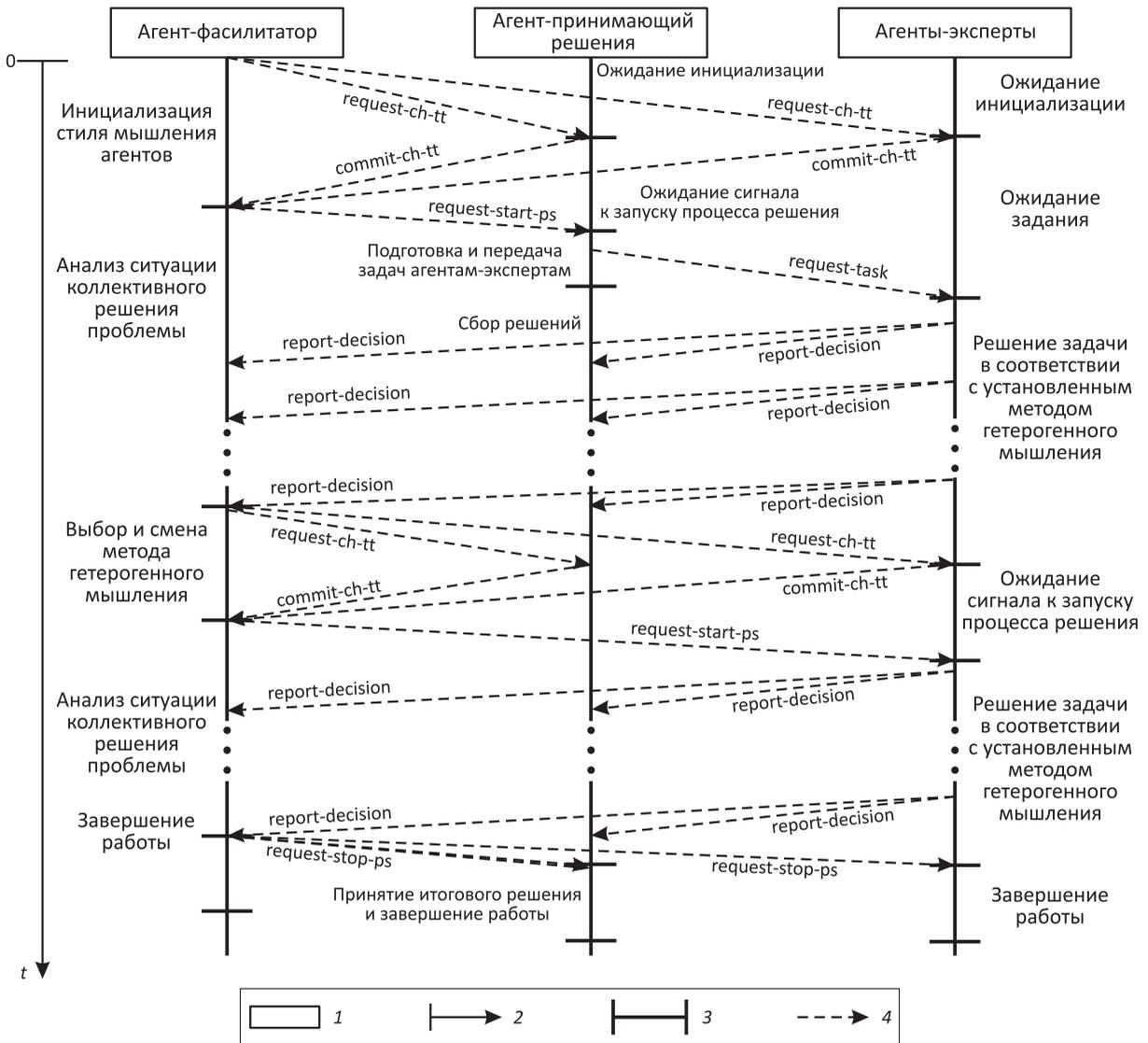


Рис. 2 Схема работы гибридной интеллектуальной многоагентной системы по протоколу гетерогенного мышления: 1 — участники коллектива; 2 — шкала времени; 3 — действие; 4 — процесс передачи информации между участниками коллектива; *t* — модельное время

Агенты-эксперты после получения задания от АПР начинают решать поставленные задачи в соответствии с установленным алгоритмом гетерогенного мышления. При этом в зависимости от алгоритма они могут генерировать несколько решений поставленной задачи. Все полученные альтернативы АЭ отправляют в теле сообщения типа report-decision одновременно АПР и АФ.

При получении очередного решения от АЭ агент-фасилитатор запускает функцию «анализ ситуации коллективного решения проблемы» act_{dmsa} , определяет с ее помощью напряженность конфликта в каждой паре АЭ и в ГИМСГМ в целом [4]. При достижении определенного уровня напряженности

конфликта в соответствии с нечеткой базой знаний АФ запускает функцию «выбор метода гетерогенного мышления» act_{htmc} , позволяющую ему выбрать релевантный ситуации коллективного принятия решений метод гетерогенного мышления.

Для реализации данной функции АФ имеет нечеткую базу знаний о релевантности «стилей мышления» агентов различным ситуациям принятия решений в ГИМСГМ, а методов — различным особенностям проблем и стадиям коллективного принятия решений. Для формирования такой базы знаний необходимо провести серию вычислительных экспериментов по решению проблем из различных классов [5] и установить соответствие

между классом проблем и релевантными подходами к организации гетерогенного мышления. Выбрав релевантный метод, АФ отправляет АЭ и АПР сообщения типа request-ch-tt, в теле которых указан метод гетерогенного мышления для АЭ и АПР. Дождется подтверждений от АПР и всех АЭ, после чего отправляет АЭ сообщение request-start-ps, сигнализирующее о том, что все агенты перешли на соответствующий метод гетерогенного мышления и АЭ могут продолжить работу. Агент-фасилитатор начинает прием решений от АЭ, и процесс анализа ситуации — выбора метода гетерогенного мышления — повторяется до тех пор, пока в ГИМСГМ не завершится стадия конвергентного мышления (см. рис. 1). После ее завершения АФ отправляет сигнал request-stop-ps к окончанию процесса решения проблемы АПР и АЭ и завершает свою работу. Получив такой сигнал, АЭ прерывают выполнение заданий и завершают работу, а АПР принимает окончательное коллективное решение в соответствии с установленным алгоритмом гетерогенного мышления, например решение, по которому был достигнут консенсус или отдано большинство голосов агентов на стадии конвергентного мышления. Далее передает это решение интерфейсному агенту и также завершает работу.

Число «переключений» методов мышления протоколом заранее не определено, так как стадия бурления может отсутствовать, а на стадиях дивергентного и конвергентного мышления последовательно могут применяться различные методы. Таким образом, за счет наличия у АФ нечеткой базы знаний, а также путем репрезентации неоднородной функциональной структуры сложной задачи и гетерогенного коллективного мышления интеллектуальных агентов, взаимодействующих в соответствии с предложенным протоколом, ГИМСГМ вырабатывает для каждой проблемы релевантный ей метод решения без упрощения и идеализации в условиях динамической среды.

4 Заключение

Рассмотрены особенности проблемы восстановления электроснабжения в региональной распределительной электросети после масштабных аварий и предложен новый класс интеллектуальных систем для ее решения — ГИМСГМ.

Представлено формализованное описание ГИМСГМ, ее основных составных частей.

Описан протокол организации коллективного гетерогенного мышления агентов на основе теории речевых актов. Его применение в гибридных интеллектуальных многоагентных системах, содер-

жащих разнородные интеллектуальные самоорганизующиеся агенты, позволяет релевантно моделировать эффективные практики коллективного решения проблем.

Применение ГИМСГМ позволит оперативно-диспетчерскому персоналу электроснабжающих организаций принимать релевантные решения по восстановлению электросети в условиях дефицита времени.

Литература

1. Adibi M. M., Fink L. H. Overcoming restoration challenges associated with major power system disturbances // IEEE Power Energy M., 2006. Vol. 4. Iss. 5. P. 68–77.
2. Красник В. В. Потребители электрической энергии, энергоснабжающие организации и органы Ростехнадзора. Правовые основы взаимоотношений. — М.: НЦ ЭНАС, 2005. 250 с.
3. Нариньяни А. С. Инженерия знаний и НЕ-факторы: краткий обзор-08 // Вопросы искусственного интеллекта, 2008. № 1. С. 61–77.
4. Kolesnikov A. V., Listopad S. V. Hybrid intelligent multi-agent system of heterogeneous thinking for solving the problem of restoring the distribution power grid after failures // Open Semantic Technologies for Intelligent Systems: Research Papers Collection. — Minsk: BGUIR, 2019. P. 133–138.
5. Колесников А. В. Гибридные интеллектуальные системы. Теория и технология разработки. — СПб.: СПбГТУ, 2001. 711 с.
6. Тарасов В. Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. — М.: Эдиториал УРСС, 2002. 352 с.
7. Gardner H. Multiple intelligences — the theory in practice. — New York, NY, USA: Basic Books, 1993. 320 p.
8. De Bono E. Parallel thinking: From Socratic to De Bono thinking. — Melbourne: Penguin Books, 1994. 228 p.
9. Kaner S., Lind L., Toldi C., Fisk S., Beger D. The facilitator's guide to participatory decision-making. — San Francisco, CA, USA: Jossey-Bass, 2011. 368 p.
10. Колесников А. В., Листопад С. В. Модель гибридной интеллектуальной многоагентной системы гетерогенного мышления для информационной подготовки оперативных решений в региональных электрических сетях // Системы и средства информатики, 2018. Т. 28. № 4. С. 31–41.
11. Tang A. Y. C., Basheer G. S. A Conflict Resolution Strategy Selection Method (ConfRSSM) in multi-agent systems // Int. J. Advanced Computer Sci. Appl., 2017. Vol. 8. Iss. 5. P. 398–404.
12. Колесников А. В., Листопад С. В. Функциональная структура гибридной интеллектуальной многоагентной системы гетерогенного мышления для решения проблемы восстановления распределительной электросети // Системы и средства информатики, 2019. Т. 29. № 1. С. 41–52.

13. Городецкий В. И., Грушинский М. С., Хабалов А. В. Многоагентные системы (обзор) // Новости искусственного интеллекта, 1998. № 2. С. 64–116.
14. Singh M. P., Chopra A. K. Programming multiagent systems without programming agents // Programming multiagent systems / Eds. L. Braubach, J.-P. Briot, J. Thangarajah. — Lecture notes in artificial intelligence ser. — Springer, 2010. Vol. 5919. P. 1–14.
15. Singh R., Singh A., Mukherjee S. A critical investigation of agent interaction protocols in multiagent systems // Int. J. Advancements Technology, 2014. Vol. 5. Iss. 2. P. 72–81.
16. Smith G. The Contract Net Protocol: High level communication and control in a distributed problem solver // IEEE T. Comput., 1980. Vol. 29. Iss. 12. P. 1104–1113.
17. Smith R. G. A framework for distributed problem solving. — Ann Arbor, MI, USA: UMI Research Press, 1981. 188 p.
18. Weerasooriya D., Rao A. S., Ramamohanarao K. Design of a concurrent agent-oriented language // Intelligent agents / Eds. M. J. Wooldridge, N. R. Jennings. — Lecture notes in computer science ser. — Springer, 1995. Vol. 890. P. 386–401.
19. Zlotkin G., Rosenschein J. Mechanisms for automated negotiation in state oriented domain // J. Artif. Intell. Res., 1996. Vol. 5. P. 163–238.
20. Marzougui B., Barkaoui K. Interaction protocols in multiagent systems based on agent Petri nets model // Int. J. Advanced Computer Sci. Appl., 2013. Vol. 4. Iss. 7. P. 166–173.

Поступила в редакцию 31.03.19

HETEROGENEOUS THINKING PROTOCOL OF HYBRID INTELLIGENT MULTIAGENT SYSTEM FOR SOLVING DISTRIBUTIONAL POWER GRID RECOVERY PROBLEM

A. V. Kolesnikov¹ and S. V. Listopad²

¹Immanuel Kant Baltic Federal University, 14 A. Nevskogo Str., Kaliningrad 236041, Russian Federation

²Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str, Kaliningrad 236022, Russian Federation

Abstract: The problem of power supply restoration in the regional distributional power grid after large-scale accidents is characterized by high combinatorial complexity, heterogeneity, underdetermination, inaccuracy, and ambiguity. The use of the collective problem solving mechanism to overcome the listed non-factors in the sense of A. S. Narinyani is impossible due to time constraints. To solve this problem, a new class of intelligent systems that model collective decision-making under the guidance of a facilitator is proposed, namely, hybrid intelligent multiagent systems of heterogeneous thinking. Unlike traditional hybrid intelligent systems that integrate models of expert knowledge, they additionally model group processes and effects arising from collective problem solving, adapting to the dynamic nature of the problem of restoring the regional distribution grid. The paper discusses one of the components of such systems, namely, the protocol for organizing collective heterogeneous thinking of agents.

Keywords: heterogeneous thinking; hybrid intelligent multiagent system; distributional power grid recovery problem

DOI: 10.14357/19922264190211

Acknowledgments

The reported study was funded by the Russian Foundation for Basic Research according to the research project No. 18-07-00448A.

References

1. Adibi, M. M., and L. H. Fink. 2006. Overcoming restoration challenges associated with major power system disturbances. *IEEE Power Energy M.* 4(5):68–77.
2. Krasnik, V. V. 2005. *Potrebiteli elektricheskoy energii, energosnabzhayushchie organizatsii i organy Rostekhnadzora. Pravoye osnovy vzaimootnosheniy* [Consumers of electric energy, energy supplying organizations and bodies of Rostekhnadzor. Legal basis of relationships]. Moscow: ENAS. 250 p.
3. Narinyani, A. S. 2008. Inzheneriya znaniy i NE-factory: kratkiy obzor-08 [Knowledge engineering and non-factors: A brief overview-08]. *Voprosy iskusstvennogo intellekta* [Artificial Intelligence Issues] 1:61–77.

4. Kolesnikov, A. V., and S. V. Listopad. 2019. Hybrid intelligent multiagent system of heterogeneous thinking for solving the problem of restoring the distribution power grid after failures. *Open Semantic Technologies for Intelligent Systems: Research Papers Collection*. Minsk: BGUIR. 133–138.
5. Kolesnikov, A. V. 2001. *Gibridnye intellektual'nye sistemy. Teoriya i tekhnologiya razrabotki* [Hybrid intelligent systems: Theory and technology of development]. St. Petersburg: SPbSTU Pubs. 711 p.
6. Tarasov, V. B. 2002. *Ot mnogoagentnykh sistem k intellektual'nym organizatsiyam: filosofiya, psikhologiya, informatika* [From multiagent systems to intelligent organizations: Philosophy, psychology, and informatics]. Moscow: Editorial URSS. 352 p.
7. Gardner, H. 1993. *Multiple intelligences — the theory in practice*. New York, NY: Basic Books. 320 p.
8. De Bono, E. 1994. *Parallel thinking: From Socratic to De Bono thinking*. Melbourne: Penguin Books. 228 p.
9. Kaner, S., L. Lind, C. Toldi, S. Fisk, and D. Beger. 2011. *The facilitator's guide to participatory decision-making*. San Francisco, CA: Jossey-Bass. 368 p.
10. Kolesnikov, A. V., and S. V. Listopad. 2018. Model' gibridnoy intellektual'noy mnogoagentnoy sistemy geterogenogo myshleniya dlya informatsionnoy podgotovki operativnykh resheniy v regional'nykh elektricheskikh setyakh [Model of a hybrid intelligent multiagent system of heterogeneous thinking for preparation of information about operational decisions in a regional power system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):31–41.
11. Tang, A. Y. C., and G. S. Basheer. 2017. A Conflict Resolution Strategy Selection Method (ConfRSSM) in multi-agent systems. *Int. J. Adv. Computer Sci. Appl.* 8(5):398–404.
12. Kolesnikov, A. V., and S. V. Listopad. 2019. Funktsional'naya struktura gibridnoy intellektual'noy mnogoagentnoy sistemy geterogenogo myshleniya dlya resheniya problemy vosstanovleniya raspredelitel'noy elektroseti [Functional structure of the hybrid intelligent multiagent system of heterogeneous thinking for solving the problem of restoring the distribution power grid]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):41–52.
13. Gorodetskiy, V. I., M. S. Grushinskiy, and A. V. Khabalov. 1998. Mnogoagentnye sistemy (obzor) [Multiagent systems (review)]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News] 2:64–116.
14. Singh, M. P., and A. K. Chopra. 2010. Programming multiagent systems without programming agents. *Programming multi-agent systems*. Eds. L. Braubach, J.-P. Briot, and J. Thangarajah. Lecture notes in artificial intelligence ser. Springer. 5919:1–14.
15. Singh, R., A. Singh, and S. Mukherjee. 2014. A critical investigation of agent interaction protocols in multiagent systems. *Int. J. Advancements Technology* 5(2):72–81.
16. Smith, G. 1980. The Contract Net Protocol: High level communication and control in a distributed problem solver. *IEEE T. Comput.* 29(12):1104–1113.
17. Smith, R. G. 1981. *A framework for distributed problem solving*. Ann Arbor, MI: UMI Research Press. 188 p.
18. Weerasooriya, D., A. S. Rao, and K. Ramamohanarao. 1994. Design of a concurrent agent-oriented language. *Intelligent agents*. Eds. M. J. Wooldridge and N. R. Jennings. Lecture notes in computer science ser. Springer. 890:386–401.
19. Zlotkin, G., and J. Rosenschein. 1996. Mechanisms for automated negotiation in state oriented domain. *J. Artif. Intell. Res.* 5:163–238.
20. Marzougui, B., and K. Barkaoui. 2013. Interaction protocols in multi-agent systems based on agent Petri nets model. *Int. J. Advanced Computer Sci. Appl.* 4(7):166–173.

Received March 31, 2019

Contributors

Kolesnikov Alexander V. (b. 1948) — Doctor of Science in technology, professor, Institute of Physical and Mathematical Sciences and Information Technology, Immanuel Kant Baltic Federal University, 14 A. Nevskogo Str., Kaliningrad 236041, Russian Federation; avkolesnikov@yandex.ru

Listopad Sergey V. (b. 1984) — Candidate of Science (PhD) in technology, senior scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236022, Russian Federation; ser-list-post@yandex.ru

СОЧЕТАЕМОСТЬ ЛОГИКО-СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ: КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА*

О. Ю. Инькова¹, М. Г. Кружков²

Аннотация: Рассматриваются логико-семантические отношения (ЛСО), обеспечивающие связность текста, и их эксплицитные текстовые показатели — коннекторы. Представлен обзор существующих подходов к определению и классификации ЛСО, рассмотрены различные варианты сочетаемости ЛСО и их показатели. Для решения задачи количественного анализа сочетаемости ЛСО предлагается использовать надкорпусные базы данных (НБД), в рамках которых для аннотирования ЛСО и их показателей используется двухуровневая фасетная классификация. На материале базы данных генерируется и интерпретируется количественная информация по сочетаемым характеристикам ЛСО. В частности, показано, что отношения экстенциональной генерализации гораздо чаще сочетаются с соединительными отношениями, чем, например, отношения спецификации, определяются речевые реализации (РР), которые чаще всего являются показателями более чем одного отношения, и т. д. Показано, как результаты анализа сочетаемости ЛСО могут быть использованы для дальнейшего развития методологии обратимой генерализации информационных объектов. Гибкость и доступность предлагаемого подхода позволяют по-новому подойти к пока мало изученной проблеме сочетаемости ЛСО.

Ключевые слова: надкорпусные базы данных; количественный анализ; коннекторы; логико-семантические отношения; аннотирование отношений; генерализация информационных объектов

DOI: 10.14357/19922264190212

1 Введение

Изучение ЛСО как обеспечивающих логическую связность текста имеет важное значение в областях лингвистики и информатики, связанных с обработкой естественного языка. Перед исследователями ставятся задачи, связанные с разметкой ЛСО в корпусах и автоматической обработкой текстов на естественном языке. В связи с этим в лингвистике, особенно в корпусной, активно обсуждаются среди прочих два вопроса, имеющих непосредственное отношение к связности текста: классификация ЛСО и структура и сочетаемость их показателей — коннекторов.

Эти вопросы тесно взаимосвязаны, хотя между ЛСО и его показателями нет симметрии:

- ЛСО может не иметь показателя, а само отношение выводится на основе смыслового соотношения фрагментов текста (*Петя не пришел. Он заболел*);
- оно может быть выражено несколькими показателями (ср. *Когда прозвонил будильник, Петя встал — Едва лишь только прозвонил будильник, как Петя тут же встал*);

- показатель ЛСО может быть полисемичным (ср. *Если надумаешь, то позвони* (условные ЛСО) — *Если Маша умная и красивая, то ее сестра* — просто чудовище (сопоставительные ЛСО)).

2 Существующие подходы к классификации и сочетаемости логико-семантических отношений

Что касается классификации, то на сегодняшний день их существует в опубликованном виде как минимум пять. При этом сам термин «отношение» понимается в них довольно по-разному.

В ТРС (теории риторической структуры, *англ.* RST — rhetorical structure theory [1]) под «отношением» понимается любое отношение, связывающее два фрагмента текста с пропозициональным содержанием, обеспечивающее связность текста. Таким образом, в дискурсивные отношения попа-

*Разделы 1–3 и 5 данной работы выполнены при финансовой поддержке РФ (проект 16-18-10004), раздел 4 выполнен по плановой теме Минобрнауки № 0063-2019-0010.

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, olyainkova@yandex.ru

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, magnit75@yandex.ru

дают и размечаются как таковые в основанных на этой классификации корпусах, например:

- отношения, в основе которых лежит общий принцип релевантности информации; ср. отношение «Решение (Solutionhood)» (*Я хочу есть, пойдём в «Шоколадницу»*) или «Обеспечение возможности (Enablement)» (*Можешь открыть дверь? Тут вот ключ лежит*);
- анафорические отношения, в частности ассоциативная анафора (*Карин очень фотогенична, её улыбка — просто очаровательна*), которые носят название «Детализация», или «Уточнение (Elaboration)», в его разновидности Часть—Целое;
- разного рода тема-рематические отношения, например цепочки с постоянной темой, которые также попадают в отношение «Детализация», или «Уточнение (Elaboration)», но в его разновидности Предмет—Признак (*Я — офицер Иванов, я родился в 1970 году, работаю в полиции с 1990 года*);
- чисто синтаксические явления, обеспечивающие связность текста (например, изъяснительные отношения, носящие имя «Содержание»: *Я знаю, что Петя уезжает*).

Близкий подход к термину «отношение» характеризует и классификацию, предложенную Н. Ашером и А. Ласкардесом (SDRT — segmental discourse representation theory [2]).

Несколько иной подход, в большей степени ориентированный на анализ отношений, которые могут быть выражены специальными показателями — коннекторами, представлен в трех других классификациях:

- наиболее известная в кругах корпусных лингвистов классификация, используемая для создания Penn Discourse TreeBank (PDTB) под руководством Б. Вебер;
- недавно опубликованная работа [3], опирающаяся во многом на предыдущие работы Т. Сандерса и классификацию А. Кнотта [4, 5];
- классификация, предложенная немецкими коллегами в фундаментальном многотомном труде, посвященном синтаксису и семантике коннекторов немецкого языка, под руководством Э. Брайндл [6].

Анализ этих пяти классификаций — отдельная задача. Подчеркнем лишь два момента. Первое — это то, что само понятие «отношение» имеет в них разное значение. В наиболее близких авторам по подходу двух последних классификациях, в меньшей степени в классификации команды Б. Вебер,

классифицируются именно ЛСО, которые могут быть выражены соответствующим показателем — коннектором. Остальные отношения, в частности перечисленные выше для TRC, остаются за пределами исследования и разметки.

Второе — это тот факт, что все пять классификаций между собой плохо соотносимы. Чтобы преодолеть эту проблему, в работе членов команды Б. Вебер [7] предлагается своего рода стандарт (Core Collection) дискурсивных отношений. По этому же пути пошли разработчики французского аннотированного корпуса ANNODIS, взяв для разметки 17 «наиболее распространенных дискурсивных отношений» [8, с. 2651]. Но эти два списка уже не совпадают, а сопоставление отношений, имеющих одинаковые названия, вызывает определенные трудности. Подробно эта проблема обсуждается в работе [3].

К этим трудностям можно добавить еще одну, которая часто остается за пределами внимания исследователей: речь идет о возможном существовании более одного отношения, в данном случае именно ЛСО между фрагментами текста, будь то отношения, эксплицитно выраженные коннектором, или имплицитные. Наибольший интерес для авторов, учитывая особенности разработанной НБД, представляет первый случай.

Вопрос о возможности существования между фрагментами текста нескольких ЛСО предполагает решение двух теоретических проблем. С одной стороны, необходимо понять, в какой степени ЛСО совместимы друг с другом и в какой степени их сочетание возможно и предсказуемо. С другой стороны, в плане взаимоотношения синтаксиса и семантики необходимо определить те механизмы, которые управляют сочетаемостью показателей ЛСО, коннекторов, и в какой степени можно говорить о композициональности их значения.

В трех подходах, которые легли в основу создания банков данных о дискурсивных отношениях (RST, SDRT и PDTB), этот вопрос решается по-разному, если решается.

Теория риторической структуры, которая квалифицирует себя как теория, ставящая своей целью описание организации текста, должна рассматриваться скорее как инструмент для лингвистического анализа, чем модель организации дискурса, способная предсказать эмпирические наблюдения над его функционированием. Заметим также, что RST в анализе мало опирается на коннекторы, считая их слабыми индикаторами связности дискурса.

Наконец, анализ структуры текста в RST подчиняется некоторым условиям (обоснование которых, хотя и недостаточно четко, сформулировано Манном и Томпсон в основополагающей работе [1]),

одно из которых в принципе ставит под вопрос возможность существования нескольких дискурсивных отношений между анализируемыми единицами. Это условие *единственности*, согласно которому не может быть ситуации, когда два разных отношения существуют между одинаковыми фрагментами текста. Из этого можно заключить, что наличие нескольких дискурсивных отношений между анализируемыми единицами считается скорее исключением.

Теория SDRT предусматривает возможность существования нескольких отношений между дискурсивными единицами, что является прямым следствием Принципа максимизации связности текста. Согласно этому принципу, если интерпретация И некоторой последовательности высказываний является более связной, чем интерпретация I_1 , то множество отношений, выявленных в I_1 , присутствует и в И. Основное требование, которое предъявляется к сосуществующим отношениям, — требование связности: предпочтение отдается таким отрезкам дискурса, в которых выявленные отношения не противоречат друг другу. Этот подход используется в базе данных ANNODIS (http://redac.univ-tlse2.fr/corpus/annodis/annodis_rr.html), в которой аннотированы тексты на французском языке. Аннотация производится при помощи программного обеспечения Glozz (<http://www.glozz.org>), которое необходимо установить для визуализации результатов разметки.

В рамках проекта разметки дискурсивных отношений PDTB вопрос о сосуществовании нескольких отношений между одними и теми же фрагментами текста решается с практической точки зрения. В общем случае, когда между фрагментами текста присутствуют два коннектора (например, *so for instance, but then, and furthermore*), они аннотируются по отдельности и для каждого из них указываются фрагменты текста («аргументы»), которые они связывают [9].

Единственный случай, при котором в PDTB допускается проставлять больше чем одно отношение, — это случай неоднозначности при имплицитно выраженных отношениях. Тогда проставляются все возможные для данной последовательности отношения:

- (1) This cannot be solved by provoking a further downturn; reducing the supply of goods does not solve inflation (Implicit = *so* Contingency.Cause.Result, Implicit = *instead* Exp.Alt.Chosen alt). Our advice

is this: Immediately return the government surpluses to the economy through incentive-maximizing tax cuts, and find some monetary policy target that balances both supply and demand for money.¹

Для визуализации результатов разметки в PDTB необходим платный доступ к базе данных и установка соответствующего программного обеспечения.

3 Разметка логико-семантических отношений в надкорпусной базе данных коннекторов и предлагаемые решения

В целях систематического исследования показателей ЛСО была разработана НБД коннекторов, позволяющая осуществлять аннотирование коннекторов и их переводных эквивалентов в корпусе параллельных текстов Национального корпуса русского языка (НКРЯ). В рамках данной работы не будем рассматривать общую структуру НБД (об этом см., например, [10]), а подробнее остановимся на той ее части, которая связана с аннотированием показателей ЛСО в русском языке.

Для аннотирования используется двухуровневая фасетная классификация: аннотированию подлежат, с одной стороны, форма коннекторов, в какой они встретились в тексте (в терминах базы данных эти формы получили название *речевые реализации*), с другой стороны, аннотируется контекст употребления этих коннекторов. Речевые реализации аннотируются с помощью структурных признаков — указывается структурный тип коннектора (одноэлементный, многоэлементный, двухкомпонентный или n -компонентный) и входящие в него элементы. Контексты употребления РР аннотируются по нескольким основаниям, из которых наибольший интерес представляет кластер признаков «Логико-семантические отношения». При этом если РР включает в себя несколько элементов, маркирующих различные ЛСО, то РР присваивается более одного признака из данного кластера (см. пример на рис. 1). Если же для РР эксперту представляются равновероятными несколько ЛСО, то он может использовать введенные в систему неоднозначные (комбинированные) признаки.

¹Пример из [9], перевод на русский язык: *Это невозможно устранить, не вызвав нового спада; уменьшение поставок не ведет к снижению инфляции* (имплицит. отн-ия: «результат» и «выбор альтернативы»). *Наш совет — немедленно вернуть правительственные излишки в экономику за счет стимулирующего снижения налогов и выработать для денежной политики новую цель, которая бы уравновешивала предложение и спрос на деньги.*

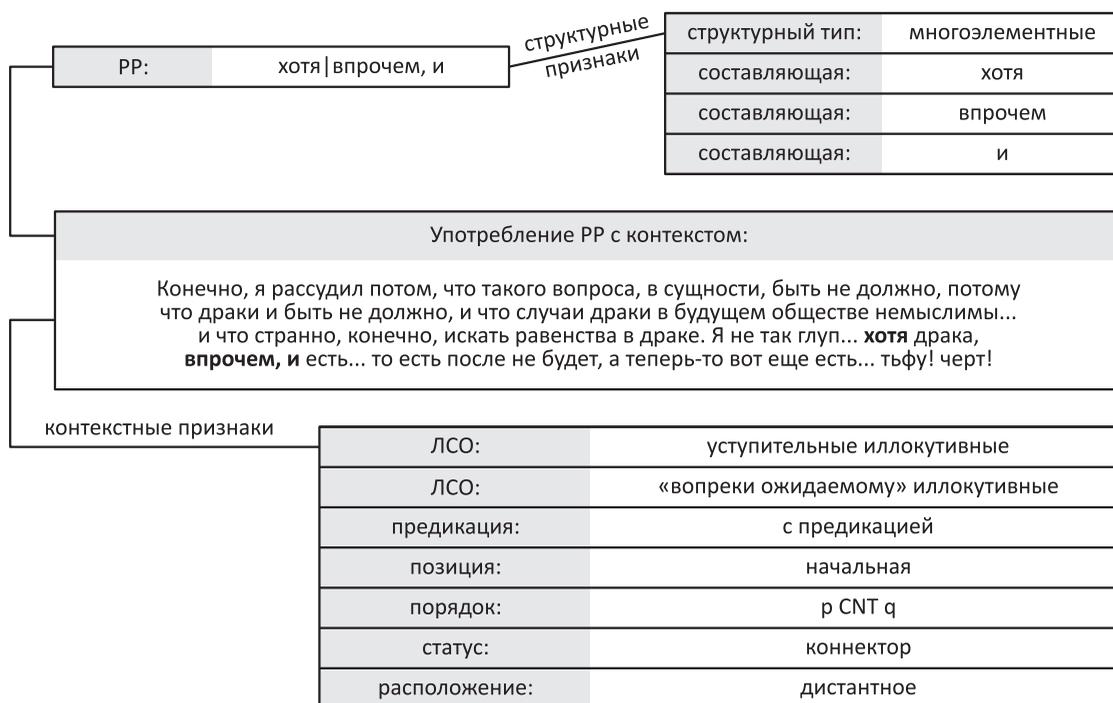


Рис. 1 Пример аннотирования контекста с несколькими ЛСО

Описанная система разметки позволяет получить количественную информацию по частотности различных сочетаний показателей ЛСО, аннотированных в корпусе.

Например, задавая то или иное сочетание признаков из кластера «Отношения», можно под-

считать, сколько раз в аннотированном корпусе встречаются употребления коннекторов, которые включают в себя элементы, маркирующие соответствующие ЛСО.

В табл. 1 представлены по мере убывания зарегистрированные в НБД сочетания. Это, одна-

Таблица 1 Сочетания показателей ЛСО в НБД

Сочетание ЛСО	Число употреблений в НБД
Соединительные; экстенциональная генерализация	120
«Вопреки ожидаемому»; уступительные	85
Пропозициональное сопутствование; соединительные	69
Соединительные; спецификация	64
Иллокутивное сопутствование; соединительные	44
Спецификация; сравнительные	41
Аддитивные иллокутивные; соединительные	38
«Вопреки ожидаемому»; пропозициональное сопутствование	36
Аддитивные пропозициональные; соединительные	35
«Вопреки ожидаемому»; возместительное противопоставление	34
Временные; соединительные	33
Аддитивные иллокутивные; экстенциональная генерализация	25
«Вопреки ожидаемому»; противительно-уступительные	22
Интенциональная генерализация; соединительные	22
«Вопреки ожидаемому» иллокутивные; уступительные иллокутивные	21
...	...
Контраст; спецификация	1
...	...

ко, абсолютные числа. Для интерпретации данной информации необходимо учитывать, сколько аннотаций для каждого ЛСО было построено. Например, для экстенциональной генерализации, т.е. генерализации, осуществляющей переход от частной ситуации к общей, создано 524 аннотаций, из них, как следует из табл. 1, в 120 аннотациях, т.е. в 23% случаев, показатель этого вида генерализации *вообще* сочетается с показателем соединительных отношений — союзом *и*; ср. (2) с (3):

- (2) Чипа угостил нас разведенным спиртом. *И вообще*, проявил услужливость. [Сергей Довлатов. Чемодан (1986)]
- (3) Никогда я не совал своего носа в литературу и в политику, не искал популярности в полемике с невеждами, не читал речей ни на обедах, ни на могилах своих товарищей. . . *Вообще* на моем ученом имени нет ни одного пятна и пожаловаться ему не на что. [А. П. Чехов. Скучная история (1889)]

Для отношения спецификации в НБД сформирована 1381 аннотация, и только в 64 из них (т.е. менее 5%) зафиксировано сочетание показателя этого ЛСО с союзом *и*:

- (4) Вчерашний день, таким образом, помаленьку высветлялся, но Степу сейчас гораздо более интересовал день сегодняшней *и, в частности*, появление в спальне неизвестного, да еще с закуской и водкой. [М. А. Булгаков. Мастер и Маргарита (1929–1940)]

Для отношения исключения сочетаемость с другими ЛСО вообще не характерна. Для него создано 356 аннотаций, и только в 5 из них (т.е. всего 1,2%) зафиксировано сочетание этого ЛСО с каким-либо другим. Этих отношений три: наименее избирательное соединительное (3 аннотации), «вопреки ожидаемому» (1 аннотация) и аддитивные иллокутивные (1 аннотация):

- (5) Весь этот мир куда-то пропал. *И только* ремень все еще цел. [Сергей Довлатов. Чемодан (1986)]
- (6) Одно это я могу сказать против своего чувства. . . Это важно < . . . > *Но, кроме этого*, сколько бы я ни искал, я ничего не найду, что бы сказать против моего чувства. [Л. Н. Толстой. Анна Каренина (1878)]

Можно также сгруппировать результаты в каждом из приведенных в табл. 1 сочетаний по конкретным РР (табл. 2, вертикальная черта «|» в РР означает, что ее элементы разделены текстом). Например, можно узнать, какие РР сочетают в себе

Таблица 2 Сочетание показателей спецификации с показателем соединительных ЛСО

РР	Число употреблений в НБД
и особенно	29
и в частности	8
и между прочим	7
и в особенности	5
и среди прочего	4
и, между прочим	1
и, например	1
и, скажем	1
и в особенности	1
и в том числе	1
и между прочим	1
и например	1
и тем более	1
особенно и особенно	1
и именно	1
и между ними	1

показатели соединительных отношений и отношений спецификации (4-я строка табл. 1) и как часто каждая из этих РР встречается в аннотированной части корпуса.

Наконец, проанализировав перекрестную таблицу отношений в кластере «Отношения», можно выяснить, с какими другими ЛСО чаще всего сочетается то или иное ЛСО. Например, в табл. 3 показано, с какими другими отношениями чаще всего сочетается отношение спецификации в оригинальных русских текстах в параллельном русско-французском подкорпусе.

Как уже отмечалось, НБД коннекторов включает в себя информацию о переводных эквивалентах русских коннекторов, которые аннотируются аналогичным образом. В результате появляется возможность сформировать аналогичную таблицу по сочетаемости отношения спецификации в переводных французских текстах (табл. 4). Как и следовало ожидать, результаты довольно близки, поскольку учтенные здесь французские коннекторы являются переводами значительной части соответствующих русских коннекторов (не считая тех, которые переводились не-коннекторами или, возможно, коннекторами, в которых часть семантики русского коннектора была утрачена). Также следует учитывать, что для некоторых русских коннекторов в НБД может быть зарегистрировано несколько переводов, выполненных разными авторами. Наиболее заметным отличием, на первый взгляд, представляется значительное уменьшение степени сочетаемости с временными и сопоставительными отношениями, а также отсутствие в перево-

Таблица 3 Сочетаемость показателей ЛСО спецификации с показателями других ЛСО в текстах на русском языке в русско-французском подкорпусе

ЛСО	Число употреблений в НБД
Соединительные	32
Сравнительные	12
Альтернатива	7
Условные	6
Сопоставительные	6
Временные	3
Аддитивные иллокутивные	3
«Вопреки ожидаемому» иллокутивные	3
Пропозициональная альтернатива	3
Причина	2
Тождество	2
Переформулирование	1
«Вопреки ожидаемому»	1
Контраст	1

Таблица 4 Сочетаемость показателей ЛСО спецификации с показателями других ЛСО в текстах переводов на французский язык

ЛСО	Число употреблений в НБД
Соединительные	32
Сравнительные	12
Альтернатива	7
Условные	6
Сопоставительные	6
«Вопреки ожидаемому» иллокутивные	3
Аддитивные иллокутивные	3
Пропозициональная альтернатива	3
Временные	3
«Вопреки ожидаемому»	1
Причина	1
Тождество	1
Контраст	1
Переформулирование	1

де сочетаемости с отношениями переформулирования и аддитивными иллокутивными отношениями.

4 Сочетаемость показателей логико-семантических отношений как объект генерализации

Результаты количественного анализа сочетаемости ЛСО могут быть использованы для дальнейшего развития методологии обратимой генерализации информационных объектов в текстовой форме. В ряде работ [11–13] эта методология разрабатывалась для моделей перевода. Здесь представлен другой пример обратимой генерализации текстовых объектов: поскольку показатели ЛСО могут маркировать различные сочетания ЛСО, аннотации РР могут быть генерализованы на этом основании по нескольким уровням обобщения. Переходы от аннотаций употреблений РР к генерализуемым объектам различных уровней и обратно обеспечиваются благодаря вычисляемым связям, реализуемым на основе структуры НБД.

На рис. 2 показан фрагмент иерархии аннотаций употреблений РР, включающий в себя 5 уровней. На нижнем уровне находятся исходные генерализуемые объекты, содержащиеся в НБД, — употребления РР русского языка, которые были найдены и аннотированы в русско-французском корпусе параллельных текстов НКРЯ. Остальные

уровни представляют собой генерализуемые объекты разной степени детализации.

Генерализуемые объекты нижнего уровня представляют собой аннотации употреблений РР русского языка, которые одновременно содержат два признака из кластера ЛСО: «соединительные отношения» и «отношения спецификации».

На первом уровне генерализации (второй уровень иерархии, начиная снизу) исходные генерализуемые объекты сгруппированы по формам РР, которые реализуют указанные отношения, наиболее частотными из которых являются формы *и особенно, и в частности, и между прочим, и в особенности*.

На втором уровне генерализуемые объекты объединены в одну группу, которая соответствует сочетаниям ЛСО {соединительные, спецификация, . . . }, которые одновременно включают в себя соединительные отношения и отношения спецификации независимо от того, какие именно РР маркируют эти сочетания.

На третьем уровне генерализации находятся сочетания ЛСО, включающие в себя по меньшей мере одно из указанных отношений (и, возможно, другие отношения). Так, группа {соединительные, . . . } объединяет 478 аннотаций, при этом в нее полностью входит группа 2-го уровня генерализации {соединительные, спецификация, . . . }. В то же время группа {спецификация, . . . } объединяет 1381 аннотацию, при этом в нее также полностью входит группа 2-го уровня генерализации {соединительные, спецификация, . . . }. Таким образом, данная иерархия допускает альтернативность

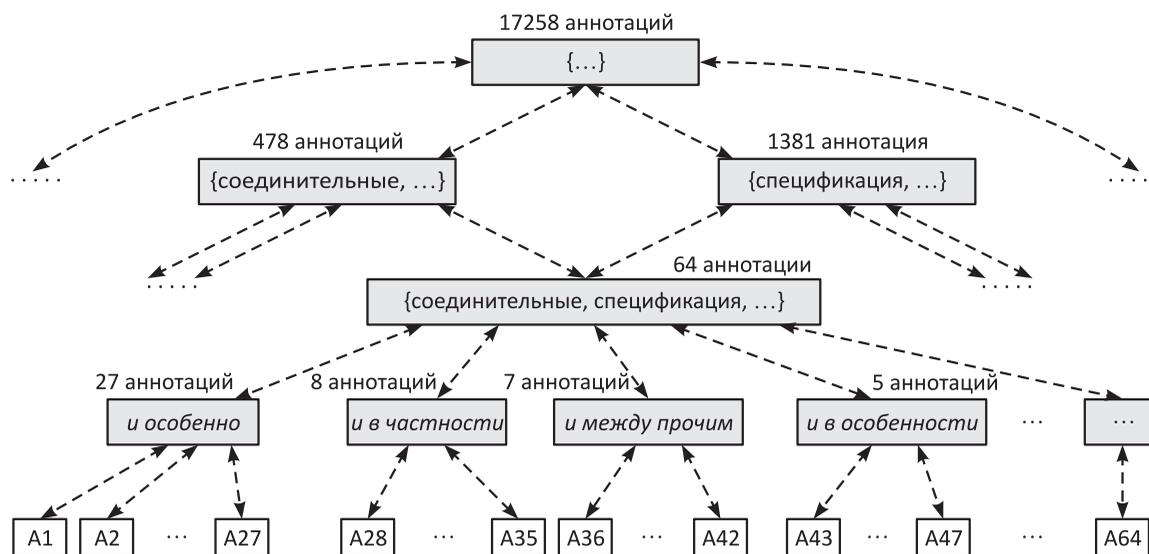


Рис. 2 Генерализация употреблений РР по маркируемым ими сочетаниям ЛСО

генерализации при переходе со второго уровня на третий.

Четвертый и последний уровень генерализации включает единственную группу {...}, которая соответствует «абстрактному» сочетанию ЛСО, т. е. всем возможным сочетаниям ЛСО, которые могут маркироваться с помощью РР. В данную группу входят все лежащие ниже группы 3-го и более низких уровней.

5 Заключительные замечания

Разработанная НБД показателей ЛСО имеет целый ряд преимуществ по сравнению с существующими информационными ресурсами, позволяющими анализировать дискурсивные, или риторические, отношения:

- НБД является доступным ресурсом, не требующим установки специального программного обеспечения и находящимся в открытом доступе (по крайней мере, ее фрагмент, который содержит 6654 аннотаций);
- применяемый для разметки ЛСО теоретический подход не только допускает сочетаемость показателей различных ЛСО, но и дает возможность аннотировать такие случаи;
- применяемая система разметки позволяет с помощью дополнительных тегов аннотировать такие случаи, когда одно из ЛСО не выражено эксплицитно;
- используемая система разметки позволяет получать количественную информацию о сочетаемости показателей различных ЛСО, а следовательно, решать этот вопрос на основе не

отдельных эмпирических наблюдений и гипотез, а верифицируемых статистических данных, полученных на представительном массиве аннотаций.

Все это позволяет по-новому подойти к пока еще мало изученной теоретической проблеме сочетаемости показателей ЛСО и в перспективе на основе дальнейшего семантического анализа позволит ответить на вопросы о том, в какой степени ЛСО совместимы друг с другом (для некоторых ЛСО ответ на этот вопрос уже получен) и в какой степени их сочетание предсказуемо, а также определить те семантические и синтаксические механизмы, которые управляют сочетаемостью показателей ЛСО.

Литература

1. Mann W. C., Thompson S. A. Rhetorical structure theory: Toward a functional theory of text organization // Text, 1988. Vol. 8. No. 3. P. 243–281.
2. Asher N., Lascarides A. Logics of conversation. — Cambridge: Cambridge University Press, 2003. 526 p.
3. Sanders T. J. M., Demberg V., Hoek J., Scholman M. C. J., Torabi A. F., Zufferey S., Evers-Vermeul J. Unifying dimensions in coherence relations: How various annotation frameworks are related // Corpus Linguist. Ling., 2018. 71 p. doi: 10.1515/cllt-2016-0078.
4. Knott A. A data-driven methodology for motivating a set of coherence relations. — Edinburgh: University of Edinburgh, 1996. Ph.D. Thesis. <https://www.era.lib.ed.ac.uk/handle/1842/583>.
5. Knott A., Sanders T. The classification of coherence relations and their linguistic markers: An exploration of two languages // J. Pragmatics, 1998. Vol. 30. No. 2. P. 135–175.

6. Breindl E., Volodina A., Waßner U. H. Handbuch der deutschen Konnektoren 2. Semantik der deutschen Satzverknüpfers. — Berlin: Walter de Gruyter, 2014. 1327 p.
7. Bunt H., Prasad R. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations // 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings. — Portoroz, 2016. P. 45–54.
8. Ho-Dac L.-M., Péry-Woodley M.-P. Annotation des structures discursives: l'expérience ANNODIS // 4e Congrès Mondial de Linguistique Française. — Berlin, 2014. P. 2647–2661.
9. Webber B. Concurrent discourse relations // Компьютерная лингвистика и интеллектуальные технологии. — М.: Изд-во РГГУ, 2016. Вып. 15. С. D. <http://www.dialog-21.ru/media/3815/webber.pdf>.
10. Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А. Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
11. Зацман И. М., Мамонова О. С., Шурова А. Ю. Обратимость и альтернативность генерализации моделей перевода коннекторов в параллельных текстах // Системы и средства информатики, 2017. Т. 27. № 2. С. 125–142.
12. Зацман И. М., Кружков М. Г., Лоцилова Е. Ю. Методы анализа частотности моделей перевода коннекторов и обратимость генерализации статистических данных // Системы и средства информатики, 2017. Т. 27. № 4. С. 164–176.
13. Зацман И. М. Методология обратимой генерализации в контексте классификации информационных трансформаций // Системы и средства информатики, 2018. Т. 28. № 2. С. 128–144.

Поступила в редакцию 29.03.19

COMPATIBILITY OF LOGICAL SEMANTIC RELATIONS: METHODS OF QUANTITATIVE ANALYSIS

O. Yu. Inkova and M. G. Kruzchkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper deals with logical semantic relations (LSR) that ensure text coherence and examines their explicit markers in text, i. e., connectives. An overview of existing approaches to definition and classification of LSRs is presented; different LSR combination patterns are examined. To make quantitative assessment of LSR compatibility, one should be able to annotate LSRs and their markers in texts. To support such annotation, a supracorpora database was developed, including customizable faceted classifications for LSRs and connectives. Based on the database data, quantitative information on relations combinations was accumulated and interpreted. For example, it was demonstrated that adjoining relations collocate with extensional generalization relations much more often than with relations of specifications; discourse realizations commonly used to mark more than one relation at a time were identified, etc. The results of LSR compatibility analysis were used to elaborate the methodology of reversible generalization of information objects. High flexibility and accessibility of the proposed approach allows researchers to address the underinvestigated problem of LSR compatibility.

Keywords: databases; quantitative analysis; connectives; logical semantic relations; annotation of relations; generalization of information objects

DOI: 10.14357/19922264190212

Acknowledgments

The work was carried out at the Institute of Informatics Problems (FRC CSC RAS); sections 1-3 and 5 were funded by the Russian Science Foundation according to the research project No. 16-18-10004; and section 4 was carried out as a part of the planned research topic No. 0063-2019-0010 of the Ministry of Education and Science of Russian Federation.

References

1. Mann, W. C., and S. A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8(3):243–281.
2. Asher, N., and A. Lascarides. 2003. *Logics of conversation*. Cambridge, Cambridge University Press. 526 p.
3. Sanders, T. J. M., V. Demberg, J. Hoek, M. C. J. Scholman, A. F. Torabi, S. Zufferey, and J. Evers-Vermeul. 2018. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguist. Ling.* 71 p. doi: 10.1515/cilt-2016-0078.
4. Knott, A. 1996. A data-driven methodology for motivating a set of coherence relations. Edinburg: University

- of Edinburgh. Ph.D. Thesis. Available at: <https://www.era.lib.ed.ac.uk/handle/1842/583> (accessed March 25, 2019).
5. Knott, A., and T. Sanders. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *J. Pragmatics* 30(2):135–175.
 6. Breindl, E., A. Volodina, and U. H. Waßner. 2014. *Handbuch der deutschen Konnektoren 2. Semantik der deutschen Satzverknüpfers*. Berlin: Walter de Gruyter. 1327 p.
 7. Bunt, H., and R. Prasad. 2016. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. *12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12) Proceedings*. Portoroz. 45–54.
 8. Ho-Dac, L.-M., and M.-P. Péry-Woodley. 2014. Annotation des structures discursives: l'expérience ANNODIS. *4e Congrès Mondial de Linguistique Française*. Berlin. 2647–2661.
 9. Webber, B. 2016. Concurrent discourse relations. *Computational linguistics and intellectual technologies*. Moscow: Russian State University for the Humanities Pubs. 15:D. Available at: <http://www.dialog-21.ru/media/3815/webber.pdf> (accessed March 25, 2019).
 10. Zatsman, I. M., O. Yu. In'kova, M. G. Kruzhhkov, and N. A. Popkova. 2016. Predstavlenie krossyazykovykh znaniy o konnektorakh v nadkorporusnykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
 11. Zatsman, I. M., O. S. Mamonova, and A. Yu. Shchurova. 2017. Obratimost' i al'ternativnost' generalizatsii modeley perevoda konnektorov v paralel'nykh tekstakh [Reversibility and alternativeness of generalization of connectives translations models in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):125–142.
 12. Zatsman, I. M., M. G. Kruzhhkov, and E. Yu. Loshchilova. 2017. Metody analiza chastotnosti modeley perevoda konnektorov i obratimost' generalizatsii statisticheskikh dannykh [Methods of frequency analysis of connectives translations and reversibility of statistical data generalization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):164–176.
 13. Zatsman, I. M. 2018. Metodologiya obratimoy generalizatsii v kontekste klassifikatsii informatsionnykh transformatsiy [Methodology of reversible generalization in context of classification of information transformations]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):128–144.

Received March 29, 2019

Contributors

Inkova Olga Yu. (b. 1965) — Doctor of Science (PhD) in philology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; olyainkova@yandex.ru

Kruzhhkov Mikhail G. (b. 1975) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ СЕТЕВОГО ПЛАНИРОВАНИЯ ПОРТФЕЛЯ ПРОЕКТОВ С НЕОДНОРОДНЫМИ РЕСУРСАМИ В УСЛОВИЯХ НЕЧЕТКОЙ ИНФОРМАЦИИ*

А. А. Зацаринный¹, В. В. Коротков², М. Г. Матвеев³

Аннотация: Рассматривается задача планирования портфеля проектов в условиях ограничений на имеющиеся трудовые ресурсы различной специализации и нечетких оценок трудозатрат выполнения работ. Был использован аппарат нечеткой W -алгебры, позволивший избежать недостатков и трудностей традиционных подходов: неоправданного расширения носителя, необходимости задания операции сравнения нечетких чисел и ряда других. Предложенная модель позволяет получить нечеткие оценки времени выполнения проектов и их работ, а также оптимальное распределение исполнителей по работам с учетом их возможной частичной занятости. Для решения задачи был реализован генетический алгоритм, основанный на кодировании особей в виде списка приоритетов работ. Приводится численный пример, демонстрирующий перспективы применения данного подхода проектными организациями при совершении оценок, необходимых для принятия решений в условиях неопределенности.

Ключевые слова: управление проектами; W -алгебра; нечеткая арифметика; комбинаторная оптимизация; генетический алгоритм

DOI: 10.14357/19922264190213

1 Введение

В условиях, когда ключевой национальной целью России стал научно-технологический прорыв и руководством страны взят курс на реализацию Программы цифровой экономики, объективно возросла значимость системного подхода к процессам планирования и управления выполнением масштабных проектов [1]. Важнейшими инструментами для реализации такого подхода должны стать результаты опережающих фундаментальных исследований, направленных в том числе и на совершенствование методических подходов к сетевому планированию [2].

Процессы управления проектами включают обоснование планов поэтапного выполнения работ в рамках комплекса проектов при заданных ресурсах. Такие планы предполагают составление сетевых план-графиков и расписания выполнения работ с учетом заданных сроков и выделенных ресурсов [3].

В основе управления лежит план-фактный анализ, который в определенные моменты времени показывает степень рассогласования планового и фактического расписания выполнения работ и является первоосновой для перераспределения

ресурсов и составления нового сетевого плана. Проблема управления особенно остро встает при реализации портфеля проектов, когда добавление нового проекта в портфель выполняемых проектов сразу требует плановой оценки времени его завершения, т. е. времени завершения последней работы в сети, и плановой оценки потребляемых работами ресурсов. Без таких оценок невозможно обеспечить эффективность проектной деятельности, так как на них основан расчет финансовых отношений заказчика проекта и проектной организации.

Таким образом, проблема сетевого планирования проектов с ограниченными ресурсами (Resource Constrained Project Scheduling Problem, RCPSP) становится одной из самых актуальных прикладных задач в деятельности практически каждой проектной организации. Параметры в задаче сетевого планирования чаще всего не являются детерминированными величинами. В частности, нельзя точно указать продолжительности выполнения работ на стадии планирования, можно говорить лишь о приблизительных сроках. Возникающую неопределенность часто пытаются описать с помощью статистических методов [4]. Но статистические выборки не всегда релевантны особенностям конкретных работ конкретной проектиру-

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-01-00251).

¹ Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

² Воронежский государственный университет, факультет компьютерных наук, chascrunk@gmail.com

³ Воронежский государственный университет, факультет компьютерных наук, mgmatveev@yandex.ru

ющей организации. Экспертная оценка — более подходящий способ задания параметров работ проекта. При этом адекватным аппаратом описания значений параметров для эксперта служит нечеткая математика [5]. Применение нечеткой математики для управления проектами давно привлекает внимание исследователей как перспективное направление [6]. В работе [7] приведена классификация методов решения задачи сетевого планирования с нечеткими параметрами с указанием их недостатков, главные из которых — громоздкость вычислений, искажение естественных свойств и отношений классических моделей, неоправданное расширение носителя результата, что сказывается на возможности его интерпретации и полезности практического применения. В работах [8, 9] для устранения указанных недостатков предлагается использовать W -алгебру и W -числа. В работе [10] приводится пример использования W -алгебры для решения задачи планирования работ проекта с фиксированными однородными ресурсами и нечетко заданными продолжительностями работ методами линейного программирования. Однако, как отмечается в [11], такой подход нельзя использовать при планировании с ограниченными переменными ресурсами. В этом случае приходится обращаться к эвристическим алгоритмам квазиоптимального выбора. В работах [12–14] предлагается использовать различные модификации генетических алгоритмов, которые существенно упрощают решение задачи даже для планирования портфеля проектов с переменными разнородными ресурсами, но при детерминированном задании параметров работ.

В статье предлагается решение задачи сетевого планирования портфеля проектов с неоднородными ограниченными ресурсами. Длительность работ задается нечеткими W -числами и зависит от выделенных ресурсов. Результатом решения является нечеткий сетевой график выполнения работ, позволяющий рассчитать нечеткую оценку длительности выполнения каждого планируемого проекта, а также нечеткую оценку плановой потребности в неисчерпаемых ресурсах и степень их загруженности.

2 Формализация задачи сетевого планирования

Под проектом понимается комплекс мероприятий, направленных на достижение к заданному сроку определенных целей в условиях ограничений на выделенные ресурсы [6]. Далее под ресурсами будем понимать трудовые ресурсы, т. е. исполнителей работ проектов.

Рассматривается портфель проектов P , состоящий из n проектов. Каждый проект $p \in P$ включает множество работ J_p . Предполагается, что работы не могут быть прерваны в процессе своего выполнения. Каждый проект $p \in P$ также имеет дату старта Start_p , раньше которой не может быть начата ни одна из его работ. Некоторые работы требуют завершения других перед своим началом, поэтому множество $\text{Pre}(j)$ задает всех предшественников работы $j \in J_p$.

Выполнение работ требует привлечения исполнителей соответствующей специализации. Введем множество всех возможных специализаций E и пусть отображение $\text{Spec} : J_p \rightarrow E \forall p \in P$ ставит в соответствие всякой работе любого проекта специализацию, необходимую для ее выполнения. На каждый проект p выделено c_{pe}^{local} исполнителей специализации $e \in E$, а кроме того возможно привлечение дополнительных ресурсов, имеющихся в количестве c_e^{global} . Дополнительные ресурсы не привязаны к какому-либо из проектов и могут привлекаться для выполнения работ любого проекта. Предполагается, что исполнители не могут временно прерывать свое участие в работе. Допускается частичная занятость исполнителя в работе.

Примем, что длительность работы $j \in J_p$ зависит от числа ее исполнителей \tilde{l}_{pj} :

$$\tilde{d}_{pj} = \frac{\tilde{q}_{pj}}{\tilde{l}_{pj}} \quad \forall p \in P, j \in J_p,$$

где \tilde{d}_{pj} — нечеткая длительность работы $j \in J_p$; \tilde{q}_{pj} — нечеткие трудозатраты; \tilde{l}_{pj} — нечеткое число исполнителей, которое рассматривается как числовая интерпретация возможности исполнителя «работать за двоих» или возможности обратного эффекта, при этом фактическое число исполнителей будет соответствовать моде нечеткого числа. Нечеткие переменные \tilde{q}_{pj} , \tilde{l}_{pj} и \tilde{d}_{pj} в соответствии с [9] представлены парами функций вида $(q_{pj}^L(\alpha); q_{pj}^R(\alpha))$, $(l_{pj}^L(\alpha); l_{pj}^R(\alpha))$ и $(d_{pj}^L(\alpha); d_{pj}^R(\alpha))$, где $\alpha \in [0; 1]$ — параметр нечеткости. Такое представление подразумевает возможность использования как треугольных, так и трапецидальных нечетких чисел. Далее для сокращения записей будем использовать только треугольные нечеткие числа.

Пусть \tilde{s}_{pj} — нечеткое время начала работы $j \in J_p$, тогда время ее окончания определяется как

$$\tilde{f}_{pj} = \tilde{s}_{pj} + \tilde{d}_{pj} \quad \forall p \in P, j \in J_p.$$

Целью является нахождение для каждой задачи $j \in J_p$ каждого из проектов $p \in P$ такого времени старта \tilde{s}_{pj} и числа выделенных трудовых ресурсов \tilde{l}_{pj} , которые не нарушают ограничения предшествования и ресурсные ограничения, а также

обеспечивают минимальное взвешенное суммарное время выполнения проектов \tilde{f}_{TPD} :

$$\tilde{f}_{\text{TPD}} = \sum_{p \in P} \omega_p \left(\tilde{f}_p - \tilde{s}_p \right).$$

Здесь $\tilde{f}_p = \max_{j \in J_p} \left(\tilde{f}_{pj} \right)$ — время завершения последней работы проекта; ω_p — весовой коэффициент, определяющий приоритетность проекта.

Таким образом, может быть сформулирована следующая оптимизационная задача:

$$\tilde{f}_{\text{TPD}} \rightarrow \min \tag{1}$$

при условиях

$$\tilde{l}_{pj} \geq 1 \quad \forall p \in P, j \in J_p; \tag{2}$$

$$\sum_{p \in P} \sum_{j \in A_{pe}(t)} \max \left(0; \tilde{l}_{pj} - c_{pe}^{\text{local}} \right) \leq c_e^{\text{global}} \quad \forall e \in E, t \in [0, T]; \tag{3}$$

$$A_{pe}(t) = \left\{ i \in J_p \mid \tilde{s}_{pj} \leq t < \tilde{s}_{pj} + \tilde{d}_{pj}, \text{Spec}(j) = e \right\}; \tag{4}$$

$$\tilde{s}_{pj} + \tilde{d}_{pj} \leq \tilde{s}_{pj'} \quad \forall p \in P, j' \in J_p, j \in \text{Pre}(j'); \tag{5}$$

$$\tilde{s}_{pj} \geq \tilde{s}_p \quad \forall p \in P, j \in J_p. \tag{6}$$

Здесь T — это верхняя оценка времени завершения всех проектов из портфеля; $A_{pe}(t)$ — множество работ проекта p , требующих специализации e и находящихся в состоянии выполнения в момент времени t . Согласно (2) у каждой задачи должен быть хотя бы один исполнитель, (3) задает ограничение на число одновременно используемых трудовых ресурсов, (5) отражает ограничения предшествования, а (6) — ограничение на дату старта проекта.

3 Описание алгоритма решения задачи сетевого планирования

В случае применения известных арифметик нечетких чисел, основанных на принципе обобщения Л. Заде [5], в дополнение к проблемам, обозначенным во введении, решение поставленной задачи потребует задания операции сравнения нечетких треугольных чисел. Известные варианты реализации операции сравнения не обеспечивают однозначного результата [5, 8]. Использование же W -алгебры при решении задачи проектного планирования, как продемонстрировано в работе [10], позволяет избежать сравнения нечетких чисел, так как результат может быть представлен как линейная комбинация решений четких задач. Представим нечеткие трудозатраты в виде выпуклой линейной комбинации

граничных значений левой $q_{pj}^L(\alpha)$ и правой $q_{pj}^R(\alpha)$ функций W -числа при $\alpha = 0$ и 1 :

$$\left. \begin{aligned} q_{pj}^L(\alpha) &= q_{pj}^L(0) + (q_{pj}^L(1) - q_{pj}^L(0)) \alpha; \\ q_{pj}^R(\alpha) &= q_{pj}^R(0) + (q_{pj}^R(1) - q_{pj}^R(0)) \alpha. \end{aligned} \right\} \tag{7}$$

Тогда для треугольных чисел необходимо решить отдельно четкие задачи для трех детерминированных значений: $q_{pj}^L(0)$; $q_{pj}^L(1) = q_{pj}^R(1)$; $q_{pj}^R(0)$. Полученные решения могут быть использованы для представления результата в виде W -чисел (7). Например, нечеткое число исполнителей \tilde{l}_{pj} , выделенных на выполнение работы, будет иметь вид:

$$\begin{aligned} l_{pj}^L(\alpha) &= l_{pj}^L(0) + (l_{pj}^L(1) - l_{pj}^L(0)) \alpha; \\ l_{pj}^R(\alpha) &= l_{pj}^R(0) + (l_{pj}^R(1) - l_{pj}^R(0)) \alpha. \end{aligned}$$

Теперь алгоритм решения задачи (1)–(6) можно рассматривать как решение трех детерминированных задач с их последующей комбинацией вида (7).

Доказано, что задача проектного планирования в условиях ресурсных ограничений относится к классу NP-трудных [15], поэтому большинство исследований в данной области рассматривают возможности применения различных эвристических подходов для поиска субоптимальных решений [16].

В данной работе для решения поставленной оптимизационной задачи был реализован генетический алгоритм, сочетающий ряд подходов из работ [12–14]. Подробности реализации приводятся далее.

3.1 Представление решения

Возможное решение задачи кодируется хромосомой, которая соответствует ровно одному возможному расписанию работ и распределению трудовых ресурсов. Каждый ген хромосомы представлен идентификатором работы и числом выделенных на нее исполнителей. При этом порядок генов в хромосоме отражает приоритет соответствующих работ. Расписание работ может быть получено по данному представлению в результате следующей последовательной процедуры. Начиная с пустого расписания, будем получать новое промежуточное расписание путем размещения следующей по приоритету работы на самое раннее время так, чтобы не нарушить ограничения предшествования и ресурсные ограничения. Такая процедура называется последовательной схемой генерации расписания. Стоит отметить, что, хотя каждому закодированному таким образом решению соответствует ровно одно возможное расписание, обратное

не верно. Одинаковое расписание может быть получено из различных решений, вследствие чего при таком подходе повторный обход уже рассмотренных решений будет происходить чаще. Однако, так как любое получаемое решение всегда будет удовлетворять всем необходимым требованиям, это позволяет избежать модификации функции приспособленности внесением штрафа за нарушение ограничений. Поэтому в качестве функции приспособленности использовалась величина, обратная суммарному времени выполнения проектов (1). Кроме того, использование списка приоритетов решает проблему непрерывного времени, которая может возникать при непосредственном кодировании времен начала работ.

3.2 Начальная популяция

Для обеспечения разнообразия особей начальная популяция генерируется случайным образом. Процедура получения случайной хромосомы начинается с пустого списка генов. На каждом шаге процедуры в список добавляется ген, кодирующий случайную работу $j \in J_p$ из списка тех работ, чьи предшественники уже были добавлены в хромосому ранее. Число исполнителей, выделенных на эту работу, определяется случайным образом в промежутке $[1; c_{pe}^{local} + c_e^{global}]$ с учетом шага дискретизации 0,1, где $Spec(j) = e$.

Получаемые экземпляры очевидным образом удовлетворяют ограничениям предшествования. Теперь необходимо таким образом определить операции скрещивания и мутации, чтобы получаемые решения также не нарушали необходимые требования.

3.3 Метод селекции

На каждой итерации алгоритма необходимо отбирать особей, подлежащих дальнейшему скрещиванию с учетом их оценки приспособленности. В работе был использован турнирный метод отбора с размером подгруппы, равным 3.

3.4 Операция скрещивания

Операция скрещивания позволяет скомбинировать два родительских решения, чтобы получить

двух потомков, наследующих признаки родителей. Операция применяется независимо к спискам работ и соответствующим спискам выделенных ресурсов.

Сначала случайным образом определяется точка k , делящая родительские хромосомы на две части. Все работы из первой части первого (второго) родителя в том же порядке перемещаются в первого (второго) наследника, а оставшиеся работы добавляются в том порядке, в каком они расположены во втором (первом) родителе.

Распределения ресурсов скрещиваются согласно равномерной схеме. Для каждой i -й работы независимым образом берется случайное число $p_i \in [0, 1]$. Если $p_i \geq 0,5$, то первый (второй) ребенок наследует число исполнителей этой работы от первого (второго) родителя. В противном случае первый (второй) ребенок наследует число исполнителей от второго (первого) родителя.

3.5 Операция мутации

Операция мутации вносит случайные изменения в решения, получаемые после скрещивания. Операция мутации также применяется отдельно к приоритетам работ и выделенным ресурсам.

Для каждой i -й работы независимо берется случайное число $p_i \in [0, 1]$. Если $p_i \leq P_{PM}$, где P_{PM} — заранее заданная вероятность мутации, то i -я и $(i + 1)$ -я работы меняются местами, если это не нарушает ограничений предшествования.

Число исполнителей каждой работы с вероятностью P_{MM} меняется на случайное, взятое аналогичным с описанным в разделе о генерации начальной популяции образом.

4 Численный пример и обсуждение результатов

Решение задачи (1)–(6) иллюстрируется на следующем тестовом примере. Пусть портфель состоит из двух одинаково приоритетных проектов. Число выделенных на проекты трудовых ресурсов трех специализаций приведено в табл. 1. В табл. 2 указаны структурные и ресурсные требования для работ обоих проектов. Нечеткие переменные в табл. 2 и 3

Таблица 1 Трудовые ресурсы проектов

Ресурсы	Число работников		
	Специализации 1	Специализации 2	Специализации 3
Проекта 1	5	3	3
Проекта 2	2	4	3
Дополнительные	2	3	1

Таблица 2 Список работ проектов

Проект	Дата запуска проекта	Работа	Специализация	Предшественники	Нечеткие трудозатраты, \tilde{q}_{pj}
№ 1	0	1	2		(14; 16; 17)
		2	1		(18; 20; 21)
		3	3		(17; 19; 20)
		4	1	1, 2	(17; 17; 20)
		5	3	4	(24; 25; 27)
		6	2	5	(23; 25; 27)
		7	1	5	(31; 33; 35)
		8	3	3, 6, 7	(30; 31; 33)
№ 2	4	1	2		(8; 9; 11)
		2	3	1	(14; 16; 17)
		3	1		(21; 22; 24)
		4	2	3	(17; 19; 20)
		5	1	2, 4	(15; 16; 18)
		6	3	4	(40; 42; 42)
		7	3	5, 6	(13; 13; 13)
		8	1	2, 4	(11; 11; 13)

Таблица 3 Результаты решения задачи

Проект	Работа	Нечеткое время старта работы			Число исполнителей	Нечеткое время окончания работы		
		s_{pj}^L	s_{pj}^m	s_{pj}^R		f_{pj}^L	f_{pj}^m	f_{pj}^R
№ 1	1	0	0	0	3	4,67	5,33	4,72
	2	0	0	0	5,3	2,69	3,77	6,18
	3	0	0	2,69	1,9	7,05	10	10,53
	4	4,67	5,33	6,18	5	8,07	8,73	9,12
	5	8,07	8,73	9,12	3	15,57	17,07	16,83
	6	15,57	17,07	16,83	5,9	19,6	21,3	22,02
	7	15,57	17,07	16,83	6,8	20,26	21,92	22,48
	8	20,26	21,92	22,48	3,9	28,16	29,87	30,73
№ 2	1	4	4	4	3,8	6,29	6,37	5,8
	2	6,29	6,37	5,8	3,2	10,29	11,37	13,19
	3	4	4	4	4	9,25	9,5	10
	4	9,25	9,5	10	6,8	11,83	12,29	13,28
	5	11,83	12,29	13,28	2,3	18,97	19,25	24,53
	6	11,83	12,29	13,28	4	22,64	22,79	25,63
	7	22,64	22,79	25,63	2,8	26,83	27,44	30,11
	8	11,83	21,92	13,28	3,3	20,29	25,25	19,19

для удобства восприятия заданы тремя точками — на границах при $\alpha = 0$ и в моде при $\alpha = 1$.

При решении использовались следующие параметры генетического алгоритма: размер популяции — 150; число поколений — 40; вероятность мутации — 0,04.

Результаты решения задачи приведены в табл. 3. Полученные значения могут быть использованы для составления нечеткой диаграммы Ганта (см. рисунок), дающей наглядное представление о сетевом графике с учетом нечеткой неопределенности. Каждый четырехугольник на диаграмме определяет

расположение во времени соответствующей работы одного из проектов. Треугольниками обозначены нечеткие времена начала и конца работ.

При анализе результатов табл. 3 следует обратить внимание на случаи, когда обе компоненты нечеткого числа отклоняются от моды в одну сторону. В этом случае, как показано в [8], рассматривается только доминирующая компонента.

Полученное решение позволяет оценить время завершения не только каждой работы, но и проектов в целом. При этом дается оценка возможности как задержки, так и опережения планиру-

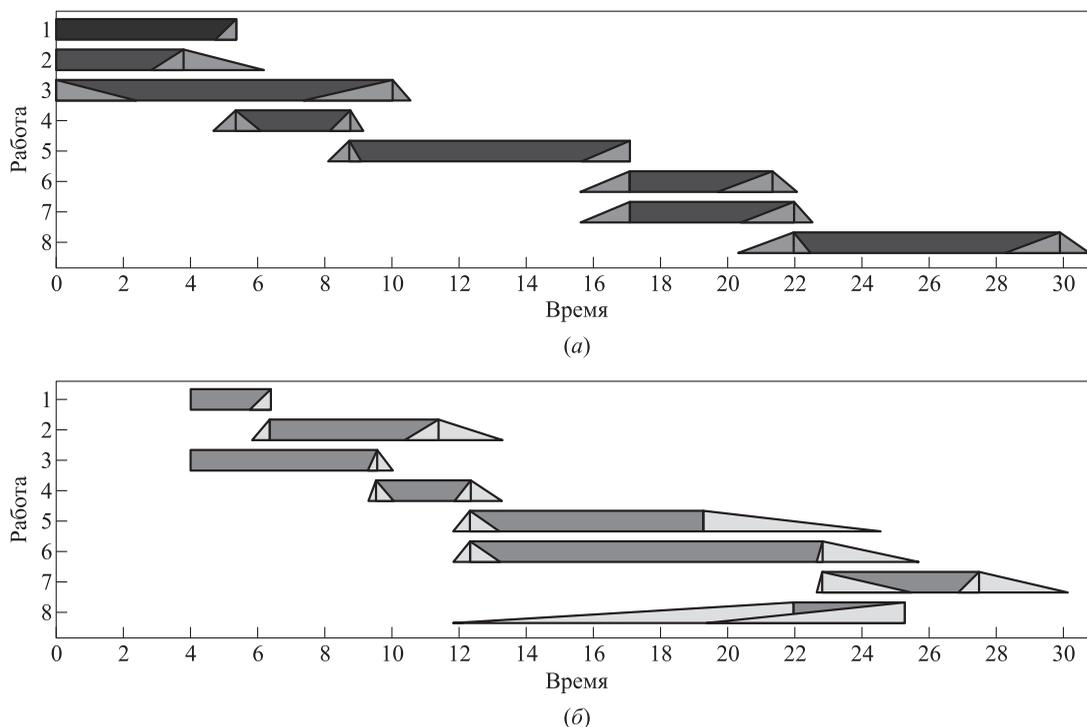


Диаграмма Ганта для полученного нечеткого решения для проектов 1 (а) и 2 (б)

емых сроков. Например, работа 1 первого проекта может быть выполнена с опережением почти на 0,7 временных единиц, а задержки не предвидятся; аналогичная ситуация со сроком выполнения проекта 2 — возможно опережение почти на 6 временных единиц. Время выполнения проекта 1 может задержаться на 0,9 временных единиц, но существует и возможность опережения срока на 1,7 временных единиц.

Кроме того, результаты решения позволяют оптимально с точки зрения занятости распределить исполнителей по работам с учетом их специализации. Дробное число исполнителей означает неполную занятость на конкретной работе.

5 Заключение

Представленные в статье методические подходы к сетевому планированию демонстрируют возможности теории нечетких множеств при описании неопределенностей, возникающих при планировании проектных работ. Такой подход может служить альтернативой, например, известному методу PERT (Project Evaluation and Review Technique), использующему β -распределение времени выполнения работ [4], отличаясь от него большими возможностями при учете неопределенностей. Так, в рассмотренном примере была учтена неопределенность эффективности работы исполнителей, не

предусматриваемая методом PERT. Разработанный подход может быть применен при ситуационном управлении одновременно проводящимися мероприятиями с ограниченным составом исполнителей.

Литература

1. Из выступления Президента России Владимира Путина на VI Международном форуме технологического развития «Технопром-2018» 27 августа 2018 г. <http://www.kremlin.ru/events/president/news/58391>.
2. О стратегии научно-технологического развития Российской Федерации: Указ Президента РФ от 01.12.2016 № 642. <http://static.kremlin.ru/media/events/files/ru/uZiATIOJiq5tZsJgqcZLY9YyL8PWTXQb.pdf>.
3. Зацаринный А. А., Киселев Э. В., Козлов С. В., Коллин К. К. Информационное пространство цифровой экономики России. Концептуальные основы и проблемы формирования. — М.: ФИЦ ИУ РАН, 2018. 236 с.
4. Taha Х. А. Введение в исследование операций / Пер. с англ. — М.: Вильямс, 2016. 912 с. (Taha H. A. Operations research: An introduction. — 8th ed. — Prentice Hall, 2006. 838 p.)
5. Piegat A. Fuzzy modeling and control. — Berlin–Heidelberg: Springer, 2001. 371 p.
6. Балашов В. Г., Заложнев А. Ю., Новиков Д. А. Механизмы управления организационными проектами. — М.: ИПУ РАН, 2003. 84 с.

7. *Matveev M. G.* Анализ и решение задач выбора с параметрической нечеткостью // Вестник Южно-Уральского государственного университета. Сер. Математическое моделирование и программирование, 2015. № 8(4). С. 14–29.
8. *Шевляков А. О., Матвеев М. Г.* Сравнение различных нечетких арифметик // Искусственный интеллект и принятие решений, 2017. № 4. С. 60–68.
9. *Шевляков А. О., Матвеев М. Г.* Алгебра трапециевидных чисел для обработки нечеткой информации // Искусственный интеллект и принятие решений, 2018. № 4. С. 53–60.
10. *Матвеев М. Г., Алейникова Н. А.* Математическое моделирование задачи сетевого планирования с помощью нечеткой математики // Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии, 2018. № 3. С. 155–162.
11. *Шевляков А. О., Матвеев М. Г.* Решение RCPSP при нечетких трудозатратах выполнения операций // Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии, 2015. № 4. С. 121–125.
12. *Blazewicz J., Lenstra J. K., Rinnooy Kan A. H. G.* Scheduling subject to resource constraints: Classification and complexity // *Discrete Appl. Math.*, 1983. Vol. 5. No. 1. P. 11–24.
13. *Hartmann S.* Project scheduling with multiple modes: A genetic algorithm // *Ann. Oper. Res.*, 2001. Vol. 102. No. 1-4. P. 111–135.
14. *Abdolshah M.* A Review of Resource-Constrained Project Scheduling Problems (RCPSP) approaches and solutions // *Int. T. J. Eng. Manage. Appl. Sci. Technol.*, 2014. Vol. 5. No. 4. P. 253–286.
15. *Kumanan S., Jose G. J., Raja K.* Multi-project scheduling using a heuristic and a genetic algorithm // *Int. J. Adv. Manuf. Tech.*, 2006. Vol. 31. No. 3-4. P. 360–366.
16. *Yannibelli V., Amandi A.* A knowledge-based evolutionary assistant to software development project scheduling // *Expert Syst. Appl.*, 2011. Vol. 38. No. 7. P. 8403–8413.

Поступила в редакцию 28.03.19

MODELING THE PROCESS OF NETWORK PLANNING OF A PORTFOLIO OF PROJECTS WITH HETEROGENEOUS RESOURCES UNDER FUZZINESS

A. A. Zatsarinny¹, V. V. Korotkov², and M. G. Matveev²

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Voronezh State University, 1 Universitetskaya Pl., Voronezh 394018, Russian Federation

Abstract: The paper discusses the problem of project portfolio scheduling subject to resource constraints and fuzzy activity durations. Employees of various specializations are considered as the only type of resource. Fuzzy W-algebra was used to avoid such difficulties and limitations of traditional fuzzy arithmetics as the support size extension, the need of fuzzy number comparison, and some others. The proposed model provides a fuzzy estimation of project execution times and optimal resource allocation. To solve the problem, the genetic algorithm based on activity list representation was implemented. The paper also provides a numerical example which demonstrates advantages of applying the approach to organizations’ workflow in the case of making estimates for decision making under uncertainty.

Keywords: project management; W-algebra; fuzzy arithmetic; combinatorial optimization; genetic algorithm

DOI: 10.14357/19922264190213

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (grant No. 17-01-00251).

References

1. Iz vystupleniya Prezidenta Rossii na VI Mezhdunarodnom forume tekhnologicheskogo razvitiya “Tekhnoprom-2018” [From the speech of the President of Russia at the 6th Forum (International) of Technological Development “Technoprom-2018”]. Available at: <http://www.kremlin.ru/events/president/news/58391> (accessed May 14, 2019).
2. O strategii nauchno-tekhnologicheskogo razvitiya Rossiyskoy Federatsii: ukaz Prezidenta ot 01.12.2016 No. 642

- [About strategy of scientific and technological development of the Russian Federation. Presidential Decree No. 642 dated 01.12.2016]. Available at: <http://static.kremlin.ru/media/events/files/ru/uZiATIOJiq5tZsJgqcZLY9YyL8PWTXQb.pdf> (accessed May 14, 2019).
3. Zatsarinny, A. A., E. V. Kiselev, S. V. Kozlov, and K. K. Kolin. 2018. *Informatsionnoe prostranstvo tsifrovoy ekonomiki Rossii. Kontseptual'nye osnovy i problemy formirovaniya* [Information space of the digital economy of Russia. Conceptual framework and problems of formation]. Moscow: FRC CSC RAS. 236 p.
 4. Taha, H. A. 2006. *Operations research: An introduction*. 8th ed. Prentice Hall. 838 p.
 5. Piegat, A. 2001. *Fuzzy modeling and control*. Berlin–Heidelberg: Springer. 371 p.
 6. Balashov, V. G., A. Yu. Zalozhnev, and D. A. Novikov. 2003. *Mekhanizmy upravleniya organizatsionnymi proektami* [Organizational project management mechanisms]. Moscow: ICS RAS. 84 p.
 7. Matveev, M. G. 2015. Analiz i reshenie zadach vybora s parametricheskoy nechetkost'yu [Analyzing and solving problems of decision making with parametric fuzzy]. *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software* 8(4):14–29.
 8. Shevlyakov, A. O., and M. G. Matveev. 2017. Sravnenie razlichnykh nechetkikh arifmetik [A comparison of different fuzzy arithmetics]. *Artificial Intelligence Decision Making* 4:60–68.
 9. Shevlyakov, A. O., and M. G. Matveev. 2018. Algebra trapetsievidnykh chisel dlya obrabotki nechetkoy informatsii [Algebra of trapezoidal fuzzy numbers for fuzzy information processing]. *Artificial Intelligence Decision Making* 4:53–60.
 10. Matveev, M. G., and N. A. Aleynikova. 2018. Matematicheskoe modelirovanie zadachi setevogo planirovaniya s pomoshch'yu nechetkoy matematiki [Mathematical modeling of the problem of network planning by means of fuzzy mathematics]. *Proceedings of Voronezh State University. Ser. Systems Analysis and Information Technologies* 3:155–162.
 11. Shevlyakov, A. O., and M. G. Matveev. 2015. Reshenie RCPSP pri nechetkikh trudozatratakh vypolneniya operatsiy [Solving RCPSP with uncertain duration times of activities]. *Proceedings of Voronezh State University. Ser. Systems Analysis and Information Technologies* 4:121–125.
 12. Blazewicz, J., J. K. Lenstra, and A. H. G. Rinnooy Kan. 1983. Scheduling subject to resource constraints: Classification and complexity. *Discrete Appl. Math.* 5(1):11–24.
 13. Hartmann, S. 2001. Project scheduling with multiple modes: A genetic algorithm. *Ann. Oper. Res.* 102(1-4): 111–135.
 14. Abdolshah, M. 2014. A Review of Resource-Constrained Project Scheduling Problems (RCPSP) approaches and solutions. *Int. T. J. Eng. Manage. Appl. Sci. Technol.* 5(4):253–286.
 15. Kumanan, S., G. J. Jose, and K. Raja. 2006. Multi-project scheduling using a heuristic and a genetic algorithm. *Int. J. Adv. Manuf. Tech.* 31(3-4):360–366
 16. Yannibelli, V., and A. Amandi. 2011. A knowledge-based evolutionary assistant to software development project scheduling. *Expert Syst. Appl.* 38(7):8403–8413.

Received March 28, 2019

Contributors

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.ru

Korotkov Vladislav V. (b. 1993) — PhD student, Department of Information Technologies of Management, Computer Science Faculty, Voronezh State University, 1 Universitetskaya Pl., Voronezh 394018, Russian Federation; chasecrunk@gmail.com

Matveev Mikhail G. (b. 1949) — Doctor of Science in technology, professor, Head of Department of Information Technologies of Management, Computer Science Faculty, Voronezh State University, 1 Universitetskaya Pl., Voronezh 394018, Russian Federation; mgmatveev@yandex.ru

О ГЕНЕЗИСЕ ИНФОРМАЦИОННОГО ОБЩЕСТВА: ИНФОРМАТИКО-КИБЕРНЕТИЧЕСКОЕ МОДЕЛЬНОЕ ПРЕДСТАВЛЕНИЕ

С. Н. Гринченко¹

Аннотация: Вводится понятие «генезис информационного общества», которое рассматривается с позиций информатико-кибернетического моделирования (ИКМ) процесса развития Человечества как самоуправляющейся иерархо-сетевой системы. На этой основе получены количественные оценки его типовых пространственно-временных характеристик, представляющих собой геометрические прогрессии со знаменателем « e в степени e » (15,15426...), а также скоординированных с ними во времени и в пространстве психико-антропологических, образовательных и информационно-коммуникационных параметров и возможностей включенного в этот процесс усложняющегося человека и его сообществ различной величины. Это позволило раздвинуть рамки существования информационного общества на всю историческую и даже археологическую эпоху такого развития. Результирующая последовательность информационных технологий (ИТ) «сигнальные позы/звуки/движения – мимика/жесты – речь/язык – письменность – тиражирование текстов – компьютеры – телекоммуникации – информационная нанотехнология – ...» позволяет рассматривать генезис информационного общества в широком контексте единой исторической ретроспективы и перспективы.

Ключевые слова: информационное общество; информационные технологии; информатико-кибернетическая модель; самоуправляющаяся иерархо-сетевая система Человечества; археологическая эпоха

DOI: 10.14357/19922264190214

В литературе, даже энциклопедической, распространена трактовка «информационного общества» как общества «современного типа», в котором общение людей опирается на компьютерные и телекоммуникационные ИТ². Такая трактовка этого понятия создает иллюзию отстраненности информационного общества от его собственного исторического прошлого, когда вышперечисленных ИТ еще не изобрели, но люди в составе сообществ как-то общались между собой, используя иные ИТ.

Поскольку от этой иллюзии недалеко до недооценки полезности соответствующего исторического опыта для современности, попытаюсь развезать ее.

Результаты ИКМ процесса развития на Земле Человечества как самоуправляющейся иерархо-сетевой³ системы [2–14] (рис. 1) позволяют раздвинуть рамки существования информационного общества на всю историческую и даже археологическую эпоху такого развития, что дает возможность выделить ту эволюционную линию этого процесса, которую логично определить как *генезис информационного общества*.

На рис. 1 используются следующие обозначения:

- восходящие стрелки (имеющие структуру «многие – к одному») отражают первую из 5 основных составляющих контура поисковой оптими-

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sgrinchenko@ipiran.ru

² В [1] дано следующее определение: «**Информационное общество**, одно из понятий, используемых в социологич. теории для обозначения обществ. систем «современного типа»... Важнейшие характеристики И. о.: 1) лавинообразное распространение информац. технологий (прежде всего компьютерных и телекоммуникационных); 2) превращение информации в важнейший социальный ресурс, необходимую предпосылку управленч. деятельности, развития экономики, образования, сферы услуг, домашнего быта, рекреационной сферы и т.д.; по некоторым данным, в наиболее развитых странах проф. деятельность более половины занятых связана исключительно с производством и обработкой информации; 3) наделение СМИ статусом «четвертой ветви власти»; 4) расширение границ и «репертуара» массовой культуры; 5) увеличение каналов вертикальной и горизонтальной мобильности; 6) изменение представлений о социальном пространстве («глобализация» пространства, мгновенная доступность даже периферийных его сегментов) и времени (расширение рамок «современности», когда даже отдаленные историч. события воспринимаются как происходящие «здесь» и «сейчас»); 7) возникновение в процессе коммуникации особой виртуальной реальности, несводимой к результатам технич. визуализации и выходящей за пределы воображения и памяти индивида; 8) превращение информац. технологий в базу для развития высоких технологий (Hi-Tech)».

³ «**Иерархо-сетевая**» структура — иерархическая структура типа «матрешки», но с существенно большим единицы числом вложений на каждом ее иерархическом уровне, которые и образуют соответствующие сетевые структуры.

- зации — *поисковую активность* представителей соответствующих ярусов в иерархии;
- нисходящие сплошные (имеющие структуру «один — ко многим») стрелки отражают вторую составляющую — *целевые критерии* поисковой оптимизации энергетики системы Человечества;
 - нисходящие пунктирные («один — ко многим») стрелки отражают третью составляющую — *оптимизационную системную память* личностно-производственно-социального (результат адаптивных влияний представителей вышележащих иерархических ярусов на структуру вложенных в них нижележащих);
 - полужирными стрелками в левой части схемы условно показана четвертая составляющая — *антропогенная активность* индивидов и их групп, трактуемая как «трудовая деятельность по созданию соответствующего инструментария и результатов его применения»;
 - пунктирными полужирными стрелками в правой части схемы условно показана пятая составляющая — *антропогенная системная память* личностно-производственно-социального (процессы вовлечения результатов антропогенной активности в структуру соответствующей иерархо-сетевой подсистемы Человечества).

Рассмотрим этот феномен поэтапно, сведя в общую таблицу расчетные данные о различных его проявлениях.

Промежутки времени между возникновением новых иерархо-сетевых подсистем Человечества (а следовательно, и между стартами новых ИТ) подчиняются, согласно ИКМ, простой математической закономерности: каждый из них в $e^e = 15,15426 \dots$ раз короче предыдущего¹ (третий столбец таблицы). В свою очередь, этой же закономерности подчиняются и размеры ареалов (радиусы кругов той же площади) устойчивых и эффективно самоуправляющихся сообществ человека как базисного элемента системы Человечества, и точности доступных усложняющемуся человеку — в конкретный момент исторического времени — антропогенных воздействий и/или производственных технологий (второй столбец таблицы) (рис. 2).

Эмпирические оценки этих времен и пространств, сделанные и опубликованные палеоантропологами, археологами и историками, — когда

они имеются! — не противоречат модельным результатам [13]. Диапазоны примерного возраста «образовываемых», приведенные в седьмом столбце таблицы, рассчитаны, исходя из «золотого сечения» (соотношения смежных членов числового ряда, равного 1,618... при увеличении ряда, либо 0,618... при его уменьшении, адекватность использования которого при выработке количественных оценок в самых различных областях знания хорошо известна²), опирающегося на ориентировочную оценку завершения человеком среднего образования к 18 годам (на сегодня).

Базируясь на ИКМ, в качестве нулевого этапа развития будущего информационного общества, как представляется, можно рассматривать процесс *цефализации* позвоночных, т. е. возникновения и усложнения у них головного мозга как основного носителя механизмов запоминания и считывания информации о результатах их адаптивного и социального поведения, начавшейся около 428 млн лет назад с кульминацией около 140,1 млн лет назад (шестой столбец таблицы) на «территории» порядка 4,2 м — т. е. в пределах отдельного многоклеточного организма.

Далее в качестве первого этапа такого развития будем рассматривать начавшуюся около 28,23 млн лет назад, с кульминацией около 9,26 млн лет назад, на территориях порядка 64 м, ИТ сигнальных поз/движений и неинтонированных звуков (типа рычания, ворчания, писка и т. п.), характерную для стадных/стайных животных, в том числе пред-пред-людей *Nominoidea* (четвертый столбец таблицы), способных обеспечивать точность своих воздействий на природу порядка 28 см. Субстрат их психики относится к иерархическому уровню органов многоклеточного организма (пятый столбец), а уровень образованности соответствует современному младенцу возрастом около 1–1,6 лет (седьмой столбец).

Следующий, второй этап развития ИТ — мимики/жестов, начавшийся около 1,86 млн лет назад, с кульминацией около 612 тыс. лет назад, на территориях порядка 1 км, реализовался далекими предками современного человека — пред-людьми *Homo ergaster/Homo erectus*, способными обеспечивать точность своих воздействий на природу порядка 1,8 см, с субстратом психики уровня тканей многоклеточного организма и уровнем образованности, соответствующим современному ребенку 1,6–2,6 лет.

¹Эту геометрическую прогрессию — как модель критических уровней развития биологических систем — выявили А. В. Жирмунский и В. И. Кузьмин [16].

²Применительно к периодизации истории Человечества в археологическую эпоху это продемонстрировано Ю. Л. Шаповой [17–19], согласование подхода к такой периодизации на основе золотого сечения и предлагаемого информатико-кибернетического подхода подробно показано в [7, 10, 12–14, 20].

Таблица 1 Свод основных характеристик генезиса информационного общества (как проявления развития самоуправляющейся и метаэволюционирующей, т. е. наращивающей число своих иерархических уровней/ярусов, системы Человечества) от прошлого до модельно прогнозируемого будущего

1	2	3	4	5	6	7
0	~ 4,2 м	~ 428 млн лет назад; ~ 140,1 млн лет назад	Цефализация позвоночных	Многочлечный организм в целом	Формирование головного мозга как основы для реализации будущих ИТ	— ~ 0,6–1,0 год
1	~ 64 м; ~ 28 см	~ 28,23 млн лет назад; ~ 9,26 млн лет назад	Пред-пред-люди Hominoidea	Органы многоклеточного организма (его нервной системы в целом)	Сигнальные позы/ движения и неинтонированные звуки (типа рычания, ворчания, писка и т. п.)	Выработка (младенцами) сигнальных поз; ~ 1,0–1,6 лет
2	~ 1 км; ~ 1,8 см	~ 1,86 млн лет назад; ~ 612 тыс. лет назад	Пред-люди Homo ergaster / Homo erectus	Ткани многоклеточного организма (сетей/ансамблей нейронов и др.)	Мимика/жесты и интонированные звуки	Овладение (ребенком) мимикой/ жестами, начальное понимание речи; ~ 1,6–2,6 лет
3	~ 15 км; ~ 1,2 мм	~ 123 тыс. лет назад; ~ 40 тыс. лет назад	Homo sapiens'	Эвкариотические клетки многоклеточного организма (отдельные нервные и глиальные клетки и др.)	Речь/язык (артикулированная устная речь)	Овладение (детьми) речью/языком (протообразование); ~ 2,6–4,2 лет
4	~ 222 км; ~ 80 мкм	~ 8,1 тыс. лет назад; ~ 2,7 тыс. лет назад	Homo sapiens''	Компартменты эвкариотической клетки (отдельные рецепторные, или постсинаптические, зоны нейронов и т. п.)	Письменность	Овладение чтением/ письмом (дошкольное образование); ~ 4,2–6,9 лет
5	~ 3370 км; ~ 5 мкм	~ 1446 г.; ~ 1806 г.	Homo sapiens'''	Субкомпартменты эвкариотической клетки	Тиражирование текстов, или книгопечатание	Начальное образование; ~ 6,9–11,1 лет
6	~ 51 тыс. км (обшепланетарный); ~ 0,35 мкм	~ 1946 г.; ~ 1970 г.	Homo sapiens''''	Ультраструктурные (прокариотические) внутриклеточные элементы эвкариотической клетки (типа клеточного ядра, деталей эндоплазматической сети и т. п. образований)	Компьютерная ИТ	Среднее образование; ~ 11,1–18 лет
7	~ 773 тыс. км (ближний космос); ~ 23 нм	~ 1979 г.; ~ 2003 г.	Homo sapiens'''''	Макромолекулы/гены (компартменты ультраструктурных– прокариотических– внутриклеточных элементов)	Телекоммуникационная ИТ	Высшее образование + «аспирантура»; ~ 18–29,1 лет
8	~ 11,7 млн км (промежуточный космос); ~ 1,5 нм	~ 1981 г.; ~ 2341 г. (?)	Homo sapiens''''''	Органические молекулы (субкомпартменты ультраструктурных– прокариотических– внутриклеточных элементов)	Нано-ИТ (возможно, «наноаппаратно поддерживаемая селективная телепатия» [15])	«Докторантура»; ~ 29,1–47,1 лет
9

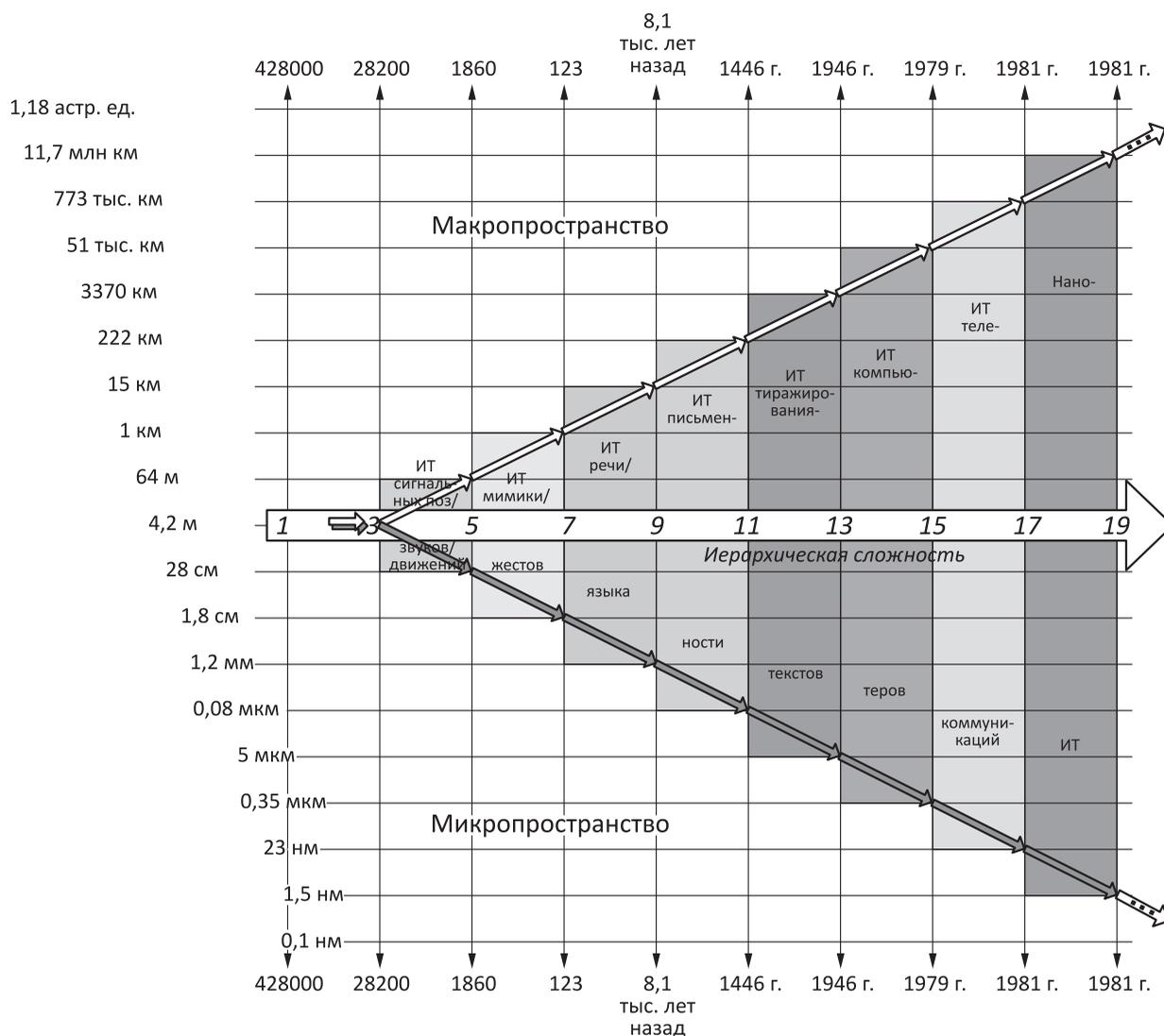


Рис. 2 Пространственно-временные характеристики и тренд ИТ в процессе генезиса информационного общества (по ИКМ, в двойном логарифмическом масштабе; иерархическая сложность — число уровней/ярусов в системной иерархии)

Все последующие этапы развития ИТ — речь/язык, письменность, тиражирование текстов (книгопечатание), компьютеры, телекоммуникации, нано-ИТ — реализовались последовательно усложняющимися формами Homo sapiens, который при этом образовывал относительно устойчивые и относительно эффективно функционирующие и самоуправляющиеся сообщества на все больших ареалах, одновременно повышая точность своих (антропогенных) действий при формировании вокруг себя «второй (рукотворной) природы».

Так, третий этап развития ИТ — речи/языка, начавшийся около 123 тыс. лет назад, с кульминацией (верхнепалеолитической революцией) около 40 тыс. лет назад, на территориях порядка 15 км,

реализовался Homo sapiens', способными обеспечивать точность своих производственных технологий порядка 1,2 мм, с субстратом психики уровня эвкариотических клеток многоклеточного организма и уровнем образованности, соответствующим современному ребенку 2,6–4,2 лет.

Четвертый этап развития ИТ — письменности, начавшийся около 8,1 тыс. лет назад, с кульминацией (городской революцией «осевого времени») около 2,7 тыс. лет назад, на территориях порядка 222 км, реализовался Homo sapiens'', способными обеспечивать точность своих производственных технологий порядка 80 мкм, с субстратом психики уровня компартментов эвкариотических клеток многоклеточного организма и уровнем образованности, со-

ответствующим современному ребенку 4,2–6,9 лет (дошкольное образование).

Пятый этап развития ИТ — тиражирования текстов (книгопечатания), начавшийся около 1446 г. н.э., с кульминацией (промышленной революцией) около 1806 г., на территориях порядка 3370 км, реализовался Homo sapiens^{IV}, способными обеспечивать точность своих производственных технологий порядка 5 мкм, с субстратом психики уровня субкомпарментов эвкариотических клеток многоклеточного организма и уровнем образованности, соответствующим современному ребенку 6,9–11,1 лет (начальное образование).

Шестой этап развития ИТ — компьютеров (локальных), начавшийся около 1946 г., с кульминацией (изобретением микропроцессоров) около 1970 г., на территориях порядка 51 тыс. км (т.е. общепланетарного, или глобального размера), реализовался Homo sapiens^V, способными обеспечивать точность своих производственных технологий порядка 0,35 мкм, с субстратом психики уровня ультраструктурных (прокариотических) внутриклеточных элементов эвкариотической клетки и уровнем образованности, соответствующим современному подростку-юноше/девушке 11,1–18 лет (среднее образование).

Седьмой этап развития ИТ — телекоммуникаций, начавшийся около 1979 г., с кульминацией (пиком скорости распространения на планете мобильной телефонии, интернета и т.п.) около 2003 г., в космическом объеме радиусом (шара) порядка 773 тыс. км (т.е. в ближнем космосе), реализовался Homo sapiens^{VI}, способными обеспечивать точность своих производственных технологий порядка 23 нм, с субстратом психики уровня макромолекул/генов (компарментов ультраструктурных—прокариотических—внутриклеточных элементов эвкариотической клетки) и уровнем образованности, соответствующим современному молодому человеку 18–29,1 лет (высшее образование + «аспирантура, с защитой диссертации кандидата наук»).

Восьмой этап развития перспективной нано-ИТ (возможно, «ИТ наноаппаратно поддерживаемой селективной телепатии» [15]), начавшийся около 1981 г., с кульминацией (пиком скорости ее распространения на планете) около 2341 г. (расчетная дата), в космическом объеме радиусом шара порядка 11,7 млн км (т.е. в промежуточном космосе [5]), реализовался Homo sapiens^{VII}, способными обеспечивать точность своих производственных технологий порядка 1,5 нм (отсюда наименование ИТ),

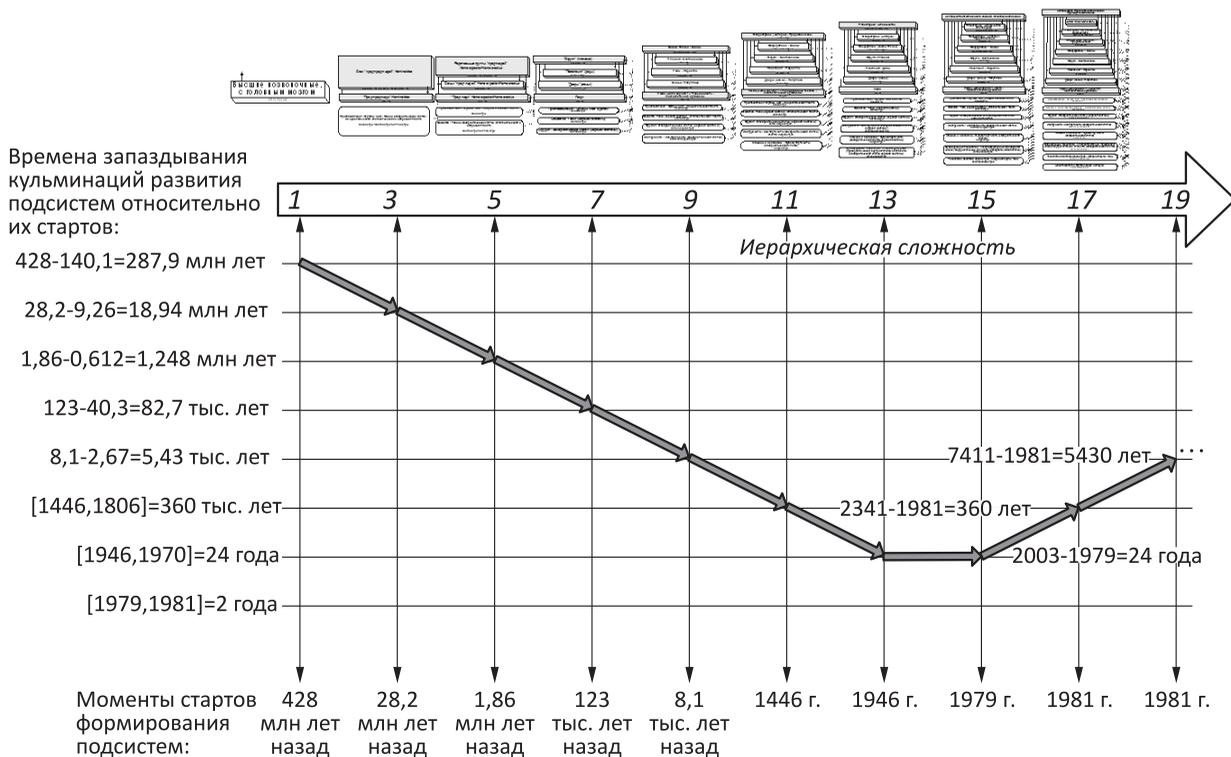


Рис. 3 Тренд изменения времен запаздывания кульминаций развития подсистем иерархо-сетевой системы Человечества относительно их стартов (по ИКМ, в двойном логарифмическом масштабе)

с субстратом психики уровня органических молекул (субкомпартов ультраструктурных—прокариотических—внутриклеточных элементов эукариотической клетки) и уровнем образованности, соответствующим современному зрелому человеку 29,1—47,1 лет («докторантура»).

Важно отметить, что процесс появления всех вышеперечисленных подсистем подчиняется кумулятивному принципу: возникновение каждой новой подсистемы не отменяет существование предыдущей: они все активно взаимодействуют между собой, коэволюционируют и т. п., но исторически более ранние, естественно, постепенно переходят на второй, третий и т. д. планы исторической сцены.

Точка сходимости этого ряда находится около 1981 г., знаменуя собой завершение этапа «детства—отрочества—юности» Человечества как целого и начало этапа его «зрелости» — достижения его максимальной иерархической сложности (числа уровней/ярусов в системной иерархии) [5, 8].

С позиции прогнозирования генезиса информационного общества на будущие времена отмечу, что, согласно ИКМ, тренд изменения времен запаздывания кульминаций развития подсистем относительно их стартов сменился прямо на наших глазах. Если во временном диапазоне с 428 млн лет назад и до 1946 г. он состоял в равномерном (в логарифмическом масштабе) укорочении согласно той же закономерности (в $e = 15, 15426 \dots$ раз), то в диапазоне от 1946 по 1979 гг. это время запаздывания не изменилось, а начиная с 1979 г. начало удлиняться (рис. 3).

Таким образом, метаэволюция системы Человечества завершилась около 1981 г. в том смысле, что все возможные ее иерархо-сетевые подсистемы в *потенции* уже созданы. Но их *актуализация*, дальнейшее усложнение, эволюция и коэволюция с ранее возникшими аналогичными подсистемами будет продолжаться неопределенно длительное время.

Выводы

1. Изучение *генезиса информационного общества* во всех его последовательных формах — от древности до современности и далее — на базе информатико-кибернетического модельного подхода и формализации процесса метаэволюционного развития в соответствующих терминах, позволило получить количественные оценки его типовых пространственно-временных характеристик, а также скоординированных с ними во времени и в пространстве психико-антропологических, образовательных

и информационно-коммуникационных параметров и возможностей включенного в этот процесс усложняющегося человека и его обществ различной величины.

2. Позиционирование ИТ локальных компьютеров и ИТ телекоммуникаций в качестве неотъемлемых составляющих совокупности монотонно усложняющихся в ходе цивилизационного развития — и информационного общества! — ИТ позволяет рассматривать их появление и функционирование в широком контексте единой исторической ретроспективы и перспективы, давая возможность делать не только теоретические, но и практические выводы.

Литература

1. Мелик-Гайгазян И. В. Информационное общество // Большая российская энциклопедия. Т. 11. — М.: Большая Российская энциклопедия, 2008. С. 490.
2. Гринченко С. Н. Социальная метаэволюция Человечества как последовательность шагов формирования механизмов его системной памяти // Исследовано в России: Электронный журнал, 2001. Т. 145. С. 1652—1681. <https://cyberleninka.ru/article/v/sotsialnaya-metaevolyutsiya-chelovechestva-kak-posledovatelnost-shagov-formirovaniya-mehanizmov-ego-sistemnoy-pamyati>.
3. Гринченко С. Н. Системная память живого (как основа его метаэволюции и периодической структуры). — М.: ИПИ РАН, Мир, 2004. 512 с.
4. Grinchenko S. N. Meta-evolution of nature system — the framework of history // Social Evolution History, 2006. Vol. 5. No. 1. P. 42—88.
5. Гринченко С. Н. Метаэволюция (систем неживой, живой и социально-технологической природы). — М.: ИПИ РАН, 2007. 456 с.
6. Гринченко С. Н. Homo eruditus (человек образованный) как элемент системы Человечества // Открытое образование, 2009. № 2. С. 48—55.
7. Гринченко С. Н., Шапова Ю. Л. История Человечества: модели периодизации // Вестник РАН, 2010. № 12. С. 1076—1084.
8. Grinchenko S. N. The pre- and post-history of Humankind: What is it? // Problems of contemporary world futurology. — Newcastle-upon-Tyne: Cambridge Scholars Publishing, 2011. P. 341—353.
9. Гринченко С. Н. Об эволюции психики как иерархической системы (кибернетическое представление) // Историческая психология и социология истории, 2012. Т. 5. № 2. С. 60—76.
10. Гринченко С. Н., Шапова Ю. Л. Информационные технологии в истории Человечества. — М.: Новые технологии, 2013. 32 с. (Приложение к журналу «Информационные технологии», 2013. № 8.)

11. *Гринченко С. Н.* Эволюция темпов жизни людей и развитие человечества // *Человек*, 2014. № 5. С. 28–36.
12. *Grinchenko S. N., Shchapova Y. L.* Archaeological epoch as the succession of generations of evolutive subject-carrier archaeological sub-epoch // *Philosophy of Nature in Cross-Cultural Dimensions: The Result of the International Symposium at the University of Vienna / Komparative Philosophie und Interdisziplinäre Bildung (KoPhil)*. Band 5. — Hamburg: Verlag Dr. Kovač, 2017. P. 478–499.
13. *Шапова Ю. Л., Гринченко С. Н.* Введение в теорию археологической эпохи: числовое моделирование и логарифмические шкалы пространственно-временных координат. — М.: Истфак МГУ, ФИЦ ИУ РАН, 2017. 236 с.
14. *Grinchenko S. N., Shchapova Yu. L.* Communications: Model representations about historical retrospective and possible perspective // *Communications Media Design Electronic J.*, 2018. Vol. 3. No. 2. P. 65–78.
15. *Гринченко С. Н.* Послесловие // Мат-лы доклада на Совместном научном семинаре ИПИ РАН и ИНИОН РАН «Методологические проблемы науки об информации». — М., 2012. С. 5–8. http://legacy.inion.ru/files/File/MPNI_9_13_12_12_posl.pdf.
16. *Жирмунский А. В., Кузьмин В. И.* Критические уровни в процессах развития биологических систем. — М.: Наука, 1982. 179 с.
17. *Шапова Ю. Л.* Хронология и периодизации древнейшей истории как числовая последовательность (ряд Фибоначчи) // *Информационный бюллетень Ассоциации «История и компьютер»*, 2000. № 25.
18. *Шапова Ю. Л.* Археологическая эпоха: хронология, периодизация, теория, модель. — М.: КомКнига, 2005. 192 с.
19. *Шапова Ю. Л.* Материальное производство в археологическую эпоху. — СПб.: Алетейя, 2011. 244 с.
20. *Гринченко С. Н., Шапова Ю. Л.* Пространство и время в археологии. Часть 3. О метрике базисной пространственной структуры человечества в археологическую эпоху // *Пространство и время*, 2014. № 1(15). С. 78–89.

Поступила в редакцию 17.10.18

ON THE GENESIS OF THE INFORMATION SOCIETY: INFORMATICS-CYBERNETIC MODEL REPRESENTATION

S. N. Grinchenko

Institute of Informatics Problems of the Federal Research Center “Informatics and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The concept of the information society genesis is introduced, which is viewed from the standpoint of informatics-cybernetic modeling of the development of Humankind as a self-controlling hierarchical-networking system. On this basis, the author obtained quantitative assessments of its typical spatial-temporal characteristics, representing geometric progressions with the denominator “ e to the degree e ” (15.15426...), as well as coordinated with them in time and space of the psychoanthropological, educational, and informational communication parameters and possibilities of a person who becomes complicated in this process and his communities of various sizes. This allowed us to push the framework of the information society for the entire historical and even archaeological epoch of such development. The resulting sequence of information technologies “signal poses/sounds/movements – mimics/gestures – speech/language – writing – replicating texts – computers – telecommunications – information nanotechnology – ...” allows us to consider the information society genesis in the broad context of a unified historical retrospective and perspective.

Keywords: information society; information technologies; informatics-cybernetic model; self-controlling hierarchical-networking system of Humankind; archaeological epoch

DOI: 10.14357/19922264190214

References

1. Melik-Gaygazyan, I. V. 2008. Informatsionnoe obshchestvo [Information society]. *Bol'shaya russiyanskaya entsiklopediya* [Great Russian Encyclopedia]. Moscow: Great Russian Encyclopedia Pubs. 11:490.
2. Grinchenko, S. N. 2001. Sotsial'naya metaevolyutsiya Chelovechestva kak posledovatel'nost' shagov formirovaniya mekhanizmov ego sistemnoy pamyati [Social meta-evolution of Mankind as a sequence of steps for the formation of the mechanisms of its system memory]. *Elektronnyy zhurnal «Issledovano v Rossii»* [Electronical J. “Investigated in Russia”]. 145:1652–1681. Available at: <https://cyberleninka.ru/article/v/sotsialnaya-metaevolyutsiya-chelovechestva-kak-posledovatelnost-shagov-formirovaniya-mekhanizmov-ego-sistemnoy-pamyati> (accessed October 5, 2018).
3. Grinchenko, S. N. 2004. *Sistemnaya pamyat' zhivogo (kak osnova ego metaevolyutsii i periodicheskoy struktury)* [Sys-

- tem memory of the life (as the basis of its meta-evolution and periodic structure)]. Moscow: IPIRAN, MIR. 512 p.
4. Grinchenko, S. N. 2006. Meta-evolution of nature system — the framework of history. *Social Evolution History* 5(1):42–88.
 5. Grinchenko, S. N. 2007. *Metaevolyutsiya (sistem nezhiyov, zhiyov i sotsial'no-tekhnologicheskoy prirody)* [Meta-evolution (of inanimate, animate, and socio-technological nature systems)]. Moscow: IPIRAN. 456 p.
 6. Grinchenko, S. N. 2009. Homo eruditus (chelovek obrazovanny) kak element sistemy Chelovechestva [Homo eruditus (educated human) as an element of the Humankind's system]. *Otkrytoe obrazovanie* [Open Education] 2:48–55.
 7. Grinchenko, S. N., and Yu. I. Shchapova. 2010. Human history periodization models. *Her. Russ. Acad. Sci.* 80(6):498–506.
 8. Grinchenko, S. N. 2011. The pre- and post-history of Humankind: What is it? *Problems of contemporary world futurology*. Newcastle-upon-Tyne: Cambridge Scholars Publishing. 341–353.
 9. Grinchenko, S. N. 2012. Ob evolyutsii psikhiki kak ierarkhicheskoy sistemy (kiberneticheskoe predstavlenie) [On the evolution of mind as a hierarchical system (a cybernetic approach)]. *Istoricheskaya psikhologiya i sotsiologiya istorii* [Historical Psychology & Sociology of History] 6(2): 60–77.
 10. Grinchenko, S. N., and Y. I. Shchapova. 2013. *Informatsionnye tekhnologii v istorii Chelovechestva* [Information technology in the history of Humankind]. Moscow: Novye tekhnologii. 32 p. (Prilozhenie k zhurnalu «Informatsionnye tekhnologii» [Supplement to J. Information Technology] 8.
 11. Grinchenko, S. N. 2014. Evolyutsiya tempov zhizni lyudey i razvitie chelovechestva [The evolution of the pace of human life and human development]. *Human Being* 5:28–36.
 12. Grinchenko, S. N., and Y. I. Shchapova. 2017. Archaeological epoch as the succession of generations of evolutive subject-carrier archaeological sub-epoch. *Philosophy of Nature in Cross-Cultural Dimensions: The Result of the International Symposium at the University of Vienna / Komparative Philosophie und Interdisziplinäre Bildung (KoPhil), Band 5*. Hamburg: Verlag Dr. Kovač. 478–499.
 13. Shchapova, Y. L., and S. N. Grinchenko. 2017. *Vvedenie v teoriyu arkheologicheskoy epokhi: chislovoe modelirovanie i logarifmicheskie shkaly prostranstvenno-vremennykh koordinat* [Introduction to the theory of the archaeological epoch: Numerical modeling and logarithmic scales of space–time coordinates]. Moscow: Faculty of History MSU, FRC CSC RAS]. 236 p.
 14. Grinchenko, S. N., and Y. I. Shchapova. 2018. Communications: Model representations about historical retrospective and possible perspective. *Communications Media Design Electronic J.* 3(2):65–78. Available at: <https://elibrary.ru/item.asp?id=36272286> (accessed October 5, 2018).
 15. Grinchenko, S. N. 2012. Posleslovie [Afterword]. *Mat-ly doklada na Sovmestnom nauchnom seminare IPI INION RAN “Metodologicheskie problemy nauk ob informatsii”* [Report materials at the Joint Scientific Seminar of the Institute of Informatics Problems of the Russian Academy of Sciences and the Institute of Scientific Information on Social Sciences of the Russian Academy of Sciences “Methodological problems of information sciences”]. Moscow. 5–8. Available at: http://legacy.inion.ru/files/File/MPNI_9_13_12_12_posl.pdf (accessed October 5, 2018).
 16. Zhirmunskiy, A. V., and V. I. Kuz'min. 1982. *Kriticheskie urovni v protsessakh razvitiya biologicheskikh sistem* [Critical levels in the development of biological systems]. Moscow: Nauka. 179 p.
 17. Shchapova, Y. L. 2000. Khronologiya i periodizatsii drevneyshey istorii kak chislovaya posledovatel'nost' (ryad Fibonachchi) [Chronology and periodization of ancient history as a numerical sequence (Fibonacci's series)]. *Informatsionnyy byulleten' Assotsiatsii “Istoriya i komp'yuter”* [Newsletter of the Association “History and Computer”] 25.
 18. Shchapova, Y. L. 2005. *Arkheologicheskaya epokha: khronologiya, periodizatsiya, teoriya, model'* [Archaeological epoch: Chronology, periodization, theory, model]. Moscow: KomKniga, 192 p.
 19. Shchapova, Y. L. 2011. *Material'noe proizvodstvo v arkheologicheskuyu epokhu* [Material production in the archaeological epoch]. St. Petersburg: Aleteyya. 244 p.
 20. Grinchenko, S. N., and Yu. I. Shchapova. 2014. Prostranstvo i vremya v arheologii. Chast' 3. O metrike bazisnoy prostranstvennoy struktury chelovechestva v arkheologicheskuyu epokhu [Space and time in archeology. Part 3. About the metric of Humankind basic spatial structure in archaeological epoch]. *Space and Time* 1(15): 78–89.

Received October 17, 2018

Contributor

Grinchenko Sergey N. (b. 1946) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sgrinchenko@ipiran.ru

A GAUSSIAN APPROXIMATION OF THE DISTRIBUTED COMPUTING PROCESS

O. V. Lukashenko¹, E. V. Morozov², and M. Pagano³

Abstract: The authors propose a refinement of the stochastic model describing the dynamics of the Desktop Grid (DG) project with many hosts and many workunits to be performed, originally proposed by Morozov *et al.* in 2017. The target performance measure is the mean duration of the runtime of the project. To this end, the authors derive an asymptotic expression for the amount of the accumulated work to be done by means of limit theorems for superposed on-off sources that lead to a Gaussian approximation. In more detail, depending on the distribution of active and idle periods, Brownian or fractional Brownian processes are obtained. The authors present the analytic results related to the hitting time of the considered processes (including the case in which the overall amount of work is only known in a probabilistic way), and highlight how the runtime tail distribution could be estimated by simulation. Taking advantage of the properties of Gaussian processes and the Conditional Monte-Carlo (CMC) approach, the authors present a theoretical framework for evaluating the runtime tail distribution.

Keywords: Gaussian approximation; distributed computing; fractional Brownian motion

DOI: 10.14357/19922264190215

1 Introduction

Gaussian processes are widely used in the performance analysis of telecommunication systems for their analytic tractability and arguments based on the central-limit theorem that make them suitable in case of a large number of independent contributions. For instance, these models are able to capture, in a simple and parsimonious way, the properties of self-similarity and long-range dependence, inherent to multimedia network traffic [1, 2]. These properties dramatically increase the difficulty of the probabilistic analysis and, as a consequence, in many cases only Monte-Carlo simulation can be used. The *fractional Brownian motion* (FBM) is one of the most studied self-similar long-range dependent Gaussian processes due to its simplicity. Its use as traffic model is supported by the following theoretical analysis [3]: the sum of an increasing number of the so-called on-off inputs, with either on-times or off-times having a heavy-tailed distribution with infinite variance, converges weakly to an FBM, after an appropriate time scaling.

In this paper, the applicability of FBM for high-performance computing is considered. In that framework, computing clusters and computational Grid systems are the main tools: computing clusters are based on computing nodes connected by a high-speed network,

while computational Grid systems include geographically dispersed computing nodes connected by a relatively slow network.

Desktop Grid belongs to the latter class. The DG combines nondedicated *hosts* (typically, desktops/laptops owned by volunteers) over the Internet to process loosely coupled *workunits* (computational tasks). Desktop Grids utilize the idle host resources, providing potentially huge, although highly variable, computing power. (For example, the DG project EinsteinHOME aggregates peak performance at about 1 PetaFLOPS [4].) Typically, DGs are managed by a scientific community that utilizes the resources to complete a *DG project* which consists of a (usually finite) number of workunits. Thus, the *runtime* of the DG project is the time to complete all the workunits and it is desirable to minimize it.

Minimization of the DG project runtime may be performed by means of scheduling optimization [5–7]. Additional information on the hosts, such as reliability and availability, can be used to improve the efficiency of DGs [8, 9]. In [10, 11], the focus is placed on the so-called workunit replication mechanism for reliability purposes. However, to the best of our knowledge, the estimation of the runtime of a DG project remains generally an unsolved issue, and it is the main motivation of this paper.

¹Institute of Applied Mathematical Research of Karelian Research Centre of RAS, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation, lukashenko@krc.karelia.ru

²Institute of Applied Mathematical Research of Karelian Research Centre of RAS, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation, emorozov@karelia.ru

³University of Pisa, 43 Lungarno Pacinotti, Pisa 56126, Italy, m.pagano@iet.unipi.it

Desktop Grids have several important distinctive features when compared to computational Grids or computing clusters. First of all, hosts, being nondedicated, possess individual availability periods. Moreover, the management server of a DG is not able to obtain information on the current state of the hosts (such as “computing,” “suspended,” etc.). These two issues make the estimation of the runtime of a DG project a hard problem.

The execution of a DG project can be divided into two stages. During the first phase, the number of workunits is greater than the number of hosts and, thus, each host will receive at least one workunit. In the second stage, all the available workunits are dispatched and there are available (idle) hosts. In this paper, the focus is on the duration of the first phase which is studied by means of a Gaussian approximation of the overall work. The study of the second stage requires a completely different probabilistic technique, which relies on the theory of order statistics and the asymptotic properties of renewal processes, and is postponed for a future work. Thus, in what follows, runtime will relate to the first stage of the project solely.

We describe the availability patterns of the hosts by treating each of them as an individual on-off source which processes workunits during on periods. Our approach is based on the asymptotics of the (properly scaled) superposition of a large number of independent on-off sources. It is well-known [3] that after an appropriate scaling, the limiting process describing the summary workload in the system turns out to be *Brownian motion* (BM), when the sojourn times are light-tailed, while it becomes *fractional Brownian motion* in case of heavy-tailed sojourn times. Then, the problem reduces to the calculation of the hitting time of the given threshold D by the process of accumulated work which is a well-known topic in probability theory.

The paper is organized as follows. Section 2 presents the theoretical background related to FBM, including functional limit theorems for the cumulative work performed by an increasing number of on-off sources. Then, Section 3 describes the model and summarizes the available analytic results, while Section 4 is devoted to the evaluation of the runtime tail distribution by means of the CMC method which potentially leads to variance reduction of the estimate of the runtime. Finally, in Section 5, the main contributions of the paper are presented and some future research issues are discussed.

2 Theoretical Background

In this section, let us recall the basic definitions about FBM and how it is related to the limiting theorems for the superposition of independent *on-off sources*.

2.1 Fractional Brownian motion

The FBM $\{B_H(t), t \in \mathbb{R}\}$ is a Gaussian centered process with $B_H(0) = 0$, stationary increments, and the following covariance function:

$$K_H(t, s) := \mathbb{E}[B_H(t) B_H(s)] = \frac{1}{2} [|t|^{2H} - |t - s|^{2H} + |s|^{2H}], \quad s, t \geq 0$$

where $H \in (0, 1)$ is the so-called *Hurst parameter*. It is easy to verify that $B_H(t)$ is a self-similar process with self-similarity parameter H , i. e., for each $c > 0$,

$$c^{-H} B_H(ct) \stackrel{d}{=} B_H(t)$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Fractional Brownian motion is widely used for modeling purposes due to its Gaussianity (that typically arises under aggregation conditions) and parsimonious description (apart from mean and variance, its behavior is unambiguously determined by H).

When $H > 1/2$, FBM is a long-range dependent process since the autocorrelation of the corresponding increment process is nonsummable. For more details on FBM and its properties, see [12].

2.2 Limit theorems for distributed computing processes

Let us assume that the DG consists of N heterogeneous hosts which can be considered as independent *on-off sources*. In more detail, let us suppose that there are n types of hosts ($n < N$) and denote by N_i the number of i -type hosts, i. e., $\sum_{i=1}^n N_i = N$. Moreover, let R_i denote the amount of processed work per unit time for i -type hosts and let $\{I^{(i)}(t), t \geq 0\}$,

$$I^{(i)}(t) = \begin{cases} R_i, & t \in \text{on-period}; \\ 0, & t \in \text{off-period}, \end{cases}$$

be the *on-off* process that characterizes the activity/silent periods of the corresponding hosts (Fig. 1). For sake of simplicity, it is assumed that for each host, both *on* and *off* periods are sequences of i.i.d. (independent and identically distributed) random variables (RVs) and mutually independent. Moreover, as already stated, the

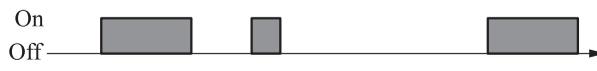


Figure 1 On-off model

on-off processes modeling the contribution of different hosts are assumed to be independent.

The *cumulative processed work*, i. e., the aggregated amount of work provided by all N hosts, during the time interval $[0, t]$ is given by

$$A(t) = \int_0^t \left(\sum_{i=1}^n \sum_{k=1}^{N_i} I_k^{(i)}(u) \right) du$$

where $I_k^{(i)}$ are the independent copies of $I^{(i)}$, $i = 1, \dots, n$. Moreover, for the i -type ($i = 1, \dots, n$) hosts, let us denote by μ_{on}^i , σ_{on}^i , μ_{off}^i , and σ_{off}^i the mean length and standard deviation (that may be infinite) of the duration of the on and off periods, respectively.

The statistical behavior of $A(t)$ is determined by the distributions F_{on}^i and F_{off}^i of the on and off periods for each type of hosts, namely, by their tail. In more detail, in case of infinite variance, let us assume that as $x \rightarrow \infty$,

$$\begin{aligned} 1 - F_{\text{on}}^i(x) &\sim \ell_{\text{on}}^i x^{-\alpha_{\text{on}}^i} L_{\text{on}}^i(x); \\ 1 - F_{\text{off}}^i(x) &\sim \ell_{\text{off}}^i x^{-\alpha_{\text{off}}^i} L_{\text{off}}^i(x) \end{aligned}$$

where $a \sim b$ means that $a/b \rightarrow 1$; ℓ_{on}^i and ℓ_{off}^i are the positive constants; the exponents α_{on}^i and $\alpha_{\text{off}}^i \in (1, 2)$; and the functions L_{on}^i and L_{off}^i are slowly varying at infinity, i. e., for any $t > 0$,

$$\lim_{x \rightarrow \infty} \frac{L^i(tx)}{L^i(x)} = 1, \quad i = 1, \dots, n.$$

Instead, if σ_{on}^i and $\sigma_{\text{off}}^i < \infty$, we set $\alpha_{\text{on}}^i = \alpha_{\text{off}}^i = 2$.

It has been shown in [3] that the scaled process of cumulative work arrived during interval $[0, Tt]$ converges weakly to a sum of the i.i.d. FBM's, provided that

- (1) $N_i \rightarrow \infty$ such that $\lim_{N \rightarrow \infty} N_i/N > 0$, $i = 1, \dots, n$; and
- (2) the scaling factor $T \rightarrow \infty$.

This *functional limit theorem* leads to the following approximation:

$$\begin{aligned} A(tT) &\approx T \left(\sum_{i=1}^n R_i N_i \frac{\mu_{\text{on}}^i}{\mu_{\text{on}}^i + \mu_{\text{off}}^i} \right) t \\ &\quad + \sum_{i=1}^n T^{H_i} R_i \sqrt{L_i(T) N_i c_i} B_{H_i}(t) \end{aligned}$$

where c_i are the positive constants; L_i are the slowly varying at infinity functions (expressed in terms of the given parameters); and B_{H_i} are the independent FBM's with the Hurst parameters H_i given by

$$H_i = \frac{3 - \min(\alpha_{\text{on}}^i, \alpha_{\text{off}}^i)}{2} \in \left(\frac{1}{2}, 1 \right), \quad i = 1, \dots, n.$$

Thus, the cumulative work processed by a large number of independent hosts (with heavy-tailed distributions of the on-off periods) is approximated by a superposition of independent FBM's $\{B_{H_i}(t)\}$, $i = 1, \dots, n$, with a linear drift that depends on the rates R_i and the average duty cycle.

Instead, if for all types of hosts the variances of the sojourn times are finite (i. e., $\sigma_{\text{on}}^i, \sigma_{\text{off}}^i < \infty \forall i = 1, \dots, n$), then the limiting (scaled) process becomes

$$T \left(\sum_{i=1}^n \frac{R_i N_i \mu_{\text{on}}^i}{\mu_{\text{on}}^i + \mu_{\text{off}}^i} \right) t + \left(\sqrt{T} \sum_{i=1}^n R_i \sqrt{N_i c_i} \right) W(t)$$

where $W(t)$ is the Wiener process, and the constants c_i are given by

$$c_i = \sqrt{\frac{(\mu_{\text{off}}^i \sigma_{\text{on}}^i)^2 + (\mu_{\text{on}}^i \sigma_{\text{off}}^i)^2}{(\mu_{\text{on}}^i + \mu_{\text{off}}^i)^3}}.$$

Finally, it is worth mentioning that taking the limits in reverse order, the (scaled) process of cumulative work converges to a *Levy stable motion*, an infinite variance process with stationary and independent increments [13]; however, such model is beyond the scope of this paper as in DG, the experimental data confirmed the convergence to processes with finite variance.

3 Model Description and Performance Measures

The above functional limit theorems provide a theoretical motivation to consider the following model for the cumulative processed work:

$$A(t) = mt + X(t) \quad (1)$$

where X is the centered Gaussian process with stationary increments (FBM or the sum of independent FBM, in case of heterogeneous systems), which describes random fluctuations around the linearly increasing mean. Such type of stochastic process was previously suggested as the model of network traffic (see [14] for more details).

Let us denote by τ_D the runtime of the DG project where D denotes the required amount of work. Thus, τ_D represents the *hitting time* of the process $\{A(t)\}$:

$$\tau_D = \min\{t : A(t) \geq D\},$$

i. e., the first time the process $\{A(t)\}$ hits the threshold D . Then, the original problem is reduced to the calculation (or estimation) of some useful performance characteristics, such as the mean hitting time.

3.1 Available analytic results

Let us recall the available analytic results for different types of Gaussian processes, corresponding to the different limiting cases.

3.1.1 Wiener case

When X is a Wiener process (i.e., $X = \sigma B_{1/2}$), the density of τ_D is available in explicit form [15]:

$$\mathbb{P}(\tau_D \in dt) = \frac{D}{\sqrt{2\pi}\sigma t^{3/2}} \exp\left(-\frac{(D - mt)^2}{2\sigma^2 t}\right) dt =: f_\tau(t|D) dt. \quad (2)$$

In this case, the corresponding expected value $\mathbb{E}[\tau_D]$ is the ratio between the given amount of the work D and the mean processing rate [15]:

$$\mathbb{E}[\tau_D] = \frac{D}{m}.$$

3.1.2 Fractional Brownian motion case

When the limiting process is an FBM, only asymptotic results and some bounds for the distribution of τ_D are available.

In [16], the following bounds (quite inaccurate when H is close to 1, see Fig. 2) for the moments of the hitting time were obtained for $1/2 \leq H < 1$:

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \left(\frac{2HD}{n-H} L_n(D, H, m) - \frac{(2H-1)m}{n+1-H} L_{n+1}(D, H, m) \right) \leq \mathbb{E}[\tau_D^n] \\ & \leq \frac{1}{\sqrt{2\pi}} \left(\frac{HD}{n-H} L_n(D, H, m) + \frac{(1-H)m}{n+1-H} L_{n+1}(D, H, m) \right) \end{aligned}$$

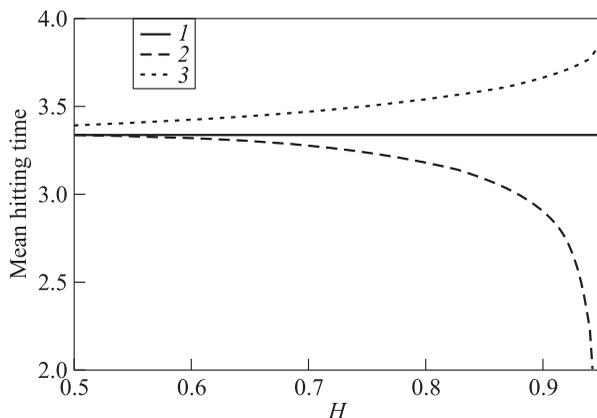


Figure 2 Bounds for the mean hitting time ($D = 10, m = 3$): 1 – D/m ; 2 – lower bound; and 3 – upper bound

where

$$L_n(D, H, m) = \int_0^\infty \exp\left\{-\frac{1}{2}\left(Dt^{-H/(n-H)} - mt^{(1-H)/(n-H)}\right)^2\right\} dt.$$

Additionally, the following asymptotic was derived for the large values of level D :

$$\lim_{D \rightarrow \infty} \frac{\mathbb{E}[\tau_D^n]}{D^n} = m^{-n}$$

for all $n \geq 1, m > 0$, from which it is quite straightforward to show that for all $n \geq 1$,

$$\frac{\tau_D}{D} \xrightarrow{L_n} \frac{1}{m} \text{ as } D \rightarrow \infty$$

where $\xrightarrow{L_n}$ means convergence in L_n space.

3.1.3 General case

In the general case, to derive asymptotic (for large values of D) for the distribution of τ_D , it is possible to take advantage of the following identity:

$$\mathbb{P}(\tau_D \leq T) = \mathbb{P}\left(\sup_{t \in [0, T]} A(t) \geq D\right).$$

The distribution of the maximum of Gaussian processes over a finite interval is a well-studied problem. In more detail, for any Gaussian process with stationary increments and strictly monotonically increasing and convex variance such that $\lim_{t \rightarrow 0} \text{Var}(X(t))/t = 0$, the following asymptotic holds [17]:

$$\mathbb{P}\left(\sup_{t \in [0, T]} A(t) \geq D\right) \sim \Phi\left(\frac{D - mT}{\sqrt{\text{Var}(X(T))}}\right) \text{ as } D \rightarrow \infty$$

where Φ denotes the tail distribution of the standard normal RV $N(0, 1)$.

3.2 A possible generalization

It seems quite natural to consider the setting in which the threshold D is an RV which is independent of the process X in (1). Such a setting seems to be highly motivated by practice because it is more realistic that the exact value of the quantity D is not available, and it is known in part. This incomplete information can be reflected by introducing the probability density function (PDF) f_D of D , which is assumed to be predefined. Provided that X in (1) is a Wiener process and, hence,

the conditional density $f_\tau(t|D)$ in (2) is known, one can write the density of the RV τ_D as

$$f_\tau(x) = \int_{y=0}^{\infty} f_\tau(x|y) f_D(y) dy.$$

In general, one can calculate this density only by numerical methods but for some cases, it is possible to derive its expression in terms of special functions. For example, when D is exponential with parameter λ , one can obtain the following expression:

$$f_\tau(x) = \frac{\lambda}{\sqrt{2\pi\sigma x^{3/2}}} \exp\left(-\frac{m^2 x}{2\sigma^2} + \frac{\gamma^2}{8\beta(x)}\right) \times (2\beta(x))^{-1/2} D_{-1}\left(\frac{\gamma}{\sqrt{2\beta(x)}}\right) \quad (3)$$

where

$$\gamma = \lambda - \frac{m}{\sigma^2}; \quad \beta(x) = \frac{1}{2\sigma^2 x};$$

and D_p , $\text{Re } p < 0$, is the parabolic cylinder function [18]. Numeric calculation of the expression (3) is shown in Fig. 3.

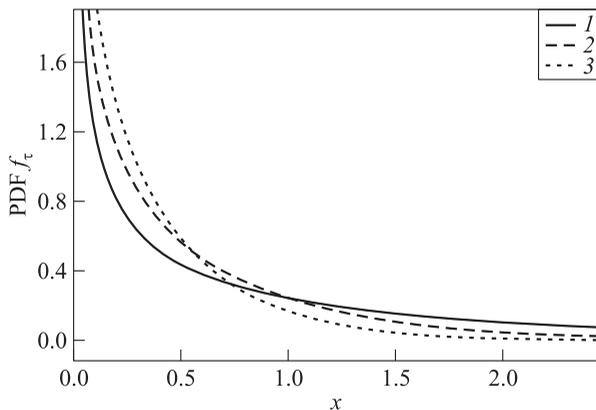


Figure 3 Probability density function of τ_D for different values of m ($\lambda = 1$): 1 – $m = 1$; 2 – $m = 2$; and 3 – $m = 3$

4 Estimation via Monte Carlo

A more flexible alternative to analytic results is represented by simulation that in our case can be used to estimate

$$\pi(T) := \mathbb{P}(\tau_D > T).$$

Such probability could be extremely small for large values of T ; thus, its estimation with a given accuracy requires to generate a large number of sample paths of the process X . However, for such type of rare events, it is possible to apply a special case of the well-known CMC method which always leads to variance reduction.

The method, originally proposed by some of the authors in [19–21] and named Bridge Monte Carlo (BMC), is based on the idea of expressing the target probability as the expectation of a function of the Bridge $Y := \{Y_t\}$ of the Gaussian process X , i. e., the process obtained by conditioning X to reach a certain level at some prefixed time instant τ :

$$Y(t) = X(t) - \psi(t)X(\tau)$$

where ψ can be easily expressed in terms of the the covariance function $\Gamma(s, t)$ of the process X

$$\psi(t) := \frac{\Gamma(t, \tau)}{\Gamma(\tau, \tau)}.$$

Since the variance of X is an increasing function of t in all models we consider, it is easy to see that $\psi(t) > 0$ for all $t \geq 0$. Moreover, for any t , $Y(t)$ is independent of $X(\tau)$ since

$$\mathbb{E}[X(\tau)Y(t)] = \Gamma(\tau, t) - \frac{\Gamma(t, \tau)}{\Gamma(\tau, \tau)} \Gamma(\tau, \tau) = 0$$

and $(X(\tau), Y(t))$ has bivariate normal distribution.

Let $\mathbb{T} = [0, T]$, then the target probability can be expressed in the following way:

$$\begin{aligned} \pi(T) &= \mathbb{P}\left(\sup_{t \in [0, T]} A(t) \geq D\right) \\ &= \mathbb{P}(\forall t \in \mathbb{T} : mt + X(t) \leq D) \\ &= \mathbb{P}\left(\forall t \in \mathbb{T} : X(\tau) \leq \frac{D - Y(t) - mt}{\psi(t)}\right) \\ &= \mathbb{P}\left(X(\tau) \leq \inf_{t \in \mathbb{T}} \frac{D - Y(t) - mt}{\psi(t)}\right) \\ &= \mathbb{P}(X(\tau) \leq \bar{Y}) \end{aligned}$$

where

$$\bar{Y} := \inf_{t \in \mathbb{T}} \frac{D - Y(t) - mt}{\psi(t)}.$$

Finally, the considered probability can be rewritten as follows:

$$\pi(T) = \mathbb{P}(X_\tau \leq \bar{Y}) = \mathbb{E}\left[\Psi\left(\frac{\bar{Y}}{\sqrt{\Gamma(\tau, \tau)}}\right)\right]$$

where independence between \bar{Y} and X_τ is used and Ψ denotes the cumulative distribution function of a standard normal variable.

Hence, given N samples $\{\bar{Y}^{(n)}, n = 1, \dots, N\}$ of \bar{Y} , the estimator of $\pi(T)$ is defined as follows:

$$\hat{\pi}_N^{\text{BMC}} := \frac{1}{N} \sum_{n=1}^N \Psi\left(\frac{\bar{Y}^{(n)}}{\sqrt{\Gamma(\tau, \tau)}}\right).$$

Note that

$$\Psi \left(\frac{\bar{Y}}{\sqrt{\Gamma(\tau, \tau)}} \right) = \mathbb{E} [I(X(\tau) \leq \bar{Y}) | \bar{Y}]$$

and, therefore, the BMC approach is actually a special case of the CMC method; so, one can expect that the BMC estimator implies variance reduction (with regard to crude Monte-Carlo simulation) in the estimation of the target probability $\pi(T)$ as also justified by the previous experience when such a method was successfully applied for estimation some other rare-event probabilities related to Gaussian processes [22].

5 Concluding Remarks and Future Research

In this paper, a stochastic model describing the dynamics of a DG project with many hosts and many workunits to be performed, originally proposed in [23], is presented. It is assumed that the project can be described by the so-called on-off model where the hosts are on-off sources of the workunits and the basic process is the completed work. It is assumed that the hosts' working sessions can have both light- and heavy-tailed distributions. Then, an approximation of the basic process, based on the asymptotics of the superposed on-off sources, is applied. The suggested approach leads to a Gaussian approximation of the process of the completed work. Finally, a simulation framework for the evaluation of the runtime of the project, using the properties of Gaussian processes and CMC simulation, is presented.

Although this note is focused on estimation of the runtime related to the 1st stage of the project completion when the number of workunits is bigger than the number of hosts, the 2nd stage could also be relevant. In more detail, it can be considered as a collection of the "tails" of the workunit remaining times. From this point of view, the completion time of the 2nd stage of the project can be interpreted as the *longest* remaining time and analyzed by means of the asymptotic results of *renewal theory*. Moreover, since the workunits are assumed to be independent, to evaluate the duration of the 2nd stage, it seems promising to apply the theory of *order statistics* and interpret the completion time as the maximal order statistics.

Acknowledgments

The study was carried out under state order to the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research KarRC RAS) and supported by the Russian Foundation

for Basic Research, projects 18-07-00187, 18-07-00147, 18-07-00156, and 19-07-00303.

References

1. Leland, W. E., M. S. Taqqu, W. Willinger., and D. V. Wilson. 1994. On the self-similar nature of ethernet traffic (extended version). *IEEE ACM T. Network*. 2(1):1–15.
2. Willinger, W., M. S. Taqqu, W. E. Leland, and D. Wilson. 1995. Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements. *Stat. Sci.* 10(1):67–85.
3. Taqqu, M. S., W. Willinger, and R. Sherman. 1997. Proof of a fundamental result in self-similar traffic modeling. *Comp. Comm. R.* 27:5–23.
4. BOINCstats. 2017. Available at: <https://boincstats.com> (accessed May 7, 2019).
5. Kondo, D., D. P. Anderson, and J. McLeod VII. 2007. Performance evaluation of scheduling policies for volunteer computing. *3rd IEEE Conference (International) on e-Science and Grid Computing Proceedings*. IEEE. 221–227.
6. Estrada, T., and M. Taufer. 2012. Challenges in designing scheduling policies in volunteer computing. *Desktop grid computing*. Eds C. Cérin and G. Fedak. CRC Press. 167–190.
7. Durrani, N., and J. Shamsi. 2014. Volunteer computing: Requirements, challenges, and solutions. *J. Netw. Comput. Appl.* 39:369–380.
8. Sonnek, J., M. Nathan, A. Chandra, and J. Weissman. 2006. Reputation-based scheduling on unreliable distributed infrastructures in distributed computing systems. *26th IEEE Conference (International) on Distributed Computing Systems Proceedings*. IEEE. Art. No. 30. P. 1–8.
9. Watanabe, K., M. Fukushi, and M. Kameyama. 2011. Adaptive group-based job scheduling for high performance and reliable volunteer computing. *J. Information Processing* 19:39–51.
10. Xavier, E., R. Peixoto, and J. da Silveira. 2013. Scheduling with task replication on desktop grids: Theoretical and experimental analysis. *J. Comb. Optim.* 30(3):520–544.
11. Chernov, I. A., and N. N. Nikitina. 2015. Virtual screening in a Desktop Grid: Replication and the optimal quorum. *Parallel computing technologies*. Ed. V. Malyshekin. Lecture notes in computer science ser. Springer. 9251:258–267.
12. Samorodnitsky, G., and M. S. Taqqu. 1994. *Stable non-Gaussian random processes: Stochastic models with infinite variance*. Chapman & Hall. 632 p.
13. Mikosch, T., S. Resnick, H. Rootzén, and A. Stegeman. 2002. Is network traffic approximated by stable Levy motion or fractional Brownian motion? *Ann. Appl. Probab.* 12(1):23–68.
14. Norros, I. 1994. A storage model with self-similar input. *Queueing Syst.* 16:387–396.
15. Borodin, A. N., and P. Salminen. 2002. *Handbook of Brownian motion — facts and formulae*. Birkhäuser. 685 p.

16. Michna, Z. 1999. On tail probabilities and first passage times for fractional Brownian motion. *Math. Method. Oper. Res.* 49(2):335–354.
17. Caglar, M., and C. Vardar. 2013. Distribution of maximum loss of fractional Brownian motion with drift. *Stat. Probabil. Lett.* 83:2729–2734.
18. Gradshteyn, I. S., I. M. Ryzhik, and A. Jeffrey, eds. 2015. *Table of integrals, series and products*. 8th ed. San Diego, CA: Academic Press. 1220 p.
19. Giordano, S., M. Gubinelli, and M. Pagano. 2005. Bridge Monte-Carlo: A novel approach to rare events of Gaussian processes. *5th St. Petersburg Workshop on Simulation Proceedings*. St. Petersburg: St. Petersburg State University. 281–286.
20. Giordano, S., M. Gubinelli, and M. Pagano. 2007. Rare events of Gaussian processes: A performance comparison between bridge Monte-Carlo and importance sampling. *Next generation teletraffic and wired/wireless advanced networking*. Eds. Y. Koucheryavy, J. Harju, and A. Sayenko. Lecture notes in computer science ser. Springer. 4712:269–280.
21. Lukashenko, O. V., E. V. Morozov, and M. Pagano. 2012. Performance analysis of Bridge Monte-Carlo estimator. *Transactions of KarRC RAS* 5:54–60.
22. Lukashenko, O. V., E. V. Morozov, and M. Pagano. 2017. On the efficiency of bridge Monte-Carlo estimator. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):16–24.
23. Morozov, E., O. Lukashenko, A. Rumyantsev, and E. Ivashko. 2017. A Gaussian approximation of runtime estimation in a desktop grid project. *9th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops*. IEEE. 107–111.

Received April 15, 2019

Contributors

Lukashenko Oleg V. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, scientist, Institute of Applied Mathematical Research of Karelian Research Centre of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; associate professor, Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; lukashenko@krc.karelia.ru

Morozov Evsei V. (b. 1947) — Doctor of Science in physics and mathematics, professor, leading scientist, Institute of Applied Mathematical Research of Karelian Research Centre of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; professor, Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; emorozov@karelia.ru

Pagano Michele (b. 1968) — PhD in Information Engineering, associate professor, University of Pisa, 43 Lungarno Pacinotti, Pisa 56126, Italy; m.pagano@iet.unipi.it

ГАУССОВСКАЯ АППРОКСИМАЦИЯ ПРОЦЕССА РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ*

О. В. Лукашенко^{1,2}, Е. В. Морозов^{1,2}, М. Пагано³

¹Институт прикладных математических исследований Карельского научного центра Российской академии наук

²Петрозаводский государственный университет

³Университет г. Пиза, Италия

Аннотация: Продолжено изучение стохастической модели процесса динамики выполнения задачи в системе Desktop Grid при наличии многих пользователей, предложенной в 2017 г. Морозовым с соавт. Требуемой характеристикой выступает средняя продолжительность времени выполнения проекта. Гауссовская аппроксимация искомого процесса производится на основе предельных теорем для суперпозиции on-off источников. Приведен обзор известных аналитических результатов для требуемой характеристики, включая результаты для броуновского и дробного броуновского движения. Также показывается, как с помощью условного метода Монте-Карло оценить хвост распределения времени выполнения проекта.

Ключевые слова: гауссовская аппроксимация; распределенные вычисления; дробное броуновское движение

DOI: 10.14357/19922264190215

*Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН) и при финансовой поддержке РФФИ (проекты 18-07-00187, 18-07-00147, 18-07-00156 и 19-07-00303).

Литература

1. *Leland W.E., Taqqu M. S., Willinger W., Wilson D. V.* On the self-similar nature of Ethernet traffic (extended version) // IEEE ACM T. Network., 1994. Vol. 2. Iss. 1. P. 1–15.
2. *Willinger W., Taqqu M. S., Leland W. E., Wilson D.* Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements // Stat. Sci., 1995. Vol. 10. Iss. 1. P. 67–85.
3. *Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling // Comp. Comm. R., 1997. Vol. 27. P. 5–23.
4. BOINCstats, 2017. <https://boincstats.com>.
5. *Kondo D., Anderson D. P., McLeod VII J.* Performance evaluation of scheduling policies for volunteer computing // 3rd IEEE Conference (International) on e-Science and Grid Computing Proceedings. — IEEE, 2007. P. 221–227.
6. *Estrada T., Taufer M.* Challenges in designing scheduling policies in volunteer computing // Desktop grid computing / Eds. C. Cérin, G. Fedak. — CRC Press, 2012. P. 167–190.
7. *Durrani N., Shamsi J.* Volunteer computing: Requirements, challenges, and solutions // J. Netw. Comput. Appl., 2014. Vol. 39. P. 369–380.
8. *Sonnek J., Nathan M., Chandra A., Weissman J.* Reputation-based scheduling on unreliable distributed infrastructures in distributed computing systems // 26th IEEE Conference (International) on Distributed Computing Systems Proceedings. — IEEE, 2006. Art. No. 30. P. 1–8.
9. *Watanabe K., Fukushi M., Kameyama M.* Adaptive group-based job scheduling for high performance and reliable volunteer computing // J. Information Processing, 2011. Vol. 19. P. 39–51.
10. *Xavier E., Peixoto R., da Silveira J.* Scheduling with task replication on desktop grids: Theoretical and experimental analysis // J. Comb. Optim., 2013. Vol. 30. Iss. 3. P. 520–544.
11. *Chernov I. A., Nikitina N. N.* Virtual screening in a desktop grid: Replication and the optimal quorum // Parallel computing technologies / Ed. V. Malyskin. — Lecture notes in computer science ser. — Springer, 2015. Vol. 9251. P. 258–267.
12. *Samorodnitsky G., Taqqu M. S.* Stable non-Gaussian random processes: Stochastic models with infinite variance. — Chapman & Hall, 1994. 632 p.
13. *Mikosch T., Resnick S., Rootzén H., Stegeman A.* Is network traffic approximated by stable Levy motion or fractional Brownian motion? // Ann. Appl. Probab., 2002. Vol. 12. Iss. 1. P. 23–68.
14. *Norros I.* A storage model with self-similar input // Queueing Syst., 1994. Vol. 16. P. 387–396.
15. *Borodin A. N., Salminen P.* Handbook of Brownian motion — facts and formulae. — Birkhäuser, 2002. 685 p.
16. *Michna Z.* On tail probabilities and first passage times for fractional Brownian motion // Math. Method. Oper. Res., 1999. Vol. 49. Iss. 2. P. 335–354.
17. *Caglar M., Vardar C.* Distribution of maximum loss of fractional Brownian motion with drift // Stat. Probabil. Lett., 2013. Vol. 83. P. 2729–2734.
18. Table of integrals, series and products / Eds. I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey. — 8 ed. — San Diego, CA, USA: Academic Press, 2015. 1220 p.
19. *Giordano S., Gubinelli M., Pagano M.* Bridge Monte-Carlo: A novel approach to rare events of Gaussian processes // 5th St. Petersburg Workshop on Simulation Proceedings. — St. Petersburg: St. Petersburg State University, 2005. P. 281–286.
20. *Giordano S., Gubinelli M., Pagano M.* Rare events of Gaussian processes: A performance comparison between bridge Monte-Carlo and importance sampling // Next generation teletraffic and wired/wireless advanced networking / Eds. Y. Koucheryavy, J. Harju, A. Sayenko. — Lecture notes in computer science ser. — Springer, 2007. Vol. 4712. P. 269–280.
21. *Lukashenko O. V., Morozov E. V., Pagano M.* Performance analysis of bridge Monte-Carlo estimator // Труды Карельского научного центра Российской академии наук, 2012. Т. 5. С. 54–60.
22. *Lukashenko O. V., Morozov E. V., Pagano M.* On the efficiency of bridge Monte-Carlo estimator // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 16–24.
23. *Morozov E., Lukashenko O., Romyantsev A., Ivashko E.* A Gaussian approximation of runtime estimation in a desktop grid project // 9th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops. — IEEE, 2017. P. 107–111.

Поступила в редакцию 15.04.2019

VIRTUAL EXPERIMENTS IN DATA INTENSIVE RESEARCH

D. Y. Kovalev¹ and E. A. Tarasov²

Abstract: Organization and management of virtual experiments (VE) in data intensive research (DIR) has been widely studied in the several past years. The authors survey existing approaches to deal with VEs and hypotheses and analyze VE management in a real use-case from the astronomy domain. A review of existing systems that can act as platforms for conducting a VE has been carried out. Requirements for a system to organize VEs in data intensive domains have been gathered and overall structure and functionality for system running VEs are presented. The relationships between hypotheses and models in VE are discussed. The authors also illustrate how to model conceptually VEs, respective hypotheses, and models. Potential benefits and drawbacks of such approach are discussed, including maintenance of experiment consistency and reduction of potential number of experiments.

Keywords: virtual experiment; hypothesis; conceptual modeling; data intensive research

DOI: 10.14357/19922264190216

1 Introduction

Data intensive research is evolving according to the 4th paradigm of scientific development and reflects the fact that modern science is highly dependent on knowledge extraction from massive datasets [1]. Data intensive research is multidisciplinary in its nature, bringing in many separate principles and techniques to handle complex data analysis and management. Researcher's time is split between management of raw and analytical data, including data collection, curation and integration, and scientific work itself, which the article focuses on. It requires knowledge inference from collected data in order to test proposed hypotheses, gather novel information, and correctly integrate it.

Large-scale scientific experiments besides data processing issues are highly sophisticated — they include workflows, models, and analytical methods. Together, they compose a VE. In [2], a survey is presented discussing different approaches to experiment modeling and how its core artifacts — hypotheses — can be specified. The use of conceptual representation of hypotheses and their corresponding implementation is emphasized, thus leading to the need of proper tools.

This particular paper is aimed at developing methods and tools to support the execution and conceptual modeling of VEs and discusses several approaches to manage it.

The paper is structured as follows. Section 2 provides an analysis of existing systems and the system

requirements for managing VE. In Section 3, main core elements of VE are defined. In Sections 4 and 5, the two modes of execution of VE based on two DIR use-cases are discussed. Section 6 concludes the paper.

This work is a continuation of [3, 4]. Comparing it with the previous work, this paper presents unified methodology to work with VE. The emphasis is made on working with two components: research lattices and real-time model/hypothesis testing. Research is carried out in the frame of Research Data Alliance³ activity.

2 Analysis of Existing Systems

Systems with explicit representation of hypotheses are being rapidly developed during last several years [5–10]. Authors analyzed 4 different systems for managing VE and hypotheses: Hephaestus, Upsilon-DB, SDI (Scientific Data Infrastructure), and FCCE (Features Collection and Correlation Engine). Some requirements for organizing and managing VEs were retrieved during the analysis.

2.1 Hephaestus

Hephaestus is a system for running VEs over existing collections of data. It provides independence from resources and the system rewrites its queries into data source queries. The system hides underlying implementation details from user, letting him/her to work only with the Hephaestus language. This language itself is

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., 119333 Moscow, Russian Federation, dkovalev@ipiran.ru

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., 119333 Moscow, Russian Federation, e.tarasov@outlook.com

³“The Research Data Alliance (RDA) was launched as a community-driven organization in 2013 by the European Commission, the United States National Science Foundation and National Institute of Standards and Technology, and the Australian Government’s Department of Innovation with the goal of building the social and technical infrastructure . . . that promotes data-sharing and data-driven research. . .” (<https://www.rd-alliance.org/about-rda>).

a declarative language and is used to specify VEs and underlying hypotheses.

Hephaestus separates two different classes of hypotheses: top-down and bottom-up. Top-down hypotheses are the one introduced by the researcher, while bottom-up hypotheses are derived from data. The system supports the discovery of bottom-up hypotheses by looking for the correlation in data. These hypotheses are then ranked by some score (e.g., p -value of some statistical test) and the one with the highest are passed to the researcher. Yet, the system does not support automatically finding of causality. Hephaestus emphasizes the role of an expert in understanding which relationships should be further studied and which should not be chased. Hephaestus also computes metrics about experiments to estimate significance adequate to abandon further computation. The system is used in testing clinical trials. The system does not catch the evolution of hypotheses or experiments yet.

2.2 Upsilon-DB

The system enables researcher to code and manage deterministic scientific hypotheses as uncertain data. It uses internal database to form hypotheses as relations and adds uncertainty parameter. Later, this uncertainty parameter is used to rank hypotheses using Bayes rule. The provided approach can be treated as complementary to classical statistical approach. The system allows to work with two types of uncertainty — theoretical, which is brought by competing hypotheses, and empirical, which appears because of alternative datasets used. The system introduces algorithm to rank hypotheses using observed data. This is done because several competing hypotheses can explain the same observation well and some score to distinguish them is needed. When new data become available, this score can be adjusted accordingly.

Hypotheses have mathematical representation and authors provide method to translate its mathematical representation into relations in database. The simulations are also treated as data and respective relations are put inside the same database as hypotheses. Authors emphasize the need to support and develop the extraction of hypotheses from data and methods to sample both hypotheses and data. They illustrate that systems such as Eureqa [11] can be used to learn formula representation from data.

An example presented in the paper is based on three different laws describing free fall and some simulated data. They rank hypotheses accordingly.

2.3 Other

Several other platforms are analyzed. We describe them separately since they are not explicitly designed with

hypotheses or VEs as their core element. However, some of the ideas that they implement are worth mentioning.

The SDI platform [5] is used to support scientific experiments. The system can integrate open data and reuse observed and simulation data in the further development of experiments. The system enables multiple groups of researchers to access data and experiments simultaneously. Components of the framework are developed in such a way that they could be deployed, adapted, and accessed in individual research projects fast. The SDI platform requires the support of lineage, provenance, classification, and indexing of experiments and data, the whole cycle of obtaining data, curating and cleaning it, building experiments to test hypotheses over massive data, and aggregating results is supported over prolonged periods of time. The use of semantics is required by the system.

The FCCE platform [10] is designed to search for correlations in heterogeneous data sets spanning long time ranges. It is oriented to work with minimal delays and access to unprocessed source data. The key component of the data model is the concept of features. The features define the relationship between the key and some value, each element of which can contain several features. The FCCE platform presents a simplified relational data model for the user, where each table stores one type of characteristic. Each line is identifiable by a key and can contain several features. The platform provides an application program interface for storing, obtaining, and calculating correlation over features. The developers offer the original approach of integrating the correlation criterion evaluation module into the query execution engine over the repository, which allows to speed up the response time to the query and to reduce the overhead of computation and input/output operations. The FCCE platform uses two distinctive mechanisms to support the effectiveness of correlations between the characteristics: the query pipe and the query modifier.

2.4 Requirements for a virtual experiment management system

Although platforms mentioned above provide important insights into defining and handling hypotheses, they miss some notable features. In this subsection, missing features are overviewed and general requirements for a VE Management System (VEMS) are defined.

First, existing platforms do not describe the perception of automatically derived hypotheses by domain experts, do not track their evolution, and do not discuss experiment design principles. In a series of experiment run, it is important to keep track on evolution of models, hypotheses, and experiments, as well as identifying new data sources. Operations to manipulate VEs and their components need to be defined. Next, a VEMS

needs to capture dependencies (like *competes*, *derived by*) between hypotheses, invariants in single hypothesis. Correlations between parameters of several hypotheses should also be considered.

Second, a VEMS should contain components responsible for automatic extraction of dependencies between hypotheses, parameters in single and multiple hypotheses. Obtained data are used in deciding which experiments should be abandoned and used in keeping hypotheses consistent in a single experiment.

Third, one needs components for maintaining experiment consistency and constraining the number of possible experiments as well as defining the metric, which is used to define if experiment poorly explains phenomena and abandon further computations. Methods for removing poor experiments based on previous experiments runs are also required. Experiments and hypotheses should stay consistent when parameters of hypotheses are changed.

As soon as several hypotheses in some experiments could explain some phenomena well and due to errors in data, researcher needs to deal with uncertainty and needs methods to rank experiments and competing hypotheses on massive datasets.

While experiment could change slightly from a previous experiment run (e.g., one hypothesis parameter changes), system should store some data about previous executions. Methods for understanding, which parts of experiments should be recomputed and which are not, should be developed as well. Defining structures to store results of previous experiments and to query these results is important. Since there could be thousands of possible experiments, the system should use a method to form a plan to execute experiments in such a way that stored results are reused and no additional computations are made.

During the execution of the VE, the researcher manipulates by the parameters of hypotheses, i.e., a set of variables that in some cases can be correlated with each other as well as with the parameters of the other hypotheses [2, 12]. Since the number of potential hypotheses in a VE can be enormous and their interaction is nontrivial, a space with many VEs is formed, some of which poorly describe observations and need to be omitted before the experiment is performed. Consequently, a researcher needs a tool that allows to preidentify and filter out VEs with a predictably bad result. At the same time, the presence of complex dependencies in the data makes their understanding difficult for the researcher and does not allow doing this manually [9]. Correlation estimation tools allow to automatically divide VEs into groups with a predictably good and poor experimental result [13].

Thus, reducing the number of experiments is achieved by:

- *correlations search* — this allows to combine the features of the phenomenon under investigation into certain groups that affect it in some aggregate;
- *analysis of features* — it is necessary to select a set of parameter values within a selected group that, with certain accuracy indicators, would describe the actual observational data; and
- *ranking* of hypotheses according to the degree of accuracy of the VE. This will help the researcher to pay attention to the most probable hypotheses without the need for a complete search through all parameters combinations.

To reduce the space of hypotheses parameters and VEs, methods for featuring selection are widely used [14]. They allow to increase the speed of data processing and to obtain the result without reducing the accuracy [15], by isolating only those informative features that are required for performing a VE. The selection of a set of characteristics makes it possible to simplify the understanding of the model by the researcher and, consequently, to use them as input data for the widely known algorithms of machine learning [16]. In addition, these methods allow to reduce noise in the data and to reveal the interaction between the parameters.

3 Hypotheses and Models in Virtual Experiment

Extracted information needs to be formally specified. For that, authors define additional artifact — VE. It is a tuple $\langle O, H, M, R, W, C \rangle$, where:

- O is the domain ontology that is a set of concepts and relationships in applied domain formally specified with some language;
- H is the set of hypotheses specifications and relationships between them. H is a part of ontology and uses concepts from it. Together they form the ontology of VE. Hypothesis is a proposed explanation of a phenomenon that still has to be rigorously tested;
- M is the set of models. Each model is a set of functions. Every model implements a hypothesis specification. If model generates expected behavior of some phenomenon, it is said that model and respective hypothesis are supported by observations;
- $R : H \rightarrow M$ is the mapping from the set of hypotheses and into the models;
- W is the workflow, the set of tasks, orchestrated by specific constructs (workflow patterns — split, join, etc.). Each task represents a function with predefined signature, which invokes models from M .

Workflow implements experiment specifying when each model that conforms to related hypotheses should be invoked; and

- C is the configuration for each experiment run. It consists of a total mapping from workflow tasks into sets of function parameter values.

There exist many possible hypotheses representations — mathematical models, Boolean networks, ontologies, predicates in first-order logic, etc. Authors use ontologies to specify hypotheses.

Possible relationships between hypotheses are *competes_with*, which is used to relate competing hypotheses and *derived_by* to relate two hypotheses, one of which was used to derive another. *Derived_by* can be used to form hypotheses lattice [17] — algebraic structure with partial order relation. Hypotheses derived from a single hypothesis are atomic, otherwise — complex (see Fig. 2 below).

Model, which implements hypothesis, should conform to the hypothesis specification. If model generates expected behavior of some phenomenon, it is said that model and respective hypothesis are supported by observations.

Since hypotheses become the core artifact of VE, there is a shift in treating data to manage it successfully. Figure 1 depicts the process of specifying VE.

First, hypotheses are extracted from scientific articles. Usually, they are represented by text or formulas. Sometimes, there is a need to provide external hypotheses and to substitute the existing ones. Next step is to define mapping between hypotheses and models, which implement these hypotheses, and build some workflow

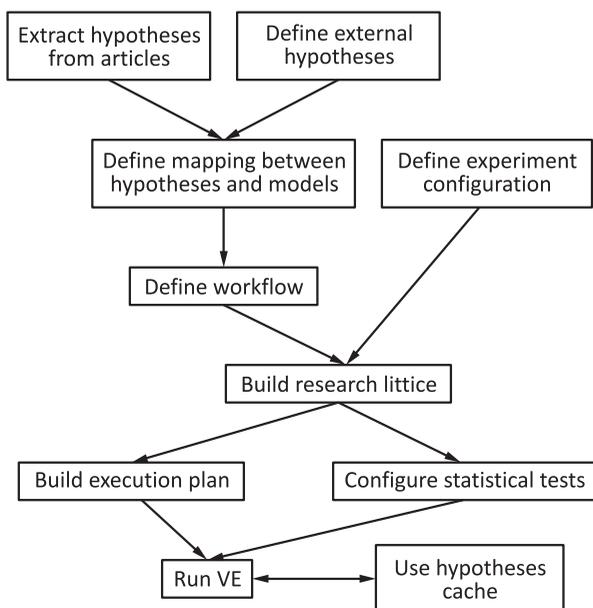


Figure 1 Methodology to form VE

specifying the sequence of tasks. Forming a research lattice is the next step. Virtual experiment needs configuration and execution plan. After that, one can launch VE.

Statistical testing is the process of deciding whether hypothesis contradicts observations. There exist two main approaches for statistical testing: *frequentist* and *Bayesian approach* discussed in the next section.

Formal specifications of VE, hypotheses, their relationship, and further clarification can be found in [4].

4 Hypotheses Quality Estimation

Important part of the methodology to form VEs is the configuration of statistical tests and hypotheses ranking. These tasks are interconnected, e. g., p -value from statistical test or Bayesian probability can be used as a metric to rank and range hypotheses.

The method for estimating the quality of a VE is based on the classical theory of signal processing [18]. This theory serves as a mean of quantifying the possibility of distinguishing the information component of a signal from noise [19].

The quality assessment tells when to rebuild a model and is represented with a statistical hypothesis:

H_0 : Model is correct.

H_1 : There is a malfunction in the model.

To make an assumption about the quality of the model, it is necessary to perform statistical test with incoming observations. One of the parameters transmitted to the input of our system once every time period is the measured value Y . During execution of the VE, the model provides the calculated value of the \hat{Y} . Estimating their difference $\hat{Y} - Y$ for n periods, we evaluate the quality of the model under the given experimental conditions.

As an example, often the difference $\hat{Y} - Y$ obeys the normal distribution law [20]. Both classical and frequency-based and the Bayesian approaches are used.

Frequentist Approach. In the framework of the classical approach, the probability is the relative frequency of occurrence of specific events. All events are independent. The general methodology for carrying out statistical testing is widely discussed in the literature.

The residuals obey the normal distribution law and the signal variance is unknown. Thus, within the framework of the frequency approach, a one-sample T -test with a two-way alternative is used.

As the random variable studied, $X_n = \hat{Y} - Y$ is the difference between the reference value of the flow rate coming from the sensors with the value of the flow velocity obtained as a result of the operation of the model:

$$X = (x_1, \dots, x_n) \in R, X_n \sim N(\mu, \sigma^2).$$

The hypothesis that the sample mean is equal to a given number against the alternative hypothesis that it is not equal:

$$\begin{aligned} H_0 : \bar{X} &= \mu ; \\ H_1 : \bar{X} &\neq \mu . \end{aligned}$$

In our case, $\mu = 0$ is assumed. Student's test is used. The statistics of the criterion has a Student's distribution with $n - 1$ degrees of freedom and a given level of significance $\alpha = 0.05$.

Bayesian Approach. The Bayesian approach differs from the classical one in treating the probability. In general, the probability is seen as the degree of confidence. Some *a priori* knowledge about the observation is refined during the experiment. A more detailed description of the application of the Bayesian approach in the theory of signal processing can be found in a survey [18].

Thus, the Bayesian approach verifies the likelihood coefficient. The empirical scale [2] of the evidential strength of the Bayesian criterion is provided in [3].

Evidence strength tells whether hypothesis H_0 can be rejected or not. Since hypothesis H_0 assumes that the model is seen as correct, its rejection serves as a quality assessment and indicates that the model begins to poorly describe observations.

As the random variable studied in the same way as in the frequency approach, $X_n = \hat{Y} - Y$ is the difference in the reference value of the flow rate coming from the sensors with the value of the flow velocity obtained as a result of the operation of the model. Thus, a sample is specified:

$$X = (x_1, \dots, x_n) \in R, \quad X_n \sim N(\mu, \sigma^2) .$$

The hypothesis is checked that the difference between the average observed and simulated signal is zero plus there is an additional component of some noise against the alternative hypothesis that in addition to noise, there is a nonzero component that informs that the signal has left the baseline:

$$\begin{aligned} H_0 : r_i &= 0 + n_i = n_i, \quad i = 1, 2, \dots, N ; \\ H_1 : r_i &= \mu + n_i, \quad i = 1, 2, \dots, N . \end{aligned}$$

The noise component obeys the normal distribution law. The final formula for calculating the likelihood ratio is given below. A more detailed derivation of the formula can be found in [3, 18].

5 Besancon Galaxy Model Use-Case

5.1 Main components

Authors' further experience on how to deal with batch execution of VEs and hypotheses is based on Besan-

con Galaxy Model (BGM). Besancon Galaxy Model is based on "the population synthesis approach ... aims at assembling current scenarios of galaxy formation and evolution, theories of stellar formation and evolution, models of stellar atmospheres and dynamical constraints, in order to make a consistent picture explaining currently available observations of different types (photometry, astrometry, spectroscopy) at different wavelengths" [21].

Besancon Galaxy Model which is being developed for more than 35 years represents a complex computational artifact, described in a series of [21–23] and presented in several major releases. Such a development represents a unique experience for catching the evolution scenarios for the model, changes to the model introduced both by using new observations (e. g., Hipparcos and Tycho-2 surveys) and by the theoretical progress in the field. Both minor changes to parameters of the model and huge improvements of the entire process were also made during the lifetime of the model. Also, the BGM authors enabled the community to change some parts of the model.

Due to the broad experience collected by the BGM authors in the respective articles and associated code, now there is a possibility to collect the requirements for the system to supports experiments and provide rationale to choosing the appropriate methods and adequate techniques for the infrastructure.

Besancon Galaxy Model takes as input hypotheses and their parameters. The examples of such hypotheses are star formation rate (SFR), initial mass function (IMF), density laws, evolutionary tracks, and so on [21]. As the model is evolving, new values for hypotheses parameters, even new parameters have been introduced into the BGM, e. g., for the IMF hypothesis in the last realization, there has not only been tests of several new values of the hypothesis but also separation of 2- and 3-slope instances of IMF has been done.

It is very important to explicitly catch the relationship between several hypotheses in VE. Hypotheses and their parameters can be interrelated. For example, stellar birthrate function is derived from both IMF and SFR functions and local volume density function is based on the provided density law. The relationships between hypotheses put constraints on the tuning of their parameters. Besancon Galaxy Model hypothesis lattice is depicted in Fig. 2.

The parameters of a single hypothesis can be linked to each other directly through equations. There are also indirect connections of parameters of several hypotheses, e. g., SFR parameter correlates with the slopes of IMF. This implies that one could not give the best solution for a particular variable without correlating it with others. So, there is a need to support for a correlation search

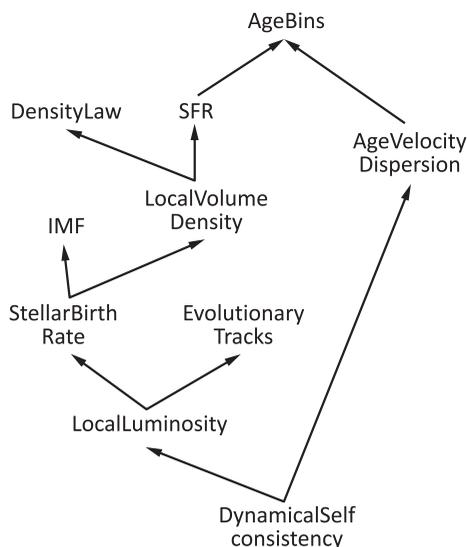


Figure 2 The BGM hypotheses lattice with derived by relationship

between hypotheses parameters and to store relationships between parameters of a single hypothesis.

Not all model ingredients are allowed to be changed by the user. This is done because if some hypothesis is changed in the model and no further adjustments for the dependent hypotheses are made, a model consistency is broken. Furthermore, the model has a property of being self-consistent meaning that when input values change, if it is possible, hypotheses derived by the one changed are properly adjusted in order not to break fundamental equations of astronomy. Therefore, derived by relationship needs to be modeled. Also, system component should enable the adjustment and calibration of any hypothesis available in the model.

Apart from explicit hypotheses, there are also implicit hypotheses in the model. They are not described in the articles and are tacit. The example of such hypothesis is that no stars come from outside of the Galaxy. It is important to explicitly store such hypotheses and understand how to extract such hypotheses from publications and data sources.

Workflow is used to implement BGM experiment specifying when each model which conforms to related hypotheses should be invoked. The workflow has also evolved since the first version, e. g., for thin disk treatment, new activities dependent on IMF and SFR hypotheses are introduced. This development can only be tracked using publications. Some activities in the model structure require the usage of statistical methods, tests, and tools which are used on both local hypotheses and on the general simulations from the whole experiment.

As the number of experiments is huge due to the increasing size of competing hypotheses family, now not

all the possible are run against the whole sky. Studying the ways to reduce the number of experiments which give the best fit and to choose when and if to abandon further computations of experiment is a major part of requirements to the new system. Using the information from experiment run done both locally and by other research groups can be helpful in achieving that goal.

Some research of data-intensive analysis emphasize the role of error bars. As the data in astronomy are provided usually with errors, the BGM uses special methods to work with such type of uncertainty. A component supporting statistical tools, which works with error bars, is a major requirement for the infrastructure.

5.2 Hypotheses specification

Examples of hypotheses and their relationships come from BGM. For the sake of clarity, not all hypotheses in BGM are specified. Every BGM hypothesis is treated as subclass of Hypothesis class. Example instances from each subclass are provided in [4].

Initial mass function is the mass distribution of a given population of stars and is represented by standard power law. Due to construction of the hypothesis in the BGM, IMF has a mathematical representation as a piecewise function with 2 or 3 pieces (slopes) where it is defined for mass regions. As there are just 2 possible sizes of the piecewise function, we put this into two disjoint subclasses. There are restrictions on available mass to Sol mass ratio. For IMF, there are 10 different versions of a hypothesis, 4 of them are 2-slope functions and 6 of them are 3-slope function. All of hypotheses are competing.

Star formation rate, $\Psi(t)$, represents the total mass of stars born per unit time per unit mass of Galaxy. Star formation rate has subclasses for representing constant $\Psi(t) = C$ and exponential function $\Psi(t) = \exp\{-\gamma t\}$ where γ is the parameter. Several competing hypotheses are specified as there are two possible values for γ (0.12 and 0.25) and one constant value. They can be stated as instances of respective classes.

Besancon Galaxy Model apart from the model ingredients has also implicit hypotheses which are not marked as ingredients. For example, (i) thin disk is divided into 7 age bins; and (ii) no stellar population comes from the outside of the galaxy.

For the first example, we can specify additional class *AgeBins* that has exactly 7 age bins.

It is more difficult to deal with the second one. As a possible solution, additional hypothesis could later be specified.

Hypotheses lattice is modeled with *derivedByobject* property. Some classes can be specified using *EquivalentClasses* construction. Hypotheses lattice for BGM is created manually but later, it should be constructed

automatically by system for executing experiments. Part of hypotheses lattice for BGM is shown in Fig. 2.

For IMF class, there are relations between slopes, *outputMass*, and *availableMass*. Based on *availableMass* parameter alpha is chosen and then *outputMass* is computed. If *availableMass* is inside the respective interval, alpha is taken and *outputMass* is computed. Next, postcondition for *ExponentSFR* is written. It says that born stars should have mass respective to the exponential equation. Other pre- and postconditions are specified in the same manner.

Since some hypotheses can take quite a few values, the number of possible models can reach thousands. This poses a question about the order of model execution and how to make these executions effective (and not to recompute previous unchanged results). For that, we use special structures to cache and store results. The system can put model execution in some order and use the results of previous executions. This could drastically increase the speed of model computation, especially on big amount of data. To implement this, we use the properties of hypotheses lattices.

The researcher can run several experiments finding the probability of each, which can be later queried by other researchers. For example, query described in [4] takes two experiments, which have underlying models best explaining observed data, and fixed values for hypothesis SFR and workflow specified by uniform resource identifier. Since there could be thousands of possible experiments, there is a need to order them by their probability. As in [9], we do not want the researched to bury in thousands of possible models and just take several the best ones.

6 Concluding Remarks

The article is aimed at development of a new approach for managing VE and its core components such as hypotheses, research lattices, ontologies, and workflows in a single manner. Analysis of existing systems, which explicitly state to manage VE or hypotheses, was carried out to extract and understand requirements for such systems.

Not all of these requirements come from the survey of existing systems but appear from studying astronomy use-case which add several important requirements for catching model evolution and real-time processing.

Future work is be concentrated on developing meta-system for handling hypotheses, models, and other metadata in VE.

Acknowledgments

This research was partially supported by the Russian Foundation for Basic Research (project 18-07-01434 A).

References

1. Hey, T., S. Tansley, and K. Tolle. 2009. *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research. 284 p.
2. Kalinichenko, L., D. Kovalev, D. Kovaleva, and O. Malkov. 2015. Methods and tools for hypothesis-driven research support: A survey. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):28–54.
3. Tarasov, E., and D. Kovalev. 2017. Otsenka kachestva nauchnykh gipotez v virtual'nykh eksperimentakh v oblastiakh s intensivnym ispol'zovaniem dannykh [Estimation of scientific hypotheses quality in virtual experiments in data intensive domains]. *CEUR Workshop Proceedings: 19th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers*. 2022:272–278.
4. Kovalev, D., L. Kalinichenko, and S. Stupnikov. 2017. Organization of virtual experiments in data-intensive domains: Hypotheses and workflow specification. *CEUR Workshop Proceedings: 19th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers*. 2022:293–300.
5. Demchenko, Y., P. Grosso, C. Laat, and P. Membrey. 2013. Addressing big data issues in scientific data infrastructure. *Conference (International) on Collaboration Technologies and Systems*. San Diego, CA: IEEE. 48–55.
6. Porto, F., and B. Schulze. 2013. Data management for eScience in Brazil. *Concurr. Comp. Pract. E.* 25(16):2307–2309.
7. Goncalves, B., F. Silva, and F. Porto. 2014. Upsilon-DB: A system for data-driven hypothesis management and analytics. arXiv:1411.7419 [cs.DB]. 6 p.
8. Kalinichenko, L., S. Stupnikov, A. Vovchenko, and D. Kovalev. 2014. Rule-based multi-dialect infrastructure for conceptual problem solving over heterogeneous distributed information resources. *New trends in databases and information systems*. Eds. B. Catania, T. Cerquitelli, S. Chiusano, et al. Advances in intelligent systems and computing ser. Springer. 241:61–68.
9. Duggan, J., and M. Brodie. 2015. Hephaestus: Data reuse for accelerating scientific discovery. *7th Biennial Conference on Innovative Data Systems Research*. Asilomar, CA. Paper 29. 12 p. Available at: http://users.eecs.northwestern.edu/~jennie/pubs/hephaestus_full.pdf (accessed June 11, 2019).
10. Schales, D., X. Hu, J. Jang, et al. 2015. FCCE: Highly scalable distributed feature collection and correlation engine for low latency big data analytics. *31st Conference (International) on Data Engineering*. Seoul, South Korea: IEEE. 1316–1327.
11. Ly, D., and H. Lipson. 2012. Learning symbolic representations of hybrid dynamical systems. *J. Mach. Learn. Res.* 13:3585–3618.
12. Tarasov, E. 2016. Sokrashchenie chisla virtual'nykh eksperimentov s pomoshch'yu otsenki korrelyatsii parametrov

- vzaimodeystvuyushchikh gipotez [Reducing the number of virtual experiments by estimating the correlation parameters of interacting hypotheses]. *CEUR Workshop Proceedings: 18th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers*. 1752:272–278.
13. Goncalves, B., and F. Porto. 2015. Managing scientific hypotheses as data with support for predictive analytics. *Comput. Sci. Eng.* 17(5):35–43.
 14. Molina, L., L. Belanche, and A. Nebot. 2002. Feature selection algorithms: A survey and experimental evaluation. *Conference (International) on Data Mining Proceedings*. IEEE. 306–313.
 15. Williams, N., S. Zander, and G. Armitage. 2006. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *ACM SIGCOMM Comp. Com.* 36(5):5–16.
 16. Hall, M. 1999. Correlation-based feature selection for machine learning. Hamilton, New Zealand: Waikato University, Department of Computer Science. PhD Thesis.
 17. Porto, F., A. Moura, B. Goncalves, R. Costa, and S. Spacapietra. 2012. A scientific hypothesis conceptual model. *Advances in conceptual modeling*. Eds. S. Castano, P. Vasiliadis, L. V. S. Lakshmanan, and M.-L. Lee. Lecture notes in computer science ser. Springer. 7518:101–110.
 18. Trees, H. 2013. *Detection, estimation, and modulation theory. Part I: Detection, estimation, and linear modulation theory*. 2nd ed. Hoboken, NJ: John Wiley & Sons. 1176 p.
 19. Rysak, A., G. Litak, and R. Mosdorf. 2016. Analysis of non-stationary signals by recurrence dissimilarity. *Recurrence plots and their quantifications: Expanding horizons*. Eds. C. L. Webber, Jr., C. Ioana, and N. Marwan. Springer proceedings in physics ser. Springer. 180:65–90.
 20. Brennen, C. 2005. *Fundamentals of multiphase flow*. New York, NY: Cambridge University Press. 410 p.
 21. Czekaj, M., A. Robin, F. Figueras, X. Luri, and M. Haywood. 2014. The Besancon Galaxy Model renewed. Constraints on the local star formation history from Tycho data. *Astron. Astrophys.* 564:A102. 20 p.
 22. Robin, A., and M. Creze. 1986. Stellar populations in the Milky Way — a synthetic model. *Astron. Astrophys.* 157(1):71–90.
 23. Robin, A., C. Reyle, S. Derriere, and S. Picaud. 2003. A synthetic view on structure and evolution of the Milky Way. *Astron. Astrophys.* 409(2):523–540.

Received November 13, 2018

Contributors

Kovalev Dmitry Y. (b. 1988) — junior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., 119333 Moscow, Russian Federation; dkovalev@ipiran.ru

Tarasov Evgeny A. (b. 1987) — programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., 119333 Moscow, Russian Federation; e.tarasov@outlook.com

ВИРТУАЛЬНЫЕ ЭКСПЕРИМЕНТЫ В ИССЛЕДОВАНИЯХ С ИНТЕНСИВНЫМ ИСПОЛЬЗОВАНИЕМ ДАННЫХ*

Д. Ю. Ковалёв, Е. А. Тарасов

Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Аннотация: Организация и управление виртуальными экспериментами в исследованиях с интенсивным использованием данных широко изучались в последние годы. Авторы рассматривают существующие подходы к работе с виртуальными экспериментами и гипотезами и анализируют управление виртуальными экспериментами на примере из области астрономии. Проведен обзор существующих систем, которые могут выступать в качестве платформ для проведения виртуального эксперимента. Были представлены требования к системе для организации виртуальных экспериментов в областях с интенсивным использованием данных, а также представлена общая структура и функциональность для систем, выполняющих виртуальные эксперименты. Обсуждаются взаимосвязи между гипотезами и моделями в виртуальном эксперименте. Также авторы иллюстрируют, как концептуально моделировать виртуальные эксперименты, соответствующие гипотезы и модели. Обсуждаются потенциальные преимущества и недостатки предлагаемого подхода, включая поддержание согласованности эксперимента и уменьшение потенциального числа экспериментов.

Ключевые слова: виртуальный эксперимент; гипотезы; концептуальное моделирование; исследования с интенсивным использованием данных

DOI: 10.14357/19922264190216

* Исследование выполнено при частичной финансовой поддержке РФФИ (проект 18-07-01434 А).

Литература

1. *Hey T., Tansley S., Tolle K.* The fourth paradigm: Data-intensive scientific discovery. — Redmond, WA, USA: Microsoft Research, 2009. 284 p.
2. *Kalinichenko L., Kovalev D., Kovaleva D., Malkov O.* Methods and tools for hypothesis-driven research support: A survey // Информатика и ее применения, 2015. Т. 9. Вып. 1. С. 28–54.
3. *Тарасов Е., Ковалёв Д.* Оценка качества научных гипотез в виртуальных экспериментах в областях с интенсивным использованием данных // CEUR Workshop Proceedings: 19th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers, 2017. Vol. 2022. P. 272–278.
4. *Kovalev D., Kalinichenko L., Stupnikov S.* Organization of virtual experiments in data-intensive domains: Hypotheses and workflow specification // CEUR Workshop Proceedings: 19th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers, 2017. Vol. 2022. P. 293–300.
5. *Demchenko Y., Grosso P., Laat C., Membrey P.* Addressing big data issues in scientific data infrastructure // Conference (International) on Collaboration Technologies and Systems. — San Diego, CA, USA: IEEE, 2013. P. 48–55.
6. *Porto F., Schulze B.* Data management for eScience in Brazil // Concurr. Comp. Pract. E., 2013. Vol. 25. Iss. 16. P. 2307–2309.
7. *Goncalves B., Silva F., Porto F.* Upsilon-DB: A system for data-driven hypothesis management and analytics // arXiv:1411.7419 [cs.DB], 2014. 6 p.
8. *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 / Eds. B. Catania, T. Cerquitelli, S. Chiusano, et al. — Advances in intelligent systems and computing ser. — Springer, 2014. Vol. 241. P. 61–68.
9. *Duggan J., Brodie M.* Hephaestus: Data reuse for accelerating scientific discovery // 7th Biennial Conference on Innovative Data Systems Research. — Asilomar, CA, USA, 2015. Paper 29. 12 p. http://users.eecs.northwestern.edu/~jennie/pubs/hephaestus_full.pdf.
10. *Schales D., Hu X., Jang J., et al.* FCCE: Highly scalable distributed feature collection and correlation engine for low latency big data analytics // 31st Conference (International) on Data Engineering. — Seoul, South Korea: IEEE, 2015. P. 1316–1327.
11. *Ly D., Lipson H.* Learning symbolic representations of hybrid dynamical systems // J. Mach. Learn. Res., 2012. Vol. 13. P. 3585–3618.
12. *Тарасов Е.* Сокращение числа виртуальных экспериментов с помощью оценки корреляции параметров взаимодействующих гипотез // CEUR Workshop Proceedings: 18th Conference (International) on Data Analytics and Management in Data Intensive Domains Selected Papers, 2016. Vol. 1752. P. 272–278.
13. *Goncalves B., Porto F.* Managing scientific hypotheses as data with support for predictive analytics // Comput. Sci. Eng., 2015. Vol. 17. Iss. 5. P. 35–43.
14. *Molina L., Belanche L., Nebot A.* Feature selection algorithms: A survey and experimental evaluation // Conference (International) on Data Mining Proceedings. — IEEE, 2002. P. 306–313.
15. *Williams N., Zander S., Armitage G.* A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification // ACM SIGCOMM Comp. Com., 2006. Vol. 36. Iss. 5. P. 5–16.
16. *Hall M.* Correlation-based feature selection for machine learning. — Hamilton, New Zealand: Waikato University, Department of Computer Science, 1999. PhD Thesis.
17. *Porto F., Moura A., Goncalves B., Costa R., Spaccapietra S.* A scientific hypothesis conceptual model // Advances in conceptual modeling / Eds. S. Castano, P. Vassiliadis, L. V. S. Lakshmanan, M.-L. Lee. — Lecture notes in computer science ser. — Springer, 2012. Vol. 7518. P. 101–110.
18. *Trees H.* Detection, estimation, and modulation theory. Part I: Detection, estimation, and linear modulation theory. — 2nd ed. — Hoboken, NJ, USA: John Wiley & Sons, 2013. 1176 p.
19. *Rysak A., Litak G., Mosdorf R.* Analysis of non-stationary signals by recurrence dissimilarity // Recurrence plots and their quantifications: Expanding horizons. — Eds. C. L. Webber, Jr., C. Ioana, N. Marwan. — Springer proceedings in physics ser. — Springer, 2016. Vol. 180. P. 65–90.
20. *Brennen C.* Fundamentals of multiphase flow. — New York, NY, USA: Cambridge University Press, 2005. 410 p.
21. *Czekaj M., Robin A., Figuera F., Luri X., Haywood M.* The Besancon Galaxy model renewed. Constraints on the local star formation history from Tycho data // Astron. Astrophys., 2014. Vol. 564. Art. A102. 20 p.
22. *Robin A., Creze M.* Stellar populations in the Milky Way — a synthetic model // Astron. Astrophys., 1986. Vol. 157. No. 1. P. 71–90.
23. *Robin A., Reyle C., Derriere S., Picaud S.* A synthetic view on structure and evolution of the Milky Way // Astron. Astrophys., 2003. Vol. 409. No. 2. P. 523–540.

Поступила в редакцию 13.11.2018

Абгарян Каринэ Карленовна (р. 1963) — доктор физико-математических наук, заведующая отделом Вычислительного центра имени А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; заведующая кафедрой Московского авиационного института (национального исследовательского университета)

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Анашин Владимир Сергеевич (р. 1951) — доктор физико-математических наук, профессор факультета Вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Арутюнов Евгений Наполеонович (р. 1952) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бахтеев Олег Юрьевич (р. 1991) — аспирант Московского физико-технического института

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Вахтанов Никита Андреевич (р. 1997) — студент магистратуры Московского института электроники и математики Национального исследовательского университета «Высшая школа экономики»

Грабовой Андрей Валериевич (р. 1997) — студент Московского физико-технического института

Гринченко Сергей Николаевич (р. 1946) — доктор технических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем инфор-

матики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забжайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

Инькова Ольга Юрьевна — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ковалёв Дмитрий Юрьевич (р. 1988) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Колесников Александр Васильевич (р. 1948) — доктор технических наук, профессор Института физико-математических наук и информационных технологий Балтийского федерального университета имени Иммануила Канта

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Коротков Владислав Владимирович (р. 1993) — аспирант факультета компьютерных наук Воронежского государственного университета

Кружков Михаил Григорьевич — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Листопад Сергей Викторович (р. 1984) — кандидат технических наук, старший научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Логачев Олег Алексеевич (р. 1950) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Лукашенко Олег Викторович (р. 1986) — кандидат физико-математических наук, научный сотрудник Института прикладных математических исследований Карельского научного центра Российской академии наук; доцент Петрозаводского государственного университета

Матвеев Михаил Григорьевич (р. 1949) — доктор технических наук, профессор, заведующий кафедрой факультета компьютерных наук Воронежского государственного университета

Миллер Григорий Борисович (р. 1980) — кандидат физико-математических наук, научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Морозов Евсей Викторович (р. 1947) — доктор физико-математических наук, профессор, ведущий научный сотрудник Института прикладных математических исследований Карельского научного центра Российской академии наук; профессор Петрозаводского государственного университета

Осипова Виктория Аркадьевна (р. 1945) — кандидат физико-математических наук, профессор Московского авиационного института (национального исследовательского университета)

Пагано Микеле (р. 1968) — доктор наук (PhD) по информационным технологиям, профессор Университета г. Пиза (Италия)

Стризов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Московского физико-технического института

Сукаев Альберт Атласович (р. 1994) — студент Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова

Тарасов Евгений Александрович (р. 1987) — программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Титова Анастасия Игоревна (р. 1995) — студентка факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Федоров Сергей Николаевич (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шнурков Петр Викторович (р. 1953) — кандидат физико-математических наук, доцент Национального исследовательского университета «Высшая школа экономики»

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 предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной 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 15 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

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;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

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.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/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 hozyaystvo [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:

Laturov, 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 modelirovaniye 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, ФИЦ ИУ 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 simultaneous 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 15 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

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;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/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 neftegazoidachi" [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 Publs. 352 p.
- Dissertation and Thesis:**
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.
- 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 Publs. 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:

FRC CSC RAS, 44, block 2, Vavilov Str., 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>