

Информатика и её применения

Том 17 Выпуск 1 Год 2023

СОДЕРЖАНИЕ

Аналитическое моделирование распределений с инвариантной мерой в стохастических системах, не разрешенных относительно производных И. Н. Синицын	2
An axiomatic viewpoint on the Rogers–Veraart and Suzuki–Elsinger models of systemic risk Yu. M. Kabanov and A. P. Sidorenko	11
Критерии нормальности вероятностного распределения при округленных данных В. Г. Ушаков, Н. Г. Ушаков	18
Среднее расстояние в конфигурационных графах со степенным распределением М. М. Лери	28
Восстановление матрицы суперпозиции в задаче символьной регрессии Р. Г. Нейчев, И. А. Шибеев, В. В. Стрижов	35
Причинно-следственные связи в задачах классификации А. А. Грушо, Н. А. Грушо, М. И. Забежайло, В. В. Кульченков, Е. Е. Тимонина, С. Я. Шоргин	43
Разработка новой модели ступенчатой сверточной нейронной сети для классификации аномалий на панорамах П. О. Архипов, С. Л. Филиппских, М. В. Цуканов	50
Моделирование структуры интероперабельности средствами структурной согласованности И. Н. Розенберг, С. К. Дулин, Н. Г. Дулина	57
Метод на основе нечетких правил для управления конфликтами агентов в гибридных интеллектуальных многоагентных системах С. В. Листопад, И. А. Кириков	66
О задаче оценки и анализа риска транспортных происшествий на рельсовом транспорте А. В. Босов, А. Н. Игнатов	73
Оценки распределения ресурсов в многопользовательской сети при равных межузловых нагрузках Ю. Е. Малашенко, И. А. Назарова	83
Об оптимизации работы резервного прибора в многолинейной системе массового обслуживания Я. М. Агаларов	89
Анализ схемы доступа с прерыванием при нарезке радиоресурсов сети пятого поколения К. И. Б. Аду, Е. В. Маркова, Ю. В. Гайдамака, С. Я. Шоргин	96
Многомерные баттерфляи в задачах оптимизации по CC-VaR Г. А. Агасандян	107
Данные, информация и знание в научной парадигме информатики И. М. Зацман	116
Об авторах	126
Правила подготовки рукописей	128
Requirements for manuscripts	131

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ С ИНВАРИАНТНОЙ МЕРОЙ В СТОХАСТИЧЕСКИХ СИСТЕМАХ, НЕ РАЗРЕШЕННЫХ ОТНОСИТЕЛЬНО ПРОИЗВОДНЫХ

И. Н. Сеницын¹

Аннотация: Рассматриваются точные и приближенные методы аналитического моделирования гауссовских и негауссовских стационарных и нестационарных стохастических процессов (СтП) с инвариантной мерой в стохастических системах, не разрешенных относительно производных (СтСНРОП). Для скалярных и векторных СтСНРОП, допускающих линейную регрессионную аппроксимацию нелинейных функций, содержащих старшие производные, разработаны методы сведения уравнений СтСНРОП к уравнениям дифференциальных стохастических систем (СтС). Предложены два точных метода аналитического моделирования СтС с инвариантной мерой. Разработаны приближенные методы аналитического моделирования, основанные на параметризации одно- и многомерных распределений. Особое внимание уделено гауссовским СтП с инвариантной мерой. В качестве примера рассмотрен нелинейный осциллятор Дуффинга, в котором вместо второй производной присутствует нелинейная функция от второй производной. Приведены уравнения нелинейной корреляционной теории. Изучены стационарные СтП, совпадающие с СтП с инвариантной мерой. Обсуждаются полученные результаты и формулируются направления дальнейших исследований в области аналитического моделирования СтСНРОП.

Ключевые слова: аналитическое моделирование; параметризация распределений; распределение с инвариантной мерой; стохастическая система (СтС); стохастическая система, не разрешенная относительно производной (СтСНРОП); стохастический процесс (СтП)

DOI: 10.14357/19922264230101

1 Введение

В [1] дан обзор точных методов аналитического моделирования одно- и многомерных распределений стационарных и нестационарных процессов, основанных на построении интегральных инвариантов специально подобранных обыкновенных дифференциальных уравнений. Особое внимание в [1] уделено случаю СтС с автокоррелированными шумами. В [2] на основе теории потенциала предложены методы расчета распределений с инвариантной мерой для гамильтоновых СтС. В [3] дано обобщение [1, 2] на случай разрывных характеристик в дифференциальных СтС. Дальнейшее развитие точных методов аналитического моделирования распределений в дифференциальных СтС, в том числе и в условиях автокоррелированных шумов, представлено в [4].

Вопросы аналитического моделирования СтП в СтСНРОП рассмотрены в [5–8]. Особое внимание в них уделено нормальным (гауссовским) СтП. В [9] предложены методы нормализации систем, стохастически не разрешенных относительно производных.

Рассмотрим обобщение [1–8] на случай, когда СтСНРОП приводятся к уравнениям дифференциальной СтС с гауссовскими и негауссовскими шумами. Представлены точные и приближенные методы аналитического моделирования гауссовских и негауссовских стационарных и нестационарных СтП с инвариантной мерой в СтСНРОП, допускающих линейную регрессионную аппроксимацию нелинейных функций, содержащих старшие производные. Разработаны методы сведения уравнений СтСНРОП к уравнениям дифференциальных СтС.

Предложены два точных метода аналитического моделирования СтС с инвариантной мерой. Разработаны приближенные методы аналитического моделирования, основанные на параметризации одно- и многомерных распределений. Особое внимание уделено гауссовским СтП с инвариантной мерой.

В качестве примера рассмотрен нелинейный осциллятор Дуффинга, в котором вместо второй производной присутствует нелинейная функция от второй производной. Приведены уравнения нелинейной корреляционной теории. Изучены стационарные СтП, совпадающие с СтП с инвариантной

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Московский авиационный институт, sinitin@dol.ru

мерой. В заключении статьи обсуждаются полученные результаты и формулируются направления дальнейших исследований в области аналитического моделирования СтСНРОП.

2 Приведение стохастических систем, не разрешенных относительно производных, к дифференциальным стохастическим системам

Рассмотрим сначала скалярную СтСНРОП следующего вида [8]:

$$\begin{aligned} \varphi &= \varphi(t, X_t, pX_t, \dots, p^{(l)}X_t, U_t) = 0; \\ dU_t &= a^U(U_t, t) dt + b^U(U_t, t) dW_0 + \\ &+ \int_{R_0^q} c^U(U_t, t, v) dP^0(t, dt). \end{aligned} \quad (1) \quad (2)$$

Здесь $p = d/dt$ — оператор дифференцирования по времени; $X_t = X(t)$ и $p^{(l)}X_t$ — скалярные СтП, понимаемые в среднеквадратичном смысле; $U_t = U(t)$ — скалярный СтП, определяемый уравнением Ито (2); φ — нелинейная, в общем случае разрывная, функция отмеченных переменных, допускающая линейную регрессионную аппроксимацию относительно старшей l -й производной U_t вида

$$\varphi \approx \varphi_0 + k_l^\varphi p^{(l)}X_t + k_U^\varphi U_t. \quad (3)$$

Здесь введены следующие обозначения:

$$\begin{aligned} \varphi_0 &= \varphi_0(t, \bar{X}_t, m^l, D^l); \\ k_l^\varphi &= k_l^\varphi(t, \bar{X}_t, m^l, D^l); \\ k_U^\varphi &= k_U^\varphi(t, \bar{X}_t, m^U, D^U), \end{aligned}$$

где

$$\begin{aligned} \bar{X}_t &= [X_t^T X_t^{(l-1)}]^T; \\ m^l &= M X_t^{(l)}; \quad D^l = M |X_t^{(l)} - m^l|^2. \end{aligned}$$

В уравнении (2) принято: $W_0 = W_0(t)$ — винеровский скалярный СтП интенсивности $\nu_0 = \nu_0(t)$; $c(u, t, v)$ — скалярная функция u и t , а также вспомогательной переменной v ; $\int_R dP^0(t, A) dt$ — центрированная пуассоновская мера, удовлетворяющая условию:

$$\int_{\Delta} dP^0(t, A) = \int_{\Delta} dP(t, A) dt - \int_{\delta} \nu_P(t, A) dt,$$

где $\int_{\Delta} dP(t, A) dt$ — число скачков пуассоновского СтП $P(t, A)$ в интервале времени Δ ; $\nu_P(t, A)$ — интенсивность пуассоновского СтП $P(t, A)$; $A = R_0$ — прямая с выколотым началом. Интеграл в (2) в общем случае распространяется на R_0 . Начальное значение $U(t_0) = U_0$ представляет собой случайную величину (СВ), не зависящую от приращений СтП $W_0(t)$ и $P(t, A)$ на интервалах времени $\Delta = (t_1, t_2]$, следующих за t_0 , $t_0 \leq t_1 \leq t_2$ для любого A .

Из условия конечности дисперсий скалярных переменных $p^{(l)}X_t$ и U_t следует соотношение

$$D [p^{(l)}X_t] \sim DU_t, \quad k_l^\varphi \neq 0.$$

Теорема 2.1. Пусть СтСНРОП (1) допускает линейную регрессионную линеаризацию согласно (3). Тогда уравнение (1) приводится к виду:

$$\begin{aligned} pX_{h-1} &= X_h \quad (h = \overline{1, l-1}), \\ pX_l &= -\varphi_0 (k_l^\varphi)^{-1} - k_U^\varphi (k_l^\varphi)^{-1} U_t. \end{aligned}$$

В векторном виде имеем следующие уравнения:

$$\bar{\varphi} = \bar{\varphi}(t, \bar{X}_t, \dot{\bar{X}}_t, \bar{U}_t) = 0; \quad (4)$$

$$\begin{aligned} d\bar{U}_t &= \bar{a}^U(t, \bar{U}_t) dt + \bar{b}^U(t, \bar{U}_t) dW_0 + \\ &+ \int_{R_0^q} c^U(t, U_t, v) dP_0(t, dt); \end{aligned} \quad (5)$$

$$\bar{\varphi} \approx \bar{\varphi}_0 + k_{\dot{\bar{X}}}^{\bar{\varphi}} \dot{\bar{X}}_t + k_{\bar{U}}^{\bar{\varphi}} \bar{U}_t; \quad (6)$$

$$\dot{\bar{X}}_t \approx a^{\dot{\bar{X}}} = -\bar{\varphi}_0 \left(k_{\dot{\bar{X}}}^{\bar{\varphi}}\right)^{-1} - k_{\bar{U}}^{\bar{\varphi}} U_t \left(k_{\dot{\bar{X}}}^{\bar{\varphi}}\right)^{-1}.$$

Здесь матричные коэффициенты регрессионной линеаризации удовлетворяют условиям:

$$k_{\dot{\bar{X}}}^{\bar{\varphi}} = k_{\dot{\bar{X}}}^{\bar{\varphi}}(t, \bar{X}_t, m^{\dot{\bar{X}}}, K^{\dot{\bar{X}}}), \quad \det k_{\dot{\bar{X}}}^{\bar{\varphi}} \neq 0;$$

$$K_{\bar{U}}^{\bar{\varphi}} = k_{\bar{U}}^{\bar{\varphi}}(t, \bar{X}_t, m^U, K^U), \quad \det K_{\bar{U}}^{\bar{\varphi}} \neq 0.$$

В уравнении (5) принято: U_t — $n^{\bar{U}}$ -мерный СтП; \bar{a}^U , \bar{b}^U и c^U — известные функции отмеченных переменных; $W_0(t)$ — n^{W_0} -мерный винеровский СтП матричной интенсивности $\nu_0 = \nu_0(t)$; $P(t, A)$ — пуассоновский СтП интенсивности $\nu_P(t, A)$; $A = R_0^{n^P}$ — некоторое борелевское множество пространства с выколотым началом.

Теорема 2.2. Пусть векторная СтСНРОП (4) допускает линейную регрессию (6). Тогда векторное уравнение (4) приводится к виду (6), а уравнение для U_t в (5) должно допускать решение с конечными моментами второго порядка.

3 Точные методы аналитического моделирования распределений стохастических процессов с инвариантной мерой

Введем расширенный вектор состояния приведенной СтСНРОП $Z_t = [\bar{X}_t^T U_t^T]^T$ и векторное стохастическое дифференциальное уравнение Ито для него:

$$dZ_t = \bar{a}(Z_t, t) dt + b(Z_t, t) W_0 + \int_{R^0} c(Z_t, t, v) dP^0(t, dv). \quad (7)$$

Здесь

$$a = a(Z_t, t) = \begin{bmatrix} a^X \\ a^U \end{bmatrix}; \quad b = b(Z_t, t) = \begin{bmatrix} 0 \\ \bar{b}^U \end{bmatrix};$$

$$c = c(Z_t, t, v) = \begin{bmatrix} 0 \\ c^{\bar{v}} \end{bmatrix}.$$

Пусть существуют одно- и n -мерные плотности $f_1 = f_1(z; t)$ и $f_n = f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ и характеристические функции $g_1 = g_1(\lambda; t)$ и $g_n = g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$ ($n \geq 2$), удовлетворяющие интегро-дифференциальным уравнениям Пугачёва [10]:

$$\frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a(z, t) f_1(z; t)] =$$

$$= \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^f(\lambda, \zeta, t) e^{i\lambda^T(\zeta - z)} f_1(z; t) d\zeta d\lambda,$$

$$f_1(z; t_0) = f_0(z); \quad (8)$$

$$\frac{\partial f_n(z_1, \dots, z_n; t_1, \dots, t_n)}{\partial t_n} =$$

$$= \frac{\partial^T}{\partial z_n} [a(z_n, t_n) f_n(z_1, \dots, z_n; t_1, \dots, t_n)] =$$

$$= \frac{1}{(2\pi)^{pn}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi_n^f(\lambda_n, \zeta_n, t_n) \times$$

$$\times \exp \left\{ i \sum_{l=1}^n \lambda_l^T (\zeta_l - z_l) \right\} f_n(\zeta_1, \dots, \zeta_n;$$

$$t_1, \dots, t_n) d\zeta_1 \dots d\zeta_n d\lambda_1 \dots d\lambda_n,$$

$$f_n(z_1, \dots, z_{n-1}, z_n; t_1, \dots, t_{n-1}, t_n) =$$

$$= f_{n-1}(z_1, \dots, z_{n-1}; t_1, \dots, t_{n-1}) \delta(z_n - z_{n-1}),$$

$$t_1 \leq t_2 \leq \dots \leq t_n, \quad n = 2, 3, \dots; \quad (9)$$

$$\frac{\partial g_1(\lambda; t)}{\partial t} - \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i\lambda^T a(z, t) e^{i(\lambda^T - \mu^T)z} \times$$

$$\times g_1(\mu; t) d\mu dz = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^g(\lambda, z, t) \times$$

$$\times e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz,$$

$$g_1(\lambda; t_0) = g_0(\lambda); \quad (10)$$

$$\frac{\partial g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)}{\partial t_n} -$$

$$- \frac{1}{(2\pi)^{pn}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} i\lambda^T a(z_n, t_n) \times$$

$$\times \exp \left[i \sum_{k=1}^n (\lambda_k^T - \mu_k^T) z_k \right] g_n(\mu_1, \dots, \mu_n;$$

$$t_1, \dots, t_n) d\mu_1 \dots d\mu_n dz_1 \dots dz_n =$$

$$= \frac{1}{(2\pi)^{pn}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \chi^n(\lambda_n, z_n, t_n) \times$$

$$\times \exp \left[i \sum_{k=1}^n (\lambda_k^T - \mu_k^T) z_k \right] g_n(\mu_1, \dots, \mu_n;$$

$$t_1, \dots, t_n) d\mu_1 \dots d\mu_n dz_1 \dots dz_n,$$

$$g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_{n-1}, t_n) =$$

$$= g_{n-1}(\lambda_1, \dots, \lambda_{n-2}, \lambda_{n-1} + \lambda_n; t_1, \dots, t_{n-1})$$

$$t_1 \leq t_2 \leq \dots \leq t_n, \quad n = 2, 3, \dots$$

Здесь приняты следующие обозначения:

$$\chi^f(\lambda, \zeta, t) = -\frac{1}{2} \lambda^T b(\zeta, t) \nu_0(t) b(\zeta, t)^T +$$

$$+ \int_{R_0^q} \left\{ \exp [i\lambda^T c(\zeta, t, v)] - 1 - i\lambda^T c(\zeta, t, v) \right\} \nu_P(t, dv);$$

$$\chi_n^f(\lambda_n, \zeta_n, t_n) = -\frac{1}{2} \lambda_n^T b(\zeta_n, t) \nu_0(t) b(\zeta_n, t)^T +$$

$$+ \int_{R_0^q} \left\{ \exp [i\lambda_n^T c(\zeta_n, t_n, v)] - \right.$$

$$\left. - 1 - i\lambda_n^T c(\zeta_n, t_n, v) \right\} \nu_P(t_n, dv);$$

$$\chi^g(\lambda, z, t) = -\frac{1}{2} \lambda^T b(z, t) \nu_0(t) b(z, t)^T +$$

$$+ \int_{R_0^q} \left\{ \exp [i\lambda^T c(z, t, v)] - 1 - i\lambda^T c(z, t, v) \right\} \nu_P(t, dv);$$

$$\chi_n^g(\lambda_n, z_n, t_n) = -\frac{1}{2} \lambda_n^T b(z_n, t) \nu_0(t) b(z_n, t)^T +$$

$$+ \int_{R_0^q} \left\{ \exp [i\lambda_n^T c(z_n, t_n, v)] - \right.$$

$$\left. - 1 - i\lambda_n^T c(z_n, t_n, v) \right\} \nu_P(t_n, dv).$$

При этом одно- и n -мерные плотности и характеристические функции связаны между собой известными соотношениями [10].

Для нахождения одномерных плотностей $f_1(z, t) = f_1^*(z)$ и характеристических функций $g_1(\lambda; t) = g_1^*(\lambda)$ СтП в стационарных системах (7), когда

$$a(z, t) = a^*(z); \quad b(z, t) = b^*(z);$$

$$\chi(\mu; t) = \chi^f(\mu, \zeta, t) = \chi^{*f}(\mu, \zeta),$$

в (8) и (9) следует положить

$$\frac{\partial f_1}{\partial t} = 0; \quad \frac{\partial g_1}{\partial t} = 0.$$

Пусть функция a в системе (7) допускает представление

$$a = a(z, t) = a_1(z, t) + a_2(z, t), \quad (11)$$

при котором функция $f_1 = f_1(z; t)$ является плотностью инвариантной меры, не возмущенной шумами системы, описываемой векторным обыкновенным дифференциальным уравнением вида

$$\dot{z} = a_1(z, t), \quad (12)$$

т.е. удовлетворяет следующему условию сохранения инвариантной меры:

$$\frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a_1(z, t) f_1(z; t)] = 0. \quad (13)$$

Для гладких функций $a_1 = a_1(z, t)$ вопросы существования и основные свойства интегральных инвариантов и инвариантных мер изучены в [1, 4]. При этом функция $a_2 = a_2(z, t)$ в (11) определяется путем решения следующего интегро-дифференциального уравнения:

$$\frac{\partial^T}{\partial z} [a_2(z, t) f_1(z; t)] =$$

$$= \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^f(\lambda, \zeta, t) e^{i\lambda^T(\zeta - z)} f_1(\zeta; t) d\zeta d\lambda. \quad (14)$$

Условия сохранения инвариантной меры можно представить в следующем развернутом виде [3]:

$$\frac{\partial f_1(z; t)}{\partial t} + A_a f_1(z; t) = 0, \quad (15)$$

где

$$A_a f_1(z; t) = \frac{\partial^T}{\partial z} [a_1(z, t) f_1(z; t)] = \text{div } \pi(z; t);$$

$$\frac{\partial g_1(\lambda; t)}{\partial t} - B_a g_1(\lambda; t) = 0, \quad (16)$$

где

$$B_a g_1(\lambda; t) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i\lambda^T a_1(z, t) \times$$

$$\times e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz =$$

$$= \int_{-\infty}^{\infty} i\lambda^T a(z, t) e^{i\lambda^T z} f_1(z; t) dz = \int_{-\infty}^{\infty} e^{i\lambda^T z} i\lambda^T \pi(z; t) dz.$$

Для разрывных функций $a_1(z, t)$ в терминах характеристических функций соотношение (13) может быть записано в виде (16). При этом для составляющих $a_2(z, t)$ имеет место уравнение

$$B_{a_2} g_1(\lambda; t) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^g(\lambda, z, t) \times$$

$$\times e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz. \quad (17)$$

Отсюда вытекают точные алгоритмы аналитического моделирования распределений с инвариантной мерой. В их основе лежат следующие две теоремы.

Теорема 3.1. *Функция $f_1 = f_1(z; t)$ будет решением (8) тогда и только тогда, когда $a = a(z, t)$ допускает представление (11), такое что $f_1 = f_1(z; t)$ является плотностью инвариантной меры обыкновенного дифференциального уравнения (14), т.е. удовлетворяет условию (13). При этом составляющая a_2 определяется из решения интегро-дифференциального уравнения (14).*

Теорема 3.2. *Функция $g_1 = g_1(\lambda; t)$ будет решением (10) тогда и только тогда, когда недифференцируемая функция $a = a(z, t)$ допускает представление (11), такое что $g_1 = g_1(\lambda; t)$ является характеристической функцией инвариантной меры уравнения (12), т.е. удовлетворяет условию (15). При этом составляющая a_2 определяется из уравнения (17).*

4 Приближенные методы аналитического моделирования распределений стохастических процессов с инвариантной мерой

Пусть нелинейная система (7) допускает применение метода нормальной аппроксимации (МНА) [10]. Тогда одно- и двумерные нормальные плотности $f_1^{\text{МНА}}$ и $f_2^{\text{МНА}}$ и характеристические функции $g_1^{\text{МНА}}$ и $g_2^{\text{МНА}}$, а также вектор математического ожидания $m_t = M^{\text{МНА}} Z(t)$, ковариационная матрица $K_t = M^{\text{МНА}} Z^{0T} Z^0(t)$ ($Z^0(t) = Z(t) - m_t$)

и матрица ковариационных функций $K(t_1, t_2) = M^{\text{MHA}} Z^{0T}(t_1) Z^0(t_2)$ ($t_1 < t_2$) определяются следующими уравнениями:

$$f_1^{\text{MHA}} = f_1^{\text{MHA}}(z; t, m_t, K_t) = [(2\pi)^p |K_t|]^{-1/2} \times \exp \left\{ -\frac{1}{2} (z^T - m_t^T) K_t^{-1} (z - m_t) \right\}; \quad (18)$$

$$f_2^{\text{MHA}} = f_2^{\text{MHA}}(z_1, z_2; t_1, t_2, m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2)) = [(2\pi)^p |\bar{K}_2|]^{-1/2} \times \exp \left\{ -\left([z_1^T z_2^T] - \bar{m}_2^T \right) \bar{K}_2^{-1} \left([z_1^T z_2^T]^T - \bar{m}_2 \right) \right\}; \quad (19)$$

$$g_1^{\text{MHA}}(\lambda; t) = \exp \left\{ i\lambda^T m - \frac{1}{2} \lambda^T K_t \lambda \right\}; \quad (20)$$

$$g_2^{\text{MHA}}(\lambda_1, \lambda_2; t_1, t_2) = \exp \left\{ i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \bar{\lambda} \right\}; \quad (21)$$

$$\left. \begin{aligned} \bar{\lambda} &= [\lambda_1^T \lambda_2^T]^T; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T; \\ \bar{K}_2 &= \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix}; \\ \dot{m}_t &= \Phi_1(t, m_t, K_t) = \\ &+ \int_{-\infty}^{\infty} a(z, t) f_1^{\text{MHA}}(z; t, m_t, K_t) dz; \end{aligned} \right\} \quad (22)$$

$$\begin{aligned} \dot{K}_t &= \Phi_2(t, m_t, K_t) = \Phi_{21} + \Phi_{12} + \Phi_{22} = \\ &= \left[\int_{-\infty}^{\infty} a(z, t) (z^T - m_t^T) + (z - m_t) a^T(z, t) + \right. \\ &\quad \left. + \bar{\sigma}(z, t) \right] f_1^{\text{MHA}}(z; t, m_t, K_t) dz; \quad (23) \end{aligned}$$

$$\begin{aligned} \frac{\partial K(t_1, t_2)}{\partial t_2} &= \\ &= \Phi_3(t_1, t_2, m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2)) = \\ &= [(2\pi)^{2p} |\bar{K}_2|]^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{t_1}) a(z_2, t_2) \times \\ &\quad \times \exp \left\{ -\left([z_1^T z_2^T] - \bar{m}_2^T \right) \bar{K}_2^{-1} \times \right. \\ &\quad \left. \times \left([z_1^T z_2^T] - \bar{m}_2 \right) \right\} dz_1 dz_2 = \\ &= K(t_1, t_2) K(t_2)^{-1} \Phi_{21}(m(t_2), K(t_2), t_2)^T. \quad (24) \end{aligned}$$

Здесь введены следующие обозначения:

$$\begin{aligned} z_1 &= z_{t_1}; \quad z_2 = z_{t_2}; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T; \\ \bar{K}_2 &= \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix}; \\ \bar{\sigma}(z, t) &= \sigma(z, t) + \int_{R_0^q} c(z, t, v) c(z, t, v)^T \nu_P(t, dv), \\ \sigma(z, t) &= b(z, t) \nu_0(t) b(z, t)^T. \end{aligned}$$

Условия наличия нормального распределения с инвариантной мерой, если заменить $a(z, t)$ статистически линейризованным выражением вида

$$a(Z, t) \approx a_{10}^{\text{MHA}}(t, m_t, K_t) + a_{11}^{\text{MHA}}(t, m_t, K_t) (Z - m_t),$$

где

$$\begin{aligned} a_{10}^{\text{MHA}} &= a_{10}^{\text{MHA}}(t, m_t, K_t); \\ a_{11}^{\text{MHA}} &= a_{11}^{\text{MHA}}(t, m_t, K_t) = \left[\int_{-\infty}^{\infty} a(z, t) \times \right. \\ &\quad \left. \times (z^T - m_t^T) f_1^{\text{MHA}}(z; t, m_t, K_t) dz \right] K_t^{-1} = \\ &= \left[\frac{\partial}{\partial m_t} (a_{10}^{\text{MHA}})^T \right]^T, \end{aligned}$$

приводят к следующим соотношениям:

$$\begin{aligned} \frac{\partial f_1^{\text{MHA}}(z; t, m_t, K_t)}{\partial t} + \frac{\partial^T}{\partial z} \left\{ [a_{10}^{\text{MHA}}(t, m_t, K_t) + \right. \\ \left. + a_{11}^{\text{MHA}}(t, m_t, K_t) (z - m_t)] f_1^{\text{MHA}}(z; t, m_t, K_t) \right\} = \\ = 0; \quad (25) \end{aligned}$$

$$\begin{aligned} \frac{\partial g_1^{\text{MHA}}(\lambda; t)}{\partial t} - \int_{-\infty}^{\infty} i\lambda^T [a_{10}^{\text{MHA}}(m_t, K_t, t) + \\ + a_{11}^{\text{MHA}}(m_t, K_t, t) (z - m_t)] \times \\ \times e^{i\lambda^T z} f_1^{\text{MHA}}(z; m_t, K_t, t) dz = 0; \quad (26) \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} i\lambda^T [a_{10}^{*\text{MHA}}(m^*, K^*) + \\ + a_{11}^{*\text{MHA}}(m^*, K^*) (z - m^*)] \times \\ \times e^{i\lambda^T z} f_1^{*\text{MHA}}(z; m^*, K^*) dz = 0. \end{aligned}$$

Отсюда вытекает следующая теорема.

Теорема 4.1. Если существуют одно- и двумерные плотности СтП, а матрица a_{11}^{MHA} коэффициентов статистической линейризации асимптотически устойчива, то приближенный алгоритм аналитического моделирования МНА нестационарных СтП в системе (7) с инвариантной мерой определяется выражениями (18)–(24) и (25).

Как известно [10], одно- и двумерные нормальные распределения определяют и все n -мерные распределения ($n > 3$). Поэтому МНА и МСЛ при $b(Y, t) = b_0(t)$ и $c(Y, t, z) = c_0(t, v)$ дают приближенные алгоритмы для любых многомерных плотностей СтП, если они существуют.

Обобщением МНА являются различные приближенные методы, основанные на параметри-

зации распределений [10]. Аппроксимируя одномерную характеристическую функцию $g_1(\lambda; t)$ и соответствующую плотность $f_1(z, t)$ известными функциями $g_1^*(\lambda; \theta)$ и $f_1^*(z; \theta)$, зависящими от конечномерного векторного параметра θ , можно свести задачу приближенного определения одномерного распределения к выводу из уравнения для характеристических функций обыкновенных дифференциальных уравнений, определяющих θ как функцию времени. Это относится и к остальным многомерным распределениям [6].

5 Пример

Рассмотрим скалярную СтСНРОП второго порядка следующего вида:

$$\begin{aligned} \varphi_1(\ddot{X}) + \omega_0^2 X - \mu X^3 &= -\delta \dot{X} + U; \\ \dot{U} &= -\alpha U + \beta V, \end{aligned}$$

где φ_1 — нелинейная функция, допускающая регрессионную линеаризацию

$$\varphi_1(\ddot{X}) \approx \varphi_{10} \left(m^{\ddot{X}}, D^{\ddot{X}} \right) + k_{\ddot{X}}^{\varphi_1} \left(m^{\ddot{X}}, D^{\ddot{X}} \right) \ddot{X}$$

при условии

$$\det k_{\ddot{X}}^{\varphi} \neq 0;$$

$\omega_0, \gamma, \delta, \alpha$ и β — постоянные параметры; V — белый нормальный шум интенсивности ν_0 . В переменных $X_1 = X, \dot{X}_1 = \dot{X}_2$ и $X_3 = U$ развернутые уравнения эквивалентной дифференциальной СтС имеют вид:

$$\left. \begin{aligned} \dot{X}_1 &= X_2; \\ \dot{X}_2 &= -\bar{\varphi}_{10} - \bar{\omega}_0^2 X_1 + \bar{\mu} X_1^3 - \bar{\delta} X_2 + \bar{\gamma} X_3; \\ \dot{X}_3 &= -\alpha X_3 + \beta. \end{aligned} \right\} (27)$$

Здесь введены обозначения:

$$\begin{aligned} \bar{\varphi}_{10} &= \bar{\varphi}_{10} \left(t, m^{\dot{X}_2}, D^{\dot{X}_2} \right); \quad \bar{\omega}_0^2 = \bar{\omega}_0^2 \bar{\gamma}; \\ \bar{\mu} &= \mu \bar{\gamma}; \quad \bar{\delta} = \delta \bar{\gamma}; \\ \bar{\gamma} &= \left[k_{\dot{X}_2}^{\varphi_1} \left(t, m^{\dot{X}_2}, D^{\dot{X}_2} \right) \right]^{-1}. \end{aligned}$$

Далее выполним статистическую линеаризацию кубической нелинейности согласно [10]:

$$\begin{aligned} X_1^3 &= m_1 (m_1^2 + 3D_1) + 3 (m_1^2 + D_1) X_1^0 = \\ &= -2m_1^3 + 3 (m_1^2 + D_1) X_1. \end{aligned} \quad (28)$$

Тогда (27) приводятся к виду:

$$\left. \begin{aligned} \dot{X}_1 &= X_2; \\ \dot{X}_2 &= -\bar{\varphi}_{10} - \bar{\omega}_0^2 X_1 - \bar{\delta} X_2 + \bar{\gamma} X_3; \\ \dot{X}_3 &= -\alpha X_3 + \beta V. \end{aligned} \right\} (29)$$

Здесь дополнительно обозначено

$$\left. \begin{aligned} \bar{\varphi}_{10} &= -\bar{\varphi}_{10} - 2m_1^3; \\ \bar{\omega}_0^2 &= \omega_0^2 \left[1 - \frac{3(m_1^2 + D_1)\mu}{\omega_0^2} \right] \bar{\gamma}. \end{aligned} \right\} (30)$$

В силу (29) и (30) для вычисления $m^{\ddot{X}} = m^{X_2}$ и $D^{\ddot{X}} = D^{X_2}$ используются следующие формулы связи:

$$\begin{aligned} m^{\ddot{X}} &= m^{\dot{X}_2} = -\bar{\varphi}_{10} - \bar{\omega}_0^2 m_1 - \bar{\delta} m_2 + \bar{\gamma} m_3; \\ D^{\ddot{X}} &= D^{\dot{X}_2} = \text{M} \left[\left| -\bar{\omega}_0^2 X_1^0 - \bar{\delta} X_2^0 + \bar{\gamma} X_3^0 \right|^2 \right] = \\ &= \bar{\omega}_0^4 D_1 + \bar{\delta}^2 D^2 + \bar{\gamma}^2 D_3 + 2\bar{\omega}_0^2 \bar{\delta} K_{12} - \\ &\quad - \omega_0^2 \bar{\gamma} K_{13} - 2\bar{\delta} \bar{\gamma} K_{23}. \end{aligned}$$

Уравнения (29) для математических ожиданий m_i , дисперсий и ковариаций K_{ij} ($i, j = 1, 2, 3$) связаны между собой параметрически вследствие как нелинейности φ_1 , так и кубической нелинейности и имеют следующий вид:

$$\left. \begin{aligned} \dot{m}_1 &= m_2; \\ \dot{m}_2 &= -\bar{\varphi}_{10} - \bar{\omega}_0^2 m_1 - \bar{\delta} m_2 + \bar{\gamma} m_3; \\ \dot{m}_3 &= -\alpha m_3; \end{aligned} \right\} (31)$$

$$\left. \begin{aligned} \dot{X}_1^0 &= X_2^0; \\ \dot{X}_2^0 &= -\bar{\omega}_0^2 X_1^0 - \bar{\delta} X_2^0 + \bar{\gamma} X_3^0; \\ \dot{X}_3^0 &= -\alpha X_3^0 + V. \end{aligned} \right\} (32)$$

Из второго и третьего уравнений (31) для $\alpha > 0$ и $\bar{\delta} > 0$ при $t \gg t_0$ получаем соотношения для стационарных математических ожиданий:

$$m_3^* = 0; \quad m_2^* = 0; \quad \omega_0^2 m_1^* = -\bar{\varphi}_{10}. \quad (33)$$

Уравнения для дисперсий D_i и ковариаций K_{ij} в силу (32) можно представить в виде:

$$\dot{K} = AK + KA^T + b\nu_0 b^T,$$

где

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -\bar{\omega}_0^2 & -\bar{\delta} & \bar{\gamma} \\ 0 & 0 & -\alpha \end{bmatrix}; \quad b = \begin{bmatrix} 0 \\ 0 \\ \beta \end{bmatrix}, \quad (34)$$

или в развернутом виде:

$$\left. \begin{aligned} \dot{D}_1 &= 2K_{12}; \\ \dot{D}_2 &= 2(-\bar{\omega}_0^2 K_{12} - \bar{\delta} D_2 + \bar{\gamma}_1 D_3); \\ \dot{D}_3 &= -2\alpha D_3 + \beta^2 \nu_0; \\ \dot{K}_{12} &= D_2 - \bar{\omega}_0^2 D_1 - \bar{\delta} K_{12} + \bar{\gamma}_1 K_{13}; \\ \dot{K}_{13} &= K_{23} - \alpha K_{13}; \\ \dot{K}_{23} &= -\bar{\omega}_0^2 K_{13} - \bar{\delta} K_{23} + \bar{\gamma} D_3. \end{aligned} \right\} (35)$$

Здесь дополнительно введены обозначения:

$$\omega_{1_3}^2 = \omega^2 \bar{\gamma} \left[1 - \frac{3(m_1^2 + D_1)}{\omega^2} \right]; \quad \bar{\gamma}_1 = k_U^{\varphi_1} \bar{\gamma}.$$

В стационарном режиме для асимптотически устойчивой матрицы A в (34) в уравнениях (35) следует положить правые части равными нулю. Тогда получим искомые совместные уравнения для стационарных дисперсий и ковариаций. Решив эти уравнения, придем к следующим соотношениям:

$$\left. \begin{aligned} D_2^* &= \frac{\beta^2 \nu_0 \bar{\gamma}_1}{2\alpha \bar{\delta}^*}; \quad D_3^* = \frac{\beta^2 \nu_0}{2\alpha}; \\ K_{12}^* &= 0; \quad K_{13}^* = \frac{\beta^2 \nu_0 \bar{\gamma}_2^*}{2\alpha(\bar{\omega}_{1_3}^{*2} + \alpha \bar{\delta}^*)}; \\ K_{23}^* &= \frac{\beta^2 \nu_0 \bar{\gamma}_1^*}{2(\bar{\omega}_{1_3}^{*2} + \alpha \bar{\delta}^*)}; \end{aligned} \right\} \quad (36)$$

$$D_1^* \bar{\omega}_{1_3}^{*2} = \frac{\beta^2 \nu_0 \bar{\gamma}_1^*}{2\alpha \bar{\delta}^*} \left[1 + \frac{\bar{\gamma}_1^* \bar{\delta}^*}{(\bar{\omega}_{1_3}^{*2} + \alpha \bar{\delta}^*)} \right]. \quad (37)$$

Соотношения (36) параметрически зависят от m_1^* и D_1^* посредством $\bar{\omega}_{1_3}^{*2}$, $\bar{\delta}^*$ и $\bar{\gamma}_1^*$. Количества m_1^* и D_1^* определяются (33) и (37).

Для разрывной нелинейной функции $\varphi_1(\ddot{X})$ коэффициенты регрессионной линеаризации имеют следующий вид:

$$\begin{aligned} \varphi_1(\ddot{X}) &= \text{sgn} \ddot{X} \approx \varphi_0(\zeta) + k \ddot{X}^0; \\ \zeta &= \frac{m \ddot{X}}{\sqrt{D \ddot{X}}}; \quad \Phi(\zeta) = \frac{1}{\sqrt{2\pi}} \int_0^\zeta e^{-t^2/2} dt; \quad k = \frac{\partial \varphi_0}{\partial m \ddot{X}}. \end{aligned}$$

Для кубической нелинейности имеем формулы (28).

6 Заключение

Получено обобщение точных и приближенных методов аналитического моделирования стационарных и нестационарных СтП с инвариантной мерой в гауссовских и негауссовских СтСНРОП, приводимых к дифференциальным СтС. Для негауссовских СтСНРОП особое внимание уделено применению методов нормальной аппроксимации и статистической линеаризации для нахождения корреляционных характеристик нормальных распределений с инвариантной мерой. Полученные результаты могут быть использованы в задачах установления эквивалентности гауссовских и негауссовских СтСНРОП.

Разработан комплекс тестовых примеров для экспериментального инструментального программного обеспечения StS-IMD в среде MATLAB.

Проведенные вычислительные эксперименты подтверждают достаточную точностью для инженерных применений [11, 12].

Среди направлений повышения точности аналитического моделирования можно отметить методы, основанные на прямом численном интегрировании исходных уравнений СтСНРОП.

Литература

1. Синецын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с автокоррелированными шумами // Информатика и её применения, 2012. Т. 6. Вып. 4. С. 4–8.
2. Soize C. The Fokker–Plank equation for stochastic dynamical systems and its explicit steady state solutions. — Singapore: World Scientific, 1994. 321 p.
3. Синецын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с разрывными характеристиками // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 3–11.
4. Синецын И. Н. Развитие методов аналитического моделирования распределений с инвариантной мерой в стохастических системах // Современные проблемы прикладной математики, информатики, автоматизации, управления: Мат-лы Междунар. семинара. — Севастополь: СевНТУ, 2012. С. 24–35.
5. Синецын И. Н. Аналитическое моделирование широкополосных процессов в стохастических системах, не разрешенных относительно производных // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 3–10.
6. Синецын И. Н. Параметрическое аналитическое моделирование процессов в стохастических системах, не разрешенных относительно производных // Системы и средства информатики, 2017. Т. 27. № 1. С. 21–45.
7. Sinityn I. N. Analytical modeling and estimation of normal processes defined by stochastic differential equations with unsolved derivatives // J. Mathematics Statistics Research, 2021. Vol. 3. Iss. 1. Art. 139. 7 p. doi: 10.36266/JMSR/139.
8. Синецын И. Н. Аналитическое моделирование и оценивание нестационарных нормальных процессов в стохастических системах, не разрешенных относительно производных // Системы и средства информатики, 2022. Т. 32. № 2. С. 58–71.
9. Синецын И. Н. Нормализация систем, стохастически не разрешенных относительно производных // Информатика и её применения, 2022. Т. 16. Вып. 1. С. 32–38.
10. Пугачёв В. С., Синецын И. Н. Теория стохастических систем. — М.: Логос, 2004. 1000 с.

11. Александровская Л. Н., Аронов И. З., Круглов В. И. и др. Безопасность и надежность технических систем. — М.: Университетская книга, Логос, 2008. 348 с.
12. Синицын И. Н., Шаламов А. С. Лекции по теории систем интегрированной логистической поддержки. — 2-е изд. — М.: ТОРУС ПРЕСС, 2019. 1072 с.

Поступила в редакцию 15.01.23

ANALYTICAL MODELING OF DISTRIBUTIONS WITH INVARIANT MEASURE IN STOCHASTIC SYSTEMS WITH UNSOLVED DERIVATIVES

I. N. Sinitsyn^{1,2}

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

²Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

Abstract: Exact and approximate analytical modeling methods for stochastic processes with invariant measure in Gaussian and non-Gaussian stochastic systems with unsolved derivatives are considered. The methods are based on the linear regression approximation of nonlinear functions with unsolved derivatives and reduction to stochastic Ito differential equations. Two exact methods for analytical modeling of one- and multidimensional distributions with invariant measure are described. Special attention is paid to normal approximation and parametrization methods. A test example for Duffing equation nonlinear in second derivative is given. The stationary and nonstationary regimes and asymptotic stability are investigated. The method of normal approximation for one- and two-dimensional distributions is accurate enough for engineering applications. Some generalizations concerning numerical analytical modeling are considered.

Keywords: analytical modeling; distribution parametrization; distribution with invariant measure; stochastic system; stochastic system with unsolved derivatives; stochastic process

DOI: 10.14357/19922264230101

References

1. Sinitsyn, I. N. 2012. Analiticheskoe modelirovanie raspredeleniy s invariantnoy meroy v stokhasticheskikh sistemakh s avtokorrelirovannymi shumami [Analytical modeling of distributions with invariant measures in stochastic systems with autocorrelated noises]. *Informatika i ee Primeneniya — Inform. Appl.* 6(4):4–8.
2. Soize, C. 1994. *The Fokker–Plank equation for stochastic dynamical systems and its explicit steady state solutions*. Singapore: World Scientific. 321 p.
3. Sinitsyn, I. N. 2013. Analiticheskoe modelirovanie raspredeleniy s invariantnoy meroy v stokhasticheskikh sistemakh s razryvnymi kharakteristikami [Analytical modeling of distributions with invariant measure in stochastic systems with discontinuous characteristics]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):3–11.
4. Sinitsyn, I. N. 2012. Razvitie metodov analiticheskogo modelirovaniya raspredeleniy s invariantnoy meroy v stokhasticheskikh sistemakh [Development of analytical modeling methods for distributions with invariant measure in stochastic systems]. *Sovremennyye problemy prikladnoy matematiki, informatiki, avtomatizatsii, upravleniya: Mat-ly Mezhdunar. seminara* [Modern Problems of Applied Mathematics Informatics, Atomization and Control: Seminar (International) Proceedings]. Sevastopol': SevNTU. 24–35.
5. Sinitsyn, I. N. 2017. Analiticheskoe modelirovanie shirokopolosnykh protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel'no proizvodnykh [Analytical modeling of wide band processes in stochastic systems with unsolved derivatives]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):3–10.
6. Sinitsyn, I. N. 2017. Parametricheskoe analiticheskoe modelirovanie protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel'no proizvodnykh [Parametric analytical modeling of processes in stochastic systems that are not allowed with respect to derivatives]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(1): 21–45.
7. Sinitsyn, I. N. 2021. Analytical modeling and estimation of normal processes defined by stochastic differential equations with unsolved derivatives. *J. Mathematics Statistics Research* 3(1):139. 7 p.
8. Sinitsyn, I. N. 2022. Analiticheskoe modelirovanie i otsenivanie nestatsionarnykh normal'nykh protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel'no proizvodnykh [Analytical modeling and estimation of nonstationary normal processes with unsolved derivatives].

- tives]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):58–71.
9. Sinitsyn, I. N. 2022. Normalizatsiya sistem, stokhasticheski ne razreshennykh otnositel'no proizvodnykh [Normalization of systems with stochastically unsolved derivatives]. *Informatika i ee Primeneniya — Inform. Appl.* 16(1):32–38.
 10. Pugachev, V. S., and I. N. Sinitsyn. 2000, 2004. *Teoriya stokhasticheskikh sistem* [Stochastic systems. Theory and applications]. Moscow: Logos. 1000 p.
 11. Aleksandrovsкая, L. N., I. Z. Aronov, V. I. Kruglov, et al. 2008. *Bezopasnost' i nadezhnost' tekhnicheskikh sistem* [Security and safety of technical systems]. Moscow: Universitetskaya kniga, Logos. 348 p.
 12. Sinitsyn, I. N., and A. S. Shalamov. 2019. *Lektsii po teorii sistem integrirovannoy logisticheskoy podderzhki* [Lectures on theory of integrated logistic support systems]. 2nd ed. Moscow: TORUS PRESS. 1072 p.

Received January 15, 2023

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, 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; professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; sinitsin@dol.ru

AN AXIOMATIC VIEWPOINT ON THE ROGERS–VERAART AND SUZUKI–ELSINGER MODELS OF SYSTEMIC RISK

Yu. M. Kabanov^{1,2} and A. P. Sidorenko¹

Abstract: The authors study a model of clearing in an interbank network with crossholdings and default charges. Following the Eisenberg–Noe approach, the authors define the model via a set of natural financial regulations including those related to eventual default charges and derive a finite family of fixpoint problems. These problems are parameterized by vectors of binary variables. The model combines features of the Ararat–Meimanjanov, Rogers–Veraart, and Suzuki–Elsinger networks. The authors develop methods of computing the maximal and minimal clearing pairs using the mixed integer-linear programming and a Gaussian elimination algorithm.

Keywords: systemic risks; financial networks; clearing; crossholdings; default charges

DOI: 10.14357/19922264230102

1 Introduction

In a financial system with strongly interconnected institutions, a shock touching even a small number of entities can propagate through the network and lead to substantial losses. Clearing is a procedure that diminishes the total amount of interbank liabilities and, as such, decreases the risk of a system-wide breakdown. In the seminal paper [1] by Eisenberg and Noe, the notion of clearing vectors was introduced in an axiomatic way via limited liability and absolute priority rules. It was shown that the set of clearing vectors is the set of fixed points of a simple nonlinear mapping and this set contains the maximal and minimal elements. The further development led to more sophisticated models including crossholdings, seniority of debts, and default charges (see, e. g., [2–5] and a survey paper [6]). Dynamic versions of the Eisenberg–Noe network are suggested, for instance, in [7, 8].

The adding of default fees introduced in the paper by Rogers and Veraart [4] is of particular interest because it allows to study the problem of a rescue consortium to aid insolvent institutions. In general, default charges have a crisis amplifying effect: external payoffs may increase the number of failures. In the Rogers–Veraart model, the clearing vectors were defined directly as solutions of a fixpoint equation. As was observed by Ararat and Meimanjanov in [9], such a formulation does not coincide with a formulation defined in terms of financial regulations.

In this paper, the authors consider a model with default charges and crossholdings using the Ararat–Meimanjanov “axiomatic” approach. The aim is to derive the fixpoint equations and suggest algorithms to find their minimal and maximal solutions. In particular,

a method using the mixed integer-linear programming and a Gaussian-type dimensionality reduction algorithm is considered.

Notations: For vectors $a, b \in \mathbb{R}^N$, the symbol $a \leq b$ denotes the componentwise partial ordering $a^i \leq b^i$ for every $i = 1, \dots, N$, $[a, b] := \{x \in \mathbb{R}^N : a \leq x \leq b\}$, $\mathbf{1}_{\{a < b\}}$ is the vector with the components $I_{\{a^i < b^i\}}$, $\mathbf{1} := (1, \dots, 1)$, $\mathbf{0} := (0, \dots, 0)$. The symbol $a \circ b := (a^1 b^1, \dots, a^N b^N)$ stands for the Hadamard (componentwise) product.

2 Suzuki–Elsinger Model with Default Payments

Let us consider a financial network consisting of $N > 1$ banks. The bank i has a cash reserve e^i , a liability l^{ij} towards the bank $j \neq i$ with the total $l^i = \sum_j l^{ij}$, and possesses a share θ^{ji} of the bank j . It is convenient to introduce the relative liabilities matrix $\Pi = (\pi^{ij})$ with

$$\pi^{ij} := \begin{cases} \frac{l^{ij}}{l^i} = \frac{l^{ij}}{\sum_j l^{ij}}, & \text{if } l^i \neq 0; \\ \delta^{ij} & \text{otherwise} \end{cases}$$

where the Kronecker symbol $\delta^{ij} = 0$ for $i \neq j$ and $\delta^{ii} = 1$. Then π^{ij} describes the fraction of the value of the debtor i due to the creditor j of the total interbank debt of i . The value $\pi^{ii} = 1$ means that the bank i has no interbank debts.

The matrix $\Theta = (\theta^{ij})$ is assumed to be substochastic and such that unit is not its eigenvalue.

Fix $\alpha, \beta, \gamma \in [0, 1]$. It is assumed that there is Central Clearing Counterparty (CCP) calculating *clearing*

¹M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

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

and equity vectors, denoting p and V , and provide the settlement service for creditors and debtors.

A clearing pair $(p, V) \in [0, l] \times \mathbb{R}_+^N$ is determined by the following rules.

1. **Limited liability:** $p^i \leq (e + \Pi'p + \Theta'V)^i$ for every i .
2. **Absolute priority:** for every i , either $p^i = l^i$ or $p^i = (\alpha e + \beta \Pi'p + \gamma \Theta'V)^i$.
3. **Equity evaluation:** for every i , if $p^i = l^i$, then $V^i = (e + \Pi'p + \Theta'V - p)^i$, otherwise $V^i = 0$.

The first rule means that the clearing payment cannot exceed the available resources (cash e^i plus collected debts $\sum_j \pi^{ji} p^j$ plus the total of owned shares $\sum_j \theta^{ji} V^j$) while the second rule stipulates that either debts are paid in full or all resources are distributed with the charges paid in the case of the default, namely, the amount $(1 - \alpha)e^i + (1 - \beta)(\Pi'p)^i + (1 - \gamma)(\Theta'V)^i$. The third relation is just the definition of the equity: V^i is the total value of assets of the bank i after clearing (corresponding to the clearing vector p). By virtue of the limited liability rule, all components $V^i \geq 0$.

Denote by \mathcal{P} the set of clearing pairs.

Put

$$G(x, V) := (e + \Pi'x + \Theta'V - l) \circ \mathbf{1}_{\{x=l\}};$$

$$G_+(x, V) := (e + \Pi'x + \Theta'V - l)^+ \circ \mathbf{1}_{\{x=l\}}.$$

In this notation, if (p, V) is a clearing pair, (i. e., a clearing vector and an equity vector), then $V = G(p, V)$.

Lemma 2.1 *If $x \in \mathbb{R}^N$, then equations $V = G(x, V)$ and $V = G_+(x, V)$ have unique solutions. As functions of x , the solution of the first equation is affine and the solution of the second one is convex increasing.*

Proof. Put $\Lambda := \text{diag } \mathbf{1}_{\{x=l\}}$, $a(x) := \Lambda(e + \Pi'x - x)$, and $B := \Theta\Lambda$. In this notation,

$$G(x, V) = a(x) + B'V.$$

The linear equation $V = G(x, V)$ has a unique solution if the matrix $I - B$ is invertible, i. e., unit is not an eigenvalue of B (recall that this property is assumed).

Analogously, for $c(x) := \Lambda(e + \Pi'x - l)$, $G_+(x, V)$ can be written in the form

$$G_+(x, V) = (c(x) + B'V)^+.$$

It remains to apply Lemma 4.3 from [6] claiming that for every $y \in \mathbb{R}^N$, the equation $v = (y + \tilde{\Theta}v)^+$, where $\tilde{\Theta}$ is a substochastic matrix with invertible $I - \tilde{\Theta}$, has a unique solution $v = v(y)$ which is a monotone increasing convex function and so is the function $H(x) := v(c(x))$. \square

Remark. In the above lemma, the existence can be deduced from the Knaster–Tarski theorem. Indeed, put

$z := e + \Pi'l - l$, $K := (I - \Theta')^{-1}z^+$. Then, for any $p \in [0, l]$,

$$G_+(p, K) \leq G_+(l, K)$$

$$= (z + \Theta'K)^+ \leq z^+ + \Theta'K = K.$$

It follows that the monotone function $V \mapsto G_+(p, V)$ maps $[0, K]$ into itself.

Lemma 2.2 *If $(p, V) \in \mathcal{P}$, then $V = G_+(p, V)$.*

Proof. As we have already observed, the equity evaluation means that $V = G(p, V)$ and $p \leq e + \Pi'p + \Theta'V$ because of the limited liability. The components $G^i(p, V)$ and $G_+^i(p, V)$ coincide if $p^i = l^i$ and are equal to zero if $p^i < l^i$. \square

If $\alpha = \beta = 1$ and the matrix Θ is zero (no crossholdings), the model is reduced to that of Eisenberg–Noe. If $\alpha, \beta \in]0, 1]$ and Θ is zero, one gets the Ararat–Meimanjanov model. The model of Suzuki–Elsinger has a nontrivial substochastic matrix Θ and $\alpha = \beta = \gamma = 1$, i. e., default charges are not imposed. The relation with the Rogers–Veraart model we will discuss later.

To determine the clearing pairs, we use the absolute priority rule. For $p \in [0, l]$, we put

$$\Phi_0^i(p, V) := (\alpha e + \beta \Pi'p + \gamma \Theta'V)^i \wedge l^i;$$

$$\Phi_1^i(p, V) := (\alpha e + \beta \Pi'p + \gamma \Theta'V)^i d^i + (1 - d^i) l^i$$

where

$$d^i := \mathbf{1}_{\{(e + \Pi'p + \Theta'V)^i < l^i\}}, \quad i = 1, \dots, N.$$

For a binary vector $b = (b^1, \dots, b^N) \in \{0, 1\}^N$ on $\mathbb{R}^N \times \mathbb{R}^N$, we define the \mathbb{R}^N -valued function $(p, V) \mapsto \Phi_b(p, V)$ with

$$\Phi_b(p, V) = (\Phi_{b^1}^1(p, V), \dots, \Phi_{b^N}^N(p, V)).$$

In particular, for $\mathbf{0} := (0, \dots, 0)$ and $\mathbf{1} := (1, \dots, 1)$, we have

$$\Phi_{\mathbf{0}}(p, V) := (\alpha e + \beta \Pi'p + \gamma \Theta'V) \wedge l;$$

$$\Phi_{\mathbf{1}}(p, V) := (\alpha e + \beta \Pi'p + \gamma \Theta'V) \circ d + (\mathbf{1} - d) \circ l$$

where $d = (d^1, \dots, d^N)$.

Let $H(x)$ be the fixpoint $G_+(x, \cdot)$. By Lemma 2.2, the mapping $H : \mathbb{R}^N \rightarrow \mathbb{R}_+^N$ is monotone. Let $F_j^i(p) := \Phi_j^i(p, H(p))$, $i = 1, \dots, N$, $j = 0, 1$, and let

$$F_b(p) := \Phi_b(p, H(p)) =$$

$$= (\Phi_b^1(p, H(p)), \dots, \Phi_b^N(p, H(p))), \quad b \in \{0, 1\}^N.$$

It is easily seen that $(b, p) \mapsto F_b(p)$ is a monotone mapping of $\{0, 1\}^N \times [0, l]$ into $[0, l]$ with respect to the componentwise ordering. By the Knaster–Tarski theorem (see, e. g., [6]), the set of its fixpoints is nonempty and contains the minimal and maximal elements. By

the same reason, for every $b \in \{0, 1\}$, the set of fixpoints of the mapping $p \mapsto F_b(p)$ is nonempty and contains the minimal and maximal elements.

Let $S := \{p \in [0, l] : (p, V) \in \mathcal{P}\}$, i. e., S is the projection of the subset \mathcal{P} of the “plane” $\mathbb{R}^N \times \mathbb{R}^N$ into the “ x -axis” \mathbb{R}^N .

The following result gives a characterization of the set S of clearing vectors.

Theorem 2.3 $S = \bigcup_b \{p \in [0, l] : p = F_b(p)\}$.

Proof. Define the set $W_b := \bigcap_{i=1}^N W_b^i$, $b \in \{0, 1\}^N$, where $W_j^i := \{p \in S : p^i = F_j^i(p)\}$.

Let $p \in S$, that is, $(p, V) \in \mathcal{P}$ for some $V = H(p)$ (see Lemma 2.2). There are two cases: $p^i = l^i$ and $p^i < l^i$. In the first one,

$$l^i \leq (e + \Pi'p + \Theta'V)^i = (e + \Pi'p + \Theta'H(p))^i$$

implying that $F_1^i(p) = l^i = p^i$; hence, $p \in W_1^i$. In the second case, by the absolute priority rule, $p^i = (\alpha e + \beta \Pi'p + \gamma \Theta'H(p))^i$, that is, $p^i = F_0^i(p)$ and $p \in W_0^i$. Thus, $S = W_0^i \cup W_1^i$ whatever is i and

$$\begin{aligned} S &= \bigcap_i (W_0^i \cup W_1^i) \\ &= \bigcup_{b \in \{0, 1\}^N} W_b \subseteq \bigcup_b \{p \in [0, l] : p = F_b(p)\}. \end{aligned}$$

To prove the converse inclusion, fix some b and take $x \in [0, l]$ such that $x = F_b(x)$. Put $V = H(x)$. It is easily seen that (x, V) satisfies all the axioms. \square

Corollary 2.4 Let \underline{p}_b and \bar{p}_b be the minimal and maximal fixpoints of the mapping F_b . Then \underline{p}_0 and \bar{p}_1 are the minimal and maximal fixpoints of S .

Proof. Since functions F_b are increasing in b , so are the \underline{p}_b and \bar{p}_b (see, e. g., [6]). This implies the result. \square

Let us elaborate on the above statements. The maximal clearing vector p^* is the component of the vector $(p^*, V^*) \in [0, l] \times \mathbb{R}_+^N$ which is the maximal solution of the system

$$\begin{aligned} p &= (\alpha e + \beta \Pi'p + \gamma \Theta'V) \circ d + l \circ (\mathbf{1} - d); \\ V &= (e + \Pi'p + \Theta'V - l)^+ \circ (\mathbf{1} - d) \end{aligned}$$

where $d := \mathbf{1}_{\{e + \Pi'p + \Theta'V < l\}}$.

The minimal clearing vector p_* is the component of the vector $(p_*, V_*) \in [0, l] \times \mathbb{R}_+^N$ which is the minimal solution of the system

$$\begin{aligned} p &= (\alpha e + \beta \Pi'p + \gamma \Theta'V) \wedge l; \\ V &= (e + \Pi'p + \Theta'V - l)^+ \circ \mathbf{1}_{\{p=l\}}. \end{aligned}$$

Consider two particular cases. Let $\Theta = 0$. Then V is irrelevant. The equation for the maximal fixpoint is

$$p = (\alpha e + \beta \Pi'p) \circ d + l \circ (\mathbf{1} - d)$$

where $d := \mathbf{1}_{\{e + \Pi'p < l\}}$.

This equation coincides with that introduced in [4] as the definition of the clearing vector. It means that the maximal clearing vector in the “axiomatic” definition of [9] and in the definition via fixpoint of [4] coincide. However, in the axiomatic approach, the minimal clearing vector is the minimal solution of a different equation

$$p = (\alpha e + \beta \Pi'p) \wedge l$$

and may be different from that coming via fixpoint equation of [4].

Our arguments are intended to show that the “axiomatic” description also leads to a fixpoint problem allowing, in principle, to find all clearing vectors and corresponding equities. Unfortunately, the number of equations is exponentially growing. For example, for a small financial system with only ten banks, we have 20 480 equations. On the other hand, it seems that the largest clearing vector is of a major practical interest and to get it, one can consider the system of only $2N$ equations suggested by Rogers–Veraart.

Comparably to the Rogers–Veraart approach, the axiomatic one might seem rather perplexing because it leads to 2^N cases equations instead of just one. There is no prospect that a significant proportion of these equations is redundant and one can rule out them. Indeed, take arbitrary Π , l with all $l^i > 0$, put $\Theta = 0$, $e = l$, $\gamma = 1$, and $\beta = \alpha$. Choose α such that $\alpha(l + \Pi'l)^i < l^i$ for all i . Obviously, the solution of the equation $p = F_b(p)$ has the component $p^i = l^i$ if $b^i = 1$ and $p^i < l^i$, otherwise. In this example, the sets W_b are the disjoint singletons.

Now, let us consider the case where $\alpha = \beta = \gamma = 1$ as in the Suzuki–Elsinger model. In this case, $F_0 = F_1$ and clearing vectors are the solutions of the system

$$\begin{aligned} p &= (e + \Pi'p + \Theta'V) \wedge l; \\ V &= (e + \Pi'p + \Theta'V - l)^+ \circ \mathbf{1}_{\{p=l\}}. \end{aligned}$$

Obviously, the second equation can be replaced by $V = (e + \Pi'p + \Theta'V - l)^+$. Hence, the proposed “axiomatic” definition coincides with that of Suzuki–Elsinger model in [3].

A short remark on the interpretation of the parameter b . Assume for simplicity that $\alpha = \beta = \gamma < 1$. Let (p, V) be a clearing pair corresponding to some binary vector b and let $x(p, V) := e + \Pi'p + \Theta'V$. The component $x^i(p, V)$ is the total assets of bank i . For $z \in \mathbb{R}_+$, put $h_0^1(z) := (\alpha z) \wedge l^1$ and $h_1^1(z) := \alpha z \mathbf{1}_{\{z < l^1\}} + l^1 \mathbf{1}_{\{z \geq l^1\}}$. The value $h_{b_1}^1(x^1(p, V)) = \Phi_{b_1}^1(p, V)$ is the debt payment of bank 1. If $x^1(p, V) \in [l^1, l^1/\alpha)$, then $h_0^1(x^1(p, V)) < h_1^1(x^1(p, V)) = l^1$. Let us compare clearing pairs (p, V) and (\bar{p}, \bar{V}) corresponding to the binary vectors $(0, \bar{b})$ and $(1, \bar{b})$ with all components equal except the first

one. Then, $(p, V) \leq (\bar{p}, \bar{V})$ and $x(p, V) \leq x(\bar{p}, \bar{V})$. It may happen that $x^1(\bar{p}, \bar{V}) \in [l^1, l^1/\alpha]$ but $x^1(p, V) < l^1$. This leads to the conclusion that the banks are motivated to require CCP to calculate maximal clearing pairs using $b = 1$.

3 Clearing Pairs via Integer-Linear Programming

In this section, we introduce two integer-linear programming problems to find the maximal and the minimal clearing pairs in the models with crossholdings and default charges. To this end, we consider a linear function $f : \{0, 1\}^N \times [0, l] \times \mathbb{R}_+^N \rightarrow \mathbb{R}$ with

$$f(a, p, V) = \sum_{i=1}^N f_{1i}a^i + f_{2i}p^i + f_{3i}V^i$$

where all coefficients f_{1i} , f_{2i} , and f_{3i} are strictly positive. Put

$$\begin{aligned} x &= x(p, V) = e + \Pi'p + \Theta'V; \\ y &= y(p, V) = \alpha e + \beta \Pi'p + \gamma \Theta'V. \end{aligned}$$

3.1 The maximal clearing pair

Let $\kappa := \|K\|_\infty$ where K is the vector introduced in the remark after Lemma 2.1. Then, $H(l) \leq \kappa \mathbf{1}$ where the function H was defined in Section 2.

Problem P1: maximize f under the constraints

$$p \leq y(p, V) + a \circ l; \quad (1)$$

$$a \circ l \leq x(p, V); \quad (2)$$

$$V \leq x(p, V) - a \circ l; \quad (3)$$

$$V \leq \kappa a. \quad (4)$$

Immediately, if (p, V) is a clearing pair, then $(\mathbf{1}_{\{p=l\}}, p, V)$ is admissible for P1.

Lemma 3.1 *If $(\hat{a}, \hat{p}, \hat{V})$ solves P1, then $\hat{a} = \mathbf{1}_{\{\hat{p}=l\}}$.*

Proof. Suppose that $\hat{p}^j < l^j$ but $\hat{a}^j = 1$. In this case, we have from (1) that $\hat{p}^j \leq y^j(\hat{p}, \hat{V}) + l^j$. We can replace \hat{p}^j by a larger value $\hat{p}^j + \varepsilon$ with $\varepsilon \in (0, l^j - \hat{p}^j)$ without violating the constraints. Since the cost function f is strictly increasing, we get a contradiction with the optimality of $(\hat{a}, \hat{p}, \hat{V})$. Thus, $\hat{a}^j = 0$.

Suppose that $\hat{p}^j = l^j$ but $\hat{a}^j = 0$. Then, the component $\hat{V}^j = 0$ by virtue of (4) and $l^j \leq y^j(\hat{p}, \hat{V}) \leq x^j(\hat{p}, \hat{V})$ due to (1). Put $\tilde{a}^j = 1$ and $\tilde{a}^i = \hat{a}^i$ for $i \neq j$. Then, $(\tilde{a}, \hat{p}, \hat{V})$ is admissible and one again obtains a contradiction with the optimality. \square

Theorem 3.2 *If $(\hat{a}, \hat{p}, \hat{V})$ solves P1, then (\hat{p}, \hat{V}) is the maximal clearing pair.*

Proof. Let $(\hat{a}, \hat{p}, \hat{V})$ solve P1. It suffices to verify that (\hat{p}, \hat{V}) is a clearing pair. Due to the above lemma, we already know that $\hat{a} = \mathbf{1}_{\{\hat{p}=l\}}$. Let us check that the required rules are met.

Limited liability. If $\hat{p}^j < l^j$, then $\hat{p}^j \leq y^j(\hat{p}, \hat{V})$ from (1). Thus, $\hat{p}^j \leq x^j(\hat{p}, \hat{V})$. If $\hat{p}^j = l^j$, then $\hat{a}^j = 1$ and $l^j \leq x^j(\hat{p}, \hat{V})$ from (2).

Absolute priority. If $\hat{p}^j = l^j$, then there is nothing to prove. If $\hat{p}^j < l^j$, then $\hat{p}^j \leq y^j(\hat{p}, \hat{V})$ by (1). If the inequality were strict, one could add a small $\varepsilon > 0$ to \hat{p}^j and obtain again an admissible vector.

Equity evaluation. If $\hat{p}^j < l^j$ and $\hat{a}^j = 0$, then $\hat{V}^j = 0$ because of (4). If $\hat{p}^j = l^j$ and $\hat{a}^j = 1$, then $l^j \leq x^j(\hat{p}, \hat{V})$ and $\hat{V}^j \leq x^j(\hat{p}, \hat{V}) - l^j$ from (2). In the case where the latter inequality is strict, one can add a small $\varepsilon > 0$ to \hat{V}^j and get an admissible vector. \square

3.2 The minimal clearing pair

Put $\kappa_1 := \|e + \Pi'l + \kappa\Theta'\mathbf{1}\|_\infty$. Then $x(l, H(l)) \leq \kappa_1 \mathbf{1}$.

Problem P2: minimize f under the constraints

$$p \geq y(p, V) - \kappa_1 a; \quad (5)$$

$$p \geq a \circ l; \quad (6)$$

$$(\mathbf{1} - a) \circ l + \kappa_1 a \geq y(p, V); \quad (7)$$

$$V \geq x(p, V) - l - \kappa(\mathbf{1} - a). \quad (8)$$

We will show, under mild assumptions, that the solution $(\check{a}, \check{p}, \check{V})$ of P2 is the minimal clearing pair. More precisely, it is the minimal solution of the system

$$p = \Phi_0(p, V); \quad (9)$$

$$V = \check{G}_+(p, V) \quad (10)$$

where

$$\check{G}_+(p, V) = (x(p, V) - l)^+ \circ \mathbf{1}_{\{y(p, V) > l\}}.$$

Note that for every solution (p, V) of the above system, $(\mathbf{1}_{\{y(p, V) > l\}}, p, V)$ is admissible for P2.

Lemma 3.3 *The vector $\check{p} = (\mathbf{1} - \check{a}) \circ x(\check{p}, \check{V}) + \check{a} \circ l$.*

Proof. If $\check{a}^j = 0$, then by (5) $\check{p}^j \geq y^j(\check{p}, \check{V})$. If the inequality is strict, one can decrease \check{p}^j by a small $\varepsilon > 0$. Hence, $\check{p}^j = y^j(\check{p}, \check{V})$. If $\check{a}^j = 1$, then by (6), $\check{p}^j = l^j$. \square

Lemma 3.4 *The vector $\check{V} = \check{a} \circ (x(\check{p}, \check{V}) - l)^+$.*

Proof. Let $\check{a}^j = 0$. Then $\check{V}^j = 0$ (otherwise, $(\check{a}, \check{p}, \check{V})$ cannot be the solution of P2).

Let $\check{a}^j = 1$ and $x^j(\check{p}, \check{V}) - l^j \geq 0$. From (8), $\check{V}^j \geq x^j(\check{p}, \check{V}) - l^j$. Since f attains its minimum at $(\check{a}, \check{p}, \check{V})$, necessarily, $\check{V}^j = x^j(\check{p}, \check{V}) - l^j$.

Let $\check{a}^j = 1$, but $x^j(\check{p}, \check{V}) - l^j < 0$. Then, $\check{V}^j = 0$ by the same reasoning. \square

Lemma 3.5 *The vector $\check{a} = \mathbf{1}_{\{y(\check{p}, \check{V}) > l\}}$. In particular, $\check{p} = y(\check{p}, \check{V}) \wedge l$.*

Proof. Let $y^j(\check{p}, \check{V}) \leq l^j$. Assume for a minute that $\check{a}^j = 1$. Immediately, $\check{p}^j = l^j$. Put $\check{a}^j = 0$ and $\check{a}^i = \check{a}^i$ for $i \neq j$, then $(\check{a}, \check{p}, \check{V})$ is admissible. A contradiction. Let $y^j(\check{p}, \check{V}) > l^j$. Then $\check{a}^j = 1$ by (5). \square

Theorem 3.6 *Let $(\check{a}, \check{p}, \check{V})$ be the solution of P2. Then (\check{p}, \check{V}) is the solution of the system (9)–(10). In particular, $\check{p} \leq \underline{p}_0$ where \underline{p}_0 was defined in Corollary 2.4. If $y^j(\check{p}, \check{V}) \neq l^j$ for every j , then $\check{p} = \underline{p}_0$.*

Proof. The fact that (\check{p}, \check{V}) solves (9)–(10) follows from the above lemmata. Note that $(\underline{p}_0, H(\underline{p}_0))$ solves the system

$$\begin{aligned} p &= y(p, V) \wedge l; \\ V &= (x(p, V) - l)^+ \circ \mathbf{1}_{\{x=l\}} \end{aligned}$$

that is equivalent to the following one:

$$p = y(p, V) \wedge l; \quad (11)$$

$$V = (x(p, V) - l)^+ \circ \mathbf{1}_{\{y(p, V) \geq l\}}. \quad (12)$$

Put $\tilde{G}_+(p, V) := (x(p, V) - l)^+ \circ \mathbf{1}_{\{y(p, V) \geq l\}}$. Obviously, $\tilde{G}_+(p, V) \leq \tilde{G}_+(p, V)$ for every (p, V) . Due to the Knaster–Tarski theorem, one has $\check{p} \leq \underline{p}_0$. Furthermore, if $y^j(\check{p}, \check{V}) \neq l^j$ for every j , then $\tilde{G}_+(p, V) = \tilde{G}_+(p, V)$. Under this condition the pair (\check{p}, \check{V}) is the solution of (11)–(12) that is why $\check{p} = \underline{p}_0$. \square

4 Gaussian Method

In this section, we present a Gaussian-type algorithm to find the maximal clearing vector in a model with crossholding and default charges assuming that $\alpha = \beta = \gamma$ and $\|\Theta\|_\infty < 1$, i. e., $\sum_j \theta^{ij} < 1$ for every i .

In such a case, the mappings $\Phi_1 : [0, l] \times \mathbb{R}_+^N \rightarrow [0, l]$ and $G_+ : [0, l] \times \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ are given by the formulae:

$$\begin{aligned} \Phi_1(p, V) &:= \alpha(e + \Pi'p + \Theta'V) \circ \mathbf{1}_D + l \circ \mathbf{1}_{\bar{D}}; \\ G_+(p, V) &:= (e + \Pi'p + \Theta'V - l)^+ \circ \mathbf{1}_{\bar{D}} \end{aligned}$$

where $D := \{e + \Pi'p + \Theta'V < l\}$.

Our aim is to find the maximal solution of the system

$$\begin{aligned} p &= \Phi_1(p, V); \\ V &= G_+(p, V) \end{aligned}$$

existing by virtue of the Knaster–Tarski theorem.

In the case where $(e + \Pi'l + \Theta'H(l))^i \geq l^i$ for all i , the maximal clearing pair is $(l, H(l))$. Of course, to check the condition, one needs to compute the value $H(l)$ which is the unique solution of the system $v = (c(l) + \tilde{\Theta}v)^+$ with $c(l) := e + P'l - l$. It is known

that the latter system can be solved in a finite number of steps (e. g., by a Gaussian-type algorithms) [6].

In the opposite case where at least one inequality fails to be true, we assume, without loss of generality, that $e^1 + (\Pi'l)^1 + (\Theta'V(l))^1 < l^1$ and $V^1 = 0$. The first equation becomes linear:

$$\alpha e^1 + \alpha(\Pi'l)^1 + \alpha(\Theta'V)^1 = p^1.$$

The vector equation $p = \Phi_1(p, V)$ can be written as the system

$$\begin{aligned} p^1 &= \alpha e^1 + \alpha\pi^{11}p^1 + \alpha T'\tilde{p} + \alpha M'\tilde{V}; \\ \tilde{p} &= \alpha(\tilde{e} + p^1 R' + \tilde{\Pi}'\tilde{p} + \tilde{\Theta}'\tilde{V}) \circ \tilde{\mathbf{1}}_D + \tilde{l} \circ \tilde{\mathbf{1}}_{\bar{D}} \end{aligned}$$

where $\tilde{\Pi} := (\pi^{ij})_{i,j=2}^N$, $R := (\pi^{1j})_{j=2}^N$, $T := (\pi^{i1})_{i=2}^N$, $\tilde{\Theta} := (\theta^{ij})_{i,j=2}^N$, $M := (\theta^{i1})_{i=2}^N$, and the tilde indicates vectors with the removed first component.

If $\alpha\pi^{11} \neq 1$, then

$$p^1 = \alpha(1 - \alpha\pi^{11})^{-1} (e^1 + T'\tilde{p} + M'\tilde{V}).$$

Substituting this expression, we get that

$$\tilde{e} + p^1 R' + \tilde{\Pi}'\tilde{p} + \tilde{\Theta}'\tilde{V} = e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V}$$

where

$$\begin{aligned} \Pi_1 &:= \tilde{\Pi} + \alpha(1 - \alpha\pi^{11})^{-1} TR; \\ \Theta_1 &:= \tilde{\Theta} + \alpha(1 - \alpha\pi^{11})^{-1} MR; \\ e_1 &:= \tilde{e} + \alpha(1 - \alpha\pi^{11})^{-1} e^1 R'. \end{aligned}$$

It follows that (\tilde{p}, \tilde{V}) is the solution of the system

$$\begin{aligned} \tilde{p} &= \alpha(e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V}) \circ \tilde{\mathbf{1}}_D + \tilde{l} \circ \tilde{\mathbf{1}}_{\bar{D}}; \\ \tilde{V} &= (e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V} - \tilde{l})^+ \circ \tilde{\mathbf{1}}_{\bar{D}}. \end{aligned}$$

Note that $TR\tilde{\mathbf{1}} = T(R\tilde{\mathbf{1}}) = ((\Pi\mathbf{1})^1 - \pi^{11})T$ and, therefore,

$$\begin{aligned} \Pi_1\tilde{\mathbf{1}} &= \tilde{\Pi}\tilde{\mathbf{1}} + (\alpha(\Pi\tilde{\mathbf{1}})^1 - \alpha\pi^{11})(1 - \alpha\pi^{11})^{-1} T \\ &\leq \tilde{\Pi}\tilde{\mathbf{1}} + T = \tilde{\Pi}\tilde{\mathbf{1}} \leq \tilde{\mathbf{1}}. \end{aligned}$$

Thus, the matrix Π_1 is substochastic. By the same arguments, we get that $\|\Theta_1\|_\infty < 1$. Moreover, $\tilde{\mathbf{1}}_D = \mathbf{1}_{D_1}$ where $D_1 := \{e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V} < l\}$.

If $\alpha = 1$ and $\pi^{11} = 1$, then we can take $p^1 = l^1$ and continue with the system:

$$\begin{aligned} \tilde{p} &= (e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V}) \circ \tilde{\mathbf{1}}_D + \tilde{l} \circ \tilde{\mathbf{1}}_{\bar{D}}; \\ \tilde{V} &= (e_1 + \Pi_1'\tilde{p} + \Theta_1'\tilde{V} - \tilde{l})^+ \circ \tilde{\mathbf{1}}_{\bar{D}} \end{aligned}$$

where $e_1 := \tilde{e} + e^1 R'$; $\Pi_1 := \tilde{P}$; and $\Theta_1 := \tilde{\Theta}$.

In both cases, we have reduced the $2N$ -dimensional problem to the $2(N - 1)$ -dimensional one of the same type.

Acknowledgments

The research is funded by the grant of the Russian Science Foundation 20-68-47030 “Econometric and probabilistic methods for the analysis of financial markets with complex structure.”

References

1. Eisenberg, L., and T. H. Noe. 2001. Systemic risk in financial systems. *Manage. Sci.* 47(2):236–249.
2. Suzuki, T. 2002. Valuing corporate debt: The effect of cross-holdings of stock and debt. *J. Oper. Res. Soc. Jpn.* 45(2):123–144.
3. Elsinger, H. 2011. *Financial networks, cross holdings, and limited liability*. Wien: Oesterreichische Nationalbank. Working Paper 156. 37 p.
4. Rogers, L. C. G., and L. A. M. Veraart. 2013. Failure and rescue in an interbank network. *Manage. Sci.* 59(4):882–898.
5. Weber, S., and K. Weske. 2017. The joint impact of bankruptcy costs, fire sales and cross-holdings on systemic risk in financial networks. *Probability Uncertainty Quantitative Risk* 2(1):1–38.
6. Kabanov, Yu. M., R. Mokbel, and Kh. El Bitar. 2017. Clearing in financial networks. *Theor. Probab. Appl.* 62(2):252–277.
7. Feinstein, Z., and A. Sojmark. 2019. A dynamic default contagion model: From Eisenberg–Noe to the mean field. *arXiv.org*. 45 p. doi: 10.48550/arXiv.1912.08695.
8. Djete, M., and N. Touzi. 2021. Mean field game of mutual holding. *arXiv.org*. 38 p. doi: 10.48550/arXiv.2104.03884.
9. Ararat, C., and N. Meimanjanov. 2019. Computation of systemic risk measures: A mixed-integer linear programming approach. *arXiv.org*. 65 p. doi: 10.48550/arXiv.1903.08367.

Received August 1, 2022

Contributors

Kabanov Yuri M. (b. 1948) — Doctor of Science in physics and mathematics, professor, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; 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; your.i.kabanov@univ-fcomte.fr

Sidorenko Artur P. (b. 1998) — PhD student, Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; Artur.Sidorenko@student.msu.ru

АКСИОМАТИЧЕСКИЙ ВЗГЛЯД НА МОДЕЛИ СИСТЕМНОГО РИСКА РОДЖЕРСА–ВЕРААРТ И СУДЗУКИ–ЭЛЬСИНГЕРА*

Ю. М. Кабанов¹, А. П. Сидоренко²

¹Московский государственный университет имени М. В. Ломоносова; Федеральный исследовательский центр «Информатика и управление» Российской академии наук

²Московский государственный университет имени М. В. Ломоносова

Аннотация: Изучается модель клиринга межбанковской сети с кросс-холдингами и стоимостью дефолта. Следуя подходу Айзенберга и Ноэ, авторы формулируют модель с помощью естественных правил финансового регулирования, включающих описание выплат в случае дефолтов. Эти правила определяют конечное семейство задач о неподвижных точках, параметризованных векторами бинарных переменных. Представленная модель обобщает известные модели Арарата–Мейманджанова, Роджерса–Вераарт и Судзуки–Эльсингера. Предложены методы вычисления максимальной и минимальной клиринговых пар с использованием комбинации целочисленного и линейного программирования, а также обсуждается алгоритм последовательного исключения переменных.

Ключевые слова: системный риск; финансовые сети; клиринг; кросс-холдинг; стоимость дефолта

DOI: 10.14357/19922264230102

* Работа выполнена при поддержке Российского научного фонда, проект 20-68-47030 «Эконометрические и вероятностные методы для анализа финансовых рынков сложной структуры».

Литература

1. *Eisenberg L., Noe T. H.* Systemic risk in financial systems // *Manage. Sci.*, 2001. Vol. 47. Iss. 2. P. 236–249.
2. *Suzuki T.* Valuing corporate debt: The effect of cross-holdings of stock and debt // *J. Oper. Res. Soc. Jpn.*, 2002. Vol. 45. Iss. 2. P. 123–144.
3. *Elsinger H.* Financial networks, cross holdings, and limited liability. — Wien: Oesterreichische Nationalbank, 2011. Working Paper 156. 37 p.
4. *Rogers L. C. G., Veraart L. A. M.* Failure and rescue in an interbank network // *Manage. Sci.*, 2013. Vol. 59. Iss. 4. P. 882–898.
5. *Weber S., Weske K.* The joint impact of bankruptcy costs, fire sales and cross-holdings on systemic risk in financial networks // *Probability Uncertainty Quantitative Risk*, 2017. Vol. 2. Iss. 1. P. 1–38.
6. *Кабанов Ю. М., Мокбель Р., Эль Битар Х.* Взаимозачет в финансовых сетях // *Теория вероятностей и ее применения*, 2017. Т. 62. № 2. С. 311–344.
7. *Feinstein Z., Sojmark A.* A dynamic default contagion model: From Eisenberg–Noe to the mean field // *arXiv*, 2019. 45 p. doi: 10.48550/arXiv.1912.08695.
8. *Djete M., Touzi N.* Mean field game of mutual holding // *arXiv*, 2021. 38 p. doi: 10.48550/arXiv.2104.03884.
9. *Ararat C., Meimanjanov N.* Computation of systemic risk measures: A mixed-integer linear programming approach // *arXiv*, 2019. 65 p. doi: 10.48550/arXiv.1903.08367.

Поступила в редакцию 01.08.2022

КРИТЕРИИ НОРМАЛЬНОСТИ ВЕРОЯТНОСТНОГО РАСПРЕДЕЛЕНИЯ ПРИ ОКРУГЛЕННЫХ ДАННЫХ*

В. Г. Ушаков¹, Н. Г. Ушаков²

Аннотация: Критерии нормальности менее чувствительны к округлению данных, чем, например, критерии на экспоненциальность, но среди критериев нормальности чувствительность сильно различается. В данной статье определено, какие критерии более, а какие менее чувствительны к округлению. Показано, что критерии, основанные на выборочных моментах, гораздо более устойчивы к округлению данных, чем критерии, основанные на порядковых статистиках (в отличие от устойчивости к выбросам, где порядковые статистики значительно более устойчивы, чем выборочные моменты). Это, однако, относится только к вероятности ошибки первого рода. Вероятность ошибки второго рода мало чувствительна к округлению данных для всех критериев нормальности.

Ключевые слова: нормальное распределение; критерий нормальности распределения; округленные данные; уровень значимости; моделирование методом Монте-Карло

DOI: 10.14357/19922264230103

1 Введение

Предположение о нормальности распределения выборки важно для многих статистических методов. Статистические процедуры, такие как t -критерий Стьюдента, критерии, связанные с коэффициентом регрессии, F -критерий однородности дисперсии и т. д., основаны на предположении, что наблюдения имеют нормальное распределение. Если предположение о нормальности не выполняется, то это может повлиять на точность результата и правильность выводов. Поэтому проблема проверки того, что выборка была взята из некоторого нормального распределения с неизвестными средним и дисперсией, — одна из наиболее распространенных проблем в критериях согласия. По этой причине в литературе было предложено много критериев нормальности. Сюда входят критерии, основанные на эмпирической функции распределения, эмпирической характеристической функции, выборочных моментах и др. Проблема, тем не менее, еще не закрыта, и новые критерии продолжают появляться. Это связано с тем, что не существует наилучшего критерия, т. е. наиболее мощного критерия для всех условий.

Сравнение различных критериев нормальности проводилось во многих работах (см., в частности, [1–5] и приведенные там ссылки). В указанных работах предполагалось, что наблюдения

регистрируются точно. Однако данные обычно регистрируются в округленном виде и содержат как случайную ошибку измерения, так и ошибку округления. Округление может возникать в результате ограниченной чувствительности измерительной аппаратуры или может быть искусственным с целью сжатия данных.

Статистические критерии для округленных данных изучались рядом авторов (см. ссылки в [6]). В частности, в [7–9] исследовались влияние округления на уровень значимости и мощность четырех параметрических критериев (t -критерий, двойной t -критерий, χ^2 -критерий и F -критерий).

Критерии нормальности, хотя менее чувствительны к округлению данных, чем, например, критерии экспоненциальности, однако сильно различаются по чувствительности между собой. В данной работе выясняется, какие критерии более, а какие менее робастны по отношению к округлению. Оказывается, что критерии, основанные на выборочных моментах, предпочтительнее, чем критерии, основанные на порядковых статистиках (в отличие от робастности по отношению к выбросам, где порядковые статистики робастнее, чем выборочные моменты). Это, впрочем, относится только к вероятности ошибки первого рода. Вероятность ошибки второго рода нечувствительна к округлению данных для всех критериев нормальности. Исследование основано на предварительном теоретическом

* Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, грант № 075-15-2020-799.

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vgushakov@mail.ru

² Институт проблем технологии микроэлектроники и особо чистых материалов Российской академии наук; Норвежский научно-технологический университет, nikolai.ushakov@ntnu.no

анализе и интенсивном моделировании методом Монте-Карло. Исследуются все критерии R-пакетов `nortest` и `normtest`. Краткое описание каждого критерия приведено в следующем разделе.

Существуют различные виды округления. Для определенности будет рассматриваться округление до ближайшего. Имеется множество $\{x' : x' = kh, k = 0, \pm 1, \pm 2, \dots\}$ (решетка или сетка округления), где $h > 0$. Для вещественного числа x его округление есть ближайшая к x точка этой решетки. Обозначим его через $r_h(x)$. Для непрерывной случайной величины X ее округлением (или дискретизацией) является дискретная случайная величина $r_h(X)$; h называется шагом дискретизации. Значения $r_h(X)$ называются округленными данными. Пусть σ — стандартное отклонение случайной величины X . Уровень округления измеряется соотношением $r = h/\sigma$.

В статье используются следующие обозначения. Разность между $r_h(x)$ и x обозначается $d_h(x)$, т.е. $d_h(x) = r_h(x) - x$. Заметим, что $|d_h(x)| \leq h/2$. Кумулятивная функция распределения нормального распределения с математическим ожиданием μ и дисперсией σ^2 обозначается $\Phi_{\mu,\sigma}(z)$, $\Phi(z) = \Phi_{0,1}(z)$.

2 Краткое описание тестов нормальности и их предварительный анализ

Критерии нормальности, которые изучаются в данной работе, перечислены в табл. 1. В ней указано, какие R-пакет и функция используют-

ся для каждого критерия. Пусть X_1, \dots, X_n — случайная выборка. Положим $Y_i = X_i/\bar{X}$. Эмпирические функции распределения выборок X_1, \dots, X_n и Y_1, \dots, Y_n обозначим соответственно $F_n(x)$ и $G_n(x)$.

Критерий хи-квадрат Пирсона. Диапазон переменных, образующих выборку, разбит на m ячеек. Пусть O_j обозначает число наблюдений в ячейке j . Пусть p_j — вероятность того, что случайное наблюдение окажется в ячейке j при условии, что нулевая гипотеза верна. Определим $E_j = p_j n$ — ожидаемое число наблюдений в ячейке j . Статистика критерия:

$$T_P = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j}.$$

Критерий Колмогорова–Смирнова. Статистика критерия:

$$T_{KS} = \sup_x |F_n(x) - \Phi_{\mu,\sigma}(x)|.$$

Статистика критерия Крамера – фон Мизеса:

$$T_{CvM} = \int_{-\infty}^{\infty} (F_n(x) - \Phi_{\hat{\mu},\hat{\sigma}}(x))^2 d\Phi_{\hat{\mu},\hat{\sigma}}(x).$$

Критерий Андерсона–Дарлингга. Тестовая статистика:

$$T_{AD} = n \int_{-\infty}^{\infty} (F_n(x) - \Phi_{\hat{\mu},\hat{\sigma}}(x))^2 w(x) d\Phi_{\hat{\mu},\hat{\sigma}}(x),$$

где $w(x)$ — весовая функция:

$$w(x) = \frac{1}{\Phi_{\hat{\mu},\hat{\sigma}}(x)(1 - \Phi_{\hat{\mu},\hat{\sigma}}(x))}.$$

Таблица 1 Исследуемые критерии и используемые R-пакеты и функции

Критерий	Обозначение	Пакет	Функция
Пирсона	P	<code>nortest</code>	<code>pearson.test(x)</code>
Колмогорова–Смирнова	KS	<code>nortest</code>	<code>lillie.test(x)</code>
Крамера – фон Мизеса	CvM	<code>nortest</code>	<code>cvm.test(x)</code>
Андерсона–Дарлингга	AD	<code>nortest</code>	<code>ad.test(x)</code>
Шапиро–Уилка	SW		<code>shapiro.test(x)</code>
Шапиро–Франция	SF	<code>nortest</code>	<code>sf.test(x)</code>
Харке–Бера	JB	<code>normtest</code>	<code>jb.norm.test(x)</code>
Скорректированный Харке–Бера	AJB	<code>normtest</code>	<code>ajb.norm.test(x)</code>
Фрозини	F	<code>normtest</code>	<code>frosini.norm.test(x)</code>
Гири	G	<code>normtest</code>	<code>geary.norm.test(x)</code>
Хегази–Грина (1)	HG1	<code>normtest</code>	<code>hegazy1.norm.test(x)</code>
Хегази–Грина (2)	HG2	<code>normtest</code>	<code>hegazy2.norm.test(x)</code>
Эксцесса	K	<code>normtest</code>	<code>kurtosis.norm.test(x)</code>
Асимметрии	S	<code>normtest</code>	<code>skewness.norm.test(x)</code>
Шпигельхальтера	Sp	<code>normtest</code>	<code>spiegelhalter.norm.test(x)</code>
Вайсберга–Бингема	WB	<code>normtest</code>	<code>wb.norm.test(x)</code>

Критерий Шапиро–Уилка основан на следующей тестовой статистике:

$$T_{SW} = \frac{\left(\sum_{i=1}^k a_i X_{(i)}\right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

Здесь (a_1, \dots, a_n) – вектор весов:

$$(a_1, \dots, a_n) = mV^{-1} (mV^{-1}V^{-1}m^t)^{-1/2},$$

где m и V – средний вектор и ковариационная матрица порядковой статистики стандартного нормального распределения.

Критерий Шапиро–Франция. Статистика критерия является коэффициентом корреляции Пирсона между вектором наблюдений и вектором m , т. е.

$$T_{SF} = \frac{\sum_{i=1}^n (X_i - \bar{X})(m_i - \bar{m})}{\sqrt{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) \left(\sum_{i=1}^n (m_i - \bar{m})^2\right)}}.$$

Критерий Харке–Бера основан на статистике

$$T_{JB} = \frac{n}{6} \left(b_1 + \frac{(b_2 - 3)^2}{4} \right),$$

где b_1 и b_2 – квадраты выборочной асимметрии и выборочного эксцесса соответственно.

Скорректированный критерий Харке–Бера. Статистика критерия:

$$T_{AJB} = \frac{(n+1)(n+3)}{6(n-2)} \times \left(b_1 + \frac{(n+1)(n+5)}{4n(n-3)} \left(b_2 - 3 \frac{n-1}{n+1} \right)^2 \right),$$

где b_1 и b_2 определены выше.

Статистика **критерия Фрозини:**

$$T_F = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| \Phi \left(\frac{X_{(i)} - \bar{X}}{S} \right) - \frac{i - 1/2}{n} \right|.$$

Для **критерия Гири** в качестве статистики критерия используется следующая мера эксцесса:

$$T_G = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{\sqrt{n \sum_{i=1}^n (X_i - \bar{X})^2}}.$$

Два **критерия Хегази–Грина** основаны на следующих статистиках:

$$T_{HG1} = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_{(i)} - \bar{X}}{S} - \Phi^{-1} \left(\frac{i}{n+1} \right) \right|;$$

$$T_{HG2} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_{(i)} - \bar{X}}{S} - \Phi^{-1} \left(\frac{i}{n+1} \right) \right)^2.$$

Критерии эксцесса и асимметрии. Статистики критериев имеют следующий вид:

$$T_K = \frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)^2};$$

$$T_S = \frac{\sqrt{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)^{3/2}}$$

соответственно.

Критерий Шпигельхальтера основан на статистике

$$T_{Sp} = \left[\frac{1}{n!} \left(\frac{2n}{X_{(n)} - X_{(1)}} \right)^{n-1} + \left(\frac{\sqrt{n(n-1)}}{\sum_{i=1}^n |X_i - \bar{X}|} \right)^{n-1} \right]^{1/(n-1)} S.$$

Статистика **критерия Вайсберга–Бингема:**

$$T_{WB} = \frac{\left(\sum_{i=1}^n m_i X_{(i)}\right)^2}{\left(\sum_{i=1}^n m_i^2\right) \sum_{i=1}^n (X_i - \bar{X})^2},$$

где

$$m_i = \Phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right).$$

Рассмотрим подробнее перечисленные критерии и их тестовые статистики. Одно свойство выделяет группу из следующих 5 критериев: Харке–Бера, скорректированного Харке–Бера, Гири, эксцесса и асимметрии. Статистики этих критериев зависят только от выборочных моментов (не выше четвертого порядка). Эти критерии выделены жирным шрифтом в табл. 2–6. Статистики остальных критериев включают в себя более сложные функции, которые прямо или косвенно основаны на порядковых статистиках или на эмпирической функции распределения. Эмпирическая функция распределения может быть записана в виде

$$F_n(x) = \begin{cases} 0 & \text{при } x < X_{(1)}; \\ \frac{k}{n} & \text{при } X_{(k)} \leq x < X_{(k+1)}, \\ & k = 1, 2, \dots, n-1; \\ 1 & \text{при } x \geq X_{(n)}, \end{cases}$$

т. е. это функция порядковых статистик. Поэтому свойства критериев, основанных на эмпирической функции распределения, аналогичны свойствам критериев, основанных непосредственно на порядковых статистиках.

Таблица 2 Оцененная вероятность ошибки первого рода, $n = 50$

Критерий	r					
	0	0,01	0,03	0,07	0,1	0,2
P	0,0406	0,0417	0,0478	0,0744	0,0991	0,2457
KS	0,0457	0,0414	0,0419	0,0548	0,0711	0,1518
CvM	0,0442	0,0421	0,0546	0,0578	0,0542	0,0859
AD	0,0488	0,0463	0,0578	0,0481	0,0571	0,0859
SW	0,0525	0,0469	0,0517	0,0422	0,0601	0,0651
SF	0,0587	0,0492	0,0521	0,0449	0,0689	0,0621
JB	0,0477	0,0464	0,0485	0,0433	0,0643	0,0441
AJB	0,0494	0,0482	0,0463	0,0438	0,0581	0,0447
F	0,0467	0,0463	0,0534	0,0543	0,0482	0,0783
G	0,0558	0,0429	0,0433	0,0501	0,0444	0,0532
HG1	0,0543	0,0548	0,0511	0,0487	0,0577	0,0743
HG2	0,0564	0,0463	0,0472	0,0433	0,0568	0,0629
K	0,0551	0,0487	0,0557	0,0442	0,0546	0,0397
S	0,0526	0,0454	0,0484	0,0513	0,0585	0,0482
Sp	0,0656	0,0311	0,0433	0,0343	0,0551	0,0384
WB	0,0561	0,0478	0,0449	0,0435	0,0628	0,0609

Таблица 3 Оцененная вероятность ошибки первого рода, $n = 100$

Критерий	r					
	0	0,01	0,03	0,07	0,1	0,2
P	0,0486	0,0498	0,0541	0,1048	0,1228	0,2857
KS	0,0383	0,0413	0,0601	0,0849	0,1308	0,3626
CvM	0,0501	0,0463	0,0463	0,0581	0,0664	0,1398
AD	0,0451	0,0458	0,0479	0,0595	0,0708	0,1418
SW	0,0436	0,0478	0,0463	0,0566	0,0594	0,0828
SF	0,0418	0,0456	0,0491	0,0634	0,0630	0,0909
JB	0,0386	0,0449	0,0489	0,0559	0,0478	0,0473
AJB	0,0372	0,0426	0,0505	0,0534	0,0453	0,0456
F	0,0492	0,0441	0,0478	0,0557	0,0645	0,1098
G	0,0525	0,0517	0,0488	0,0521	0,0482	0,0416
HG1	0,0503	0,0444	0,0505	0,0586	0,0622	0,1058
HG2	0,0423	0,0429	0,0508	0,0593	0,0582	0,0876
K	0,0450	0,0384	0,0393	0,0556	0,0472	0,0449
S	0,0419	0,0503	0,0533	0,0471	0,0561	0,0499
Sp	0,0544	0,0419	0,0524	0,0504	0,0499	0,0520
WB	0,0439	0,0437	0,0456	0,0508	0,0514	0,0859

3 Неравенства

Связь между моментами (теоретическими или выборочными) непрерывных случайных величин и их дискретизаций изучалась в ряде работ (см., в частности, [6, 10–14] и ссылки в этих работах). Коротко говоря, моменты близки, если шаг дискретизации не слишком велик. В этом разделе приведены новые количественные оценки этой близости.

Теорема 1. Пусть $h < 1$. Тогда для любого натурального m

$$\begin{aligned} |\mathbf{E}[r_h(X)]^m - \mathbf{E}X^m| &\leq \\ &\leq \left[\left(\frac{3}{2}\right)^m - 1 \right] \max\{1, \mathbf{E}|X|^{m-1}\} h. \end{aligned}$$

Доказательство. Поскольку

$$\begin{aligned} \mathbf{E}[r_h(X)]^m &= \mathbf{E}[X + d_h(X)]^m = \\ &= \mathbf{E}X^m + \sum_{j=1}^m \binom{m}{j} \mathbf{E}[X^{m-j} d_h(X)^j], \end{aligned}$$

используя неравенство Ляпунова, получаем

$$\begin{aligned} \mathbf{E}[r_h(X)]^m - \mathbf{E}X^m &= \\ &= \left| \sum_{j=1}^m \binom{m}{j} \mathbf{E}[X^{m-j} d_h(X)^j] \right| \leq \\ &\leq \sum_{j=1}^m \binom{m}{j} \mathbf{E}|X^{m-j} d_h(X)^j| \leq \end{aligned}$$

Таблица 4 Оцененная вероятность ошибки первого рода, $n = 500$

Критерий	r					
	0	0,01	0,03	0,07	0,1	0,2
P	0,0453	0,0531	0,1928	0,9360	0,8721	1
KS	0,0503	0,0599	0,0991	0,2897	0,5114	1
CvM	0,0578	0,0516	0,0547	0,0959	0,1618	1
AD	0,0563	0,0491	0,0512	0,0919	0,1538	1
SW	0,0486	0,0512	0,0489	0,0919	0,1008	0,6383
SF	0,0497	0,0536	0,0501	0,0786	0,0986	0,5644
JB	0,0514	0,0479	0,0478	0,0535	0,0438	0,0457
AJB	0,0510	0,0506	0,0481	0,0574	0,0457	0,0459
F	0,0542	0,0532	0,0505	0,0799	0,1138	1
G	0,0447	0,0409	0,0477	0,0575	0,0546	0,0362
HG1	0,0573	0,0554	0,0511	0,0819	0,1178	0,9960
HG2	0,0469	0,0471	0,0486	0,0751	0,0829	0,4325
K	0,0329	0,0443	0,0494	0,0539	0,0429	0,0519
S	0,0572	0,0495	0,0469	0,0518	0,0608	0,0499
Sp	0,0551	0,0485	0,0486	0,0552	0,0313	0,0544
WB	0,0491	0,0543	0,0487	0,0784	0,0961	0,5494

Таблица 5 Оцененная вероятность ошибки первого рода, $n = 1000$

Критерий	r					
	0	0,01	0,03	0,07	0,1	0,2
P	0,0498	0,0957	0,5254	1	1	1
KS	0,0526	0,0695	0,1428	0,5274	0,8961	1
CvM	0,0604	0,0533	0,0639	0,1618	0,5154	1
AD	0,0611	0,0562	0,0553	0,1648	0,4755	1
SW	0,0594	0,0695	0,0626	0,1088	0,1718	1
SF	0,0529	0,0699	0,0649	0,1038	0,1538	1
JB	0,0597	0,0638	0,0529	0,0548	0,0339	0,0468
AJB	0,0589	0,0619	0,0509	0,0539	0,0319	0,0509
F	0,0580	0,0591	0,0533	0,1208	0,3756	1
G	0,0515	0,0506	0,0547	0,0498	0,0479	0,0448
HG1	0,0620	0,0627	0,0548	0,1158	0,2907	1
HG2	0,0522	0,0633	0,0554	0,0899	0,1328	1
K	0,0553	0,0429	0,0528	0,0495	0,0329	0,0519
S	0,0561	0,0702	0,0453	0,0514	0,0425	0,0506
Sp	0,0577	0,0469	0,0419	0,0550	0,0489	0,0627
WB	0,0527	0,0615	0,0639	0,09521	0,1478	1

$$\begin{aligned} &\leq \sum_{j=1}^m \binom{m}{j} \mathbf{E} \left| X^{m-j} \left(\frac{h}{2}\right)^j \right| \leq \\ &\leq h \sum_{j=1}^m \binom{m}{j} \frac{1}{2^j} \mathbf{E} |X|^{m-j} \leq \\ &\leq h \sum_{j=1}^m \binom{m}{j} \frac{1}{2^j} (\mathbf{E} |X|^{m-1})^{(m-j)/(m-1)} \leq \\ &\leq \left[\left(\frac{3}{2}\right)^m - 1 \right] \max \{1, \mathbf{E} |X|^{m-1}\} h. \end{aligned}$$

Здесь использовано, что

$$\sum_{j=1}^m \binom{m}{j} \frac{1}{2^j} = \left(\frac{3}{2}\right)^m - 1$$

и что если $\mathbf{E} |X|^{m-1} > 1$, то

$$(\mathbf{E} |X|^{m-1})^{(m-j)/(m-1)} < \mathbf{E} |X|^{m-1}$$

для каждого $j > 1$.

Теорема 2. Пусть X имеет симметричное, абсолютно непрерывное, унимодальное распределение с конечным моментом порядка $m + 1$, где m — натуральное число, и пусть $h < 1$. Тогда

$$|\mathbf{E} [r_h(X)]^m - \mathbf{E} X^m| \leq Ch^m,$$

где C не зависит от h (но, возможно, зависит от m и распределения случайной величины X).

Таблица 6 Оцененная вероятность ошибки первого рода, $n = 2000$

Критерий	r					
	0	0,01	0,03	0,07	0,1	0,2
P	0,0419	0,1278	0,8901	1	1	1
KS	0,0470	0,0652	0,2487	0,9181	1	1
CvM	0,0581	0,0533	0,0849	0,5054	1	1
AD	0,0572	0,0528	0,0819	0,4635	1	1
SW	0,0508	0,0547	0,0678	0,1798	0,5904	1
SF	0,0522	0,0574	0,0719	0,1718	0,5694	1
JB	0,0541	0,0572	0,0534	0,0593	0,0626	0,0478
AJB	0,0572	0,0563	0,0564	0,0601	0,0604	0,0458
F	0,0601	0,0489	0,0727	0,3796	1	1
G	0,0552	0,0534	0,0478	0,0469	0,0343	0,0254
HG1	0,0631	0,0518	0,0799	0,3016	0,9990	1
HG2	0,0537	0,0541	0,0648	0,1568	0,4475	1
K	0,0587	0,0612	0,0404	0,0489	0,0509	0,0539
S	0,0459	0,0507	0,0562	0,0541	0,0527	0,0408
Sp	0,0580	0,0539	0,0512	0,0519	0,0521	0,0938
WB	0,0488	0,0567	0,0651	0,1658	0,5414	1

Доказательство. Без ограничения общности предположим, что распределение X симметрично относительно нуля. Пусть $f(x)$ — плотность распределения случайной величины X . Обозначим

$$p_{j,h} = \mathbf{P}(jh < X < (j+1)h) = \int_{jh}^{(j+1)h} f(x) dx, \\ j = 0, 1, 2, \dots,$$

и рассмотрим следующие распределения: симметричное относительно нуля дискретное распределение G , дающее вероятности $p_{j,h}$ точкам jh , где $j = 0, 1, 2, \dots$; симметричное относительно нуля абсолютно непрерывное распределение H с кусочно-постоянной плотностью, равной $p_{j,h}/h$ на каждом интервале $(jh, (j+1)h)$. Обозначим m -е моменты этих распределений ν_m и μ_m соответственно. Легко видеть, что $\nu_m < \mathbf{E} X^m < \mu_m$ и $\nu_m < |\mathbf{E}[r_h(X)]^m| < \mu_m$, поэтому

$$|\mathbf{E}[r_h(X)]^m - \mathbf{E} X^m| \leq \mu_m - \nu_m.$$

Оценим правую часть этого неравенства. Имеем

$$\nu_m = 2 \sum_{j=0}^{\infty} (jh)^m p_{j,h} = 2h^m \sum_{j=0}^{\infty} j^m p_{j,h}; \\ \mu_m = 2 \sum_{j=0}^{\infty} \int_{jh}^{(j+1)h} x^m \frac{p_{j,h}}{h} dx = \\ = 2h^m \sum_{j=0}^{\infty} \frac{(j+1)^{m+1} - j^{m+1}}{m+1} p_{j,h}.$$

Следовательно,

$$\mu_m - \nu_m = \\ = 2h^m \sum_{j=0}^{\infty} \left(\frac{(j+1)^{m+1} - j^{m+1}}{m+1} - j^m \right) p_{j,h} \leq \\ \leq 2h^m \sum_{j=0}^{\infty} \frac{(j+1)^{m+1}}{m+1} p_{j,h} \leq \\ \leq 2h^m \sum_{j=0}^{\infty} \frac{(j+1)^{m+1}}{m+1} p_{j,1} = Ch^m.$$

Близость моментов по точным и по округленным данным позволяет надеяться, что 5 критериев первой группы (JB, AJB, G, K и S) будут более стабильными по отношению к округлению данных, чем критерии второй группы. Это предположение проверяется с помощью метода статистического моделирования Монте-Карло.

4 Вероятность ошибки первого рода

В этом разделе изучено, насколько хорошо рассматриваемые критерии контролируют вероятность ошибки первого рода. Показано, что критерии первой группы намного надежнее, чем критерии второй группы. Вероятность ошибки первого рода оценивается для разных объемов выборки и разных уровней округления. Методом Монте-Карло 10 000 случайных выборок генерируются из нормального распределения для каждого объема выборки $n = 50, 100, 500, 1000$ и 2000. Наблюдения округлены. Все критерии применяются к одним

и тем же выборкам. Уровень значимости — 5%. Для каждой выборки и каждого из 16 критериев гипотеза о нормальности отклоняется, если полученное Р-значение меньше 0,05. Для заданного критерия и фиксированных r и n доля 10 000 выборок, для которых гипотеза отвергнута, служит оценочной вероятностью ошибки первого рода. Результаты представлены в табл. 2–6. Чтобы упростить анализ результатов моделирования в этих таблицах, примем следующее соглашение. Выберем определенное пороговое значение вероятности ошибки первого рода и выделим жирным шрифтом вероятности, превышающие этот порог. Такие значения могут рассматриваться как неприемлемые. Для определенности считаем критерий неприемлемым, если оцениваемая вероятность ошибки первого рода на 50% выше уровня значимости, т. е. 0,075 (так как уровень значимости равен 0,05). Это соглашение весьма условно и сделано только для удобства.

Из табл. 2–6 видно, что при большом объеме выборки критерии первой группы существенно более устойчивы не только к грубому округлению, но и к достаточно мягкому округлению. Критерий Шпигельхальтера — специальный случай. Он не принадлежит к первой группе, но его робастность почти не уступает робастности критериев первой группы. С одной стороны, это связано с тем, что тестовая статистика критерия основана на сочетании выборочных моментов и порядковых статистик и зависимость от порядковых статистик очень слабая: только от разницы между максимальной и минимальной статистиками. С другой стороны,

дальнейшее моделирование показывает, что при более грубом округлении критерий Шпигельхальтера становится существенно менее робастным, чем критерии первой группы.

5 Эмпирическая мощность

В этом разделе сравнивается эмпирическая мощность критериев в случае точных данных ($r = 0$) и в случае округленных данных (для определенности $r = 0,1$). Показано, что, в отличие от вероятности ошибки первого рода, округление практически не влияет на вероятность ошибки второго рода. Мощность исследуется для четырех альтернативных распределений (Лапласа, Стьюдента, равномерного и Вейбулла) и трех различных объемов выборки ($n = 50, 100$ и 1000). Альтернативные распределения имеют следующие параметры. Распределение Лапласа: нулевое математическое ожидание, среднеквадратичное отклонение — $1/(\sqrt{2}r)$; распределение Стьюдента: величина $X/(\sqrt{2}r)$, где X имеет распределение Стьюдента с 4 степенями свободы; равномерное распределение: на интервале $[-\sqrt{3}/r, \sqrt{3}/r]$; распределение Вейбулла: параметр формы — 2, параметр масштаба — $1/\sqrt{1-\pi/4r}$. Выбор параметров диктуется степенью округления r при шаге дискретизации, равном единице, т. е. при округлении сформированных данных до ближайшего целого числа.

Из каждого из четырех приведенных выше распределений генерируются 10 000 случайных выбо-

Таблица 7 Эмпирическая мощность для неокругленных и округленных данных для четырех распределений ($n = 50$)

Критерий	Распределение							
	Лапласа		Стьюдента		Равномерное		Вейбулла	
	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$
P	0,2645	0,3576	0,1791	0,2462	0,1940	0,2284	0,1318	0,1789
KS	0,4281	0,5265	0,2889	0,3803	0,2561	0,3546	0,2163	0,2572
CvM	0,5494	0,582	0,3695	0,4041	0,4352	0,4762	0,2696	0,2759
AD	0,5453	0,5836	0,4038	0,4436	0,5732	0,6058	0,3124	0,3313
SW	0,5164	0,5417	0,4597	0,4875	0,7499	0,7655	0,4077	0,4290
SF	0,5892	0,6153	0,5225	0,5454	0,4697	0,5075	0,3627	0,3724
JB	0,5532	0,5721	0,5381	0,5425	0,0091	0,0124	0,2545	0,2714
AJB	0,5673	0,5885	0,5457	0,5516	0,0001	0,0002	0,2235	0,2383
F	0,5181	0,5535	0,3562	0,3875	0,5284	0,5525	0,2783	0,2855
G	0,0009	0,0017	0,0025	0,0031	0,7604	0,7566	0,0895	0,0874
HG1	0,5523	0,5911	0,4193	0,4475	0,5824	0,5975	0,3335	0,3457
HG2	0,6182	0,6364	0,5456	0,5663	0,2217	0,2375	0,302	0,3148
K	0,5795	0,5967	0,5446	0,554	0,7114	0,7224	0,1175	0,1338
S	0,3577	0,3516	0,3915	0,3971	0,0019	0,0021	0,3657	0,3686
Sp	0,6918	0,6945	0,5261	0,5283	0,8955	0,9027	0,0752	0,0782
WB	0,5874	0,6095	0,5172	0,5383	0,4551	0,4947	0,3521	0,3656

Таблица 8 Эмпирическая мощность для неокругленных и округленных данных для четырех распределений ($n = 100$)

Критерий	Распределение							
	Лапласа		Стьюдента		Равномерное		Вейбулла	
	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$
P	0,4743	0,5931	0,3052	0,4511	0,4728	0,5147	0,2714	0,3676
KS	0,6922	0,8298	0,5117	0,6553	0,5946	0,7382	0,3784	0,5611
CvM	0,8152	0,8463	0,6084	0,6757	0,8388	0,8596	0,5017	0,5623
AD	0,8136	0,8423	0,6614	0,7231	0,9412	0,9593	0,6081	0,6524
SW	0,7790	0,7968	0,7264	0,7586	0,9945	0,9984	0,8026	0,8057
SF	0,8254	0,8331	0,7732	0,8133	0,9654	0,9765	0,7145	0,7232
JB	0,7823	0,7847	0,7772	0,7906	0,7212	0,7261	0,5483	0,5442
AJB	0,7884	0,7973	0,7836	0,7925	0,5487	0,5464	0,5024	0,4771
F	0,8072	0,8261	0,6028	0,6572	0,8893	0,9051	0,5285	0,5691
G	0	0	0,0007	0,0010	0,9691	0,9739	0,1052	0,1057
HG1	0,8322	0,8428	0,7033	0,7387	0,9546	0,9654	0,6521	0,6848
HG2	0,8443	0,8532	0,7914	0,8238	0,8485	0,8593	0,6028	0,6356
K	0,8071	0,8112	0,7914	0,8325	0,9917	0,9921	0,1781	0,1532
S	0,4167	0,4125	0,5312	0,5146	0,0024	0,0019	0,6781	0,6871
Sp	0,9360	0,9372	0,8078	0,8161	0,9969	0,9981	0,0452	0,0591
WB	0,8217	0,8262	0,7691	0,8033	0,9622	0,9691	0,7084	0,7135

Таблица 9 Эмпирическая мощность для неокругленных и округленных данных для четырех распределений ($n = 1000$)

Критерий	Распределение							
	Лапласа		Стьюдента		Равномерное		Вейбулла	
	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$	$r = 0$	$r = 0,1$
P	1	1	0,9901	1	1	1	1	1
KS	1	1	1	1	1	1	1	1
CvM	1	1	1	1	1	1	1	1
AD	1	1	1	1	1	1	1	1
SW	1	1	1	1	1	1	1	1
SF	1	1	1	1	1	1	1	1
JB	1	1	1	1	1	1	1	1
AJB	1	1	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1
G	0	0	0	0	1	1	0,2690	0,2677
HG1	1	1	1	1	1	1	1	1
HG2	1	1	1	1	1	1	1	1
K	1	1	1	1	1	1	0,3812	0,3677
S	0,5013	0,5031	0,7644	0,7582	0,0012	0,0008	1	1
Sp	1	1	1	1	0	0	0,0212	0,0198
WB	1	1	1	1	1	1	1	1

рок объема n . Для каждой выборки и каждого из 16 критериев гипотеза о нормальности не отвергается, если полученное P -значение больше 0,05, и вычисляется доля случаев, когда гипотеза отвергается. Другими словами, вычисляется доля случаев, когда P -значение меньше 0,05. Это эмпирическая мощность критерия для данного альтернативного распределения и данного объема выборки.

Результаты представлены в табл. 7–9. Видно, что либо мощность для округленных данных (практически) такая же, как мощность для точных данных,

либо мощность для округленных данных больше, чем мощность для точных данных (критерии Пирсона и Колмогорова–Смирнова).

6 Выводы

1. Тесты на нормальность, основанные на моментах выборки, гораздо более устойчивы к округлению данных, чем тесты, основанные на порядковой статистике: они существенно лучше

контролируют вероятность ошибки первого рода.

2. Вероятность ошибки второго рода практически не зависит от округления данных для всех критериев нормальности.
3. Если объем выборки большой, то критерии нормальности, основанные на порядковой статистике, не следует использовать даже при умеренных степенях округления: контроль вероятности ошибки первого рода плохой.

Литература

1. *Farrell P. J., Rogers-Stewart K.* Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test // *J. Stat. Comput. Sim.*, 2006. Vol. 76. P. 803–816.
2. *Keskin S.* Comparison of several univariate normality tests regarding Type I error rate and power of the test in simulation based on small samples // *J. Applied Science Research*, 2006. Vol. 2. P. 296–300.
3. *Yazici B., Yolacan S.* A comparison of various tests of normality // *J. Stat. Comput. Sim.*, 2007. Vol. 7. P. 175–183.
4. *Romão X., Delgado R., Costa A.* An empirical power comparison of univariate goodness-of-fit tests for normality // *J. Stat. Comput. Sim.*, 2010. Vol. 80. P. 545–591.
5. *Yap B. W., Sim C. H.* Comparisons of various types of normality tests // *J. Stat. Comput. Sim.*, 2011. Vol. 81. P. 2141–2155.
6. *Schneeweiss H., Komlos J., Ahmad A. S.* Symmetric and asymmetric rounding: A review and some new results // *ASTA — Adv. Stat. Anal.*, 2010. Vol. 94. P. 247–271.
7. *Tricker A. R.* The effect of rounding on the significance level of certain normal test statistics // *J. Appl. Stat.*, 1990. Vol. 17. P. 31–38.
8. *Tricker A. R.* The effect of rounding on the power level of certain normal test statistics // *J. Appl. Stat.*, 1990. Vol. 17. P. 219–228.
9. *Tricker A. R.* The effect of rounding on the significance level and power of certain test statistics for non-normal data // *J. Appl. Stat.*, 1990. Vol. 17. P. 329–340.
10. *Tricker A. R.* Effects of rounding on the moments of a probability distribution // *J. Roy. Stat. Soc. D — Sta.*, 1984. Vol. 33. P. 381–390.
11. *Janson S.* Rounding of continuous random variables and oscillatory asymptotics // *Ann. Probab.*, 2006. Vol. 34. P. 1807–1826.
12. *Ushakov N. G., Ushakov V. G.* Statistical analysis of rounded data: Measurement errors vs rounding errors // *J. Math. Sci.*, 2018. Vol. 234. P. 770–773.
13. *Samsonov S. V., Ushakov N. G., Ushakov V. G.* Estimation of the second moment based on rounded data // *J. Mathematical Sciences*, 2019. Vol. 237. P. 819–825.
14. *Ushakov N. G., Ushakov V. G.* Accuracy of estimating the mean from rounded data // *J. Mathematical Sciences*, 2020. Vol. 246. P. 565–568.

Поступила в редакцию 20.03.22

TESTS FOR NORMALITY OF THE PROBABILISTIC DISTRIBUTION WHEN DATA ARE ROUNDED

V. G. Ushakov^{1,2} and N. G. Ushakov^{3,4}

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

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

³Institute of Microelectronics Technology and High-Purity Materials of the Russian Academy of Sciences, 6 Academician Osipyan Str., Chernogolovka, Moscow Region 142432, Russian Federation

⁴Norwegian University of Science and Technology, 15A S. P. Andersensvei, Trondheim 7491, Norway

Abstract: Tests for normality are less sensitive to the data rounding than, for example, tests for exponentiality but among normality tests, the sensitivity is very different. In this paper, the authors find out which tests are more and which ones are less sensitive. The authors show that tests based on sample moments are much more robust with respect to the data rounding than tests based on order statistics (in contrast to the robustness with respect to outliers where order statistics are more robust than sample moments). This, however, only applies to the probability of Type I error. The probability of Type II error is very insensitive to the data rounding for all normality tests.

Keywords: normal distribution; test for normality; rounded data; significance level; Monte-Carlo simulation

DOI: 10.14357/19922264230103

Acknowledgments

The research was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2020-799.

References

1. Farrell, P. J., and K. Rogers-Stewart. 2006. Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test. *J. Stat. Comput. Sim.* 76:803–816.
2. Keskin, S. 2006. Comparison of several univariate normality tests regarding Type I error rate and power of the test in simulation based on small samples. *J. Applied Science Research* 2:296–300.
3. Yazici, B., and S. Yolacan. 2007. A comparison of various tests of normality. *J. Stat. Comput. Sim.* 7:175–183.
4. Romão X., R. Delgado, and A. Costa. 2010. An empirical power comparison of univariate goodness-of-fit tests for normality. *J. Stat. Comput. Sim.* 80:545–591.
5. Yap, B. W., and C. H. Sim. 2011. Comparisons of various types of normality tests. *J. Stat. Comput. Sim.* 81:2141–2155.
6. Schneeweiss, H., J. Komlos, and A. S. Ahmad. 2010. Symmetric and asymmetric rounding: A review and some new results. *ASTA — Adv. Stat. Anal.* 94:247–271.
7. Tricker, A. R. 1990. The effect of rounding on the significance level of certain normal test statistics. *J. Appl. Stat.* 17:31–38.
8. Tricker, A. R. 1990. The effect of rounding on the power level of certain normal test statistics. *J. Appl. Stat.* 17:219–228.
9. Tricker, A. R. 1990. The effect of rounding on the significance level and power of certain test statistics for non-normal data. *J. Appl. Stat.* 17:329–340.
10. Tricker, A. R. 1984. Effects of rounding on the moments of a probability distribution. *J. Roy. Stat. Soc. D — Sta.* 33:381–390.
11. Janson, S. 2006. Rounding of continuous random variables and oscillatory asymptotics. *Ann. Probab.* 34:1807–1826.
12. Ushakov, N. G., and V. G. Ushakov. 2018. Statistical analysis of rounded data: Measurement errors vs rounding errors. *J. Math. Sci.* 234:770–773.
13. Samsonov, S. V., N. G. Ushakov, and V. G. Ushakov. 2019. Estimation of the second moment based on rounded data. *J. Mathematical Sciences* 237:819–825.
14. Ushakov, N. G., and V. G. Ushakov. 2020. Accuracy of estimating the mean from rounded data. *J. Mathematical Sciences* 246:565–568.

Received March 20, 2022

Contributors

Ushakov Vladimir G. (b. 1952) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie 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; vgushakov@mail.ru

Ushakov Nikolai G. (b. 1952) — Doctor of Science in physics and mathematics, leading scientist, Institute of Microelectronics Technology and High-Purity Materials of the Russian Academy of Sciences, 6 Academician Osipyan Str., Chernogolovka, Moscow Region 142432, Russian Federation; professor, Norwegian University of Science and Technology, 15A S. P. Andersensvei, Trondheim 7491, Norway; ushakov@math.ntnu.no

СРЕДНЕЕ РАССТОЯНИЕ В КОНФИГУРАЦИОННЫХ ГРАФАХ СО СТЕПЕННЫМ РАСПРЕДЕЛЕНИЕМ*

М. М. Лери¹

Аннотация: В случайных конфигурационных графах с дискретным степенным распределением степеней вершин с фиксированным параметром рассматривается среднее расстояние в графе, которое вычисляется как среднее арифметическое расстояний между всеми парами вершин графа. Эта характеристика оценивается с помощью методов имитационного моделирования. В силу вычислительных ограничений рассматриваются графы в доасимптотической области (в настоящей работе это графы объемом до 7000 вершин). Построены модели зависимостей среднего расстояния от объема графа и параметра распределения степеней вершин. Проведено сравнение полученных результатов с результатами теоретических исследований типичного расстояния в графе в асимптотике (т. е. когда число вершин графа стремится к бесконечности), приведенными в работах Р. Хофстада.

Ключевые слова: конфигурационные графы; степенное распределение; среднее расстояние в графе; имитационное моделирование

DOI: 10.14357/19922264230104

1 Введение

Изучение структуры и функционирования сложных сетей продолжает оставаться одним из важных направлений исследований в науке и технике [1, 2]. Примерами таких сетей, окружающих нас в повседневной жизни, служат интернет, электрические и телекоммуникационные сети, сети социальных отношений, соавторства и цитирования и др. Их быстрое и динамичное развитие и нарастающая популярность легли в основу многих фундаментальных исследований в области топологии таких сетей (см., например, [1–5]). В качестве моделей сложных сетей широко используются случайные графы, причем их разнообразие касается как определения степеней вершин графа, так и установления связей между этими вершинами. В частности, было показано (см., например, [6, 7]), что модели случайных графов с независимыми одинаково распределенными степенями вершин с общим дискретным законом распределения подходят для моделирования сети Интернет в случае, когда в качестве узлов сети рассматриваются автономные системы.

Увеличение размеров сетей и изменчивость сетевой структуры дают понять, что для адекватного отражения их топологии и функционирования в ходе построения их математических моделей необходимо учитывать не только распределение степеней вершин в соответствующей модели случайного

графа, но также принимать во внимание и другие не менее важные характеристики исследуемых сетей [2, 3]. В связи с этим различные структурные характеристики сложных сетей представляют определенный интерес как при моделировании их топологии, так и при изучении динамических процессов, происходящих в таких сетях по мере их роста или под внешними воздействиями. Одна из таких характеристик — расстояние между двумя произвольными вершинами графа [1, 3, 5].

Можно рассматривать различные виды расстояний в графе: расстояние между двумя заданными вершинами, расстояние между произвольно выбранными вершинами, все расстояния между каждой парой вершин, наименьшее возможное расстояние, наибольшее, или диаметр графа, и т. д.

В настоящей работе рассматривается среднее расстояние в графе, которое вычисляется как среднее арифметическое расстояний между всеми парами вершин графа. Цель работы состояла в нахождении зависимостей среднего расстояния в графе от числа его вершин и параметра распределения степеней вершин, а также в сравнении результатов работы с результатами теоретических исследований расстояния в графе, приведенными в [5]. Исследование проводилось посредством методов имитационного моделирования с последующей статистической обработкой данных с помощью программного обеспечения Statistica 10 и Wolfram Mathematica 9.0.

* Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания Карельского научного центра Российской академии наук (Институт прикладных математических исследований КарНЦ РАН).

¹ Институт прикладных математических исследований Карельского научного центра Российской академии наук, leri@krc.karelia.ru

2 Описание модели

Рассматривается случайный граф, состоящий из N вершин. Через $\xi_1, \xi_2, \dots, \xi_N$ обозначим степени вершин, которые являются независимыми одинаково распределенными случайными величинами со следующим дискретным степенным распределением [7]:

$$\mathbf{P}\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \quad k = 1, 2, \dots, \quad (1)$$

с фиксированным параметром $\tau > 1$. Легко показать, что математическое ожидание распределения (1) равно $\mathbf{E}\xi = \zeta(\tau)$, где $\zeta(\tau)$ — значение дзета-функции Римана в точке τ . Что касается дисперсии, то при $\tau > 2$ она конечна, а при $\tau \in (1, 2]$ — бесконечна. В работе рассматриваются конфигурационные графы [8], построение которых происходит следующим образом. Для каждой из N вершин графа задаются степени в соответствии с распределением (1) с выбранным значением параметра τ . Степени определяют число различных полуребер [7] (под полуребром понимают ребро, инцидентное данной вершине графа, для которого смежная вершина еще не определена), занумерованных в произвольном порядке. Для формирования ребер все полуребра попарно и равновероятно соединяют между собой. Сумма степеней вершин рассматриваемого графа является случайной величиной. Очевидно, что она должна быть четной, поэтому, если это не так, для построения недостающего ребра к равновероятно выбранной вершине добавляется одно полуребро, увеличивая степень этой вершины на 1. Граф, построенный таким образом, может иметь петли, циклы и кратные ребра.

Известно (см., например, [1, 2, 7]), что степенной конфигурационный граф, значение параметра распределения степеней вершин которого $\tau > 1$, асимптотически почти наверное содержит больше одной компоненты связности, причем при $\tau \in (1, 2)$ в таком графе существует, и она единственна, так называемая гигантская компонента связности, число вершин в которой пропорционально N при $N \rightarrow \infty$, а объем любой другой компоненты такого графа бесконечно мал по сравнению с объемом гигантской компоненты.

3 Среднее расстояние в графе

Расстояния между узлами сложной сети служат важными числовыми характеристиками сетевой топологии (см., например, [5, 9]). Пусть $G = (V, E)$ — неориентированный граф, в котором V — множество вершин, а E — множество ребер. Обозначим через $l(v, u)$ число ребер простой цепи, соединяющей вершины v и u графа G ($v, u \in V$ и $v \neq u$).

Если вершины v и u принадлежат разным компонентам связности, то $l(v, u)$ полагают равным ∞ . Длину цепи от вершины v до вершины u наименьшей длины называют расстоянием между этими вершинами:

$$d(v, u) = \min_{l(v, u) \neq \infty} l(v, u).$$

Пусть k — число расстояний $d(v, u) \neq \infty$ между всеми парами вершин v и u ($v \neq u$). Среднее расстояние вычисляется как среднее арифметическое всех расстояний $d(v, u)$ графа G :

$$\text{dist} = \text{dist}(G) = \frac{\sum_{v, u \in V (v \neq u)} d(v, u)}{k}.$$

Из теорем 7.2 и 7.1 в [5] следует, что при $N \rightarrow \infty$

$$d(v, u) \sim \frac{2 \ln \ln N}{|\ln(\tau - 1)|}, \quad (2)$$

если $1 < \tau < 2$, и

$$d(v, u) \sim \frac{\ln N}{\ln \nu}, \quad (3)$$

если $\tau > 2$, где $\nu = \mathbf{E}\xi(\xi - 1)/(\mathbf{E}\xi)$ и $\nu > 1$. Выражения (2) и (3) носят асимптотический характер и получены для «типичных расстояний» (где под типичным расстоянием понимается математическое ожидание расстояния) [5] в конфигурационных графах с бесконечной и с конечной дисперсиями соответственно. Легко показать, что для конфигурационных графов с распределением (1)

$$\nu = \frac{2\zeta(\tau - 1)}{\zeta(\tau)} - 2,$$

причем $\nu > 1$ при $2 < \tau \leq 2, 8106 \dots$

Для нахождения зависимости среднего расстояния от числа вершин графа N и параметра распределения степеней вершин τ были получены оценки средних расстояний в конфигурационных графах различных размерностей с разными параметрами распределения степеней вершин. По полученным результатам были построены зависимости dist от числа вершин графа N при конкретных значениях параметра распределения степеней вершин τ , а также зависимости dist от N и τ на интервалах $\tau \in (1, 2)$ и $\tau \in (2, 2, 8]$. Рассматривались конфигурационные графы следующих размерностей: $10 \leq N \leq 100$ с шагом 10, $100 \leq N \leq 1000$ с шагом 50, $1000 \leq N \leq 7000$ с шагом 500. Значения параметра τ изменялись с шагом 0,1 в двух интервалах: $1,1 \leq \tau < 2$ и $2 < \tau \leq 2,8$, а также были взяты два дополнительных значения: $\tau = 1,99$ и $2,01$. Для каждой пары значений (N, τ) генерировалось

по 100 графов, т. е. 40 000 и 36 000 графов на интервалах $1,1 \leq \tau \leq 1,99$ и $2,01 \leq \tau \leq 2,8$ соответственно. Расстояния в графе находились с применением алгоритма Дейкстры [10].

3.1 Результаты для интервала $\tau \in (1, 2)$

Сначала для рассматриваемых в настоящей работе графов были построены зависимости среднего расстояния dist от числа вершин N при фиксированных значениях параметра τ , т. е. для каждого из рассмотренных значений $1,1 \leq \tau \leq 1,99$ были получены регрессионные зависимости вида

$$\text{dist} = a \ln \ln N + b. \quad (4)$$

Здесь и далее коэффициенты всех регрессионных уравнений находили посредством метода наименьших квадратов, значимость коэффициентов проверяли с помощью критерия Стьюдента. Для оценки степени подгонки регрессионной модели к данным вычисляли коэффициент детерминации этой модели и с помощью критерия Фишера проверяли гипотезу $H_0 : R^2 = 0$. Проверка всех статистических гипотез осуществлялась на 5%-ном уровне значимости.

Обозначим $a_f = 2/|\ln(\tau - 1)|$ и сравним эти значения с коэффициентами a уравнений (4) для каждого τ . В табл. 1 приведены значения a_f , коэффициенты a и b регрессионных уравнений вида (4) и соответствующие коэффициенты детерминации R^2 этих уравнений. Все коэффициенты a и b в табл. 1 значимы, а гипотезы $H_0 : R^2 = 0$ отвергаются для всех уравнений.

Таким образом, при фиксированных τ среднее расстояние в графе с ростом числа его вершин растет как $\ln \ln N$, так же как и расстояние в асимптотике (2). Из табл. 1 видно, что значения коэффициента a ниже значений a_f на всем интервале изменения τ , однако эта разница не остается неизменной, а возрастает с ростом τ . Более того, сам угловой коэффициент a возрастает с ростом пара-

Таблица 1 Значения a_f , коэффициенты a и b зависимостей вида (4) и коэффициенты детерминации R^2 этих уравнений

τ	a_f	a	b	R^2
1,1	0,869	0,745	1,702	0,88
1,2	1,243	0,975	1,513	0,91
1,3	1,661	1,255	1,244	0,94
1,4	2,183	1,596	0,869	0,96
1,5	2,885	1,959	0,494	0,96
1,6	3,915	2,421	-0,065	0,98
1,7	5,607	3,048	-0,909	0,98
1,8	8,963	3,668	-1,710	0,98
1,9	18,982	4,417	-2,806	0,98
1,99	198,998	5,262	-4,045	0,96

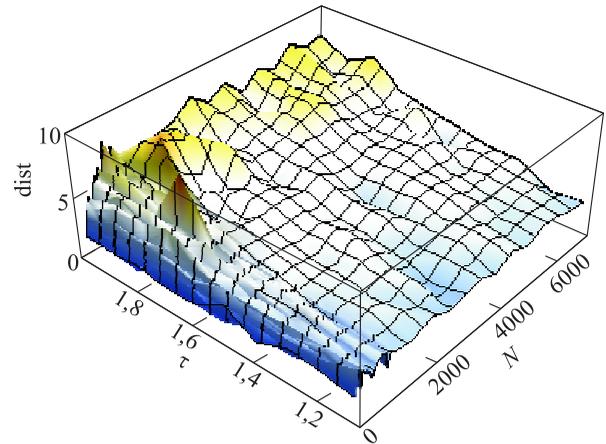


Рис. 1 График экспериментальных значений dist от N и $1,1 \leq \tau < 2$

метра τ , и это наглядно видно на рис. 1, где показана зависимость экспериментальных значений dist от N и τ .

Далее задача состояла в том, чтобы найти зависимость среднего расстояния dist от обеих переменных: N и τ . Сначала была построена зависимость в виде $\text{dist} = 2 \ln \ln N / (|\ln(\tau - 1)|) + b$. Получено следующее регрессионное уравнение:

$$\text{dist} = \frac{2 \ln \ln N}{|\ln(\tau - 1)|} - 39,604.$$

К сожалению, коэффициент детерминации полученной зависимости оказался очень низким ($R^2 = 0,01$). Гипотеза о равенстве R^2 нулю не отвергается, коэффициент b статистически не значим, поэтому такую модель использовать для прогноза не имеет смысла.

Далее была построена регрессия вида $\text{dist} = a \ln \ln N / (|\ln(\tau - 1)|) + b$ и получена зависимость

$$\text{dist} = \frac{0,0104 \ln \ln N}{|\ln(\tau - 1)|} + 3,922 \quad (5)$$

с коэффициентом детерминации $R^2 = 0,21$. В данном случае гипотеза $H_0 : R^2 = 0$ отвергается, а что касается коэффициентов a и b регрессионного уравнения (5), то коэффициент a оказался статистически значим, а коэффициент b нет. Поиск наилучшей регрессионной зависимости был продолжен и привел к получению следующей модели зависимости среднего расстояния конфигурационного графа dist от объема графа N и параметра распределения степеней вершин τ :

$$\text{dist} = \frac{2(4,488 - 3,077\tau + 0,417\tau^2) \ln \ln N}{|\ln(\tau - 1)|} \quad (6)$$

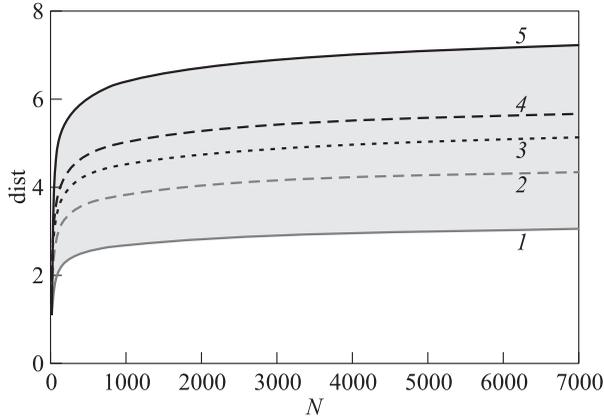


Рис. 2 Регрессионная зависимость (6) среднего расстояния dist от N при фиксированных значениях $1,1 \leq \tau < 2$: 1 – $\tau = 1,1$; 2 – $1,3$; 3 – $1,5$; 4 – $1,7$; 5 – $\tau = 1,99$

с коэффициентом детерминации $R^2 = 0,88$ и значимыми коэффициентами регрессии. Графически зависимость (6) представлена на рис. 2 и 3.

Оценка значимости различия между коэффициентами множественной корреляции $r = \sqrt{R^2}$ регрессионных моделей (5) и (6) на уровне значимости 0,05 показала, что нулевая гипотеза $H_0 : r_{(5)} = r_{(6)}$ ($r_{(5)}$ и $r_{(6)}$ – коэффициенты множественной корреляции зависимостей (5) и (6) соответственно) отвергается; следовательно, различие между коэффициентами корреляции значимо. Остатки обеих моделей распределены нормально, но сравнение сумм квадратов остатков $SSR_{(5)} = 62951,1$ и $SSR_{(6)} = 24786,9$ показывает, что $SSR_{(5)} > SSR_{(6)}$, т.е. модель (6) «лучше» в смысле описания изучаемого явления и для прогнозирования. Таким образом, в качестве наиболее подходящей модели зависимости dist от N и τ для доасимптотической области предлагается зависимость, описываемая уравнением (6).

На рис. 2 и 3 линии внутри затененных областей соответствуют зависимостям dist от N (см. рис. 2) и от τ (см. рис. 3) при некоторых (отраженных в легендах) значениях параметра τ или объема графа N соответственно. Заметим, что кривые зависимостей dist от N на рис. 2 расположены одна над другой по мере роста значения параметра $1,1 \leq \tau < 2$ в пределах его граничных значений. Аналогично кривые зависимостей dist от τ на рис. 3 также расположены друг над другом по мере возрастания числа вершин графа $10 \leq N \leq 7000$.

3.2 Результаты для интервала $\tau \in (2, 2,8]$

Исследование зависимости среднего расстояния от N и $\tau \in (2, 2,8]$ в доасимптотической

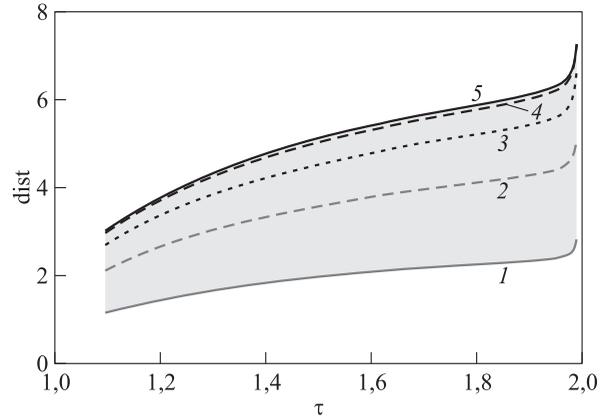


Рис. 3 Регрессионная зависимость (6) среднего расстояния dist от τ при фиксированных значениях $10 \leq N \leq 7000$: 1 – $N = 10$; 2 – 100 ; 3 – 1000 ; 4 – 5000 ; 5 – $N = 7000$

области было проведено аналогично предыдущему исследованию для $\tau \in (1, 2)$. Сначала для фиксированных значений параметра $2,01 \leq \tau \leq 2,8$ были построены зависимости среднего расстояния dist от числа вершин графа N следующего вида:

$$\text{dist} = a \ln N + b. \tag{7}$$

Обозначим

$$a_f = \frac{1}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)}.$$

Для сравнения этих значений с коэффициентами a уравнений (7) для каждого τ все они приведены в табл. 2 наряду с коэффициентами b и соответствующими коэффициентами детерминации R^2 этих уравнений. Все коэффициенты a и b значимы, а гипотезы о равенстве нулю коэффициентов детерминации полученных моделей отвергаются.

Таким образом, при фиксированных τ из интервала $(2, 2,8]$ среднее расстояние в графе возрастает логарифмически с ростом числа его вершин N , так же как и расстояние в асимптотике (см. выраже-

Таблица 2 Значения a_f , коэффициенты a и b зависимостей вида (7) и коэффициенты детерминации R^2 этих уравнений

τ	a_f	a	b	R^2
2,01	0,209	0,967	-0,684	0,96
2,1	0,408	1,151	-1,661	0,98
2,2	0,586	1,344	-2,777	0,98
2,3	0,800	1,541	-3,994	0,95
2,4	1,096	1,573	-4,428	0,91
2,5	1,565	1,430	-4,078	0,88
2,6	2,459	1,076	-2,673	0,89
2,7	4,942	0,760	-1,387	0,91
2,8	53,870	0,507	-0,352	0,94

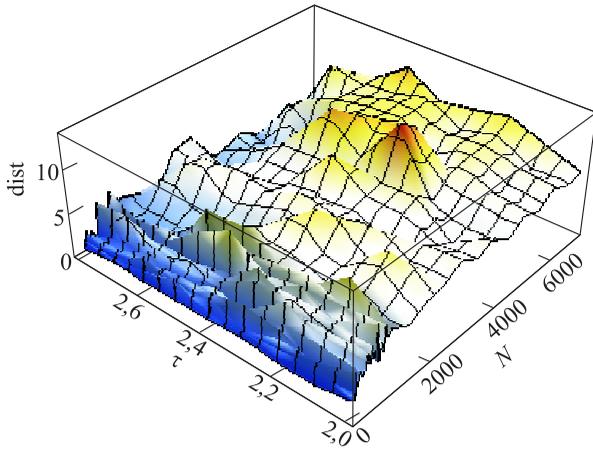


Рис. 4 График экспериментальных значений dist от N и $2 < \tau \leq 2,8$

ние (3)). Сравнение значений коэффициентов a с a_f показывает, что для $2,01 \leq \tau \leq 2,4$ значения a выше значений a_f , а при $2,5 \leq \tau \leq 2,8$ ниже (см. табл. 2). Изменение углового коэффициента a в данном случае показывает, что среднее расстояние в графе с ростом значения τ сначала возрастает, достигая максимума в промежутке от 2,2 до 2,4, а затем убывает. Наглядно это можно видеть на рис. 4, где показана зависимость экспериментальных значений dist от N и τ .

Поиск зависимости среднего расстояния dist от переменных N и τ проходил по аналогии с предыдущим интервалом изменения параметра распределения степеней вершин. Сначала была построена зависимость в виде

$$\text{dist} = \frac{\ln N}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)} + b$$

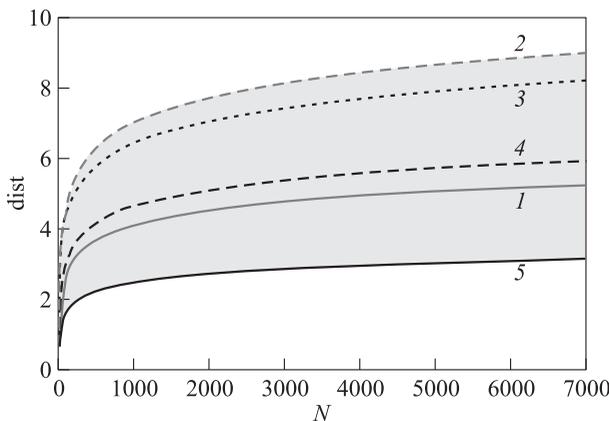


Рис. 5 Регрессионная зависимость (9) среднего расстояния dist от N при фиксированных значениях $2 < \tau \leq 2,8$: 1 – $\tau = 2,01$; 2 – 2,2288...; 3 – 2,4; 4 – 2,6; 5 – $\tau = 2,8$

и получено следующее регрессионное уравнение:

$$\text{dist} = \frac{\ln N}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)} - 40,936.$$

К сожалению, на этом интервале коэффициент детерминации полученной зависимости оказался очень низким ($R^2 = 0,0005$), гипотеза о равенстве R^2 нулю не отвергается и коэффициент b не значим. Следовательно, такую модель не имеет смысла использовать для прогноза. Поэтому была осуществлена попытка построить регрессию вида

$$\text{dist} = \frac{a \ln N}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)} + b$$

и была получена зависимость

$$\text{dist} = 4,962 - \frac{0,005 \ln N}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)} \quad (8)$$

с коэффициентом детерминации $R^2 = 0,06$. Несмотря на столь низкое значение R^2 , гипотеза о его равенстве нулю отвергается, однако оценка значимости коэффициентов a и b регрессионного уравнения (8) показала, что коэффициент a статистически значим, тогда как b – нет. Дальнейший поиск наилучшей регрессии привел к получению следующей зависимости среднего расстояния конфигурационного графа dist от N и τ :

$$\text{dist} = \frac{(31,706 - 22,076\tau + 3,841\tau^2) \ln N}{\ln(2\zeta(\tau - 1)/\zeta(\tau) - 2)}, \quad (9)$$

где все коэффициенты модели значимы, а $R^2 = 0,74$. Зависимость (9) отражена графически на рис. 5 и 6.

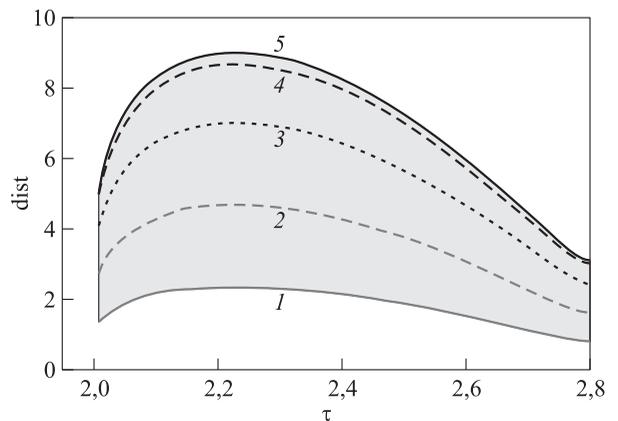


Рис. 6 Регрессионная зависимость (9) среднего расстояния dist от τ при фиксированных значениях $10 \leq N \leq 7000$: 1 – $N = 10$; 2 – 100; 3 – 1000; 4 – 5000; 5 – $N = 7000$

Для моделей (8) и (9) была оценена значимость различия между коэффициентами множественной корреляции этих моделей при 5%-ном уровне значимости. В результате $H_0 : r_{(8)} = r_{(9)}$ была отвергнута, т.е. различие между коэффициентами корреляции оказалось значимым. Проверка остатков регрессий (8) и (9) на нормальность показала, что нормальное распределение имеют только остатки модели (9). Кроме того, остаточная сумма квадратов модели (8) $SSR_{(8)} = 245793,9$ больше, чем $SSR_{(9)} = 102369,9$. Поэтому можно сделать вывод о том, что модель (9) лучше подходит для прогноза, чем модель (8). Таким образом, при значениях параметра $2 < \tau \leq 2,8$ в качестве наиболее подходящей модели зависимости среднего расстояния dist от N и τ в доасимптотической области предлагается зависимость, описываемая уравнением (9).

На рис. 5 и 6 линии, находящиеся внутри затененных областей и отраженные в легендах, соответствуют зависимостям dist от N (см. рис. 5) и от τ (см. рис. 6) при некоторых значениях параметра τ или объема графа N соответственно. На рис. 5 нижняя граница области соответствует $\tau = 2,01$, верхняя — максимуму функции (9) по параметру τ : $\tau^* = 2,2288 \dots$, а кривые зависимостей dist от N внутри затененной области расположены следующим образом: по возрастанию значений dist при увеличении значений τ от 1,1 до τ^* и по убыванию dist при росте τ от τ^* до 2,8. А на рис. 6 кривые зависимостей dist от τ расположены друг над другом по мере возрастания числа вершин графа $10 \leq N \leq 7000$ в пределах граничных значений.

3.3 Результаты при $\tau = 2$

Заметим, что модели (6) и (9) не охватывают значение параметра распределения степеней вершин $\tau = 2$. Однако по экспериментальным данным была построена следующая регрессионная зависимость dist от N при фиксированном $\tau = 2$ (все коэффициенты модели значимы):

$$\text{dist} = 5,262 \ln \ln N - 4,045 \quad (R^2 = 0,73). \quad (10)$$

4 Выводы

Итак, экспериментальные результаты на степенных конфигурационных графах с фиксированным параметром τ распределения (1) степеней вершин показывают, что на интервале (1, 2) с ростом объема N среднее расстояние dist в графе растет

как $\ln \ln N$, а на интервале $2 < \tau \leq 2,8$ растет логарифмически в доасимптотической области (при $N \leq 7000$) так же, как это было показано Р. Хофстадом [5] для типичного расстояния в графах при $N \rightarrow \infty$. Однако что касается зависимости среднего расстояния от переменных N и τ , то при малых объемах графа предлагается использовать модели (6) и (9) в соответствующих интервалах изменения параметра τ , так как они лучше описывают данную зависимость, что было подтверждено в настоящей работе с помощью методов статистического анализа и предложена модель (10) зависимости среднего расстояния от числа вершин N при $\tau = 2$ также для графов в доасимптотической области.

Литература

1. *Durrett R.* Random graph dynamics. — Cambridge: Cambridge University Press, 2007. 221 p. doi: 10.1017/CBO9780511546594.
2. *Hofstad R.* Random graphs and complex networks. — Cambridge: Cambridge University Press, 2017. Vol. 1. 337 p. doi: 10.1017/9781316779422.
3. *Newman M. E. J.* Networks. An introduction. — Oxford: Oxford University Press, 2010. 772 p. doi: 10.1093/acprof:oso/9780199206650.001.0001.
4. *Newman M. E. J.* Networks. — 2nd ed. — Oxford: Oxford University Press, 2018. 800 p. doi: 10.1093/oso/9780198805090.001.0001.
5. *Hofstad R.* Random graphs and complex networks // Notes RGCNII, 2020. Vol. 2. 314 p. <https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>.
6. *Faloutsos C., Faloutsos P., Faloutsos M.* On power-law relationships of the internet topology // Comput. Commun. Rev., 1999. Vol. 29. P. 251–262. doi: 10.1145/316194.316229.
7. *Reittu H., Norros I.* On the power-law random graph model of massive data networks // Perform. Evaluation, 2004. Vol. 55. Iss. 1-2. P. 3–23. doi: 10.1016/S0166-5316(03)00097-X.
8. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // Eur. J. Combin., 1980. Vol. 1. Iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8.
9. *Chung F., Lu L.* The average distances in random graphs with given expected degrees // P. Natl. Acad. Sci. USA, 2002. Vol. 99. Iss. 25. P. 15879–15882. doi: 10.1073/pnas.252631999.
10. *Dijkstra E. W.* A note on two problems in connexion with graphs // Numer. Math., 1959. Vol. 1. Iss. 1. P. 269–271. doi: 10.1007/BF01386390.

Поступила в редакцию 21.03.22

AN AVERAGE DISTANCE IN THE POWER-LAW CONFIGURATION GRAPHS

M. M. Leri

Institute of Applied Mathematical Research of the Karelian Research Center of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation

Abstract: In random configuration graphs with a discrete power-law vertex degree distribution with a fixed parameter, the average distance in the graph is considered, i. e., the arithmetic mean of distances between all pairs of graph nodes. This characteristic is estimated using simulation methods. Due to computational constraints, the author considers graphs in the pre-asymptotic domain (in this paper, these are the graphs up to 7000 nodes). The models of dependencies of the average distance on the graph size and the parameter of vertex degree distribution are received. The obtained results are compared with the results of theoretical studies of the typical distance in a graph in the asymptotics (i. e., when the number of graph vertices tends to infinity), given in the works by R. Hofstad.

Keywords: configuration graph; power-law distribution; average distance in a graph; simulation

DOI: 10.14357/19922264230104

Acknowledgments

The study was carried out under state order to the Karelian Research Center of the Russian Academy of Sciences (Institute of Applied Mathematical Research KarRC RAS).

References

1. Durrett, R. 2007. *Random graph dynamics*. Cambridge: Cambridge University Press. 221 p. doi: 10.1017/CBO9780511546594.
2. Hofstad, R. 2017. *Random graphs and complex networks*. Cambridge: Cambridge University Press. Vol. 1. 337 p. doi: 10.1017/9781316779422.
3. Newman, M. E. J. 2010. *Networks. An introduction*. Oxford: Oxford University Press. 772 p. doi: 10.1093/acprof:oso/9780199206650.001.0001.
4. Newman, M. E. J. 2018. *Networks*. 2nd ed. Oxford: Oxford University Press. 800 p. doi: 10.1093/oso/9780198805090.001.0001.
5. Hofstad, R. 2020. Random graphs and complex networks. *Notes RGCNII*. Vol. 2. 314 p. Available at: <https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf> (accessed January 18, 2023)
6. Faloutsos, C., P. Faloutsos, and M. Faloutsos. 1999. On power-law relationships of the internet topology. *Comput. Commun. Rev.* 29:251–262. doi: 10.1145/316194.316229.
7. Reittu, H., and I. Norros. 2004. On the power-law random graph model of massive data networks. *Perform. Evaluation* 5(1-2)5:3–23. doi: 10.1016/S0166-5316(03)00097-X.
8. Bollobas, B. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* 1(4):311–316. doi: 10.1016/S0195-6698(80)80030-8.
9. Chung, F., and L. Lu. 2002. The average distances in random graphs with given expected degrees. *P. Natl. Acad. Sci. USA* 99(25):15879–15882. doi: 10.1073/pnas.252631999.
10. Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1(1):269–271. doi: 10.1007/BF01386390.

Received March 21, 2022

Contributor

Leri Marina M. (b. 1969) — Candidate of Science (PhD) in technology, scientist, Institute of Applied Mathematical Research of the Karelian Research Center of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation; leri@krc.karelia.ru

ВОССТАНОВЛЕНИЕ МАТРИЦЫ СУПЕРПОЗИЦИИ В ЗАДАЧЕ СИМВОЛЬНОЙ РЕГРЕССИИ*

Р. Г. Нейчев¹, И. А. Шибаетов², В. В. Стрижов³

Аннотация: Исследуется проблема порождения структуры регрессионной модели. Модель представляет собой суперпозицию базовых функций. Структура модели описывается взвешенным цветным графом. Каждая вершина графа соответствует некоторой базовой функции. Ребро задает суперпозицию двух функций. Вес ребра равен вероятности суперпозиции. Для создания оптимальной модели необходимо восстановить ее структуру по матрице смежности графа. Предлагаемый алгоритм восстанавливает минимальное остовное дерево из взвешенного цветного графа. Представлено новое решение, основанное на алгоритме дерева Штейнера. Алгоритм сравнивается с альтернативами.

Ключевые слова: символьная регрессия; линейное программирование; суперпозиция; минимальное остовное дерево; матрица смежности

DOI: 10.14357/19922264230105

1 Введение

Символьная регрессия — это метод построения нелинейной модели, аппроксимирующей выборку. Структура модели определяется суперпозицией базовых функций. Набор базовых функций фиксируется для конкретной прикладной задачи. Структуры альтернативных моделей генерируются алгоритмом оптимизации для выбора оптимальной модели. В данной статье предлагается определять структуру модели с помощью вероятностного графа. Остовное дерево в графе определяет некоторую суперпозицию. Для выбора оптимальной модели необходимо реконструировать минимальное остовное дерево по графу.

Методы генетического программирования [1] находят оптимальное подмножество в наборе суперпозиций базовых функций, но имеют высокую вычислительную сложность. В [2] описаны методы, понижающие сложность. Они используют дополнительные ограничения на суперпозиции, например используют линейные комбинации базовых функций. Символьная регрессия, описанная в [3], используется для оптимизации структуры суперпозиции. Методы решения задачи символьной регрессии основаны на матричном представлении структуры модели [4]. Однако эти методы не содержат ограничений на число аргументов базовых функций и на структуру графа, обеспечивающую допустимую суперпозицию. В данной работе решается задача построения модели с помощью символьной регрессии.

Требуется восстановить допустимую суперпозицию из предсказанной матрицы смежности с вероятностями ребер. Решается задача восстановления k -минимального остовного дерева k -MST (*англ.* Minimum-cost Spanning Tree). Эта задача NP-сложная, поэтому применимы только приближенные решения [5]. Алгоритм k -MST эквивалентен проблеме дерева Штейнера PCST (*англ.* Prize-Collecting Steiner Tree) из-за его эквивалентности ослабленной формулировке постановки задачи линейного программирования [6]. В работах [5, 7, 8] представлены приближенные решения задачи k -MST.

Предлагаемое решение основано на упрощенной версии задачи k -MST, которая трансформируется в задачу PCST с постоянными призами, одинаковыми для всех вершин. Быстрый алгоритм PSCT описан в [9]. Альтернативное решение основано на алгоритме $(2 - \varepsilon)$ -аппроксимации для задачи PSCT. Она сравнивается с другими алгоритмами, включая алгоритмы обхода дерева в глубину, обхода дерева в ширину, алгоритмы Прима.

2 Задача выбора регрессионной модели

Требуется выбрать регрессионную модель φ из набора альтернативных моделей. Модель описывает выборку $D = \{(x_i, y_i)\}$ и минимизирует ошибку

$$\hat{\varphi}(D) = \arg \min_{\varphi} \sum_{i=1}^m (\varphi(x_i) - y_i)^2. \quad (1)$$

Модель представляет собой суперпозицию базовых функций из некоторого заданного набора. На рис. 1

* Работа выполнена при поддержке РФФИ (проекты 20-37-90050 и 20-07-00990).

¹ Московский физико-технический институт, neychevr@gmail.com

² Московский физико-технический институт, shibaev.kesha@gmail.com

³ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, strijov@phystech.edu

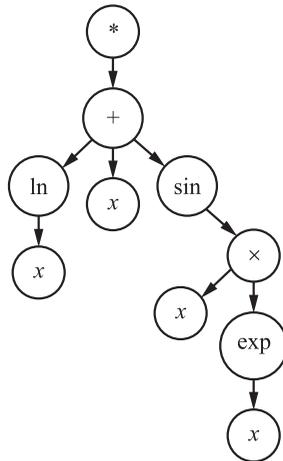


Рис. 1 Структура регрессионной модели представляет собой ориентированный граф

показан ее пример. Структура модели φ , суперпозиция, соответствует графу $G = (V, E)$, где базовые функции находятся в вершинах V . Корневая вершина обозначается через $*$. Модель:

$$\varphi(D) = \ln(x) + x + \sin(x \times \exp(x)).$$

Ее структура в виде матрицы смежности графа представлена в табл. 1. Базовые функции перечислены в первой строке. Элементами матрицы являются вероятности ребер E дерева. Жирным шрифтом выделены ребра восстановленного дерева M , образующие суперпозицию φ . Для восстановления структуры модели φ как суперпозиции, заданной деревом M , необходимы только графовое представление G и базовые функции.

Поставим задачу восстановления структуры модели. Задано множество выборок $\{D_k\}$. Каждой выборке D_k соответствует своя модель. Эта модель имеет структуру M_k . Таким образом, имеется набор пар $\{(D_k, M_k)\}$, выборка и структура. Обозначим через P отображение, которое предсказывает вероятности узлов в графе G по выборке D . Для выбора модели $\varphi(D)$ необходимо восстановить структуру модели M по графу G . Обозначим алгоритм восстановления дерева через R . Регрессионная модель $\hat{\varphi}(D)$, которая решает задачу (1), определяется

формулой $\hat{M} = R(P(D))$. Поскольку дерево M играет центральную роль в этой работе, критерий качества алгоритма восстановления дерева имеет вид:

$$\min_{M_k \in G} \frac{1}{K} \sum_{k=1}^K [\hat{M}_k = M_k].$$

Восстановленное дерево должно быть эквивалентно заданному дереву, следовательно, выбранная модель регрессии приближает выборку.

3 Задача восстановления дерева суперпозиции

Требуется восстановить дерево M_k , задающее суперпозицию и решающее задачу (1). Задан ориентированный взвешенный граф $G = (V, E)$ с раскрашенными вершинами v_i и корневой вершиной r . Каждая вершина $v_i \in V$ имеет свой цвет $t(v_i) = t_i$. Каждое ребро $e_i \in E$ имеет свой вес $w(e_i) = c_i \in [0, 1]$.

Требуется восстановить ориентированное дерево минимального веса с корнем r . Оно должно покрывать не менее k вершин в заданном графе G . Число ребер, выходящих из вершины v_i дерева, должно быть меньше или равно t_i . Корень r имеет одно ребро, $t_r = 1$.

Сформулируем это условие в виде задачи линейного программирования с целочисленными ограничениями:

$$\begin{aligned} & \underset{\substack{x_e, z_S \\ e \in E, \\ S \subseteq V \setminus \{r\}}}{\text{minimize}} \sum_{e \in E} c_e x_e \\ & \text{s.t.} \quad \sum_{\substack{e \in \delta(S): \\ e = (*, v_i), \\ v_i \in \delta(S)}} x_e + \sum_{T: T \supseteq S} z_T \geq 1, \quad S \subseteq V \setminus \{r\}; \\ & \quad \sum_{e \in E: e = (*, v)} x_e \leq 1, \quad v \in V; \\ & \quad \sum_{e \in E: e = (v, *)} x_e \leq t_i, \quad v \in V; \\ & \quad \sum_{S \subseteq V \setminus \{r\}} |S| z_S \leq n - k, \quad x_e \in \{0, 1\}, \quad z_S \in \{0, 1\}, \\ & \quad e \in E, \quad S \subseteq V \setminus \{r\}, \quad (2) \end{aligned}$$

Таблица 1 Вероятности суперпозиций в матрице смежности порождают ориентированный граф

Арность	Функция	*	+	ln	sin	×	exp	x
1	*	0,2	0,7	0,5	0,4	0,5	0,3	0,2
3	+	0,3	0,2	1,0	0,8	0,6	0,3	0,7
1	ln	0,3	0,2	0,0	0,0	0,1	0,5	0,5
1	sin	0,1	0,4	0,0	0,5	0,9	0,2	0,5
2	×	0,3	0,0	0,3	0,5	0,0	0,8	0,6
1	exp	0,3	0,3	0,4	0,1	0,5	0,4	0,4

где

$$x_e = \begin{cases} 1, & \text{если ребро } e \text{ входит в финальную} \\ & \text{суперпозицию;} \\ 0 & \text{в противном случае;} \end{cases}$$

$z_S = 1$ для всех вершин, исключенных из финальной суперпозиции. Обозначим через $e = (*, v)$ ориентированное ребро с листом v . Обозначим через $e = (v, *)$ ориентированное ребро с вершиной v .

Первое ограничение (2) определяет структуру графа решения в виде дерева с корнем r . Второе ограничение определяет ориентацию дерева: каждая вершина имеет не более одного входящего ребра. Третье ограничение определяет арность используемых базовых функций, поэтому число ребер, имеющих определенную вершину в качестве источника, фиксировано. Четвертое ограничение говорит, что итоговое дерево имеет не менее k вершин. Если все веса неотрицательны, то четвертое ограничение на минимальное число вершин принимает более строгий вид: число вершин должно быть равно k . Однако более слабое ограничение позволяет найти возможные связи с другими оптимизационными задачами. Ограничения в (2) преследуют ту же цель.

4 Алгоритмы восстановления дерева k -MST и PCST

Определение 1 (k -минимальное остовное дерево, k -MST). Задан взвешенный граф $G = (V, E)$ с корнем r и весами ребер $w(e_i) = c_i \geq 0, e_i \in E$. Требуется построить ориентированное дерево минимального веса с корнем r , покрывающее не менее k вершин в G .

Если та же задача ставится для ориентированных графов, то конечное дерево с корнем r должно быть ориентированным. Задача линейного программирования для направленного k -MST исключает третье условие в (2). В таком виде задача k -MST отличается от исходной задачи восстановления дерева суперпозиций (2) отсутствием третьего ограничения на арность базовых функций. Это эквивалентно ограничению на число ребер, выходящих из вершины.

Определение 2 (призовое дерево Штейнера, PCST). Задан взвешенный граф $G = (V, E)$ с корнем r и весами ребер $w(e_i) = c_i \geq 0, e_i \in E$, где каждой вершине $v_i \in V$ присвоен приз $\pi(v_i) = \pi_i \geq 0$. Требуется построить дерево T с корнем r , которое минимизирует функционал $\sum_{e \in E} c_e x_e + \sum_{S \subseteq V \setminus \{r\}} \pi(S) z_S$, где $x_e \in \{0, 1\}, x_e = 1$, если $e \in E$ входит в тройку T ; $z_S \in \{0, 1\}, z_S = 1$ для всех

вершин, исключенных из дерева T ; $S = V \setminus V(T)$; $\pi(S) = \sum_{v \in S} \pi(v)$.

В случае ориентированных графов эта задача обобщается до асимметричной задачи A-PCST. Задача линейного программирования для A-PCST принимает вид:

$$\begin{aligned} & \text{minimize} \sum_{\substack{x_e, z_S \\ e \in E, \\ S \subseteq V \setminus \{r\}}} c_e x_e + \sum_{S \subseteq V \setminus \{r\}} \pi(S) z_S \\ & \text{s.t.} \sum_{\substack{e \in \delta(S): \\ e = (*, v_i), \\ v_i \in \delta(S)}} x_e + \sum_{T: T \supseteq S} z_T \geq 1, \quad S \subseteq V \setminus \{r\}; \\ & \sum_{e \in E: e = (*, v)} x_e \leq 1, \quad x_e \in \{0, 1\}, \quad z_S \in \{0, 1\}, \quad v \in V, \\ & \quad \quad \quad e \in E, \quad S \subseteq V \setminus \{r\}. \quad (3) \end{aligned}$$

Если последнее ограничение из (2) входит в оптимизируемый функционал, задачи k -MST и A-PCST имеют эквивалентные ограничения и отличаются только оптимизируемым функционалом. Такое преобразование возможно согласно условиям Каруша–Куна–Таккера [10]. Если значения призов эквивалентны $\pi(v) = \lambda$, единственное отличие состоит в постоянном члене $\lambda(n - k)$. Таким образом, задачи оптимизации k -MST и A-PCST принимают вид:

$$\begin{aligned} & \text{minimize} \sum_{\substack{x_e, z_S \\ e \in E, \\ S \subseteq V \setminus \{r\}}} c_e x_e + \lambda \left(\sum_{S \subseteq V \setminus \{r\}} |S| z_S - (n - k) \right); \\ & \text{minimize} \sum_{\substack{x_e, z_S \\ e \in E, \\ S \subseteq V \setminus \{r\}}} c_e x_e + \lambda \sum_{S \subseteq V \setminus \{r\}} |S| z_S. \end{aligned}$$

Константа λ обозначает неотрицательный множитель Лагранжа в задаче k -MST и приз за вершину в задаче A-PCST. Существуют несколько алгоритмов для решения проблемы PCST, но не для решения проблемы A-PCST. Возможное решение — снять ограничения на ориентацию графа, чтобы алгоритм PCST мог позже восстановить ориентацию дерева.

5 Решение задачи восстановления ограниченного леса с помощью алгоритма $(2 - \varepsilon)$ -приближения

Обзор методов решения задачи восстановления ограниченного леса представлен в [11]. Задан взвешенный неориентированный граф $G = (V, E)$. Все его веса $w(e_i) = c_i \geq 0, e_i \in E$. Задана некоторая

функция $f : 2^V \rightarrow \{0, 1\}$. Требуется решить задачу линейного программирования с целочисленными ограничениями:

$$\begin{aligned} & \text{minimize}_{x_e: e \in E} \sum_{e \in E} c_e x_e \\ & \text{s.t. } x(\delta(S)) \geq f(S), \quad S \subset V, \quad S \neq \emptyset, \\ & \quad \quad \quad x_e \in \{0, 1\}, \quad e \in E. \end{aligned} \quad (4)$$

Здесь

$$x(\delta(S)) = \sum_{e \in \delta(S)} x_e,$$

где $x_e = 1$, если ребро e входит в финальное решение. Функция $\delta(S)$ обозначает все ребра из E такие, что только одна из смежных вершин входит в S .

Предположим, что отображение f удовлетворяет условиям

$$\begin{aligned} & f(V) = 0, \\ & \underbrace{f(S) = f(V \setminus S)}_{\text{симметричность}}, \\ & \underbrace{A, B \subset V: A \cap B = \emptyset, f(A) = f(B) = 0 \rightarrow f(A \cup B) = 0}_{\text{дизъюнктивность}}. \end{aligned}$$

При выполнении этих условий f задает число ребер, начинающихся в множестве вершин S .

Лемма 1. Пусть $B \subseteq S \subset V$. Тогда $f(S) = 0$ и $f(B) = 0$ приводит к $f(S \setminus B) = 0$.

Задача с таким описанием относится к *задачам поиска оптимального леса с ограничениями*. Такая постановка задачи (4) с соответствующим отображением f подходит для многих известных задач взвешенных графов, например: минимальный магистральный поиск, st -путь, задача Штейнера на минимальном дереве. Последняя задача является NP-полной, поэтому применим приближенный алгоритм.

Определение 3 (алгоритм α -аппроксимации). Эвристический полиномиальный алгоритм, дающий решение некоторой задачи оптимизации, называется α -аппроксимацией, если он гарантирует удовлетворяющее ограничениям решение этой задачи оптимизации с коэффициентом, меньшим или равным α , так что решение отличается от оптимального не более чем в α раз по оптимизируемому функционалу.

Чтобы предложить приближенный алгоритм, целочисленные ограничения в (4) должны быть ослаблены:

$$\begin{aligned} & \text{minimize}_{x_e: e \in E} \sum_{e \in E} c_e x_e \\ & \text{s.t. } \sum_{e \in \delta(S)} x_e \geq f(S), \quad S \subset V, \quad S \neq \emptyset, \\ & \quad \quad \quad x_e > 0, \quad e \in E, \end{aligned}$$

Двойственная задача принимает вид:

$$\begin{aligned} & \text{maximize}_{y_S: S \subset V, S \neq \emptyset} \sum_{S \subset V} f(S) y_S \\ & \text{s.t. } \sum_{S: e \in \delta(S)} y_S \leq c_e, \quad e \in E, \\ & \quad \quad \quad y_S > 0, \quad S \subset V, \quad S \neq \emptyset, \end{aligned} \quad (5)$$

относительно дополнительного условия

$$y_S \left(\sum_{e \in \delta(S)} x_e - f(S) \right) = 0, \quad S \subset V.$$

Обозначим множество вершин $A = \{v \in V : f(\{v\}) = 1\}$. Предлагается адаптивный жадный алгоритм $(2 - 2/|A|)$ -аппроксимации для задач вида (4). Алгоритм состоит из двух этапов. На первом этапе он жадно объединяет кластеры вершин, увеличивая двойственные переменные y_S . Изначально каждая вершина принадлежит своему кластеру. Если следующее ребро e достигает равенства в ограничениях в (5), это ребро добавляется к множеству S и связанные кластеры объединяются. Этот этап аналогичен алгоритму минимального остовного дерева Крускала. На втором этапе из конечного множества S удаляются некоторые ребра. Если обрезка ребра не нарушает ограничений, то это ребро должно быть удалено.

Индекс Z_{DRLP} в алгоритме 1 обозначает линейное программирование с двойной релаксацией. Начальное значение $F := \emptyset$ в алгоритме 1 эквивалентно предположению $x_e = 0, e \in E$. По условиям нежесткости $y_S = 0, S \subset V, S \neq \emptyset$.

На каждом шаге алгоритма кластер \mathcal{C} содержит две компоненты $\mathcal{C} = \mathcal{C}_i \cup \mathcal{C}_a$, где $\mathcal{C} \in \mathcal{C}_a$, если $f(\mathcal{C}) = 1$, и $\mathcal{C} \in \mathcal{C}_i$ в противном случае. Назовем \mathcal{C}_a активным компонентом. Переменные $d(v)$ в этом алгоритме связаны с переменными y_S из (5) соотношением

$$d(i) = \sum_{S: i \in S} y_S.$$

Рассмотрим две различные компоненты \mathcal{C}_q и \mathcal{C}_p , $\mathcal{C}_q \cap \mathcal{C}_p = \emptyset$, на некоторой итерации первого этапа алгоритма. Все y_S должны быть равномерно распределены по некоторому ε без нарушения ограничений

$$\sum_{S: e \in \delta(S)} y_S \leq c_e.$$

В терминах $d(v)$ это условие принимает вид:

$$\sum_{S: e \in \delta(S)} y_S = d(v_1) + d(v_2), \quad e = (v_1, v_2),$$

поэтому $y_S = 0$ для любого S такого, что $v_1, v_2 \in S$, потому что компоненты растут только на первом этапе. Увеличение некоторых компонент на ε приводит к уравнению

$$d(v_1) + d(v_2) + \varepsilon (f(C_q) + f(C_p)) \leq c_e, \quad e = (v_1, v_2),$$

что приводит к формуле, используемой в строке 10 алгоритма 1. В случае когда в состав входит следующее ребро, сумма $\sum_{S: e \in \delta(S)} y_S$ не будет увеличиваться, поэтому ограничения выполняются.

Ребра, которые можно удалить из F без добавления новых активных компонентов, удаляются на втором этапе алгоритма. Следующая лемма определяет свойства компонент связности в F' .

Лемма 2. Для каждой компоненты связности N из F' выполняется равенство: $f(N) = 0$.

Следующая теорема утверждает, что решение, полученное с помощью описанного алгоритма, удовлетворяет ограничениям исходной задачи линейного программирования.

Теорема 1. Набор ребер F' , полученный алгоритмом 1, удовлетворяет всем ограничениям исходной задачи (4).

Лемма 3. Обозначим граф H , каждая вершина которого соответствует одной из компонент связности $C \in \mathcal{C}$ на фиксированном шаге алгоритма. Ребро (v_1, v_2) присутствует, если существует ребро \hat{e} исходного графа, входящее в F' : $\hat{e} \in F'$, поэтому граф H — это лес. Внутри H нет листовых вершин, соответствующих неактивным вершинам исходного графа.

Теорема 2. Алгоритм 1 представляет собой α -приближенный алгоритм для задачи (4) с $\alpha = 2 - 2/|A|$, где $A = \{v \in V : f(\{v\}) = 1\}$.

Несмотря на эту теоретическую основу, не существует подходящей функции f для постановки задачи PCST, указанной в (4). Чтобы быть применимым в этих условиях, алгоритм 1 нуждается в нескольких модификациях.

6 Модифицированная постановка задачи для PCST

Как и в случае A-PCST, упрощенный вид задачи линейного программирования PCST принимает вид:

$$\begin{aligned} & \text{minimize}_{x_e, s_v} \sum_{e \in E} c_e x_e + \sum_{v \in V \setminus \{r\}} (1 - s_v) \pi_v \\ & \text{s.t.} \quad \sum_{e \in \delta(S)} x_e \geq s_v, \quad S \subseteq V \setminus \{r\}, \quad v \in S, \\ & \quad \quad x_e \geq 0, \quad e \in E, \quad s_v \geq 0, \quad v \in V \setminus \{r\}. \end{aligned}$$

Эта постановка задачи отличается от исходной (3) тем, что с ней возможно согласовать задачу k -MST. Индикаторы s_v показывают, что вершина v включена в дерево.

Двойственная задача принимает вид:

$$\begin{aligned} & \text{maximize}_{y_S} \sum_{S \subseteq V \setminus \{r\}} y_S \\ & \text{s.t.} \quad \sum_{S: e \in \delta(S)} y_S \leq c_e, \quad e \in E; \\ & \quad \quad \sum_{S \subseteq T} y_S \leq \sum_{v \in T} \pi_v, \quad T \subseteq V \setminus \{r\}, \\ & \quad \quad y_S \geq 0, \quad S \subseteq V \setminus \{r\}. \end{aligned}$$

Алгоритм 2 решает эту задачу. Он похож на алгоритм 1. Двойные переменные должны обновляться равномерно с дополнительными ограничениями. Тогда ε примет минимальное из двух значений в соответствии с обеими группами ограничений. Более широкий анализ аппроксимационных свойств обновленного алгоритма представлен в [11]. Алгоритм 2 представляет собой α -приближенный алгоритм для задачи PCST с $\alpha = 2 - 2/(n - 1)$, где n — число вершин в графе G .

7 Вычислительный эксперимент

Основная цель эксперимента — восстановить дерево суперпозиции. Алгоритмы, используемые для восстановления, перечислены ниже.

DFS, BFS. Алгоритмы жадного дерева обхода в глубину и жадного дерева обхода в ширину. Обход ребер с наибольшим весом эквивалентен выбору наиболее вероятного пути. Алгоритм обхода останавливается, когда число ребер, исходящих из некоторой вершины, становится равным арности соответствующей функции.

Алгоритм Прима. Алгоритм восстанавливает минимальное остовное дерево для графа с дополнительными ограничениями на арность базовых функций. Эти ограничения задают минимальный вес ребра. После добавления вершины все листовые ребра этой вершины исключаются, чтобы сохранить направление дерева. Если число ребер, начинающихся в какой-либо вершине, превышает соответствующую арность, то остальные ребра исключаются из множества возможных ребер в этой вершине. Алгоритм не зависит от процедуры обхода. В случае небольшого шума в матрице смежности этот алгоритм способен восстановить дерево суперпозиции без ошибок.

Алгоритмы на основе PCST. Матрица смежности M должна быть приведена к неориентиро-

ванному виду. Использована квадратная матрица M' без последнего столбца. PCST принимает матрицу смежности $1 - (1/2)(M' + M'^T)$ с призовым значением 0,5 для каждой вершины. Призовое значение равно 0,5, поскольку при меньших значениях дерево будет обрезано: если шум равен 0,5, некоторые вершины могут быть обрезаны по ошибке. В случае больших призовых значений дерево PCST может содержать ненужные вершины. Дерево восстанавливается по одному из опи-

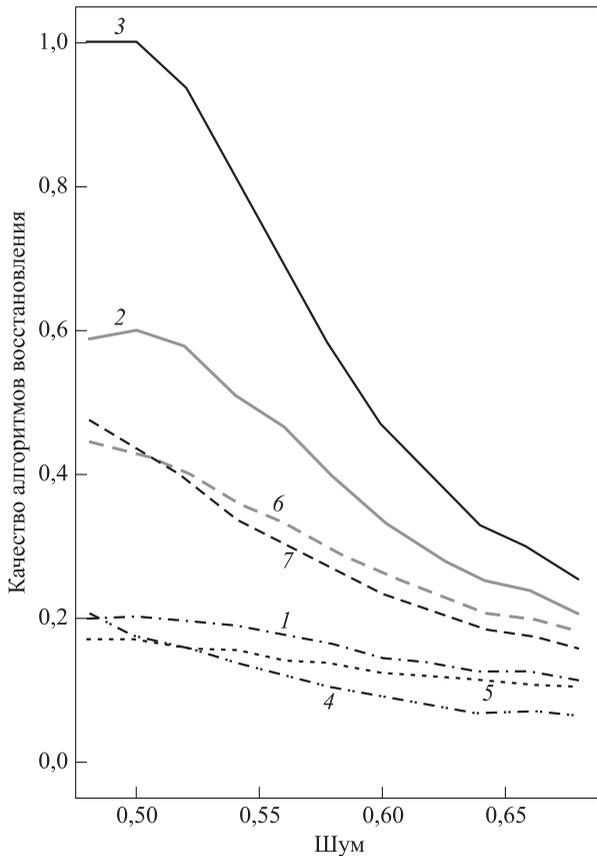


Рис. 2 Качество алгоритмов восстановления с базовыми функциями небольших арностей: 1 – DFS; 2 – BFS; 3 – алгоритм Прима; 4 – k -MST; 5 – k -MST–DFS; 6 – h -MST–BFS; 7 – k -MST – алгоритм Прима

Таблица 2 Качество алгоритмов реконструкции с равномерным шумом, близким к 0,5

Алгоритм	Шум				
	0,50	0,52	0,54	0,56	0,58
DFS	0,20	0,20	0,19	0,18	0,16
BFS	0,60	0,58	0,51	0,46	0,40
Прима	1,00	0,94	0,81	0,69	0,57
k -MST	0,17	0,16	0,14	0,12	0,10
k -MST–DFS	0,17	0,16	0,16	0,14	0,14
k -MST–BFS	0,43	0,40	0,36	0,33	0,29
k -MST–Прима	0,44	0,39	0,34	0,33	0,27

санных алгоритмов. Результаты PCST можно использовать в качестве априорных для других подходов, $M' := (1/2)(M'_{PCST} + M')$, поэтому результаты PCST обновляются M' .

Процедура генерации данных имеет следующие допущения: арности функций генерируются биномиальным распределением, поэтому существуют много функций с малой арностью, все базовые функции имеют только один вход. Любой случай с частичной реконструкцией считается ошибкой. Качество алгоритмов реконструкции:

$$\frac{1}{K} \sum_{k=1}^K [R(\bar{N}(M_k)) = M_k],$$

где R – алгоритм реконструкции; $\bar{N} = (N - \min(N)) / (\max(N) - \min(N))$ – нормированная матрица шума. Матрица N генерируется как $N(M) = M + U(-\alpha, \alpha)$. Генератор случайных чисел возвращает матрицу того же вида, что и M , где каждый элемент является независимой переменной из равномерного распределения в сегменте $[-\alpha, \alpha]$.

Вот список из семи сравниваемых алгоритмов: DFS, BFS, алгоритм Прима, k -MST через PCST, k -MST + DFS, k -MST + BFS, k -MST + алгоритм Прима. На рис. 2 показана ошибка алгоритмов реконструкции с шумом, близким к порогу 0,5. Наилучшие результаты дает алгоритм Прима. Второе по точности решение основано на BFS. Таблица 2 соответствует рис. 2 и показывает качество реконструкции семи алгоритмов для значений граничного шума 0,50–0,58.

8 Заключение

Предлагаются и сравниваются алгоритмы восстановления суперпозиции для задачи символьной регрессии. Алгоритм Прима дает наиболее точные результаты и устойчив к небольшому шуму в данных. Предлагаемый алгоритм дает точные результаты, но он более подвержен шуму в матрице суперпозиции. Алгоритмы, основанные на BFS и DFS, не могут восстановить исходную суперпозицию с зашумленными матрицами суперпозиции. Алгоритм PCST с BFS, используемый для реконструкции матрицы суперпозиции, показывает приемлемые для практического использования результаты.

Литература

1. Koza J. R. Genetic programming as a means for programming computers by natural selection // Stat. Comput., 1994. Vol. 4. P. 87–112.

2. *Searson D. P., Leahy D. E., Willis M. J.* GPTIPS: An open source genetic programming toolbox for multigene symbolic regression // *Multiconference (International) of Engineers and Computer Scientists Proceedings*, 2010. Vol. 1. P. 77–80.
3. *Stanley K. O., Miikkulainen R.* Evolving neural networks through augmenting topologies // *Evol. Comput.*, 2002. Vol. 10. Iss. 2. P. 99–127.
4. *Бочкарев А. М., Софронов И. Л., Стрижов В. В.* Порождение экспертно-интерпретируемых моделей для прогноза проницаемости горной породы // *Системы и средства информатики*, 2017. Т. 27. № 3. С. 74–87.
5. *Ravi R., Sundaram R., Marathe M. V., Rosenkrantz D. J., Ravi S. S.* Spanning trees — short or small // *SIAM J. Discrete Math.*, 1996. Vol. 9. Iss. 2. P. 178–200.
6. *Chudak F. A., Roughgarden T., Williamson D. P.* Approximate k -MSTs and k -Steiner trees via the primal-dual method and Lagrangean relaxation // *Math. Program.*, 2004. Vol. 100. Iss. 2. P. 411–421.
7. *Awerbuch B., Azar Y., Blum A., Vempala S.* New approximation guarantees for minimum-weight k -trees and prize-collecting salesmen // *SIAM J. Comput.*, 1998. Vol. 28. Iss. 1. P. 254–262.
8. *Aror S., Karakostas G.* A $2 + \varepsilon$ approximation algorithm for the k -MST problem // *Math. Program.*, 2006. Vol. 107. Iss. 3. P. 491–504.
9. *Hegde C., Indyk P., Schmidt L.* A fast, adaptive variant of the Goemans–Williamson scheme for the prize-collecting steiner tree problem // *11th DIMACS Implementation Challenge Workshop Proceedings*, 2014. P. 1–32. http://people.csail.mit.edu/ludwigs/papers/dimacs14_fastpcst.pdf.
10. *Ras C., Swanepoel K., Thomas D. A.* Approximate Euclidean Steiner trees // *J. Optimiz. Theory App.*, 2017. Vol. 172. Iss. 3. P. 845–873.
11. *Goemans M. X., Williamson D. P.* A general approximation technique for constrained forest problems // *SIAM J. Comput.*, 1995. Vol. 24. Iss. 2. P. 296–317.

Поступила в редакцию 23.01.22

OPTIMAL SPANNING TREE RECONSTRUCTION IN SYMBOLIC REGRESSION

R. G. Neychev¹, I. A. Shibaev¹, and V. V. Strijov²

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

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

Abstract: The paper investigates the problem of regression model generation. A model is a superposition of primitive functions. The model structure is described by a weighted colored graph. Each graph vertex corresponds to a primitive function. An edge assigns a superposition of two functions. The weight of an edge is equal to the probability of superposition. To generate an optimal model, one has to reconstruct its structure from its graph adjacency matrix. The proposed algorithm reconstructs the minimum spanning tree from the weighted colored graph. The paper presents a novel solution based on the prize-collecting Steiner tree algorithm. This algorithm is compared with its alternatives.

Keywords: symbolic regression; linear programming; superposition; minimum spanning tree; adjacency matrix

DOI: 10.14357/19922264230105

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, projects 20-37-90050 and 20-07-00990.

References

1. Koza, J. R. 1994. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* 4:87–112.
2. Searson, D. P., D. E. Leahy, and M. J. Willis. 2010. GPTIPS: An open source genetic programming toolbox for multigene symbolic regression. *Multiconference (International) of Engineers and Computer Scientists Proceedings*. 1:77–80.
3. Stanley, K. O., and R. Miikkulainen. 2002. Evolving neural networks through augmenting topologies. *Evol. Comput.* 10(2):99–127.
4. Bochkarev, A. M., I. L. Sofronov, and V. V. Strijov. 2017. Porozhdenie ekspertno-interpretiруемых моделей dlya prognoza pronitsaemosti gornoy porody [Generation of

- expertly-interpreted models for prediction of core permeability]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(3):74–87.
5. Ravi, R., R. Sundaram, M. V. Marathe, D. J. Rosenkrantz, and S. S. Ravi. 1996. Spanning trees — short or small. *SIAM J. Discrete Math.* 9(2):178–200.
 6. Chudak, F. A., T. Roughgarden, and D. P. Williamson. 2004. Approximate k -MSTs and k -Steiner trees via the primal-dual method and Lagrangean relaxation. *Math. Program.* 100(2):411–421.
 7. Awerbuch, B., Y. Azar, A. Blum, and S. Vempala. 1998. New approximation guarantees for minimum-weight k -trees and prize-collecting salesmen. *SIAM J. Comput.* 28(1):254–262.
 8. Arora, S., and G. Karakostas. 2006. A $2 + \varepsilon$ approximation algorithm for the k -MST problem. *Math. Program.* 107(3):491–504.
 9. Hegde, C., P. Indyk, and L. Schmidt. 2014. A fast, adaptive variant of the Goemans–Williamson scheme for the prize-collecting Steiner tree problem. *11th DIMACS Implementation Challenge Workshop Proceedings*. 1–32. Available at: http://people.csail.mit.edu/ludwigs/papers/dimacs14_fastpcst.pdf (accessed January 10, 2023).
 10. Ras, C., K. Swanepoel, and D. A. Thomas. 2017. Approximate Euclidean Steiner trees. *J. Optimiz. Theory App.* 172(3):845–873.
 11. Goemans, M. X., and D. P. Williamson. 1995. A general approximation technique for constrained forest problems. *SIAM J. Comput.* 24(2):296–317.

Received January 23, 2022

Contributors

Neychev Radoslav G. (b. 1994) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; neychev@phystech.edu

Shibaev Innokentii A. (b. 1997) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; shibaev.kesha@gmail.com

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; strijov@phystech.edu

ПРИЧИННО-СЛЕДСТВЕННЫЕ СВЯЗИ В ЗАДАЧАХ КЛАССИФИКАЦИИ

А. А. Грушо¹, Н. А. Грушо², М. И. Забейло³, В. В. Кульченков⁴, Е. Е. Тимонина⁵,
С. Я. Шоргин⁶

Аннотация: В данной работе объект классификации рассматривается как носитель причины появления одного или нескольких следствий и любой алгоритм классификации принимает решение о классе, наблюдая следствия из анализируемой причины. Рассматриваются следствия причины в задаче бинарной классификации как источники дополнительной информации, подтверждающие или отвергающие гипотезу о причине в классифицируемом объекте. При рассмотрении гипотезы о наличии или отсутствии определенной причины в классифицируемом по этому свойству объекте на основании нескольких следствий автоматически строится язык представления знаний. Тогда легко использовать доступную информацию из разных информационных пространств в задаче классификации объекта. Для использования причинно-следственных связей в задаче классификации необходимо использовать машинное обучение. В условиях обучения с учителем имеем множество прецедентов, когда известно наличие интересующей нас причины. Тогда можно статистически выделить события, которые стали следствиями этой причины. Детерминированные отношения причины и следствия порождают ошибки только за счет шума. В тех прецедентах, где нет причины, позитивная классификация появляется только за счет шума независимо от прецедента к прецеденту. Таким образом, даже слабое отклонение от равновероятного шума позволяет построить состоятельный критерий, выделяющий следствия от случайного шума. Выделение следствий можно проводить независимо друг от друга. Это следует из детерминированности отношения причина–следствие и независимости шума.

Ключевые слова: задача конечной классификации; причинно-следственные связи; машинное обучение

DOI: 10.14357/19922264230106

1 Введение

Вопросы причинно-следственных связей в построении решающих правил принятия решений исследовались в ряде научных публикаций. Обзор методов выявления причин в графических моделях можно найти в статье [1]. Большинство публикаций базируется на результатах исследований [2–4]. Особенность этих работ заключается в прогнозировании следствия в условиях зашумленной причины. Этот подход оказался достаточно успешным. Например, такой подход повысил устойчивость нейронных сетей [5].

Отметим, что рассматриваемый класс задач тесно связан с необходимостью распознавания следствий при различных модификациях причин. Чтобы сократить множество изменений причин, в работе [6] использовались специальные методы преобразования причин для повышения устойчивости работы классификаторов.

В данной статье предлагается иной подход к использованию причинно-следственных связей. В работе [7] решается задача восстановления знаний из фрагментов информации, причем эти фрагменты не всегда бывают точными. Следуя этой же идее, задачу конечной классификации можно рассматривать как выявление одной из k причин по следствиям, наблюдаемым в объектах, содержащих по одной из этих причин. Эти следствия можно рассматривать как источники дополнительной информации, доступные для уточнения искомой причины.

Таким образом, рассматривается задача классификации элементов множества D на конечное число непересекающихся классов, т.е. существует множество D и функция $F : D \rightarrow K = \{1, 2, \dots, k\}$, где k — число классов. Рассмотрим несколько примеров постановки задачи классификации.

1. Искусственные нейронные сети и машинное обучение с учителем. Пусть дана конечная учеб-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, info@itake.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

⁴Банк ВТБ (ПАО), vlad.kulchenkov@gmail.com

⁵Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

⁶Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

ная выборка S , состоящая из m элементов, и $S \subseteq D$. Обучение с учителем означает, что известна частичная функция $F_S : S \rightarrow K$, совпадающая с F на всех элементах из S . Искусственная нейронная сеть (ИНС) помнит значения функции F_S , но если $x \notin S$, то пользователи ИНС надеются, что F_S может правильно вычислить значение этой функции для данного x . Здесь возникают два вопроса:

- (а) почему F_S должна уметь вычислять что-то вне своей области определения?
- (б) если ИНС выдаст некоторый ответ при подаче на вход x , то почему этот ответ будет правильным, т. е. совпадет со значением $F(x)$?

Без ответов на эти вопросы возникает противоречие с определениями традиционной математики, а корректные ответы на них предполагают использование дополнительной информации о свойствах ИНС, функции F и свойствах множества D . Если дополнительной информации нет, то оба вопроса некорректны и любые ответы на них не имеют смысла.

2. **Риски в кредитных учреждениях.** Рассматривается поток заявок на кредит в фиксированном банке. Кредит и условия кредита определяются финансовым состоянием обратившегося за кредитом, который рассказывает о своих доходах и других аспектах своего финансового состояния. Описание состояния, данное клиентом, дополнительно оценивается по более чем 50 характеристикам, собранным из различных, чаще всего внешних для банка, источников. По собранным значениям проводится классификация обратившегося за кредитом с точки зрения возможности дефолта при выдаче кредита. Здесь, как и в предыдущем примере, обучение на опыте работы по выдаче кредитов формирует допустимые значения указанных характеристик. В работе [8] перечень источников допустимых дополнительных данных называется метаданными. Тогда заявленная обратившимся за кредитом клиентом и собранная дополнительная информация по клиенту определяют допустимость выдачи ему кредита, и причиной этого решения становится проведенная оценка его финансового состояния.
3. **Поиск причин сбоев в компьютерной системе.** Рассматривается задача удаленного администрирования информационной системы [9] в случае возникновения сбоя. Необходимо выявить первоначальную причину сбоя, с тем чтобы провести ремонт в кратчайшее время и вос-

становить рабочие процессы в информационной системе. Первопричина сбоя порождает информацию, которую администратор может запросить и получить удаленно. Такую информацию о первопричине опытный администратор умеет собирать и анализировать, чтобы локализовать место исходного сбоя. Достаточно часто такая дополнительная информация хранится в базе данных администратора и позволяет использовать накопленный опыт [10]. Заменяв блок, содержащий причину, администратор быстро восстанавливает работоспособность всей информационной системы. Как правило, этот алгоритм работает, когда первопричина единственна, хотя производных сбоев может быть несколько. В этом примере, как и в предыдущих, локализация происходит за счет дополнительной информации, поступающей из различных источников, которые могут иметь отношение к причине сбоя. Еще один интересный момент, связанный с этим примером, состоит в том, что не всегда надо искать собственно первопричину, а достаточно покрыть ее достаточно малой частью системы, которую несложно заменить [11]. В этом состоит идея приближенного причинно-следственного анализа.

4. **В задачах медицинской диагностики [12, 13] необходимо выявить причину исходного заболевания.** Врач обращается к различным источникам дополнительной информации для подтверждения диагноза. Существенной стороной этого примера является то, что врач несет повышенную ответственность за решение задачи о поставленном диагнозе.

В рассмотренных задачах важно различать дополнительную информацию о классифицируемом объекте, которая помогает правильной классификации, и информацию о предположениях, которые позволяют построить модель поиска и в рамках этой модели оценивать качество принимаемого данным методом классификации решения. Например, для ИНС это предположение о независимости и одинаковой распределенности выбора точек в множестве S , а также метрические соотношения между элементами в множестве D .

В данной работе объект классификации рассматривается с точки зрения присутствия в нем причины появления одного или нескольких следствий и любой алгоритм классификации принимает решение о классе, наблюдая следствия из анализируемой причины. В примере с ИНС предполагаемая близость x к одному из элементов x^* множества S позволяет допустить, что класс элемента x совпада-

ет с классом $F_S(x^*)$. Существует множество работ, подтверждающих или опровергающих это утверждение в различных моделях, но все эти работы основаны на предположениях, а не на достаточности информации для правильной классификации причины в данной задаче.

С другой стороны, объект классификации, отождествляемый с классифицируемой причиной, может породить множество следствий, которые несут информацию о принадлежности причины к заданному классу и могут служить подтверждением этой причины в данном классе или отрицанием этой причины в полученных новых данных. Дополнительная информация может ослабить предположения о модели классификатора.

В работе рассматриваются следствия причины в бинарной классификации как источники дополнительной информации, подтверждающие или отвергающие гипотезу о причине в классифицируемом объекте. При рассмотрении гипотезы о причине в классифицируемом по этому свойству объекте на основании нескольких следствий автоматически строится язык представления знаний при описании использования доступной информации из разных информационных пространств [14] в задаче классификации объекта.

2 Математическая модель причинно-следственных связей в задачах классификации

Сделаем два предположения о причинах и следствиях. Для простоты полагаем, что в задаче распознавания объект распознавания отождествляется только с одной причиной. Если есть причина, то каждое следствие возникает детерминированно (будем также говорить с вероятностью 1). Невозможно усовершенствовать одну процедуру (алгоритм) классификации, чтобы повысить качество и устойчивость распознавания, защищенность от атак. Для решения этих задач необходима дополнительная информация, которая содержится только в следствиях причины. Таким образом, описание процедуры распознавания одним алгоритмом описывается двумерным булевым вектором (y_1, y_2) , где y_1 соответствует наличию или отсутствию причины в распознаваемом объекте, т.е. при $y_1 = 1$ гипотеза о распознаваемой причине считается правильной, а при $y_1 = 0$ неверной. Рассматриваемый алгоритм распознавания при $y_2 = 1$ определяет распознаваемую причину, а при $y_2 = 0$ не распознает причину. Поскольку распознавание идет по следствию причины, то объясним происхождение воз-

можных ошибок распознавания. Предположим, что возможен шум в определении наличия следствия причины, т.е. алгоритм не идентифицирует появление следствия, если шум позволяет изменить значение переменной y_2 на 0 или в отсутствие причины изменить значение y_2 на 1. Для простоты будем оценивать вероятность влияния шума одной константой $\varepsilon > 0$. Тогда условные вероятности результатов классификации наличия причины данным алгоритмом можно оценить следующим образом:

$$P(y_2 = 1 | y_1 = 1) \geq 1 - \varepsilon;$$

$$P(y_2 = 0 | y_1 = 1) \leq \varepsilon;$$

$$P(y_2 = 1 | y_1 = 0) \leq \varepsilon;$$

$$P(y_2 = 0 | y_1 = 0) \geq 1 - \varepsilon.$$

В предположении, что шум появляется независимо, но, возможно, действует по-прежнему одна причина, то независимо для любого другого возможного следствия и алгоритма классификации этого следствия верны такие же формулы.

Рассмотрим важное понятие устойчивости распознавания одним алгоритмом для бинарного случая [15]. Численная оценка устойчивости в этой работе имеет следующий вид:

$$R = \frac{P(y_2 = 1 | y_1 = 1)}{P(y_2 = 1 | y_1 = 1) + P(y_2 = 1 | y_1 = 0)}.$$

Путем несложных преобразований при небольших ε получим

$$R \geq 1 - \frac{\varepsilon}{1 - \varepsilon}.$$

Предположим, что помимо основного алгоритма распознавания используется еще одно следствие для подтверждения вывода, т.е. рассматривается вектор $(y_2^{(1)}, y_2^{(2)})$. Опираясь на то, что следствия при наличии причины возникают с вероятностью 1, а шумы независимы, получаем:

$$P(y_2^{(1)} = 1, y_2^{(2)} = 1 | y_1 = 1) \geq (1 - \varepsilon)^2;$$

$$P(y_2^{(1)} = 1, y_2^{(2)} = 0 | y_1 = 1) = P(y_2^{(1)} = 0, y_2^{(2)} = 1 | y_1 = 1) \leq \varepsilon;$$

$$P(y_2^{(1)} = 1, y_2^{(2)} = 1 | y_1 = 0) \leq \varepsilon^2;$$

$$P(y_2^{(1)} = 1, y_2^{(2)} = 0 | y_1 = 0) = P(y_2^{(1)} = 0, y_2^{(2)} = 1 | y_1 = 0) \leq \varepsilon;$$

$$P(y_2^{(1)} = 0, y_2^{(2)} = 0 | y_1 = 0) \geq (1 - \varepsilon)^2.$$

Вероятность, что произойдет хотя бы одна ошибка классификации, когда нет причины, равна

$$1 - P\left(y_2^{(1)} = 0, y_2^{(2)} = 0 \mid y_1 = 0\right) \leq 1 - (1 - \varepsilon)^2.$$

Если рассматривать k следствий в качестве источников дополнительной информации, то обобщение со случая $k = 2$ проводится следующим образом. Необходимо работать с двоичным вектором $(y_2^{(1)}, y_2^{(2)}, \dots, y_2^{(k)})$ длины k , и вероятность хоть одной случайной позитивной классификации, когда нет причины, равна

$$1 - P\left(y_2^{(1)} = 0, y_2^{(2)} = 0, \dots, y_2^{(k)} = 0 \mid y_1 = 0\right) \leq 1 - (1 - \varepsilon)^k.$$

Эта формула может использоваться в оценке устойчивости, когда имеются k источников взаимно дополняющей информации.

Параметр k может трактоваться как дискретное время, тогда можно рассматривать задачу об оптимальной остановке случайного процесса с дискретным временем [16, 17].

3 Подготовительная работа для использования причинно-следственных связей в задаче конечной классификации

Для использования причинно-следственных связей в задаче классификации необходимо использовать машинное обучение. В условиях обучения с учителем имеем множество прецедентов, когда известно наличие причины. Тогда можно статистически выделить события, которые стали следствиями причины. Детерминированные отношения причины и следствия порождают ошибки только за счет шума. В тех прецедентах, где нет причины, позитивная классификация появляется только за счет шума независимо от прецедента к прецеденту. Таким образом, даже слабое отклонение ε от $1/2$ позволяет построить состоятельный критерий, выделяющий следствия среди случайного шума.

Выделение следствий можно проводить независимо друг от друга. Это следует из детерминированности отношения причина—следствие и независимости шума. Однако нельзя исключать возможность зависимости найденных следствий между собой. В силу детерминированности отношения причина—следствие, если утверждение B

является следствием утверждения A , но утверждение C является следствием утверждения B , то утверждение C является следствием утверждения A . При этом использовать утверждение C как источник дополнительной информации нельзя, так как вся информация об утверждении A уже использована в утверждении B .

Конечно, возможна ситуация, когда утверждение B является сильно зашумленным и порождает ошибку классификации, а утверждение C дает правильный ответ, но для такого рассмотрения нужна другая дополнительная информация. Поэтому необходимо на множестве следствий изучить причинно-следственные связи утверждений без привязки к основной первопричине утверждения A .

Эту задачу также можно решать с помощью обучения. Если утверждение C является следствием причины B , то на множестве прецедентов, содержащих утверждение B и не содержащих утверждение A , построенный выше алгоритм определяет, является ли утверждение C следствием утверждения B .

Рассмотрим ситуацию приближенных причинно-следственных связей, когда утверждение A содержит (покрывает) настоящую причину A^* , т.е. $A^* \subset A$, а утверждение B является следствием утверждения A^* . Тогда утверждение B является также следствием утверждения A , т.е. при появлении утверждения A следствие B появляется детерминированно. Поэтому если в классификации используется утверждение A , но не первопричина A^* , то дополнительная информация в классификации причины, содержащейся в утверждении A , также содержится в утверждении B .

Проблема может возникнуть, когда в утверждении A содержится искомая причина A^* и содержится утверждение B . Утверждение B не является следствием утверждения A^* , но у утверждения B есть следствие C , которое не является следствием A^* . В этом случае нельзя углубляться в поиск первопричины A^* и можно рассматривать утверждение C как источник дополнительной информации для классификации утверждения A . Пусть является истинным, что утверждение A покрывает причину A^* и утверждение A правильно отнесено к нужному классу за счет дополнительной информации в утверждении C . Тогда дополнительная информация в утверждении C может быть использована для классификации утверждения A .

Ситуация становится хуже, когда утверждение A не содержит причину A^* , но утверждение C является следствием A . Однако при обучении с учителем использовались только те прецеденты, которые содержат искомую причину A^* . Поэтому рассматри-

ваемый случай следствия C из утверждения A не попадает в обучающую выборку.

Проведенные рассуждения показывают, что в задаче классификации с использованием причинно-следственных связей можно пользоваться приближенным каузальным анализом.

Рассмотренный подход позволяет использовать различные множества данных, содержащих дополнительную для классификации информацию. Поиск дополнительной информации проводится в различных информационных пространствах, поэтому языки описания утверждений (событий) в разных информационных пространствах могут кардинально отличаться. Вместе с тем для использования дополнительной информации в методе причинно-следственных связей особенности языков в разных информационных пространствах не являются существенными, и для использования в описанном методе достаточно только идентификации следствия, привязанной к причине. Таким образом, построен простой язык представления знаний, объединяющий информацию, представленную в разных информационных пространствах для целей повышения достоверности решения задачи классификации.

4 Заключение

В работе построен метод использования причинно-следственных связей для повышения достоверности решения задачи конечной классификации. Основная идея работы состоит в том, что любой алгоритм решения задачи конечной классификации не может быть совершенным, так как информация для его применения ограничена множеством допустимых входных данных. В то же время качество классификации можно многократно повышать с помощью привлечения дополнительной уточняющей информации. Если представить задачу конечной классификации в виде идентификации одной из конечного множества причин и решать эту задачу по наблюдениям за следствиями этих причин, то множество следствий и есть дополнительная, уточняющая правильное решение задачи классификации, информация.

Для восстановления структуры причинно-следственных связей можно использовать машинное обучение. При этом различные следствия одной причины можно независимо друг от друга выявлять, используя одну обучающую выборку. Это позволяет строить эффективные алгоритмы решения таких задач.

Предложенный метод позволяет строить простые языки представления знаний. Такие языки позволяют использовать информацию из различ-

ных информационных пространств и возможности статистического анализа в этих пространствах. Объединяющим элементом многих информационных пространств являются причинно-следственные связи, порождаемые классифицируемой первопричиной.

Литература

1. *Glymour C., Zhang K., Spirtes P.* Review of causal discovery methods based on graphical models // *Frontiers Genetics*, 2019. Vol. 10. Art. 524. 15 p. doi: 10.3389/fgene.2019.00524.
2. *Halpern J.Y., Pearl J.* Causes and explanations: A structural-model approach. Part I: Causes. // *Brit. J. Philos. Sci.*, 2005. Vol. 56. No. 4. P. 843–887.
3. *Pearl J.* Causal inference // *Workshop on Causality Proceedings: Objectives and Assessment at NIPS* / Eds. I. Guyon, D. Janzing, B. Scholkopf. — *Proceedings of machine learning research ser.* — Whistler, Canada, 2010. Vol. 6. P. 39–58.
4. *Pearl J.* The mathematics of causal inference // *Joint Statistical Meetings Proceedings.* — ASA, 2013. P. 2515–2529.
5. *Zhang C., Zhang K., Li Y.* A causal view on robustness of neural networks // *Working Papers by Cornell University*, 2020. arXiv:2005.01095v3 [cs.LG]. 21 p.
6. *Wang B., Lyle C., Kwiatkowska M.* Provable guarantees on the robustness of decision rules to causal interventions // *30th Joint Conference (International) on Artificial Intelligence Proceedings*, 2021. P. 4258–4265. <https://www.ijcai.org/proceedings/2021/0585.pdf>.
7. *Valiant L. G.* Knowledge infusion: In pursuit of robustness in artificial intelligence // *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science* / Eds. R. Hariharan, M. Mukund, V. Vinay. — *Leibniz international proceedings in informatics ser.* — Wadern, Germany: Schloss Dagstuhl — Leibniz-Zentrum fuer Informatik, 2008. Vol. 2. P. 415–422. doi: 10.4230/LIPIcs.FSTTCS.2008.1770.
8. *Grusho A., Grusho N., Zabezhailo M., Timonina E., Senchilo V.* Metadata for root cause analysis // *Communications ECMS*, 2021. Vol. 35. Iss. 1. P. 267–271. doi: 10.7148/2021-0267.
9. *Грушо А. А., Грушо Н. А., Забежайло М. И., Тимонина Е. Е.* Удаленный мониторинг рабочих процессов // *Информатика и её применения*, 2021. Т. 15. Вып. 3. С. 2–8.
10. *Грушо А. А., Грушо Н. А., Забежайло М. И., Тимонина Е. Е.* Локализация исходной причины аномалии // *Проблемы информационной безопасности. Компьютерные системы*, 2020. № 4. С. 9–16. EDN: JXHQYU.
11. *Грушо Н. А., Грушо А. А., Забежайло М. И., Тимонина Е. Е.* Методы нахождения причин сбоев в информационных технологиях с помощью метаданных // *Информатика и её применения*, 2020. Т. 14. Вып. 2. С. 33–39. doi: 10.14357/19922264200205.

12. Richens J. G., Lee C. M., Johri S. Improving the accuracy of medical diagnosis with causal machine learning // *Nat. Commun.*, 2020. Vol. 11. Art. 3923. 9 p. doi: 10.1038/s41467-020-17419-7.
13. Забейайло М. И., Грушо А. А., Грушо Н. А., Тимонина Е. Е. Поддержка решения задач диагностического типа // *Системы и средства информатики*, 2021. Т. 31. № 1. С. 69–81.
14. Grusho A., Grusho N., Timonina E. Method of several information spaces for identification of anomalies // *Intelligent distributed computing XIII* / Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, M. Ivanovic. — *Studies in computational intelligence ser.* — Cham: Springer, 2020. Vol. 868. P. 515–520. doi: 10.1007/978-3-030-32258-8_60.
15. Xiong P., Buffett S., Iqbal S., Lamontagne P., Mamun M., Molyneaux H. Towards a robust and trustworthy machine learning system development: An engineering perspective // *J. Information Security Applications*, 2022. Vol. 65. Art. 103121. 58 p. doi: 10.1016/j.jisa.2022.103121.
16. Вальд А. Последовательный анализ / Пер. с англ. — М.: Физматлит, 1960. 328 с. (*Wald A. Sequential analysis.* — J. Wiley & Sons, 1959. 212 p.)
17. Ширяев А. Н. Вероятность: в 2 кн. — 3-е изд. — М.: МЦНМО, 2004. 521 с.

Поступила в редакцию 12.01.23

CAUSAL RELATIONSHIPS IN CLASSIFICATION PROBLEMS

A. A. Grusho¹, N. A. Grusho¹, M. I. Zabezhailo¹, V. V. Kulchenkov², E. E. Timonina¹, and S. Ya. Shorgin¹

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

²VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation

Abstract: In the present paper, a classification object is considered as the cause for the appearance of one or more consequences and any classification algorithm decides on the class observing the consequences from the analyzed cause. The paper considers the consequences of the cause in the binary classification problem as sources of additional information confirming or rejecting the hypothesis of the cause in the classified object. When considering a hypothesis about the presence or absence of a certain cause in an object classified by this property, the knowledge presentation language is automatically built based on several consequences. Then, it is easy to use the available information from different information spaces in an object classification task. To use cause-and-effect relationships in a classification task, machine learning should be used. In conditions of teaching with a teacher, there are many precedents when the presence of a cause is known. Then one can statistically single out events that are the consequences of the cause. Deterministic cause-and-effect relationships generate errors only at the expense of noise. In those precedents where there is no cause, positive classification appears only at the expense of noise regardless of precedent to precedent. Thus, even a weak deviation from equally probable noise allows one to build a consistent criterion that distinguishes consequences from random noise. Sequelae can be isolated independently of each other. This follows from the determinism of the cause-and-effect relationship and the independence of noise.

Keywords: finite classification task; cause-and-effect relationships; machine learning

DOI: 10.14357/19922264230106

References

1. Glymour, C., K. Zhang, and P. Spirtes. 2019. Review of causal discovery methods based on graphical models. *Frontiers Genetics* 10:524. 15 p. doi: 10.3389/fgene.2019.00524.
2. Halpern, J. Y., and J. Pearl. 2005. Causes and explanations: A structural-model approach. Part I: Causes. *Brit. J. Philos. Sci.* 56(4):843–887.
3. Pearl, J. 2010. Causal inference. *Workshop on Causality Proceedings: Objectives and Assessment at NIPS*. Eds. I. Guyon, D. Janzing, and B. Scholkopf. *Proceedings of machine learning research ser.* 6:39–58.
4. Pearl, J. 2013. The mathematics of causal inference. *Joint Statistical Meetings Proceedings*. ASA. 2515–2529.
5. Zhang, C., K. Zhang, and Y. Li. 2021. A causal view on robustness of neural networks. *arXiv.org*. 21 p. Available at: <https://arxiv.org/abs/2005.01095> (accessed January 20, 2023).
6. Wang, B., C. Lyle, and M. Kwiatkowska. 2021. Provable guarantees on the robustness of decision rules to causal interventions. *30th Joint Conference (International) on Artificial Intelligence Proceedings*. 4258–4265. Available at: <https://www.ijcai.org/proceedings/2021/0585.pdf> (accessed March 13, 2023).
7. Valiant, L. G. 2008. Knowledge infusion: In pursuit of robustness in artificial intelligence. *IARCS Annual Con-*

- ference on Foundations of Software Technology and Theoretical Computer Science Proceedings*. Eds. R. Hariharan, M. Mukund, and V. Vinay. Leibniz international proceedings in informatics ser. Wädern, Germany: Schloss Dagstuhl — Leibniz-Zentrum fuer Informatik. 2:415–422. doi: 10.4230/LIPIcs.FSTTCS.2008.1770.
8. Grusho, A., N. Grusho, M. Zabezhailo, E. Timonina, and V. Senchilo. 2021. Metadata for root cause analysis. *Communications ECMS* 35(1):267–271. doi: 10.7148/2021-0267.
 9. Grusho, A., N. Grusho, M. Zabezhailo, and E. Timonina. 2021. Udalennyy monitoring rabochikh protsessov [Remote monitoring of workflows]. *Informatika i ee Primeneniya — Inform. Appl.* 15(3):2–8.
 10. Grusho, A., N. Grusho, M. Zabezhailo, and E. Timonina. 2021. Localization of the root cause of the anomaly. *Autom. Control Comp. S.* 55(8):978–983.
 11. Grusho, N.A., A.A. Grusho, M.I. Zabezhailo, and E. E. Timonina. 2020. Metody nakhozhdeniya prichin sboev v informatsionnykh tekhnologiyakh s pomoshch'yu metadannykh [Methods of finding the causes of information technology failures by means of meta data]. *Informatika i ee Primeneniya — Inform. Appl.* 14(2):33–39. doi: 10.14357/19922264200205.
 12. Richens, J. G., C. M. Lee, and S. Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11(1):3923. 12 p. doi: 10.1038/s41467-020-17419-7.
 13. Zabezhailo, M. I., A. A. Grusho, N. A. Grusho, and E. E. Timonina. 2021. Podderzhka resheniya zadach diagnosticheskogo tipa [Support for solving diagnostic type problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):69–81.
 14. Grusho, A., N. Grusho, and E. Timonina. 2020. Method of several information spaces for identification of anomalies. *Intelligent distributed computing XIII*. Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, and M. Ivanovic. Studies in computational intelligence ser. Cham: Springer. 868:515–520. doi: 10.1007/978-3-030-32258-8_60.
 15. Xiong, P., S. Buffett, S. Iqbal, P. Lamontagne, M. Mammun, and H. Molyneaux. 2022. Towards a robust and trustworthy machine learning system development: An engineering perspective. *J. Information Security Applications* 65:103121. 58 p. doi: 10.1016/j.jisa.2022.103121.
 16. Wald, A. 1959. *Sequential analysis*. J. Wiley & Sons. 212 p.
 17. Shiryaev, A. N. 2004. *Veroyatnost'* [Probability]. 3rd ed. Moscow: MTsNMO. 2 vols. 521 p.

Received January 12, 2023

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, 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; grusho@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 Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; info@itake.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Kulchenkov Vladislav V. (b. 1989) — head of risk monitoring department, VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation; vlad.kulchenkov@gmail.com

Timonina Elena E. (b. 1952) — Doctor of Science in technology, 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 119133, Russian Federation; eltimon@yandex.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, 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 119133, Russian Federation; sshorgin@ipiran.ru

РАЗРАБОТКА НОВОЙ МОДЕЛИ СТУПЕНЧАТОЙ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ КЛАССИФИКАЦИИ АНОМАЛИЙ НА ПАНОРАМАХ

П. О. Архипов¹, С. Л. Филиппских², М. В. Цуканов³

Аннотация: Описывается разработанная новая модель ступенчатой сверточной нейронной сети для классификации аномалий на панорамах. Выбраны подходящие наборы данных для классификации. Сделан вывод о неполноте применявшегося ранее авторами метода поиска аномалий особых областей с высоким цветоразличием на панорамах. Поиск данных областей разработанным ранее методом не ставил перед собой задачу их классификации. Для автоматической идентификации обнаруженных объектов предлагается применить модели глубокого обучения с использованием подходящих нейросетей. Особое внимание уделено работе с данными, содержащими несбалансированные классы и изображения разного размера. Проводится сравнение результатов классификации изображений популярных архитектур нейронных сетей с разработанной ступенчатой сверточной нейронной сетью.

Ключевые слова: панорамное изображение; набор данных; многоклассовая классификация; ступенчатая сверточная нейронная сеть; ансамбль; перенос обучения

DOI: 10.14357/19922264230107

1 Введение

Применение беспилотных летательных аппаратов (БПЛА) широко представлено во многих сферах производства и строительства, таких как геодезия, картография, энергетика, инженерия, видеонаблюдение и т. п. В каждой из этих областей требуется оперативное наблюдение за технологическими процессами и своевременное выявление аномалий, возможное благодаря аэрофотоснимкам высокого разрешения, полученным с БПЛА.

В рамках выполнения работ по данному направлению [1–3] была рассмотрена задача по обнаружению аномалий на сравниваемых разновременных панорамах инспектируемой местности. Для решения поставленной задачи был разработан метод, основанный на поиске аномалий за счет обнаружения особых областей с высоким цветоразличием, представленных внешними и внутренними характеристиками [4, 5]. Поиск аномалий этим методом не ставил перед собой задачу их классификации. Поэтому дальнейшим развитием метода может стать разработка новой модели ступенчатой сверточной нейронной сети для классификации аномалий на панорамах в рамках новой парадигмы машинного обучения.

Многоклассовая классификация аномалий, полученных при анализе созданных панорамных

изображений, предполагает построение модели глубокого обучения [6]. Данная модель должна получать на вход множество примеров изображений реальных объектов, проводить их статистический анализ и выявлять скрытую структуру в данных.

Обучающая выборка должна содержать изображения разного размера, так как это соответствует реальным условиям работы модели: различная высота съемки БПЛА, реальные размеры объектов разных классов, разрешение видеопаратуры и др. Для классификации изображений с помощью нейронных сетей их нужно привести к одному размеру. Нормализация размеров приводит к потере части информации (в случае больших изображений) и появлению лишних шумов (в случае маленьких изображений). В результате точность классификации может сильно уменьшиться [7].

Проблему разных размеров входных изображений можно решить с помощью ансамбля нейронных сетей. Каждое входное изображение масштабируется сразу в несколько размеров и подается на вход нейросетевых моделей. После этого результаты работы всех сетей усредняются и выдается общий результат. Такой подход может быть очень эффективен, но он требует значительных вычислительных ресурсов, так как придется обучать на одной выборке несколько нейронных сетей [7].

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, arpaul@mail.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, philippsl@mail.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, tsukanov.m.v@yandex.ru

2 Выбор подходящего набора данных

Для экспериментов с моделями машинного обучения был выбран датасет VisDrone [8], так как он довольно объемный и включает классы, встречающиеся на анализируемых панорамах. Весь набор данных разделен на три части: train (обучающее подмножество), valid (валидационное подмножество) и test (тестовое подмножество) (табл. 1).

Из получившихся изображений сформирована следующая выборка:

Таблица 1 Сбалансированные подмножества датасета

Подмножество	Число фотографий
train	284 993
valid	47 499
test	47 499

Таблица 2 Сбалансированные соотношения классов в датасете

Класс	train, %	valid, %	test, %
0 (bus)	2,39	2,37	2,46
1 (car)	57,41	57,26	57,36
2 (people)	35,92	36,10	35,89
3 (truck)	4,28	4,27	4,29

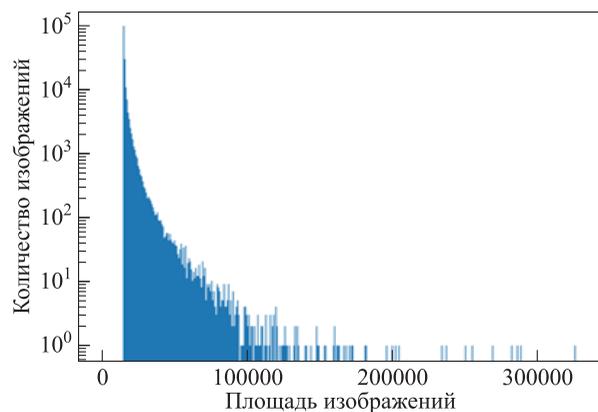


Рис. 1 Гистограмма распределения площадей изображений

класс 0 (bus): фотографии автобусов;

класс 1 (car): фотографии легковых автомобилей и микроавтобусов;

класс 2 (people): фотографии людей и пешеходов;

класс 3 (truck): фотографии грузовиков.

При работе классификатора в случае неуверенности нейронной сети, к какому классу следует отнести изображение, чаще всего модель выбирает самый вероятный класс. По этой причине очень важно для корректной оценки процесса обучения иметь одинаково сбалансированную выборку. В данном случае необходимо провести перебалансировку датасета. Число изображений в valid и test должно быть одинаковым и иметь такое же разбиение по классам, как и train (табл. 2).

Помимо разбиения на классы полезно оценить распределение размеров изображений в датасете. Для этого построим гистограмму распределения площадей изображений всей выборки (рис. 1).

Из гистограммы распределений видно, что в датасете больше всего небольших изображений размером до 50 пикселей. С увеличением размера число изображений резко уменьшается.

Для построения нейронной сети необходимо получить несколько базовых оценок выборки (табл. 3): стандартное отклонение, минимальную, максимальную, среднюю и медианную площади. У разных классов значения оценок сильно отличаются, и это еще один фактор, осложняющий построение нейросетевой модели и классификацию.

3 Выбор нейронной сети для классификации изображений

В качестве базовой оценки (baseline) попробуем классифицировать все изображения датасета VisDrone [8] с помощью одной нейронной сети. Для сравнения возьмем три базовые оценки, сделанные с помощью нейронных сетей разного размера и глубины: простая нейронная сеть, состоящая из нескольких сверточных слоев (ConvNN), VGG19 [9] и Xception [10].

Таблица 3 Оценки площадей изображений в датасете

Оценка	Для всей выборки	0 (bus)	1 (car)	2 (people)	3 (truck)
Минимальная площадь	25	25	25	25	25
Максимальная площадь	328 640	324 300	328 640	72 670	315 202
Средняя площадь	2 694	6 983	3 435	718	6 480
Стандартное отклонение	7 085	17 325	7 270	1 419	14 737
Медианная площадь	728	2 379	1 152	330	2 150

Для обучения нейронных сетей применяется распространенный алгоритм Adam [11] — метод стохастического градиентного спуска, основанный на адаптивной оценке моментов. Градиентный спуск — это один из самых эффективных методов оптимизации, поскольку вычисление частных производных первого порядка относительно всех параметров имеет ту же вычислительную сложность, что и просто вычисление функции [11]. Корректировка весов с помощью градиентного спуска вычисляется по следующей формуле:

$$\theta_t = \theta_{(t-1)} - \alpha \nabla_{\theta} f_t(\theta_{(t-1)}),$$

где θ — настраиваемый параметр; t — итерация обучения; α — шаг обучения; ∇ — градиент ошибки; f — функция потерь.

Метод Adam [11] позволяет вычислять индивидуальные скорости адаптивного обучения для разных параметров на основе оценок первого и второго моментов градиентов [11]:

$$m_t = \beta_1 m_{(t-1)} + (1 - \beta_1) \nabla_{\theta} f_t(\theta_{(t-1)});$$

$$v_t = \beta_2 v_{(t-1)} + (1 - \beta_2) (\nabla_{\theta} f_t(\theta_{(t-1)}))^2,$$

где m — первый момент; v — второй момент; $\beta_1, \beta_2 \in [0; 1]$ — экспоненциальные скорости затухания для оценок моментов.

В качестве функции потерь используется категориальная перекрестная энтропия (categorical cross-entropy). Это обусловлено тем, что метки ко всем изображениям в выборке были закодированы в векторы из 0 и 1 (one-hot encoding) [7].

Для устойчивости результатов обучения моделей применяется метод фиксации генератора случайных значений: перед каждым запуском процесса создания и обучения нейронной сети необходимо перезапускать генератор случайных чисел с фиксированным начальным значением. Это позволяет инициализировать веса моделей одинаковыми значениями, и данные при обучении будут поступать в том же порядке [7].

Сверточные слои в многослойной нейронной сети имеют два важных свойства [6]:

- (1) пространственная инвариантность признаков — любой паттерн, выученный рецептивным полем нейрона, может быть распознан в любом месте изображения;

- (2) пространственная иерархия паттернов — сложность шаблонов, которые распознают рецептивные поля нейронов, возрастает с увеличением числа слоев сети.

Эти два свойства позволяют применять очень эффективный метод построения нейронных сетей — перенос обучения (transfer learning) [7]. Если имеется уже обученная на большом наборе данных сеть, то найденные паттерны можно применить при решении других задач. Эффективность переноса обучения возрастает, если исходная обучающая выборка содержала классы, которые присутствуют в новой задаче.

В качестве первой модели для базовой оценки разработаем свою простую сверточную нейронную сеть (ConvNN), состоящую из 4 сверточных слоев (8, 16, 32 и 64 нейрона), 4 слоев MaxPooling и одного полносвязного слоя из четырех нейронов с активационной функцией softmax. Специально выбрано небольшое число слоев и нейронов. Число нейронов соответствует степеням двойки для лучшей оптимизации на GPU [7]. В сети 24 788 параметров, и все они настраиваемые.

Вторая модель — сеть VGG19 [9]. Нейронная сеть строится методом переноса обучения. Из сети VGG19 возьмем только сверточное ядро (14 сверточных слоев и 5 слоев MaxPooling), обученное на наборе данных ImageNet [6]. Общее число параметров сети: 20 156 740. Число настраиваемых параметров: 132 356.

Третья модель — сеть Xception [9]. Нейронная сеть строится методом переноса обучения. Из сети Xception возьмем только сверточное ядро, обученное на наборе данных ImageNet. В нем около 50 сверточных слоев. Общее число параметров сети: 25 581 356. Число настраиваемых параметров: 4 719 876.

Оценим для каждого класса результаты работы сетей, обученных на подмножествах train, valid и test. В табл. 4 приведены результаты работы базовых моделей на подмножестве test (метрика — точность).

Можно сделать вывод, что результаты классификации мало зависят от сложности архитектуры нейронных сетей. Модели хорошо классифицируют два больших класса (car, people) и очень

Таблица 4 Результаты классификации трех базовых моделей

Модель	Все классы, %	0 (bus), %	1 (car), %	2 (people), %	3 (truck), %
ConvNN	85,3	17,6	90,9	94,0	15,7
VGG19	82,3	27,9	87,7	89,5	12,7
Xception	81,5	28,7	82,5	95,1	11,0

плохо работают на маленьких (bus, truck). Это типичная ситуация для несбалансированной выборки с изображениями разного размера: нейронная сеть в случае спорной ситуации выбирает самые распространенные классы, тем самым максимизируя вероятность правильной классификации. Такая стратегия хорошо влияет на среднюю оценку, но точность классификации маленьких классов при этом падает.

4 Разработка модели ступенчатой сверточной нейронной сети

Для того чтобы решить проблему классификации изображений на маленьких классах, построим ансамбль нейронных сетей, который может работать с изображениями разного размера. Для уменьшения числа настраиваемых параметров новая архитектура классификатора будет строиться на основе ступенчатой нейронной сети.

Разработанная ступенчатая сверточная нейронная сеть (StepwiseNet) будет иметь два важных свойства:

- (1) способность работать с изображениями разных размеров;
- (2) использование меньших вычислительных ресурсов, чем ансамбль нейронных сетей.

При обучении ансамбля нейронных сетей первые сверточные слои каждой сети обучаются поиску самых простых паттернов на изображениях. Таких паттернов не так много, и они очень похожи во всех сетях ансамбля, т. е. при таком обучении происходит дублирование вычислений. Так как обучение ансамбля проходит на одном датасете, то дублирование паттернов встречается и в более глубоких сверточных слоях.

Основная идея ступенчатой сверточной нейронной сети — обучить все слои модели один раз, а потом при помощи переноса обучения создать ансамбль из нескольких сетей.

Ступенчатая сверточная нейронная сеть имеет модульную структуру. Число модулей и их архитектура (число слоев и число нейронов в слое) — это гиперпараметры. Они подбираются под каждую задачу. Число сетей в ансамбле зависит от числа модулей. Для примера создадим модель из

трех модулей. В дальнейшем она будет развернута в ансамбль из трех нейронных сетей.

На первом этапе необходимо создать сеть из трех модулей (рис. 2). Архитектура каждого модуля зависит от задачи и может состоять из разного числа слоев. Далее необходимо выбрать размер входного слоя. Размер должен соответствовать выбранной архитектуре сети. После этого модель обучается на обучающем подмножестве.

StepwiseNet будет состоять из следующих трех модулей:

модуль 1: входной слой 8×8 пикселей, 2 сверточных слоя, 2 слоя MaxPooling;

модуль 2: входной слой 40×40 пикселей, 4 слоя свертки с разделением по глубине (depthwise separable convolution) и пакетная нормализация (batch normalization) в конце модуля;

модуль 3: входной слой 70×70 пикселей, 7 слоев свертки с разделением по глубине (depthwise separable convolution) и пакетная нормализация (batch normalization) в конце модуля.

Классификатор состоит из двух полносвязных слоев. В сети 543 940 параметров, и все они настраиваемые. Для обучения нейронных сетей применяется алгоритм Adam [11]. В качестве функции потерь используется категориальная перекрестная энтропия.

На втором этапе из обученной модульной сети формируется ансамбль. Если в ступенчатой сверточной нейронной сети три модуля, то первый модуль используется в качестве сверточного ядра для классификации самых маленьких изображений. Первый и второй модули используются в качестве сверточного ядра для классификаций изображений среднего размера. Третья сеть формируется на основе сверточного ядра, состоящего из первого, второго и третьего модулей.

В результате этих шагов создается ансамбль нейронных сетей разной глубины, в котором повторно используются однократно обученные сверточные ядра (рис. 3). При использовании классификатора на практике изображение, в зависимости от размера, подается на вход только одной из нейронных сетей ансамбля. Такая схема работы экономит ресурсы и позволяет работать с изображениями различного размера выборки.

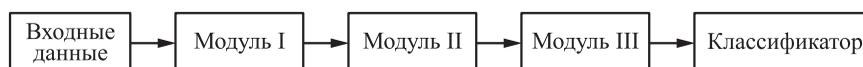


Рис. 2 Пример ступенчатой сверточной нейронной сети из трех модулей

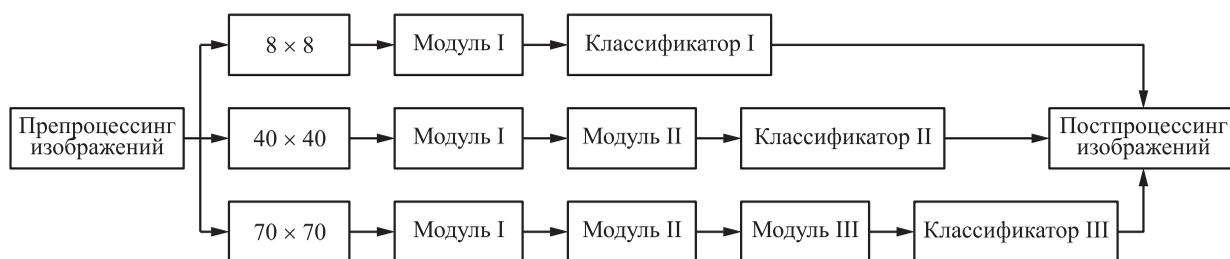


Рис. 3 Пример создания из модульной сети ансамбля нейронных сетей

Таблица 5 Результаты классификации StepwiseNet по классам

Класс	Точность, %
0 (bus)	68,3
1 (car)	90,1
2 (people)	91,0
3 (truck)	65,9
По всем классам	89,4

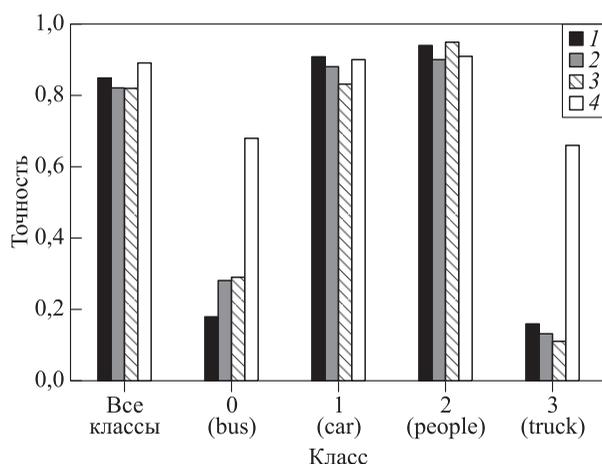


Рис. 4 Диаграмма классификации изображений с помощью всех нейронных сетей: 1 — ConvNN; 2 — VGG19; 3 — Xception; 4 — StepwiseNet

Для классификации изображений из датасета VisDrone спроектируем ступенчатую сверточную нейронную сеть из трех модулей. Число модулей выбрано в соответствии с данными, представленными ранее на рис. 1. Для каждого модуля необходимо установить размер входного слоя. Эти размеры являются гиперпараметрами и выбираются для каждой задачи отдельно. В качестве эвристики можно использовать минимальное изменение исходного размера изображения при масштабировании. Для этого нужно разбить всю выборку на несколько диапазонов в зависимости от размеров изображений. Из каждого диапазона можно брать медианы или средние значения выборки.

Входной слой для первой сети ансамбля — 8×8 пикселей (медианное значение для первого

диапазона выборки). Входной слой для второй сети ансамбля — 40×40 пикселей. Входной слой для третьей сети ансамбля — 70×70 пикселей.

В табл. 5 представлены результаты классификации ансамбля, построенного на основе ступенчатой сверточной нейронной сети. На рис. 4 представлена диаграмма классификации изображений с помощью всех нейронных сетей.

5 Заключение

Разработанная ступенчатая сверточная нейронная сеть позволяет гибко настраивать классификатор под различные задачи. Основные преимущества данного подхода:

- использование разных сетей для классификации изображений разного размера позволяет избежать потери информации и шумов при масштабировании;
- число сетей в ансамбле можно настраивать путем изменения числа модулей;
- архитектуру каждого модуля можно изменять и обучать отдельно, фиксируя веса всех остальных слоев;
- повторное использование модулей при формировании сверточных ядер позволяет экономить вычислительные ресурсы.

Результаты работы ансамбля, обученного на подмножествах train, valid и test, показывают, что с помощью разработанной ступенчатой сверточной нейронной сети можно существенно улучшить результаты классификации несбалансированной выборки с изображениями разного размера. Общая точность классификации по сравнению с моделями ConvNN, VGG19 и Xception возросла на 4%–8%. При этом точность распознавания на малочисленных классах возросла в 2–6 раз (для класса 0 (bus) — с 17%–28% до 68%, для класса 3 (truck) — с 11%–15% до 66%).

Общее число параметров в ступенчатой сверточной нейронной сети на два порядка меньше, чем в сетях VGG19 и Xception (544 тыс. против

20–25 млн). Малое число параметров ускоряет обучение, требует меньше вычислительных ресурсов и снижает риск переобучения.

Литература

1. Архипов П. О., Цуканов М. В. Алгоритмическая модель метода детектирования аномалий при аэрофотосъемке объектов жилищно-коммунального хозяйства // Информационные системы и технологии, 2019. № 6(116). С. 39–45.
2. Архипов П. О., Цуканов М. В. Информационная модель технологии коррекции яркости и цвета при создании панорамных изображений // Системы высокой доступности, 2020. Т. 16. № 3. С. 46–51. doi: 10.18127/j20729472-202003-04.
3. Архипов П. О., Цуканов М. В. Алгоритмическая модель обнаружения аномалий на разновременных панорамах // Системы высокой доступности, 2021. Т. 17. № 2. С. 5–10. doi: 10.18127/j20729472-202102-01.
4. Архипов П. О., Колесник А. В., Цуканов М. В. Программная система обнаружения аномалий на разновременных панорамах обследуемой местности (CL.22). Свидетельство о государственной регистрации программы для ЭВМ № 2022615139, опубл. 30.03.2022.
5. Trofimenkov A. K., Tsukanov M. V., Nosova N. Y. Modification of an algorithmic model for detecting anomalies on panoramas of different times obtained from unmanned aerial vehicles // Modern informatization problems in simulation and social technologies. — Yelm, WA, USA: Science Book Publishing House LLC, 2022. P. 183–186.
6. Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks // Commun. ACM, 2017. Vol. 60. P. 84–90. doi: 10.1145/3065386.
7. Chollet F. Deep learning with Python. — 2nd ed. — Shelter Island, NY, USA: Manning Publications Co., 2021. 504 p.
8. Zhu P., Wen L., Du D., Bian X., Fan H., Hu Q., Ling H. Detection and tracking meet drones challenge // arXiv.org, 2021. 20 p. doi: 10.48550/arXiv.2001.06303.
9. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv.org, 2015. 14 p. doi: 10.48550/arXiv.1409.1556.
10. Chollet F. Xception: Deep learning with depthwise separable convolutions // arXiv.org, 2017. 8 p. doi: 10.48550/arXiv.1610.02357.
11. Kingma D., Ba J. Adam: A method for stochastic optimization // arXiv.org, 2017. 15 p. doi: 10.48550/arXiv.1412.6980.

Поступила в редакцию 06.06.22

DEVELOPMENT OF A NEW MODEL OF STEP CONVOLUTIONAL NEURAL NETWORK FOR CLASSIFICATION OF ANOMALIES ON PANORAMAS

P. O. Arkhipov, S. L. Philippskih, and M. V. Tsukanov

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

Abstract: A new model of a stepped convolutional neural network for classifying anomalies in panoramas has been developed. Appropriate datasets for classification are selected. The conclusion is made about the incompleteness of the method previously used by the authors to find anomalies in special areas with high color difference in panoramas. The search for these areas by the previously developed method did not set the task of their classification. For automatic identification of detected objects, it is proposed to apply deep learning models using suitable neural networks. Particular attention is paid to work with data containing unbalanced classes and images of different sizes. The results of image classification of popular architectures of neural networks are compared with the newly developed stepped convolutional neural network.

Keywords: panoramic image; data set; multilabel classification; stepwise convolutional neural network; ensemble; transfer learning

DOI: 10.14357/19922264230107

References

1. Arkhipov, P. O., and M. V. Tsukanov. 2019. Algoritmicheskaya model' metoda detektirovaniya anomalii pri aerofotos'emke ob'ektov zhilishchno-kommunal'nogo khozyaystva [Algorithmic model of the method of detecting anomalies in the aerial photography of objects of housing and communal services]. *Informatsionnye sistemy i tekhnologii* [Information Systems and Technologies] 6(116):39–45.
2. Arkhipov, P. O., and M. V. Tsukanov. 2020. Informatsionnaya model' tekhnologii korrektsii yarkosti i tsveta pri sozdanii panoramnykh izobrazheniy [Information model of brightness and color correction technology for cre-

- ating panoramic images]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 16(3):46–51. doi: 10.18127/j20729472-202003-04.
3. Arkhipov, P. O., and M. V. Tsukanov. 2021. Algoritmicheskaya model' obnaruzheniya anomalii na raznovremennykh panoramakh [Algorithmic model of anomaly detection on different panoramas]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 17(2):5–10. doi: 10.18127/j20729472-202102-01.
 4. Arkhipov, P. O., A. V. Kolesnik, and M. V. Tsukanov. 30.03.2022. Programmnaya sistema obnaruzheniya anomalii na raznovremennykh panoramakh obsleduemoy mestnosti (CL_22) [Software system for detecting anomalies on multitemporal panoramas of the surveyed area (CL_22)]. Certificate of State Registration No. 2022615139.
 5. Trofimenkov, A. K., M. V. Tsukanov, and N. Y. Nosova. 2022. Modification of an algorithmic model for detecting anomalies on panoramas of different times obtained from unmanned aerial vehicles. *Modern informatization problems in simulation and social technologies proceedings*. Yelm, WA: Science Book Publishing House LLC. 183–186.
 6. Krizhevsky, A., I. Sutskever, and G. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60:84–90. doi: 10.1145/3065386.
 7. Chollet, F. 2021. *Deep learning with Python*. 2nd ed. Shelter Island, NY: Manning Publications Co. 504 p.
 8. Zhu, P., L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling. 2021. Detection and tracking meet drones challenge. *arXiv.org*. 20 p. doi: 10.48550/arXiv.2001.06303.
 9. Simonyan, K., and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. *arXiv.org*. 14 p. doi: 10.48550/arXiv.1409.1556.
 10. Chollet, F. 2017. Xception: Deep learning with depth-wise separable convolutions. *arXiv.org*. 8 p. doi: 10.48550/arXiv.1610.02357.
 11. Kingma, D., and J. Ba. 2017. Adam: A method for stochastic optimization. *arXiv.org*. 15 p. doi: 10.48550/arXiv.1412.6980.

Received June 6, 2022

Contributors

Arkhipov Pavel O. (b. 1979) — Candidate of Science (PhD) in technology, director, Orel Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 137 Moskovskoe Shosse, Orel 302025, Russian Federation; arpaul@mail.ru

Philippskih Sergey L. (b. 1987) — junior scientist, Orel Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 137 Moskovskoe Shosse, Orel 302025, Russian Federation; philippsl@mail.ru

Tsukanov Maxim V. (b. 1995) — junior scientist, Orel Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 137 Moskovskoe Shosse, Orel 302025, Russian Federation; tsukanov.m.v@yandex.ru

МОДЕЛИРОВАНИЕ СТРУКТУРЫ ИНТЕРОПЕРАБЕЛЬНОСТИ СРЕДСТВАМИ СТРУКТУРНОЙ СОГЛАСОВАННОСТИ

И. Н. Розенберг¹, С. К. Дулин², Н. Г. Дулина³

Аннотация: Начальный синтаксический уровень интероперабельности предполагает коммуникацию с соответствующим протоколом, аппаратные средства, программное обеспечение и необходимый уровень совместимости данных. Исследованию уровня совместимости данных, описывающих взаимодействующие элементы на основе вектора признаков, посвящена представленная работа. Для этого предлагается модель структурного соответствия, позволяющая оценивать тенденцию к установлению интероперабельности. Моделирование структурной интероперабельности на основе анализа знаков связей с помощью введенного критерия согласованности приводит к нахождению ближайшего к исходному множеству консонансного прообраза. Найденный консонансный прообраз своими подмножествами указывает на предпочтительную группировку элементов, при которой интероперабельность между ними устанавливается с наименьшей рассогласованностью относительно зафиксированных знаков связей. Поскольку рассматриваемые элементы описаны вектором параметров, из сравнения которых можно сделать вывод о сходстве между элементами, соответственно, нахождение элементов в одном подмножестве говорит о потенциальной мотивации к интероперабельности.

Ключевые слова: интероперабельность; структурная согласованность; матрица связности

DOI: 10.14357/19922264230108

1 Введение

При интеграции и глобализации информационных систем ключевым фактором становится интероперабельность как неотъемлемое свойство эффективного функционирования синтезированных систем и элементов. По определению органов стандартизации [1], «интероперабельность — способность двух или более информационных систем или компонентов к обмену информацией и к использованию информации, полученной в результате обмена». Эталонная модель интероперабельности по ГОСТ Р 55062–2012 представляет собой трехуровневую модель, включающую техническую интероперабельность, семантическую интероперабельность и организационную интероперабельность.

В настоящее время в отечественной литературе обсуждаются, главным образом, вопросы реализации технической интероперабельности. Семантический и организационный уровни интероперабельности информационных систем обсуждаются только теоретически.

В работе [2] представлена обобщенная модель интероперабельности, разработанная международным консорциумом организаций NCOIC — “Systems, Capabilities, Operations, Programs, and Enter-

prises Model for Interoperability Assessment” (SCOPE-модель), а в [3] обоснован вариант декомпозиции параметров SCOPE-модели и их приведения к эталонной модели, представленной в ГОСТ Р 55062–2012 (рис. 1).

2 Построение модели структурной согласованности

Помимо мотивации к интероперабельности активных элементов существуют определенные признаки или характеристики в структуре элементов, которые способствуют или препятствуют достижению интероперабельности. Оценивая потенциальную возможность установления интероперабельности той или иной степени в структуре взаимосвязанных элементов, можно говорить о структурной интероперабельности. Для изучения возможности информационных систем или элементов обладать тенденцией к интероперабельности в зависимости от соотношения собственных признаков или характеристик можно предложить некоторую модель структурного соответствия, позволяющую оценивать группы потенциально близких друг к другу элементов по ряду выбранных при-

¹ Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте, I.Rozenberg@vniias.ru

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте, skdulini@mail.ru

³ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ngdulina@mail.ru

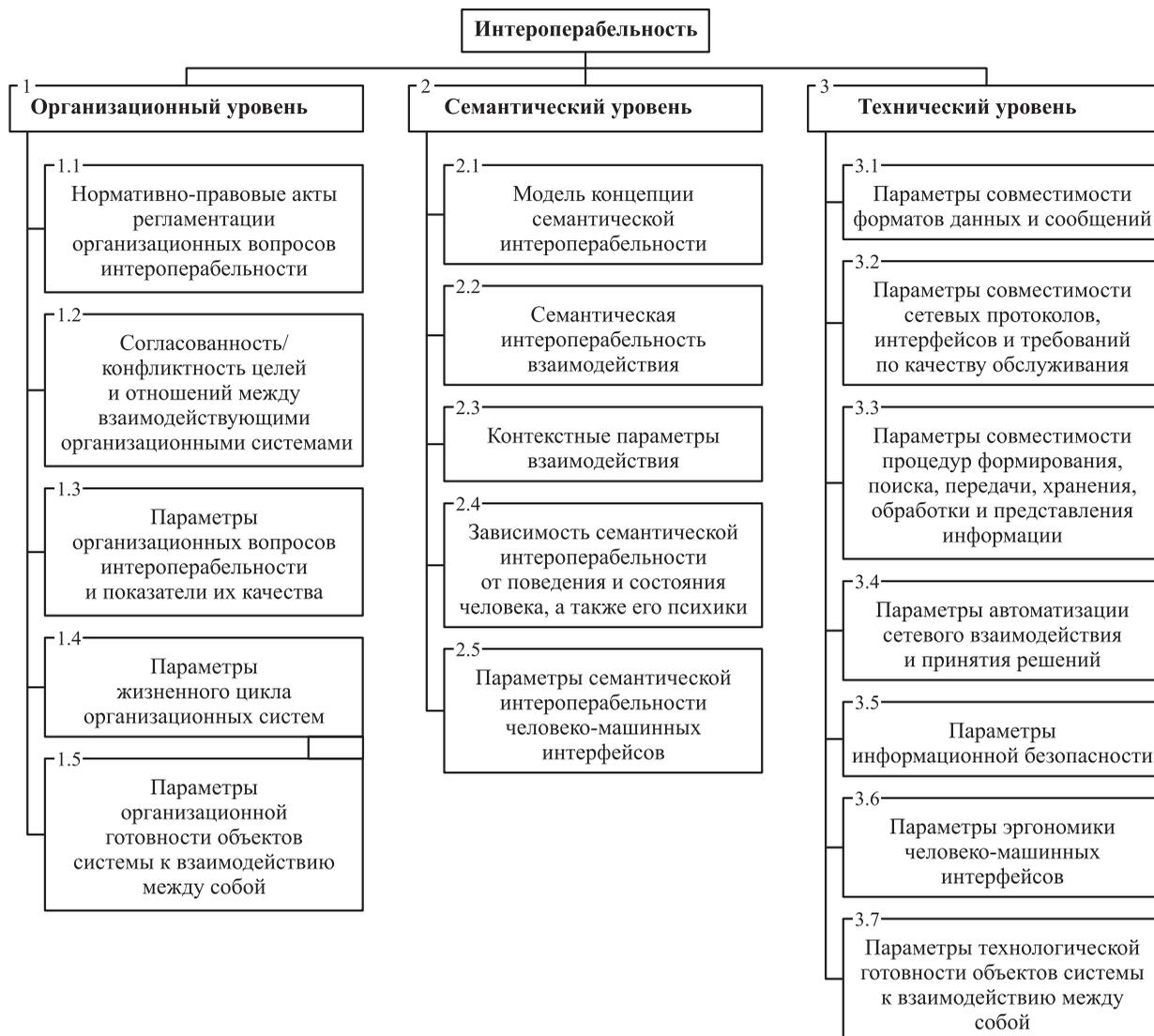


Рис. 1 Общая структура интероперабельности в соответствии с ГОСТ Р 55062-2012

знаков, важных для установления интероперабельности. Для этого рассмотрим некоторое множество из N элементов, которые могут быть представлены как агенты, информационные системы или компоненты знаний. Эти элементы оказываются вовлечены во взаимодействие, уровень которого требуется оценить на предмет интероперабельности и представить структуру интероперабельности на основании анализа предпочтительности установления отношений между элементами. Предлагается привлечь для этого аппарат структурной согласованности [4].

Поставленная задача, таким образом, определяет разбиение множества потенциально взаимодействующих элементов на наборы мотивированных к взаимодействию элементов. Рассматриваемые элементы могут быть описаны вектором параметров

(признаков), из сравнения которых можно сделать вывод о сходстве между элементами и, соответственно, потенциальной мотивации к интероперабельности.

Другими словами, рассматривается множество элементов $O = \{o_i\}$ ($i = 1, \dots, N$), где описание каждого элемента представлено в виде вектора из m признаков-атрибутов: $o_i = (p_1^i, \dots, p_m^i)$. Сравняя любые два элемента этого множества на основе этих признаков, можно оценить их сходство. Следует отметить, что допустимо использовать либо все m признаков, либо подмножество из $k \leq m$ признаков, позволяющее дать оценку для каждой пары элементов.

Поступая стандартным образом, зададим F — функцию сходства элементов по k признакам, нормированную на максимальный диапазон значений

признака. Для двух элементов o_i и o_j , сходство которых устанавливается на основе k признаков $\{p_k\}$, функция F имеет вид:

$$F(o_i, o_j) = \frac{1}{k} \sum_{i=1}^k w_{ml} \frac{|p_{ml}^i - p_{ml}^j|}{\max |p_{ml}^i - p_{ml}^j|},$$

где $0 \leq w_{ml} \leq 1$ — вес ml -го признака, а $\max |p_{ml}^i - p_{ml}^j|$ — диапазон значений ml -го признака.

Функция F принимает значения из $[0, 1]$, так что $F = 1$ — это абсолютное сходство элементов o_i и o_j , а $F = 0$ — абсолютное различие. Остальные значения функции трактуются как оцифрованные степени сходимости пар элементов по k признакам, позволяя представить множество как граф со взвешенными связями.

Далее предлагается перейти к знаковому графу, позволяющему на основе анализа согласованности его связей по выбранному критерию согласованности сгруппировать N элементов в P групп. Знаковый граф получается в результате выбора порогового значения α для функции F так, что, когда $0 \leq F(o_i, o_j) \leq \alpha$, элементы o_i и o_j считаются несходными по k признакам, а в случае $\alpha < F(o_i, o_j) \leq 1$ — сходными. Сопоставляя знак минус связям с $0 \leq F(o_i, o_j) \leq \alpha$ и знак плюс — остальным, получим знаковую структуру, представляющую собой дискретную знаковую модель множества потенциально взаимодействующих элементов.

Знаковому графу можно поставить в соответствие матрицу связности и сформулировать задачу структурной согласованности интероперабельности. Стандартный подход к анализу согласованности заключается в оценке бинарных отношений между элементами. Но взаимное влияние связей требует учитывать тернарные отношения, которые могут быть отнесены к согласованному или рассогласованному состоянию. В [5] эти состояния названы консонансным и диссонансным соответственно. Если представить знаковый граф в виде совокупности тернарных отношений, то получится система анализа структурной согласованности потенциально взаимодействующих элементов. В качестве критерия согласованности предлагается выбрать тернарный критерий [5], относящий тернарные отношения (треугольники) к консонансным, если положительная связь между любыми парами вершин задает тождественность связей этих вершин с третьей, в противном случае — к диссонансным.

При разбиении знакового графа на треугольники и исследовании согласованности множества по критерию Хайдера выясняется, что консонансное множество M_K (состоящее только из консонансных треугольников) представимо в виде двух

подмножеств M_1 и M_2 : $M_K = M_1 \cup M_2$, так что любые два элемента o_i и o_j из одного подмножества связаны положительной связью, а принадлежащие разным подмножествам — отрицательной. Если M_1 или M_2 пустое, то такое множество называется тривиальным, — оно содержит только положительные связи.

Непредсказуемое изменение знаков связей консонансного множества может привести к ассонансному множеству, так как некоторые треугольники становятся диссонансными. Инvertированное изменение вернет ассонансное множество в консонанс. Можно говорить о наборе связей, однократное изменение знаков которого преобразует ассонансное множество в консонанс. Назовем такие связи сильными диссонансными и определим поиск согласованного состояния для ассонансного множества как поиск его сильных диссонансных связей.

Существуют $[N/2]+1$ типов консонансного множества из N элементов, что указывает на множественность наборов сильных диссонансных связей у каждого ассонансного множества по отношению к типам консонанса. Среди существующих наборов сильных диссонансных связей можно найти минимальный набор, который приводит заданное ассонансное множество к некоторому консонансному типу. Соответственно, назовем консонансным прообразом ассонансного множества консонансное множество, получающееся переводом ассонансного множества посредством изменения знаков минимального набора сильных диссонансных связей [5].

Приведение ассонансного множества в консонанс по Хайдеру означает возможность представления его в виде двух классов эквивалентности элементов. Для поиска согласованного множества, состоящего более чем из двух подмножеств, необходимо расширить критерий Хайдера. Введем понятия поликонсонанса и консонанса степени P [5]. Поликонсонанс степени P соответствует согласованному состоянию множества, состоящего из не более чем P подмножеств, так что элементы внутри каждого подмножества связаны только положительными связями, а из разных подмножеств — только отрицательными. Под консонансом степени P будем понимать поликонсонанс степени P , состоящий в точности из P классов (рис. 2).

Предельный случай поликонсонанса при $P = N$, где N — число элементов множества, представляет собой тривиальное диссонансное множество. Тем самым диссонансное множество представляется как очень слабо согласованное, и вопрос преобразования в согласованное состояние можно ставить только относительно ассо-

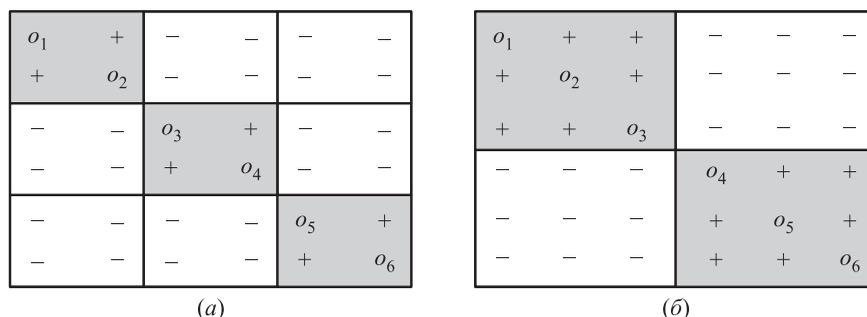


Рис. 2 Примеры матриц связности консонансных множеств: (а) консонанс степени 3; (б) консонанс степени 2

нансного множества. Эвристические методы могут уменьшить перебор при поиске поликонсонанса. Одним из таких простых методов является учет при разбиении множества на подмножества поликонсонанса тождественности структур элементов, находящихся в одном и том же подмножестве как классе эквивалентности по структурному признаку. Другими словами, любые два элемента с тождественными связями со всеми элементами из некоторого подмножества и имеющие положительную связь друг с другом, будут располагаться в этом же подмножестве. Это позволяет выявлять элементы с тождественными структурами связей, представляя их в виде интегрированного элемента, что сокращает число перебираемых элементов. Сложность такой задачи оценивается примерно как сложность задачи сортировки элементов.

Понятие поликонсонанса степени $P > 1$ корректирует понятие ассонансного множества как множества, не относящегося ни к консонансным, ни к диссонансным.

Пусть дано множество, матрица связности которого изображена на рис. 2, а. Это множество соответствует поликонсонансу степени 3, но в условиях консонанса степени 2 оно становится ассонансным. Понижение степени консонанса, как правило, переводит консонансное множество в ассонансное, а с повышением степени консонанса, наоборот, ассонансное множество может перейти в консонанс.

Центральная задача представленного исследования — это уменьшение рассогласованности посредством поиска консонансного прообраза минимальным набором сильных диссонансных связей.

Структура консонансного множества такова, что любые два элемента o_i и o_j из одного и того же подмножества обладают тождественной структурой связей и неразличимы по структуре связей среди других элементов этого подмножества. С другой стороны, если o_i и o_j взяты из разных подмножеств, то связи с элементами из противоположного подмножества у них инвертированы по знаку. Отсюда

следует правило перевода элемента из одного подмножества в другое: следует изменить знаки всех его связей с элементами подмножества, в котором он находится, на противоположные.

Перевод элемента в другое подмножество может трактоваться как переброс вершины соответствующего множества знакового графа из одного подграфа в другой. Эта операция называется *поверхинным перебросом*. Любая последовательность поверхностных перебросов не изменяет вид состояния множества: консонанс, диссонанс или ассонанс не чувствительны к этой операции.

3 Поиск ближайшей консонансной структуры

Алгоритм поиска структурного соответствия, позволяющий оценивать связность групп потенциально близких друг к другу элементов, заключается в оценке состояния ассонансного множества, моделирующего потенциальную возможность элементов к интероперабельности. Полученное в результате моделирования ассонансное множество сравнивается с некоторым консонансным прообразом и преобразуется с помощью операции поверхностного переброса так, чтобы найти ближайший по числу различаемых связей консонансный прообраз.

Пусть заданы два множества M_1 и M_2 , состоящие из одних и тех же N элементов, у которых могут различаться знаки связей между парами одних и тех же элементов. Для связей каждой пары элементов o_i и o_j в двух множествах зададим число r_{ij} :

$$r_{ij} = \begin{cases} 0, & \text{если } i = j; \\ 0, & \text{если } i \neq j \text{ и связи между } o_i \text{ и } o_j \text{ в этих} \\ & \text{множествах различны;} \\ 1, & \text{если } i \neq j \text{ и связи между } o_i \text{ и } o_j \text{ в этих} \\ & \text{множествах совпадают.} \end{cases}$$

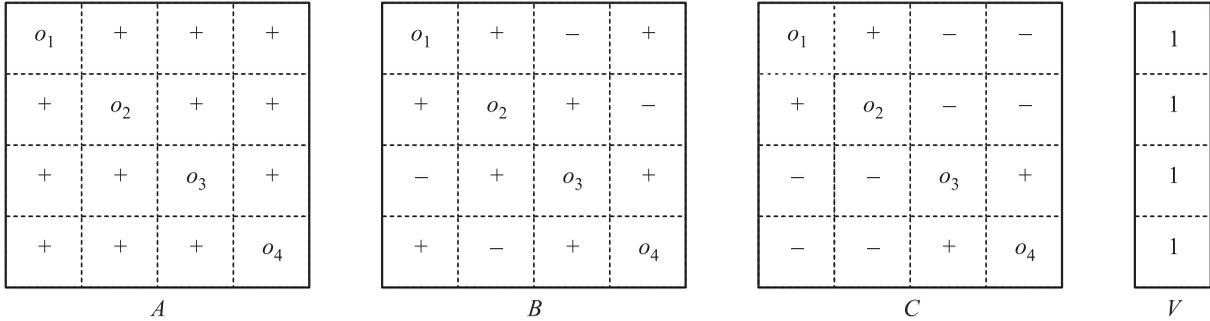


Рис. 3 Пример различных множеств A и B с одним и тем же вектором поверхностных различий от C

Для оценки различия множеств M_1 и M_2 введем последовательность из N чисел v_1, v_2, \dots, v_N , где $v_i = \sum_{j=1}^N r_{ij}$ для $i = 1, \dots, N$, и определим вектор $V = (v_1, v_2, \dots, v_N)$, характеризующий различие в знаках связей у пар одних и тех же элементов в двух множествах. Каждую компоненту вектора V можно интерпретировать как сумму различий в связях некоторой вершины в графах, соответствующих M_1 и M_2 . Благодаря такой интерпретации вектор V можно назвать вектором поверхностных различий M_1 и M_2 . Сумма v_1, v_2, \dots, v_N этого вектора $S_v = \sum_{i=1}^N v_i$ равна удвоенной сумме знаковых различий заданных множеств. Поскольку $0 \leq v_i \leq N - 1$ для любого i , то сумма поверхностных различий S_v не превышает $N(N - 1)$ и всегда четна.

Определение множества на основе вектора поверхностных различий от другого множества в общем случае неоднозначно, что можно проиллюстрировать на следующем примере (рис. 3).

Множество, заданное на основе матрицы связности B , имеет вектор поверхностных различий V , все компоненты которого — единицы, по сравнению с тривиальным консонансным множеством с соответствующей матрицей A . Этот же вектор поверхностных отличий определяет сравнение множества B с консонансным множеством с соответствующей матрицей C .

Минимальное отличие суммы инвертированных связей в множестве M выражается в минимальной сумме компонентов вектора поверхностных различий с M . Выведем соотношение для уменьшения суммы поверхностных различий при поверхностных перебросах.

Возьмем множество из N элементов и какое-нибудь консонансное множество из тех же элементов, тогда, согласно определению, вектор поверхностных различий для этих множеств:

$$V = (v_1, v_2, \dots, v_N),$$

где $v_i = \sum_{j=1}^N r_{ij}$.

Сумма поверхностных различий:

$$S_v = \sum_{i=1}^N v_i = \sum_{i=1}^N \sum_{j=1}^N r_{ij}.$$

Если существует элемент, у которого сумма его несовпадающих связей больше половины всех связей этого элемента, то, не нарушая общности, можно считать, что это элемент o_1 , и тогда $v_1 = (N - 1)/2$. Инвертирование знаков связей этого элемента изменяет вектор поверхностных различий следующим образом:

$$\begin{aligned} S'_v &= \sum_{i=1}^N v'_i = v'_1 + \sum_{i=2}^N \sum_{j=1}^N r'_{ij} = \\ &= v'_1 + \sum_{i=2}^N r'_{i1} + \sum_{i=2}^N \sum_{j=2}^N r'_{ij}. \end{aligned}$$

Поскольку при поверхностном перебросе все связи o_1 изменяются на противоположные, то $v'_1 = N - 1 - v_1$. Так как $r_{ij} = r_{ji}$, имеем

$$\sum_{i=2}^N r'_{i1} = \sum_{j=2}^N r'_{1j} = v'_1 = N - 1 - v_1.$$

При поверхностном перебросе элемента o_1 связи между o_i и o_j ($i > 1$ и $j > 1$) не изменяются; следовательно,

$$\sum_{i=2}^N \sum_{j=2}^N r'_{ij} = \sum_{i=2}^N \sum_{j=2}^N r_{ij}.$$

Подставляя в S'_v , получаем

$$S'_v = (N - 1 - v_1) + (N - 1 - v_1) + (S_v - 2v_1).$$

Отсюда

$$S'_v - S_v = 2(N - 1 - v_1) - 2v_1.$$

Значит, для того чтобы поверхностный переброс элемента привел к уменьшению суммы поверхностных

o_1	+	+	+	+	+	+	+	-	V_1	V_2
	+	o_2	+	+	-	-	+	+	3	2
	+	+	o_3	+	-	+	-	-	3	2
	+	+	+	o_4	-	+	-	-	3	2
	+	-	-	-	o_5	-	-	-	3	1
	+	-	+	+	-	o_6	+	+	3	1
	+	+	-	-	-	+	o_7	+	2	2
	+	+	-	-	-	+	+	o_8	2	2
	-	-	+	+	-	+	+	+	o_9	2
									2	2

Рис. 4 Два минимальных вектора поверхностных различий с различными суммами v_i

различий, нужно брать элементы, для которых компоненты вектора суммы поверхностных различий больше половины знаков связей:

$$v_i > \frac{N - 1}{2}.$$

Последовательные поверхностные перебросы в общем случае не обеспечивают достижение минимально удаленного консонансного прообраза. Можно продемонстрировать это, рассмотрев ассонансное множество на рис. 4.

Матрица связности этого ассонансного множества состоит из 9 элементов. Вектор поверхностных различий V_1 характеризует отличие по связям данного множества от консонансного множества ($\{o_1, o_2, o_3, o_4, o_5\}; \{o_6, o_7, o_8, o_9\}$) из тех же элементов. Ни одна из компонент V_1 не превышает половины числа связей соотносимых элементов, и, значит, любой поверхностный переброс в данном множестве будет увеличивать S_v . При этом найденное консонансное множество не является минимально удаленным от рассматриваемого ассонансного множества, так как существует множество ($\{o_5\} : \{o_2, o_3, o_4, o_1, o_6, o_7, o_8, o_9\}$) с вектором поверхностных различий V_2 , обладающим суммой S_v , меньшей, чем для вектора V_1 .

Становится ясно, что алгоритм, в основе которого лежит последовательный поверхностный переброс элемента с компонентом вектора поверхностных различий, большим, чем половина связей этого

элемента, находит только локальный минимум удаления от консонанса. Следовательно, алгоритм последовательного уменьшения поверхностных различий должен находить все возможные неулучшаемые состояния. Для реализации такой возможности нужно сформулировать некоторые специальные условия, позволяющие найти минимально удаленный консонансный прообраз.

Если вернуться к примеру ассонансного множества из 9 элементов, то можно проанализировать получение двух векторов поверхностных различий для двух консонансных прообразов. Консонансный прообраз, соответствующий вектору V_2 , предпочтительнее первого прообраза. Чтобы преобразовать первый консонансный прообраз во второй консонансный прообраз, нужно инвертировать знаки связей элементов o_1, o_2, o_3 и o_4 , что эквивалентно операции поверхностного переброса применительно к каждому из этих элементов. Возникает характерная ситуация: последовательные поверхностные перебросы элементов o_1, o_2, o_3 или o_4 , приводят к увеличению S_v , а групповой поверхностный переброс, проведенный для o_1, o_2, o_3 и o_4 одновременно, позволяет уменьшить S_v .

Значит, необходимо искать такую группу из k элементов, одновременный поверхностный переброс которой уменьшает S_v . Рассмотрим в связи с этим некое ассонансное множество из N элементов: $M = \{o_i\}, i = 1, \dots, N$. Поставим ему в соответствие какой-нибудь консонансный про-

образ с вектором попершинных различий $V = (v_1, v_2, \dots, v_N)$ с суммой элементов

$$S_v = \sum_{i=1}^N v_i = \sum_{i=1}^N \sum_{j=1}^N r_{ij}.$$

Выберем произвольные k элементов. Не нарушая общности, можно выбрать o_1, \dots, o_k . Осуществив для группы из k элементов операции попершинных перебросов, преобразуем вектор попершинных различий $V' = (v'_1, v'_2, \dots, v'_N)$ и $S'_v = \sum_{i=1}^N v'_i = \sum_{i=1}^N \sum_{j=1}^N r'_{ij}$. Распишем эту сумму в виде четырех сумм:

$$\begin{aligned} S'_v &= \sum_{i=1}^k \sum_{j=1}^N r'_{ij} + \sum_{i=k+1}^N \sum_{j=1}^N r'_{ij} = \sum_{i=1}^k \sum_{j=1}^k r'_{ij} + \\ &+ \sum_{i=k+1}^N \sum_{j=1}^k r'_{ij} + \sum_{i=1}^k \sum_{j=k+1}^N r'_{ij} + \sum_{i=k+1}^N \sum_{j=k+1}^N r'_{ij}. \end{aligned}$$

Так как при проведении попершинных перебросов связи между элементами o_i и o_j ($i > k$ и $j > k$) не затрагиваются,

$$\sum_{i=k+1}^N \sum_{j=k+1}^N r'_{ij} = \sum_{i=k+1}^N \sum_{j=k+1}^N r_{ij}.$$

В то же время связи между элементами o_i и o_j ($i \leq k$ и $j \leq k$) изменяются дважды: сначала при попершинных изменениях для элемента o_i , затем для элемента o_j ; следовательно, связи между элементами из выбранной группы останутся неизменными:

$$\sum_{i=1}^k \sum_{j=1}^k r'_{ij} = \sum_{i=1}^k \sum_{j=1}^k r_{ij}.$$

Симметричность связей между элементами позволяет полагать

$$\sum_{i=k+1}^N \sum_{j=1}^k r'_{ij} = \sum_{i=1}^k \sum_{j=k+1}^N r'_{ij}.$$

Учитывая вышесказанное, имеем

$$S'_v - S_v = 2 \sum_{i=1}^k \sum_{j=k+1}^N r'_{ij} - 2 \sum_{i=1}^k \sum_{j=k+1}^N r_{ij}.$$

Пусть $x_i = \sum_{j=k+1}^N r_{ij}$ — число диссонансных связей элемента o_i из группы в k элементов с элементами, не входящими в эту группу. Так как эти связи инвертируются, имеем $x'_i = N - k = x_i$. Отсюда для уменьшения суммы попершинных различий необходимо, чтобы

$$\begin{aligned} S'_v - S_v &= 2 \sum_{i=1}^k x'_i - 2 \sum_{i=1}^k x_i = 2 \sum_{i=1}^k (N - k - x_i) - \\ &- 2 \sum_{i=1}^k x_i = 2k(N - k) - 4 \sum_{i=1}^k x_i < 0, \end{aligned}$$

а это справедливо при условии

$$\sum_{i=1}^k x_i > \frac{k(N - k)}{2}.$$

Если группа состоит из одного элемента ($k = 1$), то приходим к сформулированному ранее условию $v_i > (N - 1)/2$.

Программная реализация алгоритма уменьшения рассогласованности осложняется тем, что его трудоемкость оценивается как 2^N [4]. Для консонансного или диссонансного множества это не станет проблемой: ввиду определенности структуры можно указать эффективный алгоритм приведения к нужному консонансу. Для ассонансного множества важно правильно задать начальный консонансный прообраз исходного множества: число итераций решающим образом зависит от его близости к исходному множеству.

На выбор начального прообраза оказывают существенное влияние многие факторы, в том числе и субъективные. Можно использовать только математические методы задания начального прообраза, но значительное влияние на эффективность алгоритма [4] оказывают и дополнительные факторы. Использование этих факторов в программной реализации алгоритма уменьшения рассогласованности обеспечило приемлемые временные характеристики.

4 Заключение

Моделирование структурной интероперабельности на основе анализа знаков связей с помощью введенного критерия согласованности приводит к нахождению ближайшего к исходному множеству консонансного прообраза. Найденный консонансный прообраз своими подмножествами указывает на предпочтительную группировку элементов, при которой интероперабельность между ними устанавливается с наименьшей рассогласованностью относительно зафиксированных знаков связей. Поскольку рассматриваемые элементы могут быть описаны вектором параметров, из сравнения которых можно сделать вывод о сходстве между элементами, соответственно, нахождение элементов в одном подмножестве говорит о потенциальной мотивации к интероперабельности.

Литература

1. ГОСТ Р 55062-2012. Информационные технологии (ИТ). Системы промышленной автоматизации и их интеграция. Интероперабельность. Основные положения. — М.: Стандартинформ, 2014. 12 с.
2. Systems, Capabilities, Operations, Programs, and Enterprises (SCOPE) Model for Interoperability Assessment. Version 1.0. — NCOIC, 2008. 154 p.
3. Макаренко С. И., Соловьева О. С. Основные положения концепции семантической интероперабельности сетцентрических систем // Ж. радиоэлектроники, 2021. № 4. 24 с.
4. Дулин С. К., Розенберг И. Н. Об одном подходе к структурной согласованности геоанных // Мир транспорта, 2005. Т. 3. № 3. С. 16–29.
5. Дулин С. К. Введение в теорию структурной согласованности. — М.: ВЦ РАН, 2005. 135 с.

Поступила в редакцию 22.02.22

MODELING THE STRUCTURE OF INTEROPERABILITY BY MEANS OF STRUCTURAL CONSISTENCY

I. N. Rozenberg¹, S. K. Dulin^{1,2}, and N. G. Dulina²

¹Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport, 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation

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

Abstract: The initial syntactic level of interoperability involves communication with the appropriate protocol, hardware, software, and necessary level of data compatibility. The work is devoted to the study of the level of compatibility of data describing interacting elements based on the feature vector. To do this, a model of structural correspondence is proposed which allows assessing the tendency to establish interoperability. Modeling structural interoperability based on the analysis of signs of connections using the introduced consistency criterion leads to finding the closest consonant pre-image to the original set. The found consonant pre-image with its subsets indicates the preferred grouping of elements in which the interoperability between them is established with the least mismatch with respect to the fixed signs of connections. Since the elements under consideration are described by a vector of parameters, from the comparison of which one can infer the similarity between the elements, respectively, the presence of elements in the same subset indicates a potential motivation for interoperability.

Keywords: interoperability; structural consistency; connectivity matrix

DOI: 10.14357/19922264230108

References

1. GOST R 55062-2012. 2012. Informatsionnye tekhnologii (IT). Sistemy promyshlennoy avtomatizatsii i ikh integratsiya. Interoperabel'nost'. Osnovnye polozheniya [Information technologies (IT). Systems of industrial automation and their integration. Interoperability. Basic statements]. Moscow: Standartinform Publ. 20 p.
2. NCOIC. 2008. Systems, Capabilities, Operations, Programs, and Enterprises (SCOPE). Model for Interoperability Assessment. Version 1. 154 p. Available at: <https://jfsowa.com/ikl/scope08.pdf> (accessed December 29, 2022).
3. Makarenko, S. I., and O. S. Solovieva. 2021. Osnovnye polozheniya kontseptsii semanticheskoy interoperabel'nosti setetsentricheskikh sistem [Basic provisions of the concept of semantic interoperability of net-centric systems]. *Zh. radioelektroniki* [J. Radio Electronics] 4:1–24.
4. Dulin, S. K., and I. N. Rozenberg. 2005. Ob odnom podkhode k strukturnoy soglasovannosti geodannykh [An approach to the structural consistency of geographical data]. *Mir transporta* [World of Transport and Transportation] 3(3):16–29.
5. Dulin, S. K. 2005. *Vvedenie v teoriyu strukturnoy soglasovannosti* [Introduction to structural consistency theory]. Moscow: CC RAS. 135 p.

Received February 22, 2022

Contributors

Rozenberg Igor N. (b. 1965) — Doctor of Science in technology, professor, Corresponding Member of the Russian Academy of Sciences, research advisor, Research & Design Institute for Information Technology, Signalling and

Telecommunications on Railway Transport, 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation; I.Rozenberg@vniias.ru

Dulin Sergey K. (b. 1950) — Doctor of Science in technology, 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; principal scientist, Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport, 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation; skdulin@mail.ru

Dulina Natalia G. (b. 1947) — Candidate of Science (PhD) in technology, leading programmer, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; ngdulina@mail.ru

МЕТОД НА ОСНОВЕ НЕЧЕТКИХ ПРАВИЛ ДЛЯ УПРАВЛЕНИЯ КОНФЛИКТАМИ АГЕНТОВ В ГИБРИДНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ МНОГОАГЕНТНЫХ СИСТЕМАХ

С. В. Листопад¹, И. А. Кириков²

Аннотация: Работа продолжает исследования по компьютерному моделированию гибридными интеллектуальными многоагентными системами (ГиИМАС) работы коллектива специалистов различных профилей, решающих проблемы за круглым столом. Агенты таких систем — автономные программные сущности, имитирующие рассуждения реальных специалистов. Моделирование в единой интеллектуальной системе разнородных знаний, целей и точек зрения агентов на поставленную проблему обуславливает их столкновение, возникновение конфликтов по аналогии с тем, как это происходит в моделируемых коллективах. Не каждый конфликт между агентами носит деструктивный характер и требует подавления: управление конфликтом в ГиИМАС, как и в коллективе, предполагает идентификацию ситуации принятия решений, при необходимости стимуляцию и последующее разрешение конструктивных форм конфликта, а также предотвращение его деструктивных форм. Для управления конфликтами между агентами в ГиИМАС предлагается метод на основе нечетких правил.

Ключевые слова: конфликт; гибридная интеллектуальная многоагентная система; коллектив специалистов; управление конфликтами

DOI: 10.14357/19922264230109

1 Введение

Как показано в [1], для множества практических проблем, традиционно решаемых коллективом специалистов под руководством лица, принимающего решения, актуальна разработка ГиИМАС, моделирующих работу такого коллектива с целью ускорения выработки решений в условиях высокой динамичности внешней среды и временных ограничений на принятие решения. Характерная особенность как указанных коллективов, так и их компьютерных моделей — возникновение внутри них конфликтов различного характера и необходимость задействовать механизмы управления ими для обеспечения эффективной работы коллектива. В последние несколько десятилетий конфликты в коллективах рассматриваются не только как препятствия, но и как потенциальные источники для повышения эффективности групповой работы [2]. Наличие противоречивых эмпирических данных о роли конфликтов в коллективной работе обуславливает необходимость дальнейших исследований [3–8].

В [9] показано, что конструктивная роль конфликтов проявляется преимущественно тогда, когда цели и потребности конфликтующих специалистов не противоположны целям, ценностям

и нормам коллектива, а социально-психологическая структура коллектива относительно гибкая, т.е. для коллектива характерна относительная открытость подгрупп и динамичные связи между ними и не включенными в них членами [9]. Конструктивные конфликты — преимущественно инструментальные конфликты, возникающие по поводу проблемы или процесса ее решения. Открытое обсуждение и споры по этим вопросам повышают результативность деятельности группы.

Деструктивный конфликт возникает, когда цели и потребности конфликтующих субъектов радикально расходятся с целями, ценностями и нормами коллектива, при этом коллектив обладает жесткой структурой, т.е. для него характерна закрытость подгрупп по отношению к коллективу и явное доминирование дезинтегративных связей между ними [9]. Чаще всего деструктивные конфликты — конфликты по поводу отношений, т.е. разногласия между специалистами по личным вопросам и не относящимся к выполняемой работе проблемам, которые обусловлены несовместимостью и враждебностью между ними.

Управление конфликтом в малых коллективах специалистов состоит в предотвращении деструк-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

тивных конфликтов, но стимулировании и разрешении конструктивных. В [1] для управления конфликтами, возникающими между агентами, по аналогии с тем, как это происходит в коллективах специалистов, предложена модель ГиИМАС с проблемно- и процессно-ориентированными конфликтами. Цель настоящей работы — разработка в рамках данной модели метода управления конфликтами между агентами на основе нечетких правил.

2 Моделирование конфликтов в гибридных интеллектуальных многоагентных системах

Подробное описание модели ГиИМАС с проблемно- и процессно-ориентированными конфликтами представлено в [10]. Рассмотрим основные ее элементы, необходимые для описания процесса управления конфликтами.

Формально ГиИМАС в целом определяется следующим образом [10]:

$$\text{himas} = \langle \text{AG}^*, \text{env}, \text{INT}^*, \text{ORG}, \text{MLP} \rangle, \quad (1)$$

где $\text{AG}^* = \{\text{ag}_1, \dots, \text{ag}_n, \text{ag}^{\text{dm}}, \text{ag}^{\text{fc}}\}$ — множество агентов, включающее n агентов-специалистов (АС) ag_i , $i \in \mathbb{N}$, $1 \leq i \leq n$, агента, принимающего решения (АПР), ag^{dm} и агента-фасилитатора (АФ), управляющего взаимодействиями агентов при решении проблемы и конфликтами между ними, ag^{fc} ; env — концептуальная модель внешней среды системы; INT^* — множество элементов структурирования взаимодействий агентов [10], содержащее среди прочего модель предметной области ont ; ORG — множество архитектур ГиИМАС; $\text{MLP} = \{\text{cnffm}, \text{gdid}\}$ — множество концептуальных моделей макроуровневых процессов в ГиИМАС, содержащее модель cnffm процесса нечеткого управления конфликтами агентов, которая представлена выражением (2), и модель gdid идентификации взаимозависимости целей агентов, описанную выражением (4).

Агент $\text{ag}_{\text{id}} \in \text{AG}^*$ из формулы (1) описывается выражением:

$$\text{ag}_{\text{id}} = \langle \text{id}_{\text{id}}^{\text{ag}}, \text{gl}_{\text{id}}^{\text{ag}}, \text{ACT}_{\text{id}}^{\text{ag}} \rangle,$$

где $\text{id}_{\text{id}}^{\text{ag}}$ — идентификатор агента; $\text{gl}_{\text{id}}^{\text{ag}}$ — нечеткая цель агента, т.е. нечеткое множество с функцией принадлежности $\mu_{\text{id}}(\text{pr}_1, \dots, \text{pr}_{\text{NPRG}})$, заданной на подмножестве целевых концептов-свойств $\text{PR}^g = \{\text{pr}_1, \dots, \text{pr}_{\text{NPRG}}\}$ множества концептов-

свойств $\text{PR}^g \subseteq \text{PR}$ модели предметной области ont ; $\text{ACT}_{\text{id}}^{\text{ag}}$ — множество действий агента.

Модель нечеткого управления конфликтами агентов описывается выражением:

$$\begin{aligned} \text{cnffm} = \\ = \langle \text{CNF}, \text{cnfcl}, \text{ACT}^{\text{afcfm}}, \text{ACT}^{\text{agcs}}, \text{ACT}^{\text{agcr}} \rangle, \quad (2) \end{aligned}$$

где CNF — матрица конфликтов (3) между парами агентов; cnfcl — классификатор конфликтов агентов [10], формирующий матрицу CNF ; $\text{ACT}^{\text{afcfm}} = \{\text{act}^{\text{cnffm}}, \text{act}^{\text{cnfi}}, \text{act}^{\text{cnfs}}, \text{act}^{\text{cnfr}}\}$ — множество функций АФ по управлению конфликтами АС, содержащее функцию «нечеткое управление конфликтом» $\text{act}^{\text{cnffm}}$, обеспечивающую идентификацию act^{cnfi} конфликтов с помощью классификатора cnfcl и инициализацию функции стимуляции act^{cnfs} или разрешения act^{cnfr} конфликтов; ACT^{agcs} — множество действий АС, выполняемых при стимулировании противоречий [11] АФ, $\text{ACT}^{\text{agcs}} \subseteq \cup_{\text{ag}_{\text{id}} \in \text{AG}^*} \text{ACT}_{\text{id}}^{\text{ag}}$; ACT^{agcr} — множество допустимых действий АС по разрешению противоречий [12], $\text{ACT}^{\text{agcr}} \subseteq \cup_{\text{ag}_{\text{id}} \in \text{AG}^*} \text{ACT}_{\text{id}}^{\text{ag}}$.

Конфликт между агентами, т.е. элемент матрицы CNF из выражения (2), описывается следующей формулой:

$$\begin{aligned} \text{cnf}_{ij}^{\text{cnft}} = \\ = \langle \text{ag}_i, \text{ag}_j, \text{cnfin}, \text{cnft}, \text{ACT}_i^{\text{agcr}}, \text{ACT}_j^{\text{agcr}} \rangle, \quad (3) \end{aligned}$$

где ag_i и ag_j — агенты-субъекты конфликта, $i, j \in \mathbb{N}$, $1 \leq i, j \leq n$, $i \neq j$; cnfin — напряженность конфликта в виде скалярной величины $\text{cnfin} \in [0, 1]$, вычисляемой классификатором конфликтов cnfcl в соответствии с реализуемой им мерой напряженности конфликта, в зависимости от того, предлагают АС частные решения проблемы или альтернативные решения проблемы в целом, могут использоваться меры на основе совместности частных решений [13] или на основе ранжирования альтернатив [14, 15]; cnft — символьная переменная «тип конфликта», определенная на множестве $\text{CNFT} = \{\text{cnfprb} = \text{«проблемно-ориентированный»}, \text{cnfprc} = \text{«процессно-ориентированный»}\}$; $\text{ACT}_i^{\text{agcr}}$ и $\text{ACT}_j^{\text{agcr}}$ — множество допустимых действий агентов ag_i и ag_j соответственно по разрешению противоречий, $\text{ACT}_i^{\text{agcr}} \subseteq \text{ACT}_i^{\text{ag}}$, $\text{ACT}_j^{\text{agcr}} \subseteq \text{ACT}_j^{\text{ag}}$, $\text{ACT}_i^{\text{agcr}}, \text{ACT}_j^{\text{agcr}} \subseteq \text{ACT}^{\text{agcr}}$.

Модель идентификации взаимозависимости целей описывается выражением

$$\text{gdid} = \langle \mathbf{GD}, \text{gdi} \rangle. \quad (4)$$

Здесь \mathbf{GD} — матрица показателей взаимозависимости целей пар агентов, $\text{gd}_{ij} \in [-1, 1]$:

$$gd_{ij} = \begin{cases} -1 & \text{— максимальная отрицательная} \\ & \text{взаимозависимость;} \\ 0 & \text{— независимость;} \\ 1 & \text{— максимальная положительная} \\ & \text{взаимозависимость целей} \\ & \text{пары агентов } ag_i \text{ и } ag_j; \end{cases}$$

gd_{ij} — показатель взаимозависимости целей агентов, формирующий матрицу **GD** в случае одномерных целей, т. е. $|PR^g| = 1$ в соответствии с выражением

$$gd_{ij} = gdi(g_i^{ag}, g_j^{ag}) = \frac{\int_{v_{min}}^{v_{max}} \mu_{i \cap j}(pr) d(pr)}{\int_{v_{min}}^{v_{max}} \mu_i(pr) d(pr)} + \frac{\int_{v_{min}}^{v_{max}} \mu_{i \cap j}(pr) d(pr)}{\int_{v_{min}}^{v_{max}} \mu_j(pr) d(pr)} - 1, \quad (5)$$

где v_{min} и v_{max} — минимальное и максимальное значение свойства pr ; $\mu_{i \cap j}(pr)$ — пересечение функций принадлежности нечетких целей агентов ag_i и ag_j .

3 Нечеткое управление конфликтами агентов в гибридных интеллектуальных многоагентных системах

Для нечеткого управления конфликтами АФ использует метод на основе нечеткого вывода Мамдани со следующими лингвистическими переменными: «конфликт», «взаимозависимость целей», «длительность», «стадия». При формировании нечетких множеств используются колоколообразные

$$bmf(u, a, b, c) = (1 + |u - c|^{2b} |a|^{-2b})^{-1}$$

и сигмоидальные функции

$$sigmf(u, b, c) = (1 + e^{-b(u-c)})^{-1},$$

где u — элемент нечеткого множества; a, b и c — параметры функции принадлежности.

Кроме того, используется детерминированная переменная «номер итерации» $it \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$, описывающая число повторений действий по анализу текущей ситуации, выполняемых АФ при управлении конфликтами.

Лингвистическая переменная «конфликт» для оценки напряженности конфликта между агентами ГиИМАС задается выражением

$$cnfl = \langle \beta_{cnfl}, T_{cnfl}, U_{cnfl}, G_{cnfl}, M_{cnfl} \rangle, \quad (6)$$

где β_{cnfl} = «конфликт» — наименование лингвистической переменной; $T_{cnfl} = \{\text{«нет»}; \text{«слабый»}; \text{«умеренный»}; \text{«острый»}\}$ — терм-множество ее значений, названий нечеткой переменной; $U_{cnfl} = [0; 1]$ — универсум нечетких переменных; $G_{cnfl} = \emptyset$ — процедура образования из элементов множества T_{cnfl} новых термов; $M_{cnfl} = \{\mu_{нет}, \mu_{слабый}, \mu_{умеренный}, \mu_{острый}\}$ — процедура, ставящая в соответствие каждому терму множества T_{cnfl} осмысленное содержание путем формирования нечеткого множества: $\mu_{нет} = bmf(u_{cnfl}; 0,15; 5; 0)$, $\mu_{слабый} = bmf(u_{cnfl}; 0,15; 5; 0,3)$, $\mu_{умеренный} = bmf(u_{cnfl}; 0,15; 5; 0,6)$, $\mu_{острый} = bmf(u_{cnfl}; 0,25; 5; 1)$. Параметры этих и других функций принадлежности, рассмотренных в работе, должны быть уточнены в ходе тестирования системы.

Лингвистическая переменная «взаимозависимость целей» описывается по аналогии с формулой (6) следующим выражением:

$$gdl = \langle \beta_{gdl}, T_{gdl}, U_{gdl}, G_{gdl}, M_{gdl} \rangle,$$

где β_{gdl} = «взаимозависимость целей»; $T_{gdl} = \{\text{«отрицательная»}; \text{«нет»}; \text{«положительная»}\}$; $U_{gdl} = [-1; 1]$; $G_{gdl} = \emptyset$; $M_{gdl} = \{\mu_{отрицательная}, \mu_{нет}, \mu_{положительная}\}$, $\mu_{отрицательная} = sigmf(u_{dcnfl}; -10; -0,3)$, $\mu_{нет} = bmf(u_{dcnfl}; 0,3; 3; 0)$, $\mu_{положительная} = sigmf(u_{dcnfl}; 10; 0,3)$.

Лингвистическая переменная «длительность», используемая для оценки продолжительности стадии управления конфликтом, представляется выражением

$$durl = \langle \beta_{durl}, T_{durl}, U_{durl}, G_{durl}, M_{durl} \rangle,$$

где β_{durl} = «длительность»; $T_{durl} = \{\text{«малая»}; \text{«большая»}\}$; $U_{durl} = \mathbb{N}_0$; $G_{durl} = \emptyset$; $M_{durl} = \{\mu_{малая}, \mu_{большая}\}$, $\mu_{малая} = sigmf(u_{durl}; -0,2; 10 \cdot n)$, $\mu_{большая} = sigmf(u_{durl}; 0,5; 10 \cdot n)$.

Лингвистическая переменная «стадия» описывается выражением

$$stgl = \langle \beta_{stgl}, T_{stgl}, U_{stgl}, G_{stgl}, M_{stgl} \rangle,$$

где β_{stgl} = «стадия»; $T_{stgl} = \{\text{«начало»}; \text{«стимуляция»}; \text{«разрешение»}; \text{«конец»}\}$; $U_{stgl} = [0; 1]$; $G_{stgl} = \emptyset$; $M_{stgl} = \{\mu_{начало}, \mu_{стимуляция}, \mu_{разрешение}, \mu_{конец}\}$, $\mu_{начало} = bmf(u_{stgl}; 6^{-1}; 5; 0)$, $\mu_{стимуляция} = bmf(u_{stgl}; 6^{-1}; 5; 3^{-1})$, $\mu_{разрешение} = bmf(u_{stgl}; 6^{-1}; 5; 2 \cdot 3^{-1})$, $\mu_{конец} = bmf(u_{stgl}; 6^{-1}; 5; 1)$.

Таким образом, функция «нечеткое управление конфликтом» act^{cnfl} АФ может быть представлена следующей последовательностью шагов:

- (1) установить начальные значения переменных $it = 0$, $u_{stgl}^{it} = 0$ и $u_{durl} = 0$;
- (2) используя выражение (5), вычислить среднее арифметическое показателей взаимозависимости целей агентов по формуле

$$gd^{himas} = \sum_{i=1}^n \sum_{j=i+1}^n 2gd_{ij}(n-2)!(n!)^{-1};$$

- (3) ожидать поступления сообщений от АС или АПР;
- (4) если получено сообщение от АПР о завершении работы, закончить работу;
- (5) если получено сообщение от АС, содержащее решение проблемы или ее части, перейти к п. 6, иначе сообщить отправителю об ошибке и перейти к п. 12;
- (6) увеличить значение счетчиков $it = it + 1$, $u_{durl} = u_{durl} + 1$;
- (7) запустить функцию идентификации конфликтов act^{cnfi} , чтобы с помощью классификатора конфликтов $cnfcl$ на основе решений, предложенных АС, сформировать матрицу конфликтов CNF между парами агентов [10] и вычислить для ГиИМАС общие показатели напряженности конфликтов по каждому типу:

$$cnf_{cnfprb}^{himas} = \sum_{i=1}^n \sum_{j=i+1}^n 2cnf_{ij}^{cnfprb}(n-2)!(n!)^{-1};$$

$$cnf_{cnfprc}^{himas} = \sum_{i=1}^n \sum_{j=i+1}^n 2cnf_{ij}^{cnfprc}(n-2)!(n!)^{-1};$$

- (8) фазсифицировать четкие значения лингвистических переменных $u_{cnfl}^{cnfprb} = cnf_{cnfprb}^{himas}$, $u_{cnfl}^{cnfprc} = cnf_{cnfprc}^{himas}$ и $u_{gdl} = gd^{himas}$;
- (9) установить стадию управления конфликтами $stgl^{it}$ в результате нечеткого вывода Мамдани [16] по следующим правилам:

правило 1: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция») И (cnf_{cnfprc} ЕСТЬ «нет» ИЛИ cnf_{cnfprc} ЕСТЬ «слабый»), ТО $stgl^{it}$ ЕСТЬ «стимуляция»;

правило 2: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция») И (cnf_{cnfprb} ЕСТЬ «нет» ИЛИ cnf_{cnfprb} ЕСТЬ «слабый»), ТО $stgl^{it}$ ЕСТЬ «стимуляция»;

правило 3: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция»)

И (cnf_{cnfprc} ЕСТЬ «умеренный» ИЛИ cnf_{cnfprc} ЕСТЬ «острый»), ТО $stgl^{it}$ ЕСТЬ «разрешение»;

правило 4: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция») И cnf_{cnfprb} ЕСТЬ «острый» ТО $stgl^{it}$ ЕСТЬ «разрешение»;

правило 5: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция») И gdl ЕСТЬ «положительная» И cnf_{cnfprb} ЕСТЬ «тумеренный», ТО $stgl^{it}$ ЕСТЬ «стимуляция»;

правило 6: ЕСЛИ ($stgl^{it-1}$ ЕСТЬ «начало» ИЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция») И gdl ЕСТЬ «отрицательная» И cnf_{cnfprb} ЕСТЬ «умеренный», ТО $stgl^{it}$ ЕСТЬ «разрешение»;

правило 7: ЕСЛИ $stgl^{it-1}$ ЕСТЬ «стимуляция» И $durl$ ЕСТЬ «большая», ТО $stgl^{it}$ ЕСТЬ «разрешение»;

правило 8: ЕСЛИ $stgl^{it-1}$ ЕСТЬ «разрешение» И $durl$ ЕСТЬ «большая», ТО $stgl^{it}$ ЕСТЬ «конец»;

- (10) если $stgl^{it-1} \neq stgl^{it}$, установить $u_{durl} = 0$;
- (11) если $stgl^{it-1} =$ «стимуляция», выполнить функцию «стимуляция конфликтов» act^{cnfs} , если же $stgl^{it} =$ «разрешение», выполнить функцию «разрешение конфликтов» act^{cnfr} ;
- (12) если $stgl^{it} =$ «конец» или после выполнения функций «стимуляция конфликтов» act^{cnfs} или «разрешение конфликтов» act^{cnfr} установлен флаг завершения работы, то отправить АПР сообщение о необходимости завершения работы ГиИМАС и закончить работу, иначе перейти к п. 3.

Как показывает анализ функции «нечеткое управление конфликтом» act^{cnfmm} АФ, порядок смены стадий не predetermined, а зависит от ситуации решения проблемы в ГиИМАС, в частности от длительности использования метода, интенсивности конфликтов и соотношения целей АС. С использованием данной функции, а также действий по стимулированию АСТ^{agcs} и разрешению АСТ^{agcr} противоречий АФ может повышать напряженность конфликтов в ГиИМАС и разнообразие рассматриваемых в ходе решения проблемы альтернатив либо переходить к этапу согласования позиций АС, если между ними возникает большое число конфликтов высокой интенсивности. Таким образом, благодаря динамической комбинации различных технологий искусственного интеллекта, реализуемых агентами, ГиИМАС при решении очередной

проблемы вырабатывает наиболее релевантный ей метод. Применение нечеткого вывода Мамдани при управлении конфликтами в ГиИМАС обеспечивает прозрачность базы знаний для интерпретации человеком, в отличие от систем нечеткого вывода Такаги–Сугено [17] или нейронечетких гибридов: ANFIS [18], GARIC [19], SONFIN [20], за счет того, что в консеквенте правила присутствуют лингвистические переменные. Кроме того, сохраняется возможность более тонкой настойки базы знаний в ходе тестирования системы с применением методов подстройки параметров функций принадлежности по алгоритму обратного распространения ошибки, рассмотренных в [21].

4 Заключение

Рассмотрены особенности возникновения и управления конфликтами в коллективах реальных специалистов и их компьютерных моделей, решающих практические проблемы за круглым столом. Представлена модель ГиИМАС с управлением проблемно- и процессно-ориентированными конфликтами на основе анализа их напряженности, типа и соотношения целей агентов. Предложен метод управления конфликтами в ГиИМАС с использованием нечеткого вывода Мамдани, что обеспечивает прозрачность базы знаний, а также возможность более тонкой ее настойки в ходе тестирования системы. Управление конфликтами в ГиИМАС заключается в релевантной ситуации стимулировании или разрешении проблемно- и процессно-ориентированных конфликтов. Стимулирование конструктивных конфликтов в ГиИМАС позволяет расширить множество рассматриваемых альтернатив, учесть различные точки зрения и интересы специалистов, моделируемых агентами системы. Последующее снижение интенсивности и разрешение конфликтов, возникших между агентами ГиИМАС на первых стадиях рассмотрения поставленной проблемы, позволяет выработать согласованную позицию с учетом различных точек зрения и интересов специалистов, моделируемых агентами системы, повышая таким образом релевантность системы реальному коллективу специалистов.

Литература

1. Листопад С. В., Кириков И. А. Моделирование конфликтов агентов в гибридных интеллектуальных многоагентных системах // Системы и средства информатики, 2019. Т. 29. № 3. С. 139–148. doi: 10.14357/08696527190312.
2. Maltarich M. A., Kukenberger M., Reilly G., Mathieu J. Conflict in teams: Modeling early and late conflict states and the interactive effects of conflict processes // Group Organ. Manage., 2018. Vol. 43. Iss. 1. P. 6–37.
3. Jehn K. A., Northcraft G. B., Neale M. A. Why differences make a difference: A field study of diversity, conflict, and performance in workgroups // Admin. Sci. Quart., 1999. Vol. 44. P. 741–763.
4. Rahim M. A. Toward a theory of managing organizational conflict // Int. J. Confl. Manage., 2002. Vol. 13. Iss. 3. P. 206–235.
5. Емельянов С. М. Практикум по конфликтологии. — СПб.: Питер, 2009. 384 с.
6. Greer L. L., Caruso H. M., Jehn K. A. The bigger they are, the harder they fall: Linking team power, team conflict, and performance // Organ. Behav. Hum. Dec., 2011. Vol. 116. P. 116–128.
7. De Wit F. R. C., Greer L. L., Jehn K. A. The paradox of intragroup conflict: A meta-analysis // J. Appl. Psychol., 2012. Vol. 97. P. 360–390.
8. Хохлов А. С. Конфликтология: История. Теория. Практика. — Самара: СФ МГПУ, 2014. 312 с.
9. Сидоренков А. В. Конфликт в малой группе: понятие, функции, виды и модель // Северо-Кавказский психологический вестник, 2008. Т. 6. № 4. С. 22–28.
10. Листопад С. В., Кириков И. А. Метод идентификации конфликтов агентов в гибридных интеллектуальных многоагентных системах // Системы и средства информатики, 2020. Т. 30. № 1. С. 56–65. doi: 10.14357/08696527200105.
11. Листопад С. В., Кириков И. А. Стимуляция конфликтов агентов в гибридных интеллектуальных многоагентных системах // Системы и средства информатики, 2021. Т. 31. № 2. С. 47–58. doi: 10.14357/08696527210205.
12. Листопад С. В., Кириков И. А. Разрешение конфликтов в гибридных интеллектуальных многоагентных системах // Информатика и её применения, 2022. Т. 16. Вып. 1. С. 54–60. doi: 10.14357/19922264220108.
13. 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.
14. Bana e Costa C. The use of multi-criteria decision analysis to support the search for less conflicting policy options in a multi-actor context: Case study // J. Multi-Criteria Decision Analysis, 2001. Vol. 10. Iss. 2. P. 111–125.
15. Fasth T., Larsson A., Ekenberg L., Danielson M. Measuring conflicts using cardinal ranking: An application to decision analytic conflict evaluations // Advances Operations Research, 2018. Vol. 2018. Art. 8290434. 14 p.
16. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. — СПб.: БХВ Петербург, 2005. 736 с.
17. Abonyi J., Nagy L., Szeifert F. Adaptive fuzzy inference system and its application in modelling and model

- based control // Chem. Eng. Res. Des., 1999. Vol. 77. Iss. 4. P. 281–290.
18. Jang J.-S. R. ANFIS: Adaptive-network-based fuzzy inference systems // IEEE Transactions on Systems, Man, and Cybernetics, 1993. Vol. 23. P. 665–685.
 19. Berenji H. R., Khedkar P. Learning and tuning fuzzy logic controllers through reinforcements // IEEE Trans. Neural Networks, 1992. Vol. 3. P. 724–740.
 20. Feng J. C., Teng L. C. An online self constructing neural fuzzy inference network and its applications // IEEE T. Fuzzy Syst., 1998. Vol. 6. Iss. 1. P. 12–32.
 21. Колесников А. В., Кириков И. А., Листопад С. В. Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор. — М.: ИПИ РАН, 2014. 189 с.

Поступила в редакцию 10.01.23

FUZZY RULES BASED METHOD FOR AGENT CONFLICT MANAGEMENT IN HYBRID INTELLIGENT MULTIAGENT SYSTEMS

S. V. Listopad and I. A. Kirikov

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

Abstract: The paper continues research on computer simulation with hybrid intelligent multiagent systems of a teamwork of specialists of various profiles who solve problems at a round table. The agents of such systems are autonomous software entities that imitate the reasoning of real specialists. Modeling of heterogeneous knowledge, goals, and points of view of agents on the problem posed within single intelligent system causes their collision, the emergence of conflicts by analogy with how it happens in simulated teams. Not every conflict between agents is destructive and requires suppression: conflict management in a hybrid intelligent multiagent system as well as in a team involves the identification of a decision-making situation, if necessary, stimulation and subsequent resolution of constructive forms of conflict as well as the prevention of its destructive forms. The paper proposes the method based on fuzzy rules to manage conflicts between agents in hybrid intelligent multiagent systems.

Keywords: conflict; hybrid intelligent multiagent system; team of specialists; conflict management

DOI: 10.14357/19922264230109

References

1. Listopad, S. V., and I. A. Kirikov. 2019. Modelirovanie konfliktov agentov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Modeling of agent conflicts in hybrid intelligent multiagent systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(3):139–148. doi: 10.14357/08696527190312.
2. Maltarich, M. A., M. Kukenberger, G. Reilly, and J. Mathieu. 2018. Conflict in teams: Modeling early and late conflict states and the interactive effects of conflict processes. *Group Organ. Manage.* 43(1):6–37.
3. Jehn, K. A., G. B. Northcraft, and M. A. Neale. 1999. Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Admin. Sci. Quart.* 44:741–763.
4. Rahim, M. A. 2002. Toward a theory of managing organizational conflict. *Int. J. Confl. Manage.* 13(3):206–235.
5. Emel'yanov, S. M. 2009. *Praktikum po konfliktologii* [Tutorial at conflictology]. St. Petersburg: Piter. 384 p.
6. Greer, L. L., H. M. Caruso, and K. A. Jehn. 2011. The bigger they are, the harder they fall: Linking team power, team conflict, and performance. *Organ. Behav. Hum. Dec.* 116:116–128.
7. De Wit, F. R. C., L. L. Greer, and K. A. Jehn. 2012. The paradox of intragroup conflict: A meta-analysis. *J. Appl. Psychol.* 97:360–390.
8. Khokhlov, A. S. 2014. *Konfliktologiya: Istoriya. Teoriya. Praktika* [Conflictology: History. Theory. Practice]. Samara: SF MGPU. 312 p.
9. Sidorenkov, A. V. 2008. Konflikt v maloy gruppe: ponyatie, funktsii, vidy i model' [Conflict in small group: Concept, functions, forms and model]. *Severo-Kavkazskiy psikhologicheskii vestnik* [North Caucasian Psychological J.] 6(4):22–28.
10. Listopad, S. V., and I. A. Kirikov. 2020. Metod identifikatsii konfliktov agentov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Agent conflict identification method in hybrid intelligent multiagent systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1):56–65. doi: 10.14357/08696527200105.
11. Listopad, S. V., and I. A. Kirikov. 2021. Stimulyatsiya konfliktov agentov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Stimulation of agent conflicts in hybrid intelligent multiagent systems]. *Sistemy i Sredstva*

- Informatiki — Systems and Means of Informatics* 31(2):47–58. doi: 10.14357/08696527210205.
12. Listopad, S. V., and I. A. Kirikov. 2022. Razreshenie konfliktov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Conflict resolution in hybrid intelligent multi-agent systems]. *Informatika i ee Primeneniya — Inform. Appl.* 16(1):54–60. doi: 10.14357/19922264220108.
 13. 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.
 14. Bana e Costa, C. 2001. The use of multi-criteria decision analysis to support the search for less conflicting policy options in a multi-actor context: Case study. *J. Multi-Criteria Decision Analysis* 10(2):111–125.
 15. Fath, T., A. Larsson, L. Ekenberg, and M. Danielson. 2018. Measuring conflicts using cardinal ranking: an application to decision analytic conflict evaluations. *Advances Operations Research* 2018:8290434. 14 p.
 16. Leonenkov, A. V. 2005. *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* [Fuzzy modeling in MATLAB and fuzzyTECH]. St. Petersburg: BHV-Peterburg. 736 p.
 17. Abonyi, J., L. Nagy, and F. Szeifert. 1999. Adaptive fuzzy inference system and its application in modelling and model based control. *Chem. Eng. Res. Des.* 77(4):281–290.
 18. Jang, J.-S. R. 1993. ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Transactions Systems, Man, and Cybernetics* 23:665–685.
 19. Berenji, H. R., and P. Khedkar. 1992. Learning and tuning fuzzy logic controllers through reinforcements. *IEEE Trans. Neural Networks* 3:724–740.
 20. Feng, J. C., and L. C. Teng. 1998. An online self constructing neural fuzzy inference network and its applications. *IEEE T. Fuzzy Syst.* 6(1):12–32.
 21. Kolesnikov, A. V., I. A. Kirikov, and S. V. Listopad. 2014. *Gibridnye intellektual'nye sistemy s samoorganizatsiyey: koordinatsiya, soglasovannost', spor* [Hybrid intelligent systems with self-organization: Coordination, consistency, and dispute]. Moscow: IPI RAN. 189 p.

Received January 10, 2023

Contributors

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 236000, Russian Federation; ser-list-post@yandex.ru

Kirikov Igor A. (b. 1955) — Candidate of Sciences (PhD) in technology; director, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

О ЗАДАЧЕ ОЦЕНКИ И АНАЛИЗА РИСКА ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ НА РЕЛЬСОВОМ ТРАНСПОРТЕ*

А. В. Босов¹, А. Н. Игнатов²

Аннотация: Рассматривается задача по оцениванию и анализу риска транспортных происшествий на рельсовом транспорте. Предлагаются две функции интегрального риска, позволяющие оценить опасность движения на всем маршруте следования транспортного средства. В качестве таких функций выбираются вероятность возникновения неблагоприятного события при транспортировке, а также средний ущерб. Предлагается концепция оценивания вероятности и ущерба от неблагоприятных событий при движении грузовых поездов. На основе ранее обработанной статистики по движению грузовых поездов и происходивших с ними неблагоприятных событий приводится содержательный пример расчета функций интегрального риска.

Ключевые слова: риск; неблагоприятное событие; рельсовый транспорт; вероятность; средний ущерб

DOI: 10.14357/19922264230110

1 Введение

Типовой интерпретацией риска служит отклонение в большую или меньшую сторону от целевого значения некоторого показателя вследствие неопределенностей [1]. В данной терминологии у понятия присутствует не только негативная коннотация, а риск ничем не отличается от убытка/прибыли от ведения некоторой деятельности. В то же время в межгосударственном стандарте [2] постулируется, что риск — это сочетание вероятности нанесения ущерба и тяжести этого ущерба. Именно этим понятием риска и будем пользоваться, не учитывая доход.

В одной из первых работ, посвященных прогнозированию и оценке последствий железнодорожных происшествий, рассматривалась задача оценки влияния номера первой сошедшей с рельсов подвижной единицы (вагонов и секций локомотива) на число сошедших с рельсов подвижных единиц [3]. Предлагалось проводить оптимизацию расположения вагонов с опасными грузами с целью уменьшения вероятности схода с рельсов именно этих вагонов. В дальнейшем [4, 5] была предложена оценка вероятности схода с рельсов, зависевшая от длины пути и числа подвижных единиц в составе поезда. В [6] для прогнозирования числа подвижных единиц в сходе с рельсов было использовано отрицательное биномиальное распределение. В [7] была рассмотрена похожая на [6] модель, но с большим числом факторов. В [8] была рас-

смотрена задача прогнозирования дополнительного негативного последствия от схода с рельсов помимо собственно повреждения подвижного состава — выхода в габарит соседнего пути хотя бы одной сошедшей с рельсов подвижной единицы. Такое событие влечет остановку движения на соседнем пути и, кроме того, может привести к столкновению со встречным поездом.

Наиболее близки к теме данной статьи работы [9, 10]. В них строится функция интегрального (общего на всем пути следования) риска с целью оптимального расположения вагонов с опасными грузами с точки зрения минимизации суммарного риска. Данная функция зависит от скорости (при итоговом расчете используется не реальная скорость поезда в пути, а максимально допустимая на участке), а также зависит от номера первой сошедшей с рельсов подвижной единицы и причины схода с рельсов. Следует отметить, что сход с рельсов по конкретной причине — случайное событие, однако учет этого обстоятельства (например, при помощи формулы полной вероятности) в [9, 10] опускается. Другие факторы, например профиль и кривизна пути, не используются, а это, в свою очередь, снижает точность полученного результата. Кроме того, в [9, 10] поезд представляется материальной точкой, для которой не учитывается, что голова поезда может находиться, например, на стрелочном переводе, а хвост еще нет. Также не рассматриваются другие помимо схода с рельсов неблагоприятные события, которые могут произойти

* Работа выполнена при поддержке РФФИ (проект 20-07-00046 А).

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Московский авиационный институт, avbosov@ipigan.ru

²Московский авиационный институт, alexei.ignatov1@gmail.com

при транспортировке, а в качестве функции риска не постулируется вероятность отсутствия всех неблагоприятных событий при движении. Устранению этих, а также некоторых других недостатков посвящена настоящая статья.

Рассматривается задача построения функций интегрального риска, характеризующих опасность возникновения различных неблагоприятных событий при движении на рельсовом транспорте. Предлагаются две функции интегрального риска: вероятность отсутствия всех неблагоприятных событий при движении и средний ущерб при транспортировке. Представлена концепция оценки вероятности и ущерба неблагоприятных событий при движении грузовых поездов. На содержательном примере показывается важность выбора скоростного режима применительно к снижению интегрального риска в виде среднего ущерба.

2 Постановка задачи

Пусть по рельсам по заданному маршруту длиной S метров движется некоторое транспортное средство (трамвай, грузовой поезд или вагонетка). При движении происходит управление скоростью: пусть v_s — скорость на s -м метре дистанции, $s = \overline{1, S}$. Для упрощения модель допускает мгновенное изменение скорости на участке. При этом скорость считается заданной детерминированной величиной. Пусть на каждом метре дистанции при движении могут произойти n неблагоприятных событий (транспортных происшествий). Обозначим через A_{si} событие, заключающееся в том, что на s -м метре маршрута произойдет i -е событие, $s = \overline{1, S}$, $i = \overline{1, n}$. Например, такими событиями могут выступать сход с рельсов по причине неисправности подвижного состава, сход с рельсов по причине неисправности железнодорожного полотна. Будем предполагать, что на одном метре дистанции может произойти максимум одно неблагоприятное событие, причем если это событие наступает, то далее движение не осуществляется. Введем обозначения: $P_{1i}(v_1) = \mathcal{P}(A_{1,i})$, $P_{si}(v_s) = \mathcal{P}(A_{s,i} | \prod_{k=1}^{s-1} \prod_{j=1}^n \overline{A_{k,j}})$, $s = \overline{2, S}$, т.е. для $s \geq 2$ $P_{si}(v_s)$ — это вероятность того, что на s -м метре маршрута произойдет i -е неблагоприятное событие при условии, что на предыдущих $s - 1$ метрах неблагоприятных событий не произошло. Если $P_{si}(v_s)$ не точные значения вероятностей, а оценки, то равенства выше понимаются в приближенном смысле. Ущерб $C_{si}(v_s)$ от возникновения i -го неблагоприятного события на s -м метре маршрута будем полагать случайной величиной с известным распределением, параметры которого определяются только скоростью. Дополнительно предположим, что ма-

тематическое ожидание $\mathbf{M}[C_{si}(v_s)]$ конечно для любого значения скорости v_s . Имеет место следующая лемма.

Лемма. Случайная величина $\Phi(v_1, \dots, v_S)$ — ущерб при осуществлении перевозок — имеет вид:

$$\Phi(v_1, \dots, v_S) = \sum_{s=1}^S \sum_{i=1}^n C_{si}(v_s) I(A_{s,i}),$$

где $I(A_{s,i})$ — индикатор события $A_{s,i}$, вероятность которого

$$\mathcal{P}(A_{s,i}) = P_{si}(v_s) \prod_{t=1}^{s-1} \left(1 - \sum_{j=1}^n P_{tj}(v_t) \right),$$

$$s = \overline{2, S}, i = \overline{1, n}.$$

Доказательство. Выберем некоторое $i \in \{1, \dots, n\}$. По условию $P_{1i}(v_1) = \mathcal{P}(A_{1,i})$. Далее найдем вероятность того, что i -е неблагоприятное событие произойдет на 2-м метре маршрута. Для того чтобы это событие произошло, на 1-м метре неблагоприятных событий произойти не должно. Поэтому

$$\mathcal{P}(A_{2,i}) = \mathcal{P}(A_{2,i} | \overline{A_{1,1}} \cdots \overline{A_{1,n}}) \mathcal{P}(\overline{A_{1,1}} \cdots \overline{A_{1,n}}).$$

Используя равенство

$$\mathcal{P}(\overline{A_{1,1}} \cdots \overline{A_{1,n}}) = 1 - \mathcal{P}(A_{1,1} + \cdots + A_{1,n}),$$

получаем

$$\begin{aligned} \mathcal{P}(A_{2,i}) &= \\ &= \mathcal{P}(A_{2,i} | \overline{A_{1,1}} \cdots \overline{A_{1,n}}) \left[1 - \mathcal{P}\left(\sum_{j=1}^n A_{1,j}\right) \right] = \\ &= P_{2i}(v_2) \left[1 - \mathcal{P}\left(\sum_{j=1}^n A_{1,j}\right) \right]. \end{aligned}$$

Так как события $A_{1,1}, \dots, A_{1,n}$ несовместны по условию, получаем

$$\mathcal{P}\left(\sum_{j=1}^n A_{1,j}\right) = \sum_{j=1}^n \mathcal{P}(A_{1,j}).$$

Отсюда

$$\begin{aligned} \mathcal{P}(A_{2,i}) &= P_{2i}(v_2) \left(1 - \sum_{j=1}^n \mathcal{P}(A_{1,j}) \right) = \\ &= P_{2i}(v_2) \left[1 - \sum_{j=1}^n P_{1j}(v_1) \right]. \end{aligned}$$

Нетрудно видеть, что аналогичные рассуждения повторяются на последующих метрах маршрута и дают по индукции

$$\mathcal{P}(A_{s,i}) = P_{si}(v_s) \prod_{t=1}^{s-1} \left(1 - \sum_{j=1}^n P_{tj}(v_t) \right),$$

$s = \overline{2, S}, i = \overline{1, n}.$

По условию ущерб при возникновении события $A_{s,i}$ составляет $C_{si}(v_s)$, поэтому

$$\Phi(v_1, \dots, v_S) = \sum_{s=1}^S \sum_{i=1}^n C_{si}(v_s) I(A_{s,i}).$$

Лемма доказана.

Найдем теперь вероятность того, что с транспортным средством при движении не произойдет ни одного неблагоприятного события, т. е.

$$\begin{aligned} \mathcal{P} \left(\prod_{s=1}^S \prod_{j=1}^n \bar{A}_{s,j} \right) &= 1 - \mathcal{P} \left(\sum_{s=1}^S \sum_{j=1}^n A_{s,j} \right) = \\ &= 1 - \sum_{s=1}^S \sum_{j=1}^n \mathcal{P}(A_{s,j}) = \\ &= 1 - \sum_{j=1}^n P_{1j}(v_1) - \sum_{s=2}^S \sum_{j=1}^n \mathcal{P}(A_{s,j}) = \\ &= 1 - \sum_{j=1}^n P_{1j}(v_1) - \sum_{j=1}^n P_{2j}(v_2) \left[1 - \sum_{j=1}^n P_{1j}(v_1) \right] - \\ &- \sum_{s=3}^S \sum_{j=1}^n \mathcal{P}(A_{s,j}) = \dots = \prod_{s=1}^S \left[1 - \sum_{j=1}^n P_{sj}(v_s) \right]. \end{aligned}$$

Теперь в качестве функции интегрального риска можно предложить следующие два варианта:

$$\begin{aligned} R_1(v_1, \dots, v_S) &= 1 - \mathcal{P} \left(\prod_{s=1}^S \prod_{j=1}^n \bar{A}_{s,j} \right) = \\ &= 1 - \prod_{s=1}^S \left[1 - \sum_{j=1}^n P_{sj}(v_s) \right]; \end{aligned}$$

$$\begin{aligned} R_2(v_1, \dots, v_S) &= \mathbf{M}[\Phi(v_1, \dots, v_S)] = \\ &= \sum_{s=1}^S \sum_{i=1}^n \mathbf{M}[C_{si}(v_s)] P_{si}(v_s) \prod_{t=1}^{s-1} \left[1 - \sum_{j=1}^n P_{tj}(v_j) \right]. \end{aligned}$$

Функция $R_1(v_1, \dots, v_S)$ при выборе режима движения, т. е. набора скоростей v_1, \dots, v_S , характеризует вероятность того, что на маршруте произойдет неблагоприятное событие. Функция $R_2(v_1, \dots, v_S)$ позволяет вычислить средний ущерб при транспортировке. В конкретных задачах целесообразным может оказаться использование как первой,

так и второй функции, а уместнее всего использовать их комбинацию. Это вызвано тем, что при выбранном наборе скоростей v_1, \dots, v_S вероятность возникновения неблагоприятного события $R_1(v_1, \dots, v_S)$ может быть небольшой, но средний ущерб $R_2(v_1, \dots, v_S)$ при выборе такого режима движения может оказаться неоправданно высоким. Напротив, может оказаться, что прогнозируемый средний ущерб при заданном наборе скоростей v_1, \dots, v_S невелик, однако неблагоприятные события возникают часто.

3 Концепция оценивания вероятности неблагоприятных событий и ущерба при движении грузовых поездов

Зафиксируем некоторый промежуток времени \mathcal{T} в прошлом. Пусть L_1 — число вагонов; L_0 — число секций локомотива; d_l — длина l -й подвижной единицы от головы поезда, округленная в большую сторону до метров. Для удобства введем величину

$$S_0 = d_1 + d_2 + \dots + d_L,$$

где L — общее число подвижных единиц грузового поезда, т. е. $L = L_0 + L_1$. Введем также обозначения:

p_1 — число сходов с рельсов / крушений за промежуток времени \mathcal{T} вне стрелочных переводов по причине неисправности подвижного состава;

p_2 — число сходов с рельсов / крушений за промежуток времени \mathcal{T} вне стрелочных переводов по причине неисправности железнодорожного полотна;

p_3 — число сходов с рельсов / крушений за промежуток времени \mathcal{T} на стрелочных переводах;

$p_{\text{общ}}$ — общее число сходов с рельсов / крушений за промежуток времени \mathcal{T} ;

Q'_1 — число вагоно-километров за промежуток времени \mathcal{T} ;

Q'_2 — число поездо-километров за промежуток времени \mathcal{T} ;

w — вес поезда, т;

$\tilde{\mu} = w / (69L_1) - 1/3$ — величина, характеризующая степень загрузки поезда [7];

\varkappa_s — кривизна кривой на s -м метре пути (величина, обратно пропорциональная радиусу кривизны, для прямой полагается равной нулю);

γ_s — профиль пути на s -м метре пути, измеряемый в тысячных, имеющий знак минус, ес-

ли уклон представляет спуск, знак плюс, если уклон представляет подъем, ед.;

λ_s — величина, характеризующая наличие стрелочного перевода на s -м метре пути (0 — нет стрелочного перевода; 1 — есть);

χ_s — величина, характеризующая наличие соседнего/соседних пути/путей на s -м метре пути (0 — нет соседнего пути; 1 — есть);

η_s — величина, характеризующая начало железнодорожного переезда на s -м метре пути (0 — железнодорожный переезд не начинается на s -м метре пути; 1 — иначе); ξ_s — величина, характеризующая начало стрелочного перевода на s -м метре пути в пределах железнодорожной станции (0 — стрелочный перевод на s -м метре пути в пределах станции; 1 — иначе), $s = \overline{2 - S_0, S}$.

Будем рассматривать следующие неблагоприятные события:

$$A_{s,i} = \left\{ \begin{array}{l} \text{сход } \lfloor (i-1)/L \rfloor + 1 \text{ подвижных единиц} \\ \text{по причине неисправности подвижного} \\ \text{состава вне стрелочного перевода} \\ \text{по прошествии } s \text{ метров пути, начиная} \\ \text{с } i - \lfloor (i-1)/L \rfloor L \text{ подвижной единицы,} \\ \quad i = \overline{1, L^2}; \\ \text{сход } \lfloor (i-1-L^2)/L \rfloor + 1 \text{ подвижных} \\ \text{единиц по причине неисправности} \\ \text{железнодорожного пути вне стрелоч-} \\ \text{ного перевода по прошествии } s \text{ метров} \\ \text{пути, начиная с } i - L^2 - \lfloor (i-1-L^2)/L \rfloor L \\ \text{подвижной единицы, } i = \overline{L^2 + 1, 2L^2}; \\ \text{сход } \lfloor (i-1-2L^2)/L \rfloor + 1 \text{ подвижных} \\ \text{единиц на стрелочном переводе по} \\ \text{прошествии } s \text{ метров пути, начиная} \\ \text{с } i - 2L^2 - \lfloor (i-1-2L^2)/L \rfloor L \\ \text{подвижной единицы, } i = \overline{2L^2 + 1, 3L^2}; \end{array} \right.$$

$A_{s,3L^2+1} = \{ \text{столкновение на железнодорожном переезде, начинающемся на } s\text{-м метре пути} \};$

$A_{s,3L^2+2} = \{ \text{столкновение на железнодорожной станции на стрелочном переводе, начинающемся на } s\text{-м метре пути} \},$

где $\lfloor x \rfloor$ обозначает целую часть x .

Такой способ нумерации позволяет одним индексом i пронумеровать все события в группе с учетом как номера подвижной единицы, с которой начнется сход, так и числа сошедших подвижных единиц в сходе, т.е. L^2 вариантов, и, кроме то-

го, позволяет учесть разницу причин, вызвавших неблагоприятное событие.

Проиллюстрируем принцип работы предложенной нумерации, введя для лаконичности обозначения

$$q_i \stackrel{\text{def}}{=} \begin{cases} \lfloor \frac{i-1}{L} \rfloor + 1, & i = \overline{1, L^2}; \\ \lfloor \frac{i-1-L^2}{L} \rfloor + 1, & i = \overline{L^2 + 1, 2L^2}; \\ \lfloor \frac{i-1-2L^2}{L} \rfloor + 1, & i = \overline{2L^2 + 1, 3L^2}. \end{cases}$$

Таблица 1 Потенциальное число подвижных единиц в сходе с рельсов и номер первой сошедшей подвижной единицы при $L = 10$

Событие	i	q_i	f_i
Сход вне стрелочного перевода по причине неисправности подвижного состава	1	1	1
	2	1	2

	10	1	10
	11	2	1
	12	2	2

	20	2	10

	91	10	1
92	10	2	
...	
100	10	10	
Сход вне стрелочного перевода по причине неисправности железнодорожного пути	101	1	1
	102	1	2

	110	1	10
	111	2	1
	112	2	2

	120	2	10

	191	10	1
192	10	2	
...	
200	10	10	
Сход на стрелочном переводе	201	1	1
	202	1	2

	210	1	10
	211	2	1
	212	2	2

	220	2	10

	291	10	1
291	10	2	
...	
300	10	10	

$$f_i \stackrel{\text{def}}{=} \begin{cases} i - \left\lfloor \frac{i-1}{L} \right\rfloor L, & i = \overline{1, L^2}; \\ i - L^2 - \left\lfloor \frac{i-1-L^2}{L} \right\rfloor L, & i = \overline{L^2+1, 2L^2}; \\ i - 2L^2 - \left\lfloor \frac{i-1-2L^2}{L} \right\rfloor L, & i = \overline{2L^2+1, 3L^2}. \end{cases}$$

Значение q_i характеризует число подвижных единиц в сходе с рельсов, f_i — номер первой сошедшей подвижной единицы.

В табл. 1 представлен расчет для $L = 10$.

4 Оценивание вероятности неблагоприятных событий

В [4, 5] была предложена следующая (с учетом используемых обозначений) оценка вероятности схода с рельсов подвижных единиц грузового поезда (при движении по конкретному классу пути):

$$\mathcal{P}(A_{\text{сх}}) = 1 - \exp \left\{ -S'' \left(\frac{p_1 + p_2 + p_3}{Q_1''} L + \frac{p_{\text{общ}} - p_1 - p_2 - p_3}{Q_2''} \right) \right\},$$

где S'' — длина пути в милях; Q_1'' — число вагоно-миль за промежуток времени T ; Q_2'' — число поездо-миль за промежуток времени T . Пусть S' — длина пути в километрах. Тогда при пересчете миль в километры и метры имеем

$$\begin{aligned} \mathcal{P}(A_{\text{сх}}) &= 1 - \exp \left\{ -\frac{S'}{1,609} \left(\frac{p_1 + p_2 + p_3}{Q_1'/1,609} L + \frac{p_{\text{общ}} - p_1 - p_2 - p_3}{Q_2'/1,609} \right) \right\} = \\ &= 1 - \exp \left\{ -S \left(\frac{p_1 + p_2 + p_3}{1000Q_1'} L + \frac{p_{\text{общ}} - p_1 - p_2 - p_3}{1000Q_2'} \right) \right\}. \end{aligned}$$

Обозначив

$$Q_1 = \frac{p_1 + p_2 + p_3}{1000Q_1'}; \quad Q_2 = \frac{p_{\text{общ}} - p_1 - p_2 - p_3}{1000Q_2'},$$

будем далее полагать

$$\mathcal{P}(A_{\text{сх}}) = 1 - \exp \{ -S(Q_1 L + Q_2) \}. \quad (1)$$

Согласно (1) вероятность схода с рельсов на 1-м метре дистанции

$$\mathcal{P}(A_{\text{сх},1}) = 1 - \exp \{ -(Q_1 L + Q_2) \}. \quad (2)$$

Условная вероятность того, что сход произойдет на 2-м метре дистанции при условии, что на 1-м

метре схода не было, определяется при решении уравнения

$$\begin{aligned} \mathcal{P}(A_{\text{сх},1}) + \mathcal{P}(A_{\text{сх},2} | \overline{A_{\text{сх},1}}) \mathcal{P}(\overline{A_{\text{сх},1}}) &= \\ &= 1 - \exp \{ -2(Q_1 L + Q_2) \}. \end{aligned}$$

Подставляя (2) в последнее уравнение, записываем:

$$\begin{aligned} 1 - \exp \{ -(Q_1 L + Q_2) \} + \\ + \mathcal{P}(A_{\text{сх},2} | \overline{A_{\text{сх},1}}) \exp \{ -(Q_1 L + Q_2) \} &= \\ = 1 - \exp \{ -2(Q_1 L + Q_2) \}. \quad (3) \end{aligned}$$

Решая (3), получаем

$$\mathcal{P}(A_{\text{сх},2} | \overline{A_{\text{сх},1}}) = \mathcal{P}(A_{\text{сх},1}) = 1 - \exp \{ -(Q_1 L + Q_2) \}$$

и далее по индукции

$$\begin{aligned} P_s \stackrel{\text{def}}{=} \mathcal{P} \left(A_{\text{сх},s} \left| \prod_{t=1}^{s-1} \overline{A_{\text{сх},t}} \right. \right) &= \\ = 1 - \exp \{ -(Q_1 L + Q_2) \}, \quad s = \overline{1, S}. \quad (4) \end{aligned}$$

Введем вспомогательные вероятности, необходимые для задания $P_{s_i}(v_s)$: $\overline{P}^{s,l}$, $\tilde{P}^{s,l}$, $\hat{P}^{s,l}$ — вероятность (оценка вероятности) того, что сход начнется с l -й по порядку следования подвижной единицы по причине:

- неисправности подвижного состава вне стрелочного перевода ($\overline{P}^{s,l}$);
- неисправности железнодорожного полотна вне стрелочного перевода ($\tilde{P}^{s,l}$);
- неисправности любого вида на стрелочном переводе ($\hat{P}^{s,l}$) при условии схода по происшествии s метров пути, $s = \overline{1, S}$, $l = \overline{1, L}$.

В предположении, что сход может равновероятно начаться с любой подвижной единицы, а также основываясь на частоте сходов по причине той или иной неисправности и учитывая наличие/отсутствие стрелочного перевода в предполагаемом месте схода, зададим

$$\begin{aligned} \overline{P}^{s,l} &= \frac{1}{L} (1 - \lambda_{s-(d_1+d_2+\dots+d_{l-1})}) \frac{p_1}{p_1 + p_2}; \\ \tilde{P}^{s,l} &= \frac{1}{L} (1 - \lambda_{s-(d_1+d_2+\dots+d_{l-1})}) \frac{p_2}{p_1 + p_2}; \\ \hat{P}^{s,l} &= \frac{1}{L} \lambda_{s-(d_1+d_2+\dots+d_{l-1})}. \end{aligned}$$

Очевидно, что вероятности $\overline{P}^{s,l}$, $\tilde{P}^{s,l}$ и $\hat{P}^{s,l}$ должны зависеть от возраста и состояния рельсов и возраста и условий эксплуатации той или иной по-

движной единицы, однако для простоты использованы оценки вероятности в форме частоты. Имеет смысл более точное оценивание, которое, однако, представляет собой отдельную задачу.

Пусть $\bar{P}_{s,l,k}(v_s)$, $\tilde{P}_{s,l,k}(v_s)$ и $\hat{P}_{s,l,k}(v_s)$ — вероятность (оценка вероятности) схода k подвижных единиц грузового поезда при условии схода по прошествии s метров пути с учетом того, что сход начнется с l -й по порядку следования подвижной единицы по причине:

- неисправности подвижного состава вне стрелочного перевода ($\bar{P}_{s,l,k}(v_s)$);
- неисправности железнодорожного полотна вне стрелочного перевода ($\tilde{P}_{s,l,k}(v_s)$);
- неисправности любого вида на стрелочном переводе ($\hat{P}_{s,l,k}(v_s)$), $s = \overline{1, S}$, $l = \overline{1, L}$, $k = \overline{1, L}$.

Следуя [6, 7], зададим

$$\bar{P}_{s,l,k}(v_s) = \frac{\Gamma(k-1+1/\theta_1)}{\Gamma(k)\Gamma(1/\theta_1)} \left(1 + \theta_1 g_1 \left(a_1, v_s, l, w, L_1, L, \tilde{\mu}, \varkappa_{s-\sum_{r=1}^{l-1} d_r}, \gamma_{s-\sum_{r=1}^{l-1} d_r} \right) \right)^{-(k-1+1/\theta_1)} \times \left(\theta_1 g_1 \left(a_1, v_s, l, w, L_1, L, \tilde{\mu}, \varkappa_{s-\sum_{r=1}^{l-1} d_r}, \gamma_{s-\sum_{r=1}^{l-1} d_r} \right) \right)^{k-1},$$

где a_1 и θ_1 — параметры; $g_1(\cdot)$ — функция, подлежащая подбору на основе метода максимального правдоподобия; $\Gamma(\cdot)$ — гамма-функция. Аналогичным (4) образом зададим $\tilde{P}_{s,l,k}(v_s)$ и $\hat{P}_{s,l,k}(v_s)$ с точностью до замены параметров a_1 и θ_1 и функции $g_1(\cdot)$ на параметры a_2 , a_3 , θ_2 и θ_3 и функции $g_2(\cdot)$ и $g_3(\cdot)$ соответственно.

Роли вероятностей P_s , $\bar{P}^{s,l}$ и $\bar{P}_{s,l,k}(v_s)$ при вычислении $P_{si}(v_s)$ иллюстрирует следующая схема:

$$P_{si}(v_s) = \mathcal{P} \left(\left\{ \begin{array}{l} \text{сход } q_i \text{ подвижных} \\ \text{единиц по причине} \\ \text{неисправности} \\ \text{пути, начиная} \\ \text{с } f_i\text{-й} \\ \text{по счету} \\ \text{вне стрелки} \\ \text{после } s \text{ метров} \end{array} \right\} \left\{ \begin{array}{l} \text{на} \\ \text{предыдущих} \\ s-1 \text{ метрах} \\ \text{не произошло} \\ \text{неблагоприятных} \\ \text{событий} \end{array} \right\} \right) = \mathcal{P} \left(\left\{ \begin{array}{l} \text{сход} \\ \text{после} \\ s \text{ метров} \end{array} \right\} \left\{ \begin{array}{l} \text{на предыдущих} \\ s-1 \text{ метрах} \\ \text{не произошло} \\ \text{неблагоприятных} \\ \text{событий} \end{array} \right\} \right) \times$$

$= P^s \approx P_s$

$$\times \mathcal{P} \left(\left\{ \begin{array}{l} \text{сход по причине} \\ \text{неисправности} \\ \text{пути, начиная} \\ \text{с } f_i\text{-й по счету} \\ \text{подвижной} \\ \text{единицы} \\ \text{вне стрелки} \end{array} \right\} \left\{ \begin{array}{l} \text{на предыдущих} \\ s-1 \text{ метрах} \\ \text{не произошло} \\ \text{неблагоприятных} \\ \text{событий,} \\ \text{сход после} \\ s \text{ метров} \end{array} \right\} \right) \times$$

$= \bar{P}^{s,f_i}$

$$\times \mathcal{P} \left(\left\{ \begin{array}{l} \text{сход } q_i \\ \text{подвижных} \\ \text{единиц} \end{array} \right\} \left\{ \begin{array}{l} \text{на предыдущих} \\ s-1 \text{ метрах} \\ \text{не произошло} \\ \text{неблагоприятных} \\ \text{событий, сход} \\ \text{после } s \text{ метров} \\ \text{по причине неис-} \\ \text{правности пути} \\ \text{начиная с } f_i\text{-й по} \\ \text{счету подвижной} \\ \text{единицы вне стрелки.} \end{array} \right\} \right).$$

$= \bar{P}_{s,f_i,q_i}(v_s)$

Прокомментируем приближенное равенство $P^s \approx P_s$. Имеем

$$P_s = \mathcal{P} \left(A_{cx,s} \prod_{t=1}^{s-1} \bar{A}_{cx,t} \right) = \mathcal{P} \left(A_{cx,s} \prod_{t=1}^{s-1} \prod_{i=1}^{3L^2} \bar{A}_{t,i} \right) = \mathcal{P} \left(A_{cx,s} \prod_{t=1}^{s-1} \prod_{i=1}^{3L^2} \bar{A}_{t,i} \left(\prod_{i=3L^2+1}^{3L^2+2} \bar{A}_{t,i} + \sum_{i=3L^2+1}^{3L^2+2} A_{t,i} \right) \right) = \mathcal{P} \left(A_{cx,s} \prod_{t=1}^{s-1} \prod_{i=1}^{3L^2+2} \bar{A}_{t,i} \right) / \underbrace{\left(\mathcal{P} \left(\prod_{t=1}^{s-1} \prod_{i=1}^{3L^2+2} \bar{A}_{t,i} \right) \right)}_{\approx 1} + \mathcal{P} \left(\prod_{t=1}^{s-1} \prod_{i=1}^{3L^2} \bar{A}_{t,i} \sum_{i=3L^2+1}^{3L^2+2} A_{t,i} \right) \approx \mathcal{P} \left(A_{cx,s} \prod_{t=1}^{s-1} \prod_{i=1}^{3L^2+2} \bar{A}_{t,i} \right) = P^s.$$

≈ 0 , так как вероятность любого неблагоприятного события должна быть около нуля — в противном случае начинать движение опасно

Учитывая вышесказанное, получаем следующие оценки для величин $P_{si}(v_s)$:

$$P_{si}(v_s) = \begin{cases} P_s \bar{P}^{s, f_i} \bar{P}_{s, f_i, q_i}(v_s), & i = \overline{1, L^2}; \\ P_s \tilde{P}^{s, f_i} \tilde{P}_{s, f_i, q_i}(v_s), & i = \overline{L^2 + 1, 2L^2}; \\ P_s \hat{P}^{s, f_i} \hat{P}_{s, f_i, q_i}(v_s), & i = \overline{2L^2 + 1, 3L^2}. \end{cases}$$

Вероятность $P_{s, 3L^2+2}(v_s)$ можно оценить, например, при помощи имитационного моделирования [11] либо на основе пуассоновских потоков [12], вероятность $P_{s, 3L^2+2}(v_s)$ может быть оценена в форме частоты на основании статистики по пересечениям грузовыми поездами переездов и столкновениям на них.

5 Оценивание ущерба

Будем рассматривать только материальный ущерб. Ущерб, связанный с задержками поездов, вызванными возникновением неблагоприятных событий, ввиду сложности пересчета в материальный ущерб учитывать не будем. Иными словами, предполагается, что ущерб при задержках отсутствует. При сходе с рельсов ущерб включает в себя расходы, связанные с ремонтом/списанием подвижных единиц грузового поезда. Поскольку стоимость различных видов ремонта подвижных единиц грузовых поездов, участвовавших в сходах и крушениях, недоступна, предложим следующую оценку среднего размера ущерба:

$$M[C_{si}(v_s)] = 10^5 P_{si}^a(v_s) + 4,5 \cdot 10^6 \cdot \frac{9}{200} v_s \cdot \begin{cases} q_i, & f_i + q_i - 1 \leq L; \\ 0 & \text{иначе,} \end{cases} \quad (5)$$

где $P_{si}^a(v_s)$ — оценка вероятности выхода в габарит соседнего пути хотя бы одной подвижной единицы грузового поезда, $i = \overline{1, 3L^2}$. Согласно [13], средний ущерб от столкновений составляет 1599 тыс. руб., а по открытым источникам, примерно каждый 16-й раз в случае выхода в габарит соседнего пути происходит столкновение со встречным поездом. Поэтому средний ущерб при выходе в габарит соседнего пути оценивается в $1\,599\,000/16 \approx 10^5$ руб., что и представляет сомножитель $P_{si}^a(v_s)$ в (5). Отметим, что данный сомножитель, вообще говоря, зависит от интенсивности встречного движения. Поэтому приведенное число следует рассматривать лишь как некоторое начальное приближение. Далее, скорость в 80 км/ч (как правило, выше этого порога скорость грузового поезда недопустима) эквивалентна 200/9 м/с, поэтому сомножитель 9/200 используется в формуле (5). Заметим,

что 4,5 млн руб. — приблизительная стоимость нового грузового вагона. При этом в сходе с рельсов может оказаться и секция локомотива, однако повреждение секции локомотива до исключения из инвентарного парка намного более редкое событие, чем исключение вагона, поэтому для простоты предполагается, что максимальный ущерб от схода секции локомотива равен стоимости нового вагона. Следует отметить, что с ростом скорости растет и кинетическая энергия, а следовательно, и последствия для подвижных единиц от схода. Поэтому в (5) предполагается линейный рост размера ущерба с ростом скорости. Значение $P_{si}^a(v_s)$ можно оценить с привлечением метода максимального правдоподобия на основе исторических данных о возникновении данного события [8]:

$$P_{si}^a(v_s) = \begin{cases} \bar{P}_s \left(v_s, f_i, q_i, b_1, w, L_1, L, \tilde{\mu}, \right. \\ \left. \varkappa_{s-\sum_{r=1}^{f_i-1} d_r}, \gamma_{s-\sum_{r=1}^{f_i-1} d_r} \right), & i = \overline{1, L^2}; \\ \tilde{P}_s \left(v_s, f_i, q_i, b_2, w, L_1, L, \tilde{\mu}, \right. \\ \left. \varkappa_{s-\sum_{r=1}^{f_i-1} d_r}, \gamma_{s-\sum_{r=1}^{f_i-1} d_r} \right), & i = \overline{L^2 + 1, 2L^2}; \\ 1, & i = \overline{2L^2 + 1, 3L^2}, \end{cases}$$

где $\bar{P}_s(\cdot)$ и $\tilde{P}_s(\cdot)$ — подлежащие оцениванию функции; b_1 и b_2 — подлежащие оцениванию параметры.

При столкновениях на железнодорожном переезде, как правило, виновным признается владелец автотранспорта, поэтому ущерб, вызванный столкновением, компенсируется, откуда

$$M[C_{si}(v_s)] = 0, \quad s = \overline{1, S}, \quad i = 3L^2 + 1.$$

Аналогично, согласно [13], можно положить

$$M[C_{si}(v_s)] = 1,599 \cdot 10^6, \quad s = \overline{1, S}, \quad i = 3L^2 + 2.$$

6 Пример

Выберем промежуток времени T с 2013 по 2016 гг. Согласно [7], $p_{\text{общ}} = 246$, $p_1 = 150$, $p_2 = 38$ и $p_3 = 46$. Данные о величинах Q_1 и Q_2 в открытых источниках за данный период отсутствуют. Предположим, что $Q'_1 = 1,8 \cdot 10^{11}$ и $Q'_2 = 3 \cdot 10^9$. Отметим, что при таких значениях чисел Q'_1 и Q'_2 вероятность (1) имеет тот же порядок, что и аналогичная вероятность из [4]. Пусть $L_0 = 2$, $L_2 = 58$, $w = 5000$ и $S = 250\,000$. Положим $d_1 = d_2 = 20$ и $d_3 = d_4 = \dots = d_{L_0+L_1} = 14$. Таким образом, длина состава $S_0 = 852$. Пусть движение происходит по двухпутной железной дороге, т. е. $\chi_s = 1$, $s = \overline{2 - S_0, S}$, железнодорожных переездов

нет, т. е. $\eta_s = 0$, $s = \overline{2 - S_0, S}$, а стрелочные переводы располагаются не в пределах станций ($\xi_s = 0$, $s = \overline{2 - S_0, S}$), но на каждом километре пути так, что

$$\lambda_s = \begin{cases} 1 & \text{если } s = \overline{(j-1) \cdot 1000 + 1, (j-1) \cdot 1000 + 30}; \\ 0 & \text{иначе,} \end{cases}$$

где $j = \overline{1, \lfloor S/1000 \rfloor}$.

Зададим в табл. 2 кривизну и уклон на пути движения.

Выберем различные режимы движения, т. е. наборы скоростей (табл. 3).

Отметим, что для всех режимов время проследования маршрута одинаково и составляет 20 000 с.

Для задания функций $g_1(\cdot)$ и $g_3(\cdot)$, а также параметров a_1, a_3, θ_1 и θ_3 воспользуемся приведенными в [7] результатами. Функцию $g_2(\cdot)$ выберем отличную от [7], а именно: положим

$$g_2(a_2, v, l, w, L_1, L, \tilde{\mu}, \varkappa, \gamma) = \exp \{ -3,7 + 0,97\tilde{\mu} + 0,2 \ln(v) \mathbf{1}_{\gamma > 0} + 1,33 \ln(L - l + 1) \}.$$

Здесь

$$\mathbf{1}_{\gamma > 0} = \begin{cases} 1, & \text{если } \gamma > 0; \\ 0 & \text{в противном случае.} \end{cases}$$

Такая функция $g_2(\cdot)$ обеспечивает большее значение функции максимального правдоподобия, нежели функция из [7].

Вычисленные значения функций интегрального риска на рассматриваемых режимах движения приведены в табл. 4.

Прокомментируем полученные в табл. 4 результаты. Значения функции $R_1(\cdot)$ на рассмотренных режимах движения ожидаемо примерно равны (в табл. 4 данные приведены с округлением), так как в предложенной концепции оценки вероятностей неблагоприятных событий величина P_s не зависит от скорости, а $\forall s \in \{1, \dots, S\} \forall v_s > 0$ имеем $\sum_{i=1}^{3L^2} P_{si}(v_s) \approx P_s$. Отметим, что здесь используется приближенное, а не точное равенство, так как у применяемого при прогнозировании числа подвижных единиц в сходе с рельсов отрицательного биномиального распределения бесконечное число значений.

Комментируя значения функции $R_2(\cdot)$, следует обратить внимание на небольшие по абсолютной величине значения. Хотя последствия схода могут исчисляться миллионами рублей, средний ущерб на рассмотренных данных составляет порядка 170 руб. Это вызвано тем, что сход с рельсов является крайне редким событием. Действительно, согласно (1) вероятность схода составляет $2,05 \cdot 10^{-5}$. Далее,

Таблица 2 Карта подъемов, спусков и кривых

s	\varkappa_s	γ_s
[1, 50000]	1/2000	-0,0001
[50001, 60000]	1/1500	-0,009
[60001, 70000]	1/1500	-0,003
[70001, 100000]	0	0,01
[100001, 110000]	1/800	-0,004
[-850, 0]	0	0
[110001, 150000]	1/1000	0,008
[150001, 200000]	0	0
[200001, 205000]	1/600	0,005
[205001, 220000]	0	-0,01
[220001, 250000]	0	0,005

Таблица 3 Значения скоростей (в м/с) при различных режимах движения

s	v_s^1	v_s^2	v_s^3
[1, 50000]	12,5	16	16,34
[50001, 60000]	12,5	10	11,5
[60001, 70000]	12,5	12,5	12,5
[70001, 100000]	12,5	10	10,5
[100001, 110000]	12,5	10	12,5
[110001, 150000]	12,5	12,5	10
[150001, 200000]	12,5	20	16,28
[200001, 205000]	12,5	5	10,5
[205001, 220000]	12,5	6	10
[220001, 250000]	12,5	16	11,692

Таблица 4 Интегральные функции риска при различных режимах движения

Режим движения	$R_1(v_1, \dots, v_s)$	$R_2(v_1, \dots, v_s)$
1	$2,046 \cdot 10^{-5}$	168,68
2	$2,046 \cdot 10^{-5}$	183,6
3	$2,046 \cdot 10^{-5}$	167,86

согласно [7] в среднем сходит 4,23 подвижные единицы. При предполагаемом максимальном ущербе от схода подвижной единицы в 4,5 млн руб. и средней скорости в 12,5 м/с получаем, что в среднем ущерб от схода составляет порядка 220 руб. Таким образом, порядок цифр верен. Однако в то же время возникает вопрос, зачем использовать рассмотренный в статье сложный аппарат для оценивания функции $R_2(\cdot)$. Ответ в том, что благодаря представленной концепции выяснилось, что *средний ущерб при различных режимах движения может отличаться на 10%*. А это значит, что выбором скоростного режима можно существенно сократить средний ущерб от схода.

7 Заключение

Разработаны функции интегрального риска возникновения различных неблагоприятных событий,

сопутствующим движению на рельсовом транспорте. Предложена концепция оценивания вероятности и ущерба от различных неблагоприятных событий, случающихся во время движения грузовых поездов. На содержательном примере показано, что выбор скоростного режима может существенно повлиять на средний ущерб при осуществлении перевозок. Результат дает импульс к дальнейшему развитию старых и исследованию новых моделей:

- прогнозирования места схода с рельсов;
- прогнозирования причины схода с рельсов;
- прогнозирования номера первой сошедшей с рельсов подвижной единицы;
- оценивания материального ущерба от схода с рельсов при различных факторах движения.

Литература

1. ГОСТ Р ИСО 31000-2019. Менеджмент риска. Принципы и руководство. — М.: Стандартинформ, 2020. 19 с.
2. ГОСТ 33433-2015. Безопасность функциональная. Управление рисками на железнодорожном транспорте. — М.: Стандартинформ, 2016. 34 с.
3. *Saccomanno F. F., El-Hage S.* Establishing derailment profiles by position for corridor shipments of dangerous goods // *Can. J. Civil Eng.*, 1991. Vol. 18. P. 67–75.
4. *Anderson R. T., Barkan C. P. L.* Derailment probability analyses and modeling of mainline freight trains // 8th Heavy Haul Conference (International) Proceedings. — Rio de Janeiro, Brazil, 2005. P. 491–497.
5. *Bagheri M., Saccomanno F., Chenouri S., Fu L.* Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist // *Accident Anal. Prev.*, 2011. Vol. 43. No. 3. P. 613–620.
6. *Liu X., Saat M. R., Qin X., Barkan C. P. L.* Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression // *Accident Anal. Prev.*, 2013. Vol. 59. P. 87–93.
7. *Zamyshliaev A. M., Ignatov A. N., Kibzun A. I., Novozhilov E. O.* Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors // *Dependability*, 2018. Vol. 18. No. 1. P. 53–60.
8. *Zamyshliaev A. M., Ignatov A. N., Kibzun A. I., Novozhilov E. O.* On traffic safety incidents caused by intrusion of derailed freight cars into the operational space of an adjacent track // *Dependability*, 2018. Vol. 18. No. 3. P. 39–45.
9. *Bagheri M., Saccomanno F., Fu L.* Effective placement of dangerous goods cars in rail yard marshaling operation // *Can. J. Civil Eng.*, 2010. Vol. 37. P. 753–762.
10. *Rahbar M., Bagheri M.* Risk assessment framework for the rail transport of hazardous materials: Formulation and solution // *Transp. Res. Rec.*, 2014. Vol. 2411. P. 90–95.
11. *Босов А. В., Игнатов А. Н., Наумов А. В.* Модель передвижения поездов и маневровых локомотивов на железнодорожной станции в приложении к оценке и анализу вероятности бокового столкновения // *Информатика и её применения*, 2018. Т. 12. Вып. 3. С. 107–114.
12. *Игнатов А. Н., Кибзун А. И., Платонов Е. Н.* Оценка вероятности столкновения составов на железнодорожных станциях на основе пуассоновской модели // *Автоматика и телемеханика*, 2016. № 11. С. 43–59. EDN: XWTSYD.
13. *Shubinsky I. B., Zamyshliaev A. M., Ignatov A. N., Kibzun A. I.* Use of automatic signalling system for reduction of the risk of transportation incidents in railway stations // *Dependability*, 2017. Vol. 17. No. 3. P. 49–57.

Поступила в редакцию 24.10.22

ON THE PROBLEM OF ASSESSING AND ANALYZING TRAFFIC ACCIDENTS RISK ON THE RAIL TRANSPORT

A. V. Bosov^{1,2} and A. N. Ignatov²

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

²Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

Abstract: The problem of assessing and analyzing traffic accidents risk on the rail transport is considered. Two functions of the integral risk are proposed that allow assessing danger of transportation along the entire route of a transport. The probability of an unfavorable event occurring during transportation and the expected damage are chosen as such functions. The concept of assessing probability and damage from unfavorable events during the freight trains transportation is proposed. A meaningful example of calculating integral risk functions is given on the basis of previously investigated statistics on the freight trains transportation and unfavorable events that occurred with them.

Keywords: risk; unfavorable event; rail transport; probability; expected damage

DOI: 10.14357/19922264230110

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (grant 20-07-00046 A).

References

1. GOST R ISO 31000-2019. 2020. Menedzhment riska. Printsipy i rukovodstvo [Risk management. Principles and guidelines]. Moscow: Standartinform. 19 p.
2. GOST 33433-2015. 2016. Bezopasnost' funktsional'naya. Upravlenie riskami na zheleznodorozhnom transporte [Functional safety. Risk control in railroad transport]. Moscow: Standartinform. 34 p.
3. Saccomanno, F. F., and S. El-Hage. 1991. Establishing derailment profiles by position for corridor shipments of dangerous goods. *Can. J. Civil Eng.* 18: 67–75.
4. Anderson, R. T., and C. P. L. Barkan. 2005. Derailment probability analyses and modeling of mainline freight trains. *8th Heavy Haul Conference (International) Proceedings*. Rio de Janeiro, Brazil. 491–497.
5. Bagheri, M., F. Saccomanno, S. Chenouri, and L. Fu. 2011. Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. *Accident Anal. Prev.* 43(3):613–620.
6. Liu, X., M. R. Saat, X. Qin, and C. P. L. Barkan. 2013. Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Anal. Prev.* 59:87–93.
7. Zamyshliaev, A. M., A. N. Ignatov, A. I. Kibzun, and E. O. Novozhilov. 2018. Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors. *Dependability* 18(1):53–60.
8. Zamyshliaev, A. M., A. N. Ignatov, A. I. Kibzun, and E. O. Novozhilov. 2018. On traffic safety incidents caused by intrusion of derailed freight cars into the operational space of an adjacent track. *Dependability* 18(3):39–45.
9. Bagheri, M., F. Saccomanno, and L. Fu. 2010. Effective placement of dangerous goods cars in rail yard marshaling operation. *Can. J. Civil Eng.* 37:753–762.
10. Rahbar, M., and M. Bagheri. 2014. Risk assessment framework for the rail transport of hazardous materials: Formulation and solution. *Transp. Res. Rec.* 2411(1):90–95.
11. Bosov, A. V., A. N. Ignatov, and A. V. Naumov. 2018. Model' peredvizheniya poezdov i manevrovyykh lokomotivov na zheleznodorozhnoy stantsii v prilozhenii k otsenke i analizu veroyatnosti bokovogo stolknoveniya [Transportation of trains and shunting locomotives at the railway station model for evaluating and analysis of side-collision probabilities]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):107–114.
12. Ignatov, A. N., A. I. Kibzun, and E. N. Platonov. 2016. Estimating collision probabilities for trains on railroad stations based on a Poisson model. *Automat. Rem. Contr.* 77:1914–1927.
13. Shubinsky, I. B., A. M. Zamyshliaev, A. N. Ignatov, and A. I. Kibzun. 2017. Use of automatic signalling system for reduction of the risk of transportation incidents in railway stations. *Dependability* 17(3):49–57.

Received October 24, 2022

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; professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; AVBosov@ipiran.ru

Ignatov Aleksei N. (b. 1991) — Candidate of Science in physics and mathematics, assistant professor, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; alexei.ignatov1@gmail.com

ОЦЕНКИ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ В МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СЕТИ ПРИ РАВНЫХ МЕЖУЗЛОВЫХ НАГРУЗКАХ

Ю. Е. Малашенко¹, И. А. Назарова²

Аннотация: Предлагается метод оценки ресурсов при уравнительном распределении межузловых нагрузок в многопользовательской сети. В рамках формальной математической модели пропускные способности ребер рассматриваются как компоненты вектора ресурсов, которые требуются для передачи потоков разных видов. Предлагается алгоритмическая процедура перераспределения и использования пропускных способностей при равных объемах ресурсов, выделяемых всем корреспондентам. При поиске соответствующих реберных загрузок определяются значения максимальных однопродуктовых потоков для каждой пары узлов. В ходе вычислительных экспериментов суммарный ресурс считается заданным для сетей с различными структурными особенностями.

Ключевые слова: многопродуктовая потоковая модель; распределение ресурсов и межузловых нагрузок; предельная нагрузка сети

DOI: 10.14357/19922264230111

1 Введение

Данная работа продолжает исследование методов анализа функциональных возможностей телекоммуникационных систем [1, 2]. В рамках многопродуктовой сетевой модели рассматривается вычислительная процедура оценки ресурсов, которые требуются при уравнительном распределении межузловых нагрузок. Под ресурсом, выделяемым некоторой паре узлов-корреспондентов, понимается суммарная величина пропускных способностей, необходимых для обеспечения передачи информационного потока определенного вида. Дуговые потоки разных видов рассматриваются как *нагрузка* на указанном ребре, возникающая при одновременном соединении корреспондентов. В ходе экспериментов осуществляется поиск распределения потоков, при котором межузловые нагрузки равны друг другу. Согласно предложенной процедуре для каждой пары узлов вычисляется максимальный однопродуктовый поток. Маршруты передачи для каждой пары прокладываются по ребрам соответствующих минимальных разрезов, и для всех корреспондентов подсчитываются реберные нагрузки. На основе найденных значений определяются требуемые пропускные способности при одновременной передаче всех межузловых потоков. Анализ результатов экспериментов позволяет оценить распределение ресурсов при равных межузловых нагрузках и предельной нагрузке в многопользовательских сетях [1, 2].

В настоящее время для создания, развития и эксплуатации телекоммуникационных систем используются потоковые модели и разрабатываются специальные методы решения [3, 4]. Математические модели передачи многопродуктового потока применяются для поиска недискриминирующих правил распределения ресурсов в системах связи [5]. В русле исследований, изложенных в указанных работах, лежит алгоритмическая схема получения уравнительных распределений межузловых нагрузок (разд. 3). В разд. 4 обсуждаются результаты экспериментов и сравниваются достижимые межузловые потоки при предельных распределениях пропускных способностей в сетях с различной структурой.

2 Математическая модель

Для описания многопользовательской сетевой системы связи воспользуемся следующей математической записью модели передачи многопродуктового потока. Сеть G задается множествами $\langle V, R, U, P \rangle$: узлов (вершин) сети $V = \{v_1, v_2, \dots, v_n, \dots, v_N\}$; неориентированных ребер $R = \{r_1, r_2, \dots, r_k, \dots, r_E\}$; ориентированных дуг $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$; пар узлов-корреспондентов $P = \{p_1, p_2, \dots, p_M\}$. Предполагается, что в сети отсутствуют петли и сдвоенные ребра. В многопользовательской сети G рассматривается $M = N(N - 1)$ независимых, невзаимозаменяемых

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, malash09@ccas.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

и равноправных межузловых потоков различных видов.

Ребро $r_k \in R$ соединяет смежные вершины v_{n_k} и v_{j_k} . Каждому ребру r_k ставятся в соответствие две ориентированные дуги u_k и u_{k+E} из множества U . Дуги $\{u_k, u_{k+E}\}$ определяют прямое и обратное направления передачи потока по ребру r_k между концевыми вершинами v_{n_k} и v_{j_k} .

Каждой паре узлов-корреспондентов p_m из множества P ставятся в соответствие: вершина-источник с номером s_m , из s_m входной поток m -го вида поступает в сеть; вершина-приемник с номером t_m , из t_m поток m -го вида покидает сеть. Обозначим через z_m величину межузлового потока m -го вида, поступающего в сеть через узел с номером s_m и покидающего сеть из узла с номером t_m ; x_{mk} , $x_{m(k+E)}$ — поток m -го вида, который передается по дугам u_k и u_{k+E} согласно направлению передачи, $x_{mk} \geq 0$, $x_{m(k+E)} \geq 0$, $m = \overline{1, M}$, $k = \overline{1, E}$; $S(v_n)$ — множество номеров исходящих дуг, по ним поток покидает узел v_n ; $T(v_n)$ — множество номеров входящих дуг, по ним поток поступает в узел v_n . Состав множеств $S(v_n)$ и $T(v_n)$ однозначно формируется в ходе выполнения следующей процедуры.

Пусть некоторое ребро $r_k \in R$ соединяет вершины с номерами n и j , такими что $n < j$. Тогда ориентированная дуга $u_k = (v_n, v_j)$, направленная из вершины v_n в v_j , считается *исходящей* из вершины v_n и ее номер k заносится в множество $S(v_n)$, а дуга u_{k+E} , направленная из v_j в v_n , — *входящей* для v_n и ее номер $k + E$ помещается в список $T(v_n)$. Дуга u_k является *входящей* для v_j , и ее номер k попадает в $T(v_j)$, а дуга u_{k+E} — *исходящей*, и номер $k + E$ вносится в список исходящих дуг $S(v_j)$.

Во всех узлах сети $v_n \in V$, $n = \overline{1, N}$, для каждого вида потока должны выполняться условия сохранения потоков:

$$\sum_{i \in S(v_n)} x_{mi} - \sum_{i \in T(v_n)} x_{mi} = \begin{cases} z_m, & \text{если } v_n = v_{s_m}; \\ -z_m, & \text{если } v_n = v_{t_m}; \\ 0 & \text{в остальных случаях,} \end{cases} \\ n = \overline{1, N}, m = \overline{1, M}, x_{mi} \geq 0, z_m \geq 0. \quad (1)$$

Величина z_m равна входному межузловому потоку m -го вида, проходящему от источника к приемнику пары p_m при распределении потоков x_{mi} по дугам сети.

Каждому ребру $r_k \in R$ приписывается неотрицательное число d_k , определяющее суммарный предельно допустимый поток, который можно передать по ребру r_k в обоих направлениях. В исходной сети компоненты вектора пропускных способно-

стей $\mathbf{d} = (d_1, d_2, \dots, d_k, \dots, d_E)$ — наперед заданные положительные числа $d_k > 0$. Вектор \mathbf{d} задает следующие ограничения на сумму потоков всех видов, передаваемых по ребру r_k одновременно:

$$\sum_{m=1}^M (x_{mk} + x_{m(k+E)}) \leq d_k, \quad x_{mk} \geq 0, \\ x_{m(k+E)} \geq 0, \quad k = \overline{1, E}. \quad (2)$$

Ограничения (1) и (2) задают множество допустимых значений компонент вектора межузловых потоков $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$:

$$\mathcal{Z}(\mathbf{d}) = \{ \mathbf{z} \geq 0 \mid \mathbf{E} \mathbf{x} \geq 0 : (\mathbf{z}, \mathbf{x}) \text{ удовлетворяют (1), (2)} \}.$$

В рамках данной модели пропускная способность ребер сети измеряется в условных единицах потока и трактуется как *ресурсное* ограничение. Суммарное значение пропускной способности сети $D(0) = \sum_{k=1}^E d_k$ считается заданным. Для каждой пары узлов-корреспондентов $p_m \in P$, для некоторого допустимого межузлового потока \tilde{z}_m и соответствующих дуговых потоков \tilde{x}_{mk} , $k = \overline{1, E}$, величина

$$\tilde{y}_m = \sum_{i=1}^{2E} \tilde{x}_{mi}, \quad m = \overline{1, M},$$

характеризует *нагрузку* на сеть при передаче межузлового потока величины \tilde{z}_m из узла-источника s_m в узел-приемник t_m . Величина \tilde{y}_m показывает, какая суммарная пропускная способность сети требуется для передачи дуговых потоков \tilde{x}_{mk} . В рамках модели отношение реберных и межузловых потоков

$$\tilde{w}_m = \frac{\tilde{y}_m}{\tilde{z}_m}, \quad m = \overline{1, M},$$

можно трактовать как удельные *затраты* ресурсов сети при передаче единичного потока m -го вида между узлами s_m и t_m при дуговых потоках \tilde{x}_{mi} . Величины $\tilde{z}_m = \tilde{z}_m / \tilde{y}_m$, $\tilde{x}_{mi} = \tilde{x}_{mi} / \tilde{y}_m$, $m = \overline{1, M}$, $i = \overline{1, E}$, соответствуют межузловому потоку при единичной нагрузке для пары p_m .

3 Метод оценки распределения ресурсов при выравнивании межузловой нагрузки

В рамках модели проводились вычислительные эксперименты для оценки распределения пропускной способности сети при равных нагрузках для

всех пар-корреспондентов. При подготовке данных формировался вектор исходных пропускных способностей $\mathbf{d}(0)$ для некоторой заданной сети $G(0)$, в которой $D(0) = \sum_{k=1}^E d_k(0)$. В сети $G(0)$ при заданных $d_k(0)$ последовательно решалась задача поиска максимального независимого однопродуктового потока [6] для каждой пары узлов $p_m \in P$, $m = \overline{1, M}$.

Задача 1. Для некоторой пары узлов p_a найти

$$z_a^0(0) = \max \{z_a | (\mathbf{z}, x) \in \mathcal{Z}(\mathbf{d}(0))\}$$

при дополнительных условиях

$$z_m = 0, m \neq a, m = \overline{1, M}.$$

При последовательном решении задачи 1 для каждой пары $p_a \in P$ вычисляются максимальный межузловой поток $z_a^0(0)$ и дуговые потоки $(x_{ak}^0(0), x_{a(k+E)}^0(0))$, $k = \overline{1, E}$. Далее подсчитываются нагрузка

$$y_a^0(0) = \sum_{k=1}^{2E} x_{ak}^0(0),$$

при $y_a^0(0) \neq 0$ нормирующий коэффициент $\theta_a^0(0) = 1/y_a^0(0)$ и дуговые потоки $x_{ak}^0 = \theta_a^0(0)x_{ak}^0(0)$, $k = \overline{1, 2E}$. При передаче дуговых потоков x_{ak}^0 соответствующая нагрузка y_a^0 равна единице при передаче межузлового потока $z_a^0 = \theta_a^0(0)z_a^0(0)$.

Задача 1 решается последовательно для всех $p_m \in P$, и для найденных дуговых потоков x_{mk}^0 определяется суммарная нагрузка на ребра сети при передаче межузлового потока z_m^0 :

$$\Delta_k^0 = \sum_{m=1}^M (x_{mk}^0 + x_{m(k+E)}^0), k = \overline{1, E}.$$

Для заданного значения $D(0)$ определяется допустимое перераспределение реберных нагрузок:

$$\beta(0) \left[\sum_{m=1}^M \theta_m^0(0) (x_{mk}^0(0) + x_{m(k+E)}^0(0)) \right] = \Delta_k^0(0), k = \overline{1, E}. \quad (3)$$

В рамках модели суммарная величина нагрузки численно равна требуемой суммарной пропускной способности, которая считается заданной:

$$\sum_{k=1}^E \Delta_k^0(0) = D(0). \quad (4)$$

Из соотношений (3) и (4) находится значение $\beta(0)$:

$$\beta(0) \sum_{k=1}^E \left[\sum_{m=1}^M \theta_m^0(0) (x_{mk}^0(0) + x_{m(k+E)}^0(0)) \right] = D(0),$$

вычисляются значения $\Delta_k^0(0)$, $k = \overline{1, E}$.

На основе найденных значений формируется сеть $G(1)$, в которой $d_k^*(1) := \Delta_k^0(0)$, $k = \overline{1, E}$. В сети $G(1)$ для всех пар узлов $p_a \in P$, $a = \overline{1, M}$, определяется максимальный однопродуктовый поток.

Задача 2. Найти

$$z_a^0(1) = \max \{z_a | (\mathbf{z}, x) \in \mathcal{Z}(\mathbf{d}^*(1))\}$$

при дополнительных условиях

$$z_m = 0, m \neq a, m = \overline{1, M}.$$

На основе последовательного решения задачи 2 вычисляются коэффициенты нормировки

$$\theta_m^0(1) = \frac{1}{y_m^0(1)} \text{ для всех } z_m^0(1) > 0, m = \overline{1, M},$$

и формируется система уравнений для поиска распределения нагрузок

$$\beta^*(1) \sum_{m=1}^M \theta_m^0(1) (x_{mk}^0(1) + x_{m(k+E)}^0(1)) = \Delta_k^0(1), k = \overline{1, E}, \sum_{k=1}^E \Delta_k^0(1) = D(0).$$

Вычисляются $\beta^*(1)$ и $\Delta_k^0(1)$, $k = \overline{1, E}$, и формируется сеть $G(2)$, в которой пропускные способности $d_k^*(2)$ полагаются равными $\Delta_k^0(1)$, т.е. $d_k^*(2) := \Delta_k^0(1)$, $k = \overline{1, E}$. Для полученных решений находим

$$z_m^*(1) = \beta^*(1)\theta_m^0(1)z_m^0(1),$$

$p_m \in P$ — распределение межузловых потоков при равных значениях нагрузок $y_m^0(1)$. Все ребра сети $G(2)$ загружены полностью, выделенный ресурс совпадает с нагрузкой на ребро при равных межузловых нагрузках.

Норма вектора требуемых пропускных способностей в «новой» сети $G(2)$:

$$\|\Delta_k^0(1)\| = \|\mathbf{d}^*(2)\| = \left[\sum_{k=1}^E d_k^*(2)^2 \right]^{1/2}.$$

4 Вычислительный эксперимент

Вычислительный эксперимент проводился на моделях сетевых систем, представленных на рис. 1. В каждой сети 69 узлов. Пропускные способности ребер — значения $d_k(0)$ — выбирались случайным образом из отрезка $[900, 999]$ и совпадали для ребер, общих для обеих сетей. В кольцевой сети пропускная способность каждого из добавленных ребер внутреннего кольца составила 900. В ходе вычислительного эксперимента проводилась нормировка,

и суммарная пропускная способность в обеих сетях была одинакова: $\sum_{k=1}^E d_k(0) = D(0) = 68\,256$.

Результаты вычислительных экспериментов представлены на рис. 2 и 3. Толщина ребер на рисунках пропорциональна результирующей нагрузке.

В базовой сети основная нагрузка — 1400 единиц — приходится на радиальные ребра, которые исходят из центрального узла, а на ребрах внешнего кольца не превышает 1250 единиц. При формировании кольцевой сети в состав внутреннего кольца

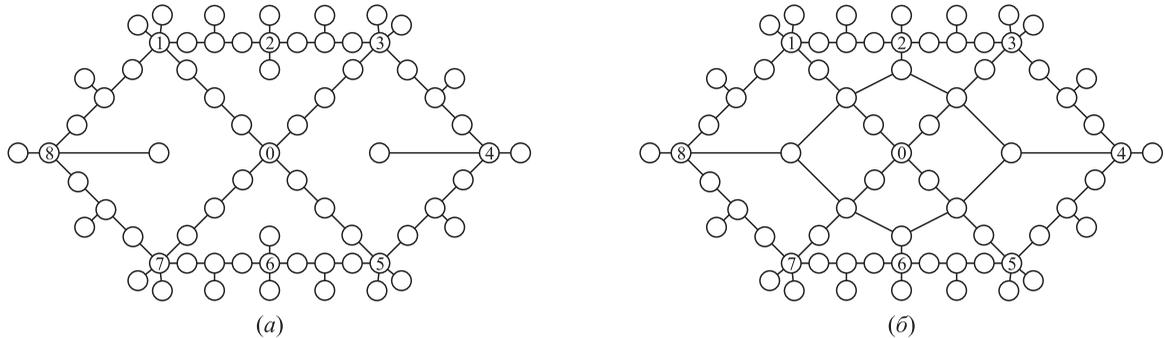


Рис. 1 Базовая (а) и кольцевая (б) сети

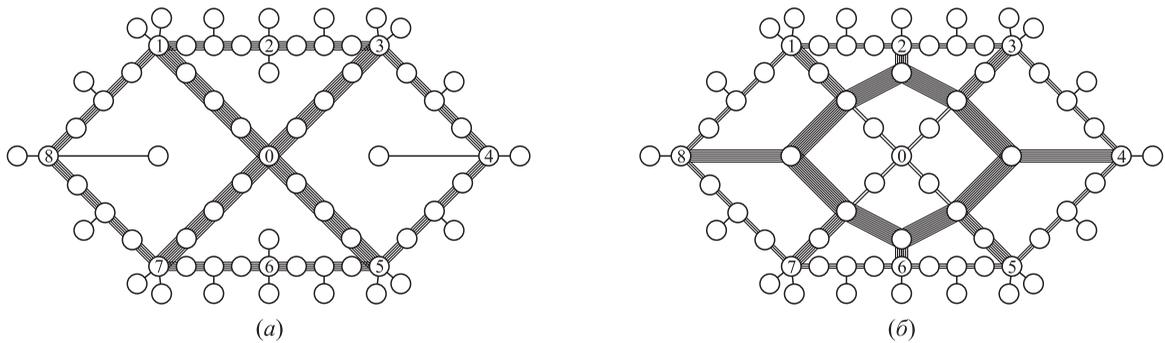


Рис. 2 Результирующие нагрузки на ребра базовой (а) и кольцевой (б) сети

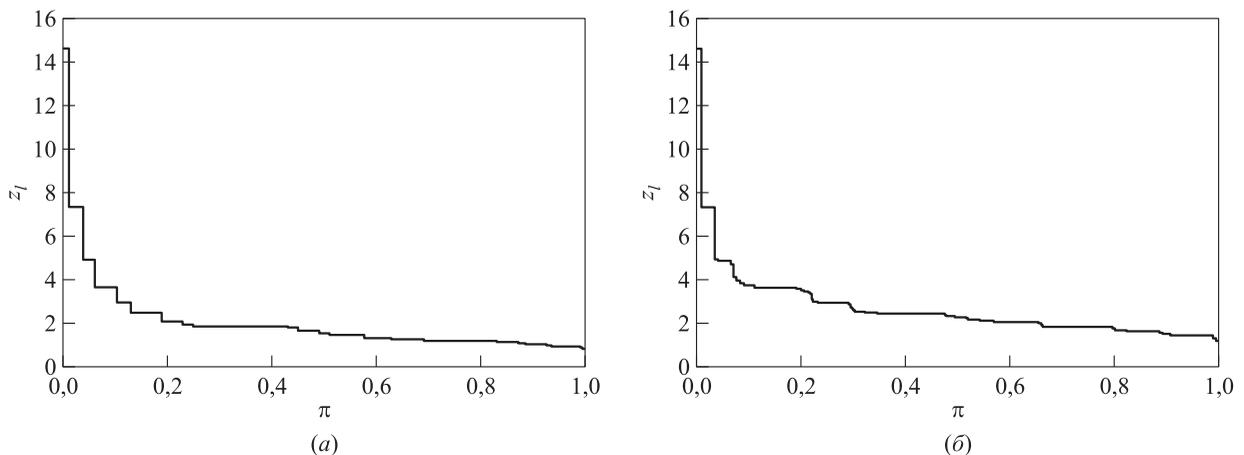


Рис. 3 Распределение межузловых потоков в базовой (а) и кольцевой (б) сети

включаются 4 узла, которые в базовой сети были *висячими*. Соответствующие ребра образуют *мостики* для передачи транзитных потоков между внутренним и внешним кольцом. Нагрузка на ребра внутреннего кольца составляет 2050 единиц, на внешнем не превышает 800, а на *ребрах-мостиках* — 1500. В обеих сетях минимальная нагрузка достигается на *висячих* ребрах: 300 единиц для базовой и 400 для кольцевой. Указанные значения определяются собственными информационными потоками, которые передаются из каждого узла. При этом транзитные нагрузки, например, на ребрах внутреннего кольца, на порядок превышают собственные потоки инцидентных узлов.

Для полученных значений нагрузок вычисляются межузловые потоки $z_m^*(1)$. Результирующие диаграммы представлены на рис. 3, на котором приведены значения $z_l, l = \overline{1, m}$, упорядоченные по величине от большего к меньшему. На вертикальной оси по невозрастанию откладываются z_l , а по горизонтальной указываются порядковые номера в данной упорядоченной последовательности

$$\pi(l) = \frac{l}{m} \text{ для } l = \overline{1, M},$$

где M — общее число элементов в множестве пар корреспондентов P . Результаты вычислительных экспериментов и диаграммы, представленные на рис. 3, свидетельствуют, что как суммарный поток, так и медианное значение межузлового потока на 50% больше в кольцевой сети при равных пропускных способностях в обеих сетях. Дело в том, что при вычислении максимального потока, а в дальнейшем при поиске равных нагрузок используются все пути, проходящие через минимальные сечения. В кольцевой сети маршруты соединения оказываются короче, чем в базовой, для большого числа пар узлов, что уменьшает межузловые нагрузки и удельные затраты ресурсов. В базовой сети только 46 ребер пригодны для транзитной передачи, а в кольцевой — 58, что только на 25% больше, однако позволяет на 30% увеличить суммарный межузловой поток и более чем на 50% медианное значение. Величина суммы межузловых потоков в базовой сети $\sum_{m=1}^M z_m^*(1) = 9400$, а в кольцевой $\sum_{m=1}^M z_m^*(1) = 12400$. Значение медианы в базовой сети $z^* = 1,5$, а в кольцевой $z^* = 2,3$.

Результаты расчетов на рис. 3 показывают, что для 80% пар межузловые потоки близки к медианному значению, а для 20% — в несколько раз превышают его. Для пар узлов, кратчайший путь между которыми состоит из одного или двух ребер,

межузловые потоки превышают медианное значение. Следует отметить, что при использовании *последовательного максиминного* правила распределения потоков и нагрузок близко расположенные корреспонденты также оказываются в привилегированном положении и получают преимущество при *уравнительных* способах *дележа* [1, 2].

5 Заключение

Предложенный метод позволяет проводить оценку и сравнение различных способов формирования сети, построенной на основе арендованных каналов связи при сохранении их общего числа. В рассмотренном примере исходные сети, представленные на рис. 1 имеют на всех ребрах почти равное число каналов. На рис. 2 представлены вторичные сети, схема построения которых позволяет выделить всем парам одинаковое число каналов. Кроме того, изменение структуры (графа) сети и переход от базовой к кольцевой сети позволяет значительно изменить межузловые потоки, хотя общее число каналов совпадает при одинаковых ресурсных ограничениях.

Литература

1. Малащенко Ю. Е., Назарова И. А. Оценка предельных распределений пропускной способности в многопользовательской сети при передаче межузловых потоков по кратчайшим маршрутам // Известия РАН. Теория и системы управления, 2022. № 5. С. 79–89.
2. Малащенко Ю. Е., Назарова И. А. Анализ распределения нагрузки и межузловых потоков при различных стратегиях маршрутизации в многопользовательской сети // Известия РАН. Теория и системы управления, 2022. № 6. С. 112–122.
3. Ogryczak W., Luss H., Pioro M., Nace D., Tomaszewski A. Fair optimization and networks: A survey // J. Appl. Math., 2014. Vol. 3. P. 1–25.
4. Salimifard K., Bigharaz S. The multicommodity network flow problem: State of the art classification, applications, and solution methods // Operational Research, 2020. Vol. 22. Iss. 2. P. 1–47.
5. Luss H. Equitable resource allocation: Models, algorithms, and applications. — Hoboken, NJ, USA: John Wiley & Sons, 2012. 420 p.
6. Йенсен П., Барнес Д. Потокосное программирование / Пер. с англ. — М.: Радио и связь, 1984. 392 с. (Jensen P. A., Barnes J. W. Network flow programming. — New York, NY, USA: Wiley, 1980. 408 p.)

Поступила в редакцию 13.10.22

ESTIMATES OF THE RESOURCE DISTRIBUTION IN THE MULTIUSER NETWORK WITH EQUAL INTERNODAL LOADS

Yu. E. Malashenko and I. A. Nazarova

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

Abstract: A method for estimating resources with an equalizing distribution of internodal loads in a multiuser network is proposed. Within the framework of a formal mathematical model, the capacity of edges is considered as components of a vector of resources that are required for the transmission of different types of flows. An algorithmic procedure for the redistribution and usage of capacity with equal quota of resources for all pairs is proposed. When searching for the corresponding edge loads, the values of the maximum single-product flows for each pair of nodes are determined. In the course of computational experiments, the total resource is considered to be set for networks with various structural features.

Keywords: multicommodity flow model; network resource distribution and internodal loads; network peak load

DOI: 10.14357/19922264230111

References

1. Malashenko, Yu. E., and I. A. Nazarova. 2022. Estimate of resource distribution with the shortest paths in the multiuser network. *J. Comput. Syst. Sci. Int.* 61(4):599–610.
2. Malashenko, Yu. E., and I. A. Nazarova. 2022. Analysis of load distribution and internodal flows under various routing strategies in the multiuser network. *J. Comput. Syst. Sci. Int.* 61(6):956–965.
3. Ogryczak, W., H. Luss, and M. Pioro. 2014. Fair optimization and networks: A survey. *J. Appl. Math.* 3:1–25.
4. Salimifard, K., and S. Bigharaz. 2020. The multicommodity network flow problem: State of the art classification, applications, and solution methods. *Operational Research* 22(2):1–47.
5. Luss, H. 2012. *Equitable resource allocation: Models, algorithms, and applications*. Hoboken, NJ: John Wiley & Sons. 420 p.
6. Jensen, P. A., and J. W. Barnes. 1980. *Network flow programming*. New York, NY: Wiley. 408 p.

Received October 13, 2022

Contributors

Malashenko Yuri E. (b. 1946) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

Nazarova Irina A. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

ОБ ОПТИМИЗАЦИИ РАБОТЫ РЕЗЕРВНОГО ПРИБОРА В МНОГОЛИНЕЙНОЙ СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ

Я. М. Агаларов¹

Аннотация: Рассматривается задача оптимизации работы (подключения или отключения) резервного прибора управляемой системы массового обслуживания (СМО) типа $G/M/s$. Она сформулирована в виде задачи нелинейного программирования со стоимостной целевой функцией, учитывающей плату за обслуживание заявок, затраты на техническое обслуживание резервного прибора, потери из-за задержки заявок в очереди и простоя резервного прибора. Процесс работы системы описан управляемой цепью Маркова, где состояние цепи определяется числом заявок в системе, а последовательность решений, принимаемых на каждом шаге цепи в зависимости от его состояния, определяет процесс управления резервным прибором. В качестве допустимого множества управлений рассмотрено множество стационарных стратегий управления цепью, решение в которой в каждом состоянии цепи о подключении или отключении резервного прибора принимается по длине очереди. Для случая пуассоновского входного потока доказана теорема о необходимых и достаточных условиях существования конечного оптимального порогового значения длины очереди и унимодальность целевой функции от порогового значения, предложен простой алгоритм решения задачи.

Ключевые слова: многолинейная система массового обслуживания; оптимизация; резервный прибор

DOI: 10.14357/19922264230112

1 Введение

Одной из практических задач, при решении которой исследователи в качестве математической модели используют многолинейные СМО с резервными приборами, является задача выбора оптимального режима подключения (отключения) резервного элемента системы с целью обеспечения требуемого качества работы системы при одновременном снижении затрат на ресурсы [1–3]. Математические постановки некоторых подобных задач и результаты их исследования приведены в работах [1–7]. Рассматриваемая в данной работе задача представляет собой обобщение задачи максимизации дохода СМО $G/M/1$ с резервным прибором, исследованной в работе [7] с помощью метода вложенных цепей Маркова, где переходы цепи определяются моментами поступления заявок, а состояние цепи — числом заявок, находящихся в системе в момент поступления. Подключение (отключение) резервного прибора системы, рассмотренной в работе [7], происходит по достижении длиной очереди порогового значения h_1 . Доход представлен функцией вида:

$$D(h_1) = \sum_{i=0}^{\infty} \frac{d_i^{h_1} \pi_i^{h_1}}{\bar{v}}, \quad (1)$$

где $\{\pi_i^{h_1}, 0 \leq i \leq \infty\}$ — стационарное распределение вероятностей цепи при пороге h_1 ($\pi_i^{h_1}$ — стационарная вероятность того, что цепь находится в состоянии i); \bar{v} — среднее время нахождения в состоянии i , $\bar{v} < (\mu + \gamma)^{-1}$, $\mu > 0$ — интенсивность обслуживания основным прибором, $\gamma > 0$ — интенсивность обслуживания резервным прибором; $d_i^{h_1}$ — средний доход, получаемый системой в состоянии i при пороге h_1 .

Доход $d_i^{h_1}$ вычисляется по формуле:

$$d_i^{h_1} = C_0 - C_1 \bar{T}_i^{h_1} - C_2 \left(\bar{v} - \bar{T}_{\text{пр},i}^{h_1} \right) - C_3 \bar{T}_{\text{пр},i}^{h_1}, \quad 0 \leq i, \quad (2)$$

где $\bar{T}_i^{h_1}$ — среднее суммарное время задержки заявок в очереди за время нахождения системы в состоянии i ; $\bar{T}_{\text{пр},i}^{h_1}$ — среднее время простоя резервного прибора в состоянии i ; $C_0 \geq 0$ — плата, получаемая системой, если поступившая заявка обслужена системой (принята в накопитель); $C_1 \geq 0$ — потери в единицу времени за ожидание заявки в системе; $C_2 \geq 0$ — потери на амортизацию резервного прибора в единицу времени, когда он подключен к системе и занят заявкой; $C_3 \geq 0$ — потери в единицу времени простоя ($C_3 \leq C_2$).

Предполагается, что доход системы в единицу времени при любом значении порога имеет некоторое конечное значение.

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, agglar@yandex.ru

Основные результаты работы [7] — доказательство унимодальности целевой функции D по h_1 и вывод условий существования конечного оптимального порогового значения в случае СМО $G/M/1$.

В данной работе показано, что аналогичные результаты имеют место и для многолинейной СМО с резервным прибором.

2 Постановка задачи и метод решения

Рассматривается СМО типа $G/M/s$ с накопителем бесконечной емкости, s основными однотипными приборами и одним резервным прибором обслуживания, на которую поступает рекуррентный поток заявок с функцией распределения вероятностей $V(t)$. Время обслуживания заявки на основном приборе распределено по экспоненциальному закону с параметром $\mu > 0$, а на резервном — с параметром $\gamma > 0$. Резервный прибор подключается к системе в момент времени, когда в системе число заявок становится равным пороговому значению h_1 , и отключается от системы, когда число заявок становится равным $(h_1 - 1)$, причем если это происходит в момент завершения обслуживания очередной заявки на основном приборе, то заявка на резервном приборе перед его отключением переходит на освободившийся прибор и заново обслуживается. По завершении обслуживания заявка уходит из системы, освободив одновременно прибор и занимаемое место в накопителе. Пусть, как и в [7], показатель эффективности работы системы представлен функцией дохода вида (1). Ставится задача целочисленного программирования следующего вида: найти порог h^* , такой что

$$\max_{s < h_1} D(h_1) = D(h^*),$$

где $D(h_1)$ определена в (1).

Подход, применяемый ниже к решению данной задачи, аналогичен подходу, изложенному в [7], суть которого заключается в проверке выполнения условий теоремы 1 из [8] для целевой функции задачи для СМО с ограниченным накопителем.

Пусть данная задача рассматривается для системы $G/M/s/h_2$ с ограниченным накопителем емкости h_2 , $h_2 = h_1 + a$, $h_1 \geq s + 1$, $a = const \geq 0$. Пусть заявка, которая застаёт накопитель полностью занятой, теряется и в этом случае система не получает плату за обслуживание. Ставится задача: найти порог h^* , такой что

$$\max_{s < h_1} D(h_1, a) = D(h^*, a). \quad (3)$$

Здесь

$$D(h_1, a) = \sum_{i=0}^{h_2} d_i^{h_1} \pi_i^{h_1},$$

где

$$d_i^{h_1} = \begin{cases} C_0 - C_1 \bar{T}_i^{h_1} - C_2 (\bar{v} - \bar{T}_{\text{пр},i}^{h_1}) - C_3 \bar{T}_{\text{пр},i}^{h_1}, & 0 \leq i \leq h_2 - 1; \\ d_{h_2-1}^{h_1} - C_0, & i = h_2; \end{cases}$$

$d_i^{h_1}$, $\pi_i^{h_1}$, C_0 , C_1 , C_2 , C_3 , \bar{v} , $\bar{T}_i^{h_1}$ и $\bar{T}_{\text{пр},i}^{h_1}$ — обозначения, использованные в разд. 1.

Обозначим:

$$\mu_i^{h_1} = \begin{cases} s\mu & \text{при } s - 1 \leq i \leq h_1 - 2; \\ s\mu + \gamma & \text{при } h_1 \leq i \leq h_2; \end{cases}$$

$r_{i,m}^{h_1}$ — вероятность того, что в состоянии i будут обслужены ровно m заявок при условии, что в этом состоянии не произойдет отключение резервного прибора; $a_{i,m}^{h_1}$ — вероятность того, что в состоянии i будут обслужены ровно m заявок при условии, что в этом состоянии произойдет отключение резервного прибора; q_{ij} — вероятность того, что в состоянии i будут обслужены ровно $(i + 1 - j)$ заявок при условиях $0 \leq i \leq s - 1$ и $0 \leq j \leq i + 1$; b_{ij} — вероятность того, что в состоянии i будут обслужены ровно $(i + 1 - j)$ заявок при условиях $h_1 - 1 \leq i \leq h_2$ и $0 \leq j \leq s - 1$.

Справедливы формулы:

$$r_{i,m}^{h_1} = \int_0^\infty \frac{(\mu_i^{h_1} v)^m}{m!} e^{-\mu_i^{h_1} v} dV(v),$$

$$h_1 - 1 \leq i \leq h_2, \quad 0 \leq m \leq i - h_1 + 1$$

или $s - 1 \leq i \leq h_1 - 2, \quad 0 \leq m \leq h_1 - s - 1;$ (4)

$$a_{i,m}^{h_1} = \int_0^\infty \int_0^v \frac{\mu_i^{h_1} (\mu_i^{h_1} t)^{i-h_1+1}}{(i - h_1 + 1)!} e^{-\mu_i^{h_1} t} \times$$

$$\times \frac{[\mu_1(v - t)]^{m-i+h_1-2}}{(m - i + h_1 - 2)!} e^{\mu_1(v-t)} dt dV(v),$$

$$h_1 - 1 \leq i \leq h_2, \quad s - 1 \leq j \leq h_1 - 1; \quad (5)$$

$$q_{ij} = \int_0^\infty \binom{i+1}{j} [1 - e^{-\mu v}]^{i+1-j} e^{-j\mu v} dV(v),$$

$$0 \leq i \leq s - 1, \quad 0 \leq j \leq i + 1; \quad (6)$$

$$p_{ij}^{h_1} = \int_0^\infty \binom{s}{j} e^{-\mu v j} \left[\int_0^v \frac{(s\mu t)^{i-s}}{(i-s)!} (e^{-\mu t} - e^{-\mu v})^{s-j} s\mu dt \right] dV(v), \quad s \leq i \leq h_1 - 2, \\ 0 \leq j \leq s - 1; \quad (7)$$

$$b_{ij}^{h_1} = \int_0^\infty \binom{s}{j} \int_0^v \int_0^y \frac{\mu_i^{h_1} (\mu_i^{h_1} t)^{i-h_1+1}}{(i-h_1+1)!} e^{-\mu_i^{h_1} t} \times \\ \times \frac{[s\mu(y-t)]^{h_1-s-2}}{(h_1-s-2)!} e^{-s\mu(y-t)} s\mu dt e^{-\mu(v-y)j} \times \\ \times (1 - e^{-\mu(v-y)})^{s-j} dy dV(v) = \\ = \int_0^\infty \binom{s}{j} e^{-\mu v} \int_0^v \int_0^y \frac{\mu_i^{h_1} (\mu_i^{h_1} t)^{i-h_1+1}}{(i-h_1+1)!} e^{\mu_2 t} \times \\ \times \frac{[s\mu(y-t)]^{h_1-s-2}}{(h_1-s-2)!} (e^{-\mu y} - e^{\mu v})^{s-j} s\mu dt dy dV(v), \\ h_1 - 1 \leq i \leq h_2, \quad 0 \leq j \leq s - 1. \quad (8)$$

Из системы уравнений равновесия для вложенной цепи Маркова и выражений (4)–(8) для параметров $r_{i,m}^{h_1}$, $a_{i,m}^{h_1}$, q_{ij} , $p_{ij}^{h_1}$ и $b_{ij}^{h_1}$ следует справедливость следующих рекуррентных формул для стационарных вероятностей состояний $\pi_j^{h_1}$, $0 \leq j \leq h_2$:

$$\pi_j^{h_1} = \frac{R_j^{h_1}}{\sum_{i=0}^{h_2} R_i^{h_1}}, \quad 0 \leq j \leq h_2, \quad (9)$$

где

$$R_{h_2}^{h_1} = 1; \quad R_{h_2-1}^{h_1} = \frac{1 - r_{h_2,0}^{h_1}}{r_{h_2,0}^{h_1}};$$

$$R_{j-1}^{h_1} = \frac{R_j^{h_1} (1 - r_{j,1}^{h_1})}{r_{j-1,0}^{h_1}} - \\ - \frac{\sum_{i=j+1}^{h_2-1} R_i^{h_1} r_{j,i+1-j}^{h_1} + R_{h_2}^{h_1} r_{h_2,h_2-j}^{h_1}}{r_{j-1,0}^{h_1}}, \\ h_1 \leq j \leq h_2 - 1;$$

$$R_{h_1-2}^{h_1} = \frac{R_{h_1-1}^{h_1} (1 - r_{h_1-1,1})}{r_{h_1-2,0}^{h_1}} - \\ - \frac{\sum_{i=h_1}^{h_2-1} R_i^{h_1} r_{i,i+2-h_1}^{h_1} + R_{h_2}^{h_1} r_{h_2,h_2-h_1+1}}{r_{h_1-2,0}^{h_1}};$$

$$R_{j-1}^{h_1} = \frac{R_j^{h_1} (1 - r_{j,1})}{r_{j-1,0}^{h_1}} - \\ - \frac{\sum_{i=j+1}^{h_1-2} R_i^{h_1} r_{i,i+1-j}^{h_1} + \sum_{i=h_1-1}^{h_2-1} R_i^{h_1} a_{i,i+1-j}}{r_{j-1,0}^{h_1}} - \\ - \frac{R_{h_2}^{h_1} a_{h_2,h_2-j}}{r_{j-1,0}^{h_1}}, \quad s \leq j \leq h_1 - 2;$$

$$R_{j-1}^{h_1} = \frac{R_j^{h_1} (1 - q_{ij})}{q_{j-1,j}} - \\ - \frac{\sum_{i=j+1}^{s-1} R_i^{h_1} q_{ij} + \sum_{i=s}^{h_1-2} R_i^{h_1} p_{ij}^{h_1}}{q_{j-1,j}} - \\ - \frac{\sum_{i=h_1-1}^{h_2-1} R_i^{h_1} b_{ij}^{h_1}}{q_{j-1,j}} - \frac{R_{h_2}^{h_1} b_{h_2,j}^{h_1}}{q_{j-1,j}}, \quad 1 \leq j \leq s - 1.$$

Из (9) следует равенство:

$$\pi_{i+1}^{k+1} = A_{k+1} \pi_i^k, \quad i = s - 1, \dots, k. \quad (10)$$

Здесь

$$A_{k+1} = \frac{1 - Q_{s-1}^{k+1}}{1 - Q_{s-2}^k},$$

где

$$Q_{s-1}^{k+1} = \sum_{i=0}^{s-1} \pi_i^{k+1}; \quad Q_{s-2}^k = \begin{cases} 0, & 1 \leq s \leq 2; \\ \sum_{i=0}^{s-2} \pi_i^k, & s \geq 2. \end{cases}$$

В дальнейшем для краткости изложения будем пользоваться обозначениями:

$$r_{i,m}^{h_1}(v) = \frac{(\mu_i^{h_1} v)^m}{m!} e^{-\mu_i^{h_1} v};$$

$$z_i^{h_1}(v, t, m) = \frac{\mu_i^{h_1} (\mu_i^{h_1} t)^{i-h_1+1}}{(i-h_1+1)!} e^{-\mu_i^{h_1} t} \times \\ \times \frac{[\mu_1(v-t)]^{m-i+h_1-2}}{(m-i+h_1-2)!} e^{-\mu_1(v-t)}.$$

Ниже приведем ряд формул, доказательства справедливости которых для краткости изложения не приводим (они аналогичны доказательствам, приведенным в работе [7] для подобных формул).

Среднее значение дохода (2), получаемого системой с ограниченным накопителем при пороге h_1 в состоянии i , равно

$$d_i^{h_1} = \begin{cases} C_0 - C_1 \bar{T}_i^{h_1} - C_2 (\bar{v} - \bar{T}_{\text{пр},i}^{h_1}) - C_3 \bar{T}_{\text{пр},i}^{h_1}, & 0 \leq i \leq h_2 - 1; \\ d_{h_2-1}^{h_1} - C_0, & i = h_2, \end{cases} \quad (11)$$

где $\bar{T}_i^{h_1}$ и $\bar{T}_{np,i}^{h_1}$ определены в (2) и в данном случае для них справедливы формулы:

$$\bar{T}_i^{h_1} = \begin{cases} \frac{1}{\mu_i^{h_1}} \left[(i+1-s) \sum_{m=1}^{i+2-s} mr_{i,m}^{h_1} - \frac{1}{2} \sum_{m=1}^{i+2-s} (m-1)mr_{i,m}^{h_1} + \frac{1}{2} (i+1-s)(i+2-s) \sum_{m=i+3-s}^{\infty} r_{i,m}^{h_1} \right] \\ \text{при } s \leq i \leq h_1 - 2 \\ \text{и } \bar{T}_i^{h_1} = 0 \text{ при } i \leq s - 1; \\ \frac{1}{\mu_i^{h_1}} \left[(i+1-s) \sum_{m=1}^{i-h_1+3} mr_{i,m}^{h_1} - \frac{1}{2} \sum_{m=1}^{i-h_1+3} (m-1)mr_{i,m}^{h_1} \right] + \\ + \sum_{m=i-h_1+3}^{i+1-s} \int_0^v \int_0^v \left[\frac{i-h_1+1}{2} t + \right. \\ \left. + (m-i+h_1-1)t + \frac{1}{2}(m-i+h_1-2) \times \right. \\ \left. \times (v-t) \right] z_i^{h_1}(v,t,m) dt dV(v) + \\ + \sum_{m=i-h_1+3}^{i-s} \int_0^v \int_0^v (i+1-m-s)v \times \\ \times z_i^{h_1}(v,t,m) dt dV(v) + \\ + \sum_{m=i+2-s}^{\infty} \int_0^v \int_0^v \left[\frac{i-h_1+1}{2} t + (h_1-1)t + \right. \\ \left. + \frac{(h_1-2)(h_1-1)(v-t)}{2(m-i+h_1-1)} \right] \times \\ \left. \times z_i^{h_1}(v,t,m) dt dV(v) \text{ при } i \geq h_1 - 1; \right. \end{cases} \quad (12)$$

$$\bar{T}_{np,i}^{h_1} = \begin{cases} \bar{v} - \int_0^v \int_0^v t \frac{\mu_i^{h_1} (\mu_i^{h_1} t)^{i-h_1+1}}{(i-h_1+1)!} e^{-\mu_i^{h_1} t} dt dV(v), \\ \quad \quad \quad i \geq h_1 - 1; \\ \bar{v}, \quad \quad \quad 0 \leq i \leq h_1 - 2. \end{cases} \quad (13)$$

Среднее время отсутствия очереди при пороге h_1 в состоянии i равно

$$\bar{T}_{оч,i}^{h_1} = \begin{cases} \sum_{m=i+1}^{\infty} \int_0^v \int_0^v \left[v-t - \frac{(h_1-1)(v-t)}{m-i+h_1-1} \right] \times \\ \times z_i^{h_1}(v,t,m) dt dV(v), \quad i \geq h_1 - 1; \\ \frac{1}{\mu_1} \sum_{m=i+2}^{\infty} (m-i-1)r_{i,m}^{h_1}, \quad i \leq h_1 - 2. \end{cases}$$

Из (11)–(13) после несложных преобразований получаем равенства:

$$d_{i+1}^{h_1+1} - d_i^{h_1} = \begin{cases} -\frac{C_1}{\mu_i^{h_1}} \sum_{m=1}^{i-s+2} mr_{i,m}^{h_1} - \\ - \frac{C_1(i+2-s)}{\mu_i^{h_1}} \sum_{m=i-s+3}^{\infty} r_{i,m}^{h_1} \\ \text{при } s-1 \leq i \leq h_1 - 2; \\ -C_1 \int_0^v v \left[\sum_{m=0}^{i-h_1+2} r_{i,m}^{h_1}(v) + \right. \\ \left. + \sum_{m=i-h_1+3}^{i-s+1} \int_0^v z_i^{h_1}(v,t,m) dt \right] dV(v) - \\ - C_1 \sum_{m=i-s+2}^{\infty} \int_0^v \int_0^v \left[t + \frac{(h_1-1)(v-t)}{m-i+h_1-1} \right] \times \\ \times z_i^{h_1}(v,t,m) dt dV(v) \text{ при } i \geq h_1 - 1. \end{cases} \quad (14)$$

Введем обозначения:

$$w(h_1, a) = \sum_{i=0}^{h_1-2} \pi_i^{h_1} \frac{1}{\mu_i^{h_1}} \sum_{m=i-s+3}^{\infty} (m-i-1)r_{i,m}^{h_1} + \\ + \sum_{i=h_1-1}^{h_2-1} \pi_i^{h_1} \sum_{m=i-s+2}^{\infty} \int_0^v \left[v - \int_0^v \left[t + \frac{(h_1-1)(v-t)}{m-i+h_1-1} \right] \times \right. \\ \left. \times z_i^{h_1}(v,t,m) dt \right] dV(v) + \\ + \pi_{h_2}^{h_1} \sum_{m=h_2}^{\infty} \int_0^v \left[v - \int_0^v \left[t + \frac{(h_1-1)(v-t)}{m-h_2+h_1} \right] \times \right. \\ \left. \times z_i^{h_1}(v,t,m) dt \right] dV(v); \quad (15)$$

$$f(h_1, a) = C_0 - C_3 \bar{v} - \\ - C_1 \frac{A_{h_1+1}}{1 - A_{h_1+1}} [\bar{v} - w(h_1, a)]. \quad (16)$$

Для разности доходов при порогах h_1 и $h_1 + 1$ выполняются равенства (следуют из формул (1), (3), (10), (14)–(16)):

$$D(h_1, a) - D(h_1 + 1, a) = \\ = \sum_{i=0}^{h_2} \pi_i^{h_1} d_i^{h_1} - \sum_{i=0}^{h_2+1} \pi_i^{h_1+1} d_i^{h_1+1} = \\ = (C_0 - C_3 \bar{v}) (Q_{s-2}^{h_1} - Q_{s-1}^{h_1+1}) + \\ + \sum_{i=s-1}^{h_2} \pi_i^{h_1} d_i^{h_1} - \sum_{i=s}^{h_2+1} \pi_i^{h_1+1} d_i^{h_1+1} =$$

$$\begin{aligned}
 &= (C_0 - C_3\bar{v}) \left(Q_{s-2}^{h_1} - Q_{s-1}^{h_1+1} \right) + \\
 &+ \frac{Q_{s-1}^{h_1+1} - Q_{s-2}^{h_1}}{1 - Q_{s-2}^{h_1}} \left[\sum_{i=0}^{h_2} \pi_i^{h_1} d_i^{h_1} - (C_0 - C_3\bar{v}) Q_{s-2}^{h_1} \right] - \\
 &- \frac{1 - Q_{s-1}^{h_1+1}}{1 - Q_{s-2}^{h_1}} \sum_{i=s-1}^{h_2} \pi_i^{h_1} \left(d_{i+1}^{h_1+1} - d_i^{h_1} \right) = \\
 &= (1 - A_{h_1+1}) \left[D(h_1, a) - (C_0 - C_3\bar{v}) - \right. \\
 &- \frac{A_{h_1+1}}{1 - A_{h_1+1}} \left(\sum_{i=s-1}^{h_2-1} \pi_i^{h_1} \left(d_{i+1}^{h_1+1} - d_i^{h_1} \right) + \right. \\
 &\quad \left. \left. + \pi_{h_2}^{h_1} \left(d_{h_2}^{h_1+1} - d_{h_2-1}^{h_1} \right) \right) \right] = \\
 &= (1 - A_{h_1+1}) [D(h_1, a) - f(h_1, a)].
 \end{aligned}$$

Как показывают многочисленные вычислительные эксперименты, проведенные при различных значениях параметров СМО, величина A_{h_1+1} возрастает по h_1 . Ниже приведем доказательство этого положения для случая пуассоновского входного потока. Отметим, что доказательство справедливости такого свойства A_{h_1+1} в случае произвольного рекуррентного входного потока остается открытым.

Для случая пуассоновского входного потока процесс работы системы описывается случайным процессом рождения и гибели, для стационарных вероятностей состояний (числа заявок в системе) которого справедлива формула:

$$\pi_i^{h_1} = \begin{cases} \frac{\rho^i}{i!} \pi_0^{h_1}, & 0 \leq i \leq s; \\ \frac{\rho^s}{s!} \left(\frac{\rho}{s} \right)^{i-s} \pi_0^{h_1}, & s < i \leq h_1 - 1; \\ \frac{\rho^s}{s!} \left(\frac{\rho}{s} \right)^{h_1-s-1} \left(\frac{\rho}{s+\alpha} \right)^{i-h_1+1} \pi_0^{h_1}, & h_1 \leq i \leq h_2, \end{cases}$$

где

$$\begin{aligned}
 \pi_0^{h_1} &= \left[\sum_{j=0}^s \frac{\rho^j}{j!} + \frac{\rho^s}{s!} \sum_{j=1}^{h_1-s-1} \left(\frac{\rho}{s} \right)^j + \right. \\
 &\left. + \frac{\rho^s}{s!} \left(\frac{\rho}{s} \right)^{h_1-s-1} \sum_{j=1}^{h_2-h_1-1} \left(\frac{\rho}{s+\alpha} \right)^j \right]^{-1}, \quad \alpha = \frac{\gamma}{\mu}.
 \end{aligned}$$

Рассмотрим разность $(A_{h_1+2} - A_{h_1+1})$, представимую в виде:

$$\begin{aligned}
 &\frac{\pi_0^{h_1+2} \left(\pi_0^{h_1+2} \right)^{-1} \left(1 - Q_{s-1}^{h_1+2} \right)}{\pi_0^{h_1+1} \left(\pi_0^{h_1+1} \right)^{-1} \left(1 - Q_{s-2}^{h_1+1} \right)} - \\
 &- \frac{\pi_0^{h_1+1} \left(\pi_0^{h_1+1} \right)^{-1} \left(1 - Q_{s-1}^{h_1+1} \right)}{\pi_0^{h_1} \left(\pi_0^{h_1} \right)^{-1} \left(1 - Q_{s-2}^{h_1} \right)}. \quad (17)
 \end{aligned}$$

Выполняется равенство:

$$\begin{aligned}
 &\left(\pi_0^{h_1+2} \right)^{-1} \left(1 - Q_{s-1}^{h_1+2} \right) \left(\pi_0^{h_1} \right)^{-1} \left(1 - Q_{s-2}^{h_1} \right) = \\
 &= \left(\pi_0^{h_1+1} \right)^{-1} \left(1 - Q_{s-1}^{h_1+1} \right) \pi_0^{h_1+1} \times \\
 &\quad \times \left(\pi_0^{h_1+1} \right)^{-1} \left(1 - Q_{s-2}^{h_1+1} \right).
 \end{aligned}$$

Отсюда следует, что знак выражения (17) такой же, что и у разности $\pi_0^{h_1+2} \pi_0^{h_1} - \left(\pi_0^{h_1+1} \right)^2$. Находим знак выражения $\left(\pi_0^{h_1+2} \pi_0^{h_1} \right)^{-1} - \left(\pi_0^{h_1+1} \right)^{-2}$, противоположный знаку выражения $\pi_0^{h_1+2} \pi_0^{h_1} - \left(\pi_0^{h_1+1} \right)^2$. Проведя упрощающие преобразования, находим:

$$\begin{aligned}
 &\left(\pi_0^{h_1+2} \pi_0^{h_1} \right)^{-1} - \left(\pi_0^{h_1+1} \right)^{-2} = \\
 &= \left(\sum_{j=0}^s \frac{\rho^j}{(j-1)!s} - \sum_{j=0}^s \frac{\rho^j}{j!} \right) (E_{h_1+1} - E_{h_1}),
 \end{aligned}$$

где

$$\begin{aligned}
 E_{h_1} &= \frac{\rho^s}{s!} \sum_{j=1}^{h_1-s-1} \left(\frac{\rho}{s} \right)^j + \\
 &+ \frac{\rho^s}{s!} \left(\frac{\rho}{s} \right)^{h_1-s-1} \sum_{j=1}^{h_2-h_1-1} \left(\frac{\rho}{s+\alpha} \right)^j.
 \end{aligned}$$

Как видим, первая скобка в правой части имеет отрицательный знак. Для второй скобки имеем:

$$\begin{aligned}
 E_{h_1+1} - E_{h_1} &= \frac{\rho^{s+1}}{s!s} \left(\frac{\rho}{s} \right)^{h_1-s-1} \times \\
 &\times \left[\left(\frac{\rho}{s} \right)^{h_2-h_1-1} + \frac{\rho/s + \rho/(s+\alpha)}{1 + \rho/s} \right] > 0.
 \end{aligned}$$

Следовательно, $A_{h_1+2} - A_{h_1+1} > 0$, так как $\left(\pi_0^{h_1+2} \pi_0^{h_1} \right)^{-1} - \left(\pi_0^{h_1+1} \right)^{-2} < 0$ (или $\pi_0^{h_1+2} \pi_0^{h_1} - \left(\pi_0^{h_1+1} \right)^2 > 0$), т. е. A_{h_1+1} возрастает по h_1 .

Из этого свойства A_{h_1+1} и из того, что среднее время отсутствия очереди также убывает по h_1 , следует, что $f(h_1, a)$ убывает по h_1 . Следовательно, функция $D(h_1, a)$ удовлетворяет всем условиям теоремы 1 из [8]. Далее, рассуждая точно так же, как и в работе [7], находим, что условия теоремы 1 из [8]

выполняются и в случае бесконечного накопителя и функция $D(h_1)$ унимодальна по пороговому значению $h_1 > s$.

В заключение отметим также следующий результат, вытекающий из свойства унимодальности функции $D(h_1)$: если существует конечный оптимальный порог, то для его поиска (решения исходной задачи) достаточно применить следующий простой алгоритм.

1. Положить $h_1 = s + 1$.
2. До тех пор пока выполняется условие

$$D(h_1 + 1) > D(h_1),$$

полагать $h_1 = h_1 + 1$.

3. Положить $h_1^* = h_1$.

Литература

1. Горцев А. М. Система массового обслуживания с произвольным числом резервных каналов и гистерезисным управлением включением и выключением резервных каналов // Автоматика и телемеханика, 1977. Вып. 10. С. 30–37.
2. Крылова Д. С., Головкин Н. И., Жук Т. А. Анализ СМО с резервным прибором и скачкообразной интенсивностью входного потока // Вестник ВГУ. Сер. Физика. Математика, 2017. № 4. С. 109–123.
3. Клименок В. И. Многолинейная система массового обслуживания с резервными приборами // Ж. Белорусского государственного университета. Математика. Информатика, 2019. № 3. С. 57–70.
4. Дудин А. Н. О задаче оптимального управления многоскоростной системой массового обслуживания // Автоматика и телемеханика, 1980. Вып. 9. С. 43–51
5. Самочернова Е. С., Петров Л. И. Оптимизация системы массового обслуживания с однотипным резервным прибором // Известия Томского политехнического университета, 2010. Т. 317. № 5. С. 28–31.
6. Агаларов Я. М. Оптимизация порогового управления переключением скорости обслуживания в системе массового обслуживания $G/M/1$ // Информатика и её применения, 2022. Т. 16. Вып. 1. С. 73–81.
7. Агаларов Я. М. Оптимальное управление подключением резервного прибора в системе массового обслуживания $G/M/1$ // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 34–41.
8. Агаларов Я. М. Признак унимодальности целочисленной функции одной переменной // Обзорение прикладной и промышленной математики, 2019. Т. 26. Вып. 1. С. 65–66.

Поступила в редакцию 05.09.22

OPTIMIZATION OF A QUEUE-LENGTH DEPENDENT ADDITIONAL SERVER IN THE MULTISERVER QUEUE

Ya. M. Agalarov

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 optimal control of an additional server in a stationary $G/M/s$ queue is considered. The additional server can be turned on and off at instants when the queue length is changed. It is formulated as the nonlinear optimization problem, in which the objective function accounts for amounts for service, losses due to the waiting of customers, maintenance, and downtime of the additional server. The functioning of the system is described as a controlled Markov chain. Only stationary control policies are considered. For Poisson arrivals, necessary and sufficient conditions are given for the existence of the optimal decision point (threshold) and it is proved that the objective function is unimodal. A simple algorithm for the computation of the threshold is provided.

Keywords: multiserver queuing system; optimization; additional server

DOI: 10.14357/19922264230112

References

1. Gortsev, A. M. 1978. A queueing system with an arbitrary number of stand-by channels and hysteresis control of their connection and disconnection. *Automat. Rem. Contr.* 38(10):1451–1457.
2. Krylova, D. S., N. I. Golovko, and T. A. Zhuk. 2017. Analiz SMO s rezervnym priborom i skachkoobraznoy intensivnost'yu vkhodnogo potoka [Analysis of SMO backup device and the abrupt intensity of the input stream]. *Vestnik VGU. Ser. Fizika. Matematika* [Proceedings of Voronezh State University. Ser. Physics. Mathematics] 4:109–123.
3. Klimenok, V. I. 2019. Mnogolineynaya sistema massovogo obsluzhivaniya s rezervnymi priborami [Multi-server queueing system with reserve servers]. *J. Belarusian State University. Mathematics Informatics* 3:57–70.

4. Dudin, A. N. 1981. On optimal control of a multi-rate service system. *Automat. Rem. Contr.* 41(9):1221–1228.
5. Samochnova, E. S., and L. I. Petrov. 2010. Optimizatsiya sistemy massovogo obsluzhivaniya s odnotipnym rezervnym priborom [Optimization of the queuing system with the same type of backup device]. *Bulletin Tomsk Polytechnic University* 317(5):28–31.
6. Agalarov, Ya. M. 2022. Optimizatsiya porogovogo upravleniya pereklyucheniem skorosti obsluzhivaniya v sisteme massovogo obsluzhivaniya $G/M/1$ [Optimization of the threshold service speed control in the $G/M/1$ queue]. *Informatika i ee Primeneniya — Inform. Appl.* 16(1):73–81.
7. Agalarov Ya. M. 2022. Optimal'noe upravlenie podklyucheniem rezervnogo pribora v sisteme massovogo obsluzhivaniya $G/M/1$ [Optimal control of a queue-length dependent additional server in $GI/M/1$ queue]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):34–41.
8. Agalarov, Ya. M. 2019. Priznak unimodal'nosti tselochislennoy funktsii odnoy peremennoy [A sign of unimodality of an integer function of one variable]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Surveys on Applied and Industrial Mathematics] 26(1):65–66.

Received September 5, 2022

Contributor

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

АНАЛИЗ СХЕМЫ ДОСТУПА С ПРЕРЫВАНИЕМ ПРИ НАРЕЗКЕ РАДИОРЕСУРСОВ СЕТИ ПЯТОГО ПОКОЛЕНИЯ*

К. И. Б. Аду¹, Е. В. Маркова², Ю. В. Гайдамака³, С. Я. Шоргин⁴

Аннотация: Активно исследуемая в последние годы технология «нарезки радиоресурсов сети» (NS — Network Slicing), основанная на представлении общей сетевой инфраструктуры в виде различных настраиваемых логических сетей, называемых слайсами, предполагает разделение операторов мобильной сети на две группы — провайдеры физической сетевой инфраструктуры InPs (Infrastructure Providers) и операторы мобильной виртуальной сети (MVNOs — Mobile Virtual Network Operators). Последние арендуют физические ресурсы InPs для создания собственных слайсов с целью предоставления своим пользователям услуг с различными требованиями к качеству обслуживания. В статье для сети с технологией NS предложена схема доступа к радиоресурсам сети, предоставляющей пользователям услуги с гарантированной скоростью передачи данных (GBR — Guaranteed Bit Rate) и приоритетным управлением, основанным на реализации механизма прерывания обслуживания пользователей. Для оценки эффективности предлагаемой схемы проведен сравнительный анализ ее характеристик с характеристиками схемы доступа, основанной на механизме резервирования ресурсов.

Ключевые слова: 5G; нарезка сети; качество обслуживания; ключевые показатели эффективности; приоритетное управление; прерывание обслуживания; итерационный метод

DOI: 10.14357/19922264230113

1 Введение

Ввиду ограниченности спектра частотного диапазона мобильных сетей для предоставления пользователям услуг с требуемым качеством обслуживания необходимо внедрение новых технологий. В последние годы активно исследуется технология «нарезки радиоресурсов сети» NS, основанная на представлении общей сетевой инфраструктуры в виде различных настраиваемых логических сетей, называемых слайсами. Одной из важнейших проблем реализации технологии NS является проблема эффективного распределения радиоресурсов — полосы пропускания или физических ресурсных блоков PRB (Physical Resource Block). Радиоресурсы должны быть распределены между несколькими слайсами в соответствии с динамически меняющимися требованиями пользователей — операторов мобильной виртуальной сети MVNOs (Mobile Virtual Network Operators), при этом должны быть выполнены ключевые требования к изоляции слайсов [1, 2], в частности трафик, обслуживаемый в рамках одного слайса, не должен оказывать негативного влияния на трафик, обслуживаемый в других слайсах. Например, в работах [3, 4] для

реализации межслайсовой изоляции предложены подходы, основанные на резервировании ресурсов.

В данной работе рассмотрена модель схемы распределения ресурсов базовой станции (БС) соты сети между несколькими слайсами, пользователям которых предоставляются услуги реального времени с GBR, например видео- и голосовая телефония. Особенность предложенной модели заключается в управлении доступом RAC (Radio Admission Control), основанном на реализации совместного доступа к части имеющихся ресурсов и введении приоритетного обслуживания. Для реализации межслайсовой изоляции приоритетное управление использует механизм прерывания обслуживания пользователей, при этом в статье разработан итерационный метод расчета числа прерванных запросов. Чтобы оценить эффективность предлагаемой схемы доступа с прерыванием обслуживания пользователей (далее — ПС, схема с прерыванием), проведен сравнительный анализ ее основных ключевых показателей эффективности (КПЭ) с КПЭ схемы доступа, основанной на реализации механизма резервирования ресурсов (далее — РС, схема с резервированием) [5, 6].

* Исследование выполнено за счет гранта Российского научного фонда № 22-79-10053, <https://rscf.ru/project/22-79-10053/>.

¹ Российский университет дружбы народов, adu-k@rudn.ru

² Российский университет дружбы народов, markova-ev@rudn.ru

³ Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaydamaka-yuv@rudn.ru

⁴ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

2 Описание модели

Рассмотрим работу БС соты сети, принадлежащей одному провайдеру InP и имеющей емкость C . Емкость БС используется несколькими мобильными операторами MVNOs, предоставляющими услуги своим пользователям, при этом под каждую услугу оператору MVNO выделяется так называемый «слайс» — часть от общей емкости C . Обозначим множество слайсов через \mathcal{S} , $\mathcal{S} \subset \mathbb{N} \setminus \{0\}$, $\mathbb{N} = \{0, 1, 2, \dots\}$. Для каждого слайса s определена максимальная емкость C_s , причем $C_s \leq C$ и $\sum_{s \in \mathcal{S}} C_s \geq C$. Емкость C_s слайса s включает в себя гарантированную часть емкостью Q_s и общедоступную часть $C_s - Q_s$, при этом $Q_s \leq C_s$ и $\sum_{s \in \mathcal{S}} Q_s \leq C$. Поток запросов пользователей на предоставление услуги s моделируется с помощью пуассоновского потока запросов типа s с интенсивностью λ_s , $s \in \mathcal{S}$.

Основные обозначения представлены в табл. 1. Указанные параметры позволяют реализовать схему ПС управления доступом с механизмом прерывания обслуживания пользователей. Отличие предлагаемой схемы ПС управления доступом с механизмом прерывания от классической неполнодоступной схемы с потолками в общей части ресурса (Sharing with Maximum Queue Length and Minimum Allocation) [7–9] заключается в отсутствии индивидуальной зоны для запросов типа s , $s \in \mathcal{S}$, в которую не допускаются запросы других типов. Отметим, что в гарантированной части незагруженного слайса s , $s \in \mathcal{S}$, может начать обслуживаться принятый

в систему запрос произвольного типа \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$, при этом с точки зрения слайса s запрос типа \hat{s} становится так называемым «нарушителем». Если при последующем поступлении запроса типа s число обслуживаемых в этом слайсе запросов окажется меньше гарантированного значения $\lfloor Q_s/b_s \rfloor$, а объем доступного ресурса БС меньше требуемого b_s , то запрос будет принят на обслуживание за счет прерывания обслуживания одного или нескольких запросов-«нарушителей» типа \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$. Для реализации механизма прерывания каждому слайсу присвоен приоритет в обслуживании, что также отличает предложенную схему ПС от классической неполнодоступной схемы с потолками в общей части ресурса. В предположении о различающихся приоритетах у слайсов перенумеруем слайсы в порядке убывания приоритета, т. е. высший приоритет в обслуживании получают запросы, находящиеся в слайсе с номером «1», низший — в слайсе с номером « $|\mathcal{S}|$ ». Введем вектор-функцию прерывания

$$\mathbf{z}(s, \mathbf{n}) = (z_{\hat{s}}(s, \mathbf{n})) = (z_1(s, \mathbf{n}), \dots, z_{|\mathcal{S}|}(s, \mathbf{n})),$$

определяющую число обслуживаемых запросов слайса \hat{s} , $\hat{s} \in \mathcal{S}$, которое необходимо прервать для приема одного запроса в слайс s , $s \in \mathcal{S}$.

Таким образом, при поступлении в слайс s , $s \in \mathcal{S}$, запроса на предоставление услуги возможны три случая:

- (1) запрос немедленно будет принят на обслуживание, если число обслуживаемых запросов в данном слайсе меньше $\lfloor C_s/b_s \rfloor$, а объем до-

Таблица 1 Основные символы системы

Обозначение	Описание	Единица измерения
\mathcal{S}	Множество слайсов в системе, $\mathcal{S} \subset \mathbb{N} \setminus \{0\}$, $\mathbb{N} = \{0, 1, 2, \dots\}$	—
C	Общая емкость системы	ед. емкости
C_s	Максимальная емкость слайса s , $s \in \mathcal{S}$, $\sum_{s \in \mathcal{S}} C_s \geq C$	ед. емкости
Q_s	Гарантированная емкость слайса s , $s \in \mathcal{S}$, $Q_s \leq C_s$ и $\sum_{s \in \mathcal{S}} Q_s \leq C$	ед. емкости
λ_s	Интенсивность поступления запросов в слайс s , $s \in \mathcal{S}$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{ \mathcal{S} })$	запросов/ед. вр.
μ_s^{-1}	Среднее время обслуживания одного запроса в слайсе s , $s \in \mathcal{S}$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{ \mathcal{S} })$	ед. вр.
$\rho_s = \lambda_s/\mu_s$	Предложенная нагрузка, создаваемая запросами в слайсе s , $s \in \mathcal{S}$	—
b_s	Требование к ресурсам, необходимым для обслуживания одного запроса в слайсе s , $s \in \mathcal{S}$, $b_s \leq Q_s$, $\mathbf{b} = (b_1, \dots, b_{ \mathcal{S} })$	ед. емкости
$\lfloor C_s/b_s \rfloor$	Максимальное число запросов в слайсе s , $s \in \mathcal{S}$, $\mathbf{N}^{\max} = (\lfloor C_1/b_1 \rfloor, \dots, \lfloor C_{ \mathcal{S} }/b_{ \mathcal{S} } \rfloor)$	—
$\lfloor Q_s/b_s \rfloor$	Максимальное число запросов, которое может быть обслужено с использованием гарантированной емкости слайса s , $s \in \mathcal{S}$, $\mathbf{N}^g = (\lfloor Q_1/b_1 \rfloor, \dots, \lfloor Q_{ \mathcal{S} }/b_{ \mathcal{S} } \rfloor)$	—
n_s	Число запросов в слайсе s , $s \in \mathcal{S}$, когда система находится в состоянии $\mathbf{n} = (n_1, \dots, n_{ \mathcal{S} })$	—
\mathbf{e}_s	Строка s , $s \in \mathcal{S}$, единичной матрицы $ \mathcal{S} \times \mathcal{S} $	—

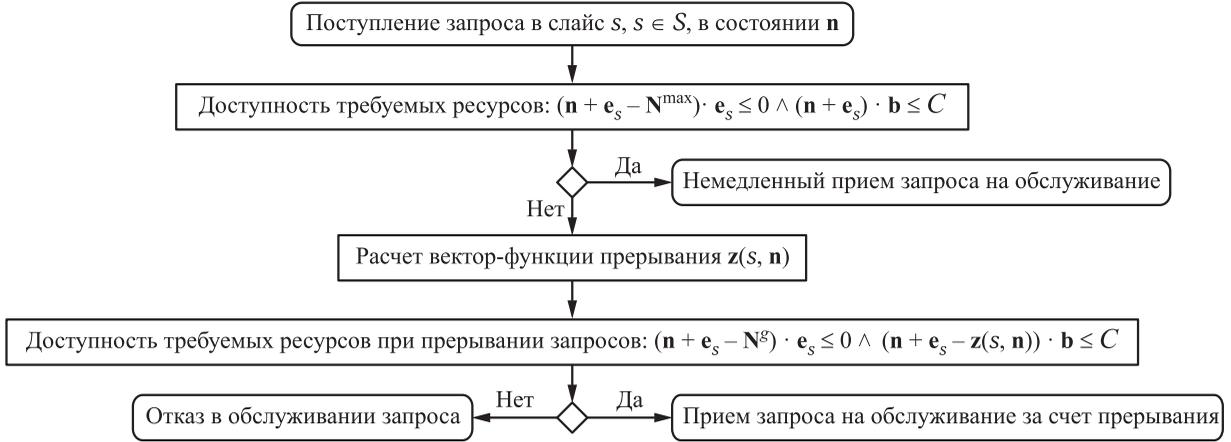


Рис. 1 Блок-схема для иллюстрации правил управления доступом к радиоресурсам системы

ступного ресурса БС больше или равен b_s , т.е. $(\mathbf{n} + \mathbf{e}_s - \mathbf{N}^{\max}) \cdot \mathbf{e}_s \leq 0 \wedge (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{b} \leq C$;

- (2) запрос будет принят на обслуживание за счет прерывания обслуживания запросов-«нарушителей» из других слайсов \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$, число которых определяется с помощью вектор-функции прерывания $\mathbf{z}(s, \mathbf{n})$, если число обслуживаемых запросов в слайсе s меньше $\lfloor Q_s/b_s \rfloor$, а объем доступных ресурсов БС меньше b_s , т.е. $(\mathbf{n} + \mathbf{e}_s - \mathbf{N}^g) \cdot \mathbf{e}_s \leq 0 \wedge (\mathbf{n} + \mathbf{e}_s - \mathbf{z}(s, \mathbf{n})) \cdot \mathbf{b} \leq C$;
- (3) запрос будет заблокирован, если число обслуживаемых запросов в слайсе s больше или равно $\lfloor Q_s/b_s \rfloor$, а объем доступного ресурса БС меньше b_s , т.е. $(\mathbf{n} + \mathbf{e}_s - \mathbf{N}^g) \cdot \mathbf{e}_s > 0 \wedge (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{b} > C$.

Отметим, что, в отличие от классических схем управления доступом с приоритизацией, предлагаемая схема ПС для приема запроса в гарантированную часть соответствующего слайса предусматривает необходимость прервать обслуживание запроса-«нарушителя» (т.е. запроса, обслуживаемого вне гарантированной части своего слайса) не только более низкого приоритета, но и более высокого приоритета.

Утверждение 1. При начальном условии $\mathbf{z}(s, \mathbf{n}) = \mathbf{0}$, числа $z_{\hat{s}}(s, \mathbf{n})$, $\hat{s} \in \mathcal{S}$, запросов, обслуживание которых необходимо будет прервать, можно вычислить с помощью рекуррентного соотношения

$$z_{\hat{s}}(s, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_{\hat{s}}), R \left(\left[\frac{(\mathbf{n} + \mathbf{e}_s - \mathbf{z}(s, \mathbf{n})) \cdot \mathbf{b} - C}{\mathbf{e}_{\hat{s}} \cdot \mathbf{b}} \right] \right) \right\}, \quad \hat{s} = |\mathcal{S}|, \dots, 1, \quad (1)$$

где $R(x) = xH(x)$ — функция Рампы¹, а $H(x)$ — функция Хевисайда².

Отметим, что число $z_s(s, \mathbf{n})$, $s \in \mathcal{S}$, всегда равно нулю, так как запрос не может быть принят на обслуживание за счет прерывания обслуживаемого(-ых) запроса(-ов) в этом же слайсе.

Для наглядности схему управления доступом можно описать с помощью блок-схемы, представленной на рис. 1.

3 Построение математической модели

В соответствии с описанной в разд. 2 схемой управления доступом к радиоресурсам сети поведение системы описывает $|\mathcal{S}|$ -мерный случайный процесс (СП) $\mathbf{X}(t) = (X_1(t), \dots, X_{|\mathcal{S}|}(t), t > 0)$, где $X_s(t)$, $s \in \mathcal{S}$, — число обслуживаемых запросов в слайсе s в момент времени t над пространством состояний

$$\Omega = \left\{ \mathbf{n} \in \mathbb{N}^{|\mathcal{S}|} : (\mathbf{n} - \mathbf{N}^{\max}) \cdot \mathbf{j} \leq 0 \wedge \mathbf{n} \cdot \mathbf{b} \leq C \right\},$$

где $\mathbb{N}^{|\mathcal{S}|}$ — множество всевозможных $|\mathcal{S}|$ -мерных векторов-строк с натуральными элементами, а \mathbf{j} — единичная матрица размера $1 \times |\mathcal{S}|$.

Схема соответствующей мультисервисной системы массового обслуживания (СМО) [10] изображена на рис. 2.

Для дальнейшего анализа модели введем следующие подмножества пространства состояний Ω системы:

- (1) Ω_s^{dad} , $s \in \mathcal{S}$, — множество состояний системы, в которых поступающий в слайс s запрос немедленно будет принят на обслуживание:

¹<https://mathworld.wolfram.com/RampFunction.html>.

²<https://mathworld.wolfram.com/HeavisideStepFunction.html>.

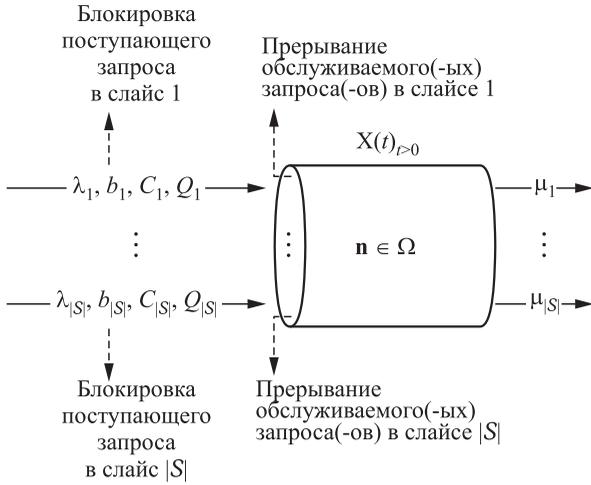


Рис. 2 Схема СМО

$$\Omega_s^{\text{dad}} = \{ \mathbf{n} \in \Omega : (\mathbf{n} - \mathbf{N}^{\text{max}}) \cdot \mathbf{e}_s < 0 \wedge (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{b} \leq C \}; \quad (2)$$

- (2) Ω_s^{vpad} , $s \in \mathcal{S}$, — множество состояний системы, в которых поступающий в слайс s запрос будет принят на обслуживание за счет прерывания обслуживаемого(-ых) запроса(-ов) в слайсах \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$:

$$\Omega_s^{\text{vpad}} = \{ \mathbf{n} \in \Omega : (\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_s < 0 \wedge (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{b} > C \}; \quad (3)$$

- (3) Ω_s^{block} , $s \in \mathcal{S}$, — множество состояний системы, в которых поступающий в слайс s запрос будет заблокирован:

$$\Omega_s^{\text{block}} = \{ \mathbf{n} \in \Omega : (\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_s \geq 0 \wedge (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{b} > C \}; \quad (4a)$$

или

$$\Omega_s^{\text{block}} = \Omega \setminus (\Omega_s^{\text{dad}} \cup \Omega_s^{\text{vpad}}), \quad (4b)$$

где $\Omega_s^{\text{dad}} \cup \Omega_s^{\text{vpad}}$ — множество состояний системы, в которых поступающий в слайс s запрос будет принят на обслуживание.

Пример для иллюстрации данных подмножеств для $\mathcal{S} = \{1, 2\}$ представлен в разд. 4.1 (см. рис. 4).

Подмножества (2)–(4) позволяют разделить все состояния \mathbf{n} , $\mathbf{n} \in \Omega$, системы на шесть групп:

- (1) $\bigcap_{s \in \mathcal{S}} \Omega_s^{\text{dad}}$ — множество состояний системы, в которых любой поступающий в систему запрос немедленно будет принят на обслуживание;
- (2) $\Omega_s^{\text{dad}} \setminus \bigcup_{\hat{s} \in \mathcal{S} \setminus \{s\}} \Omega_{\hat{s}}^{\text{dad}}$, $s \in \mathcal{S}$, — множество состояний системы, в которых только поступа-

ющий в слайс s запрос немедленно будет принят на обслуживание;

- (3) $\Omega_s^{\text{vpad}} \cap \bigcup_{\hat{s} \in \mathcal{S} \setminus \{s\}} \Omega_{\hat{s}}^{\text{vpad}}$, $s \in \mathcal{S}$, — множество состояний системы, в которых поступающие в слайсы s и \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$, запросы будут приняты на обслуживание за счет прерывания обслуживаемого(-ых) запроса(-ов) в слайсах \tilde{s} , $\tilde{s} \in \mathcal{S} \setminus \{s, \hat{s}\}$;
- (4) $\Omega_s^{\text{vpad}} \setminus \bigcup_{\hat{s} \in \mathcal{S} \setminus \{s\}} \Omega_{\hat{s}}^{\text{vpad}}$, $s \in \mathcal{S}$, — множество состояний системы, в которых только поступающий в слайс s запрос будет принят на обслуживание за счет прерывания одного или нескольких обслуживаемых запросов в слайсе \hat{s} , $\hat{s} \in \mathcal{S} \setminus \{s\}$;
- (5) $\bigcap_{s \in \mathcal{S}} \Omega_s^{\text{block}}$ — множество состояний системы, в которых любой поступающий в систему запрос будет заблокирован;
- (6) $\Omega_s^{\text{block}} \setminus \bigcup_{\hat{s} \in \mathcal{S} \setminus \{s\}} \Omega_{\hat{s}}^{\text{block}}$, $s \in \mathcal{S}$, — множество состояний системы, в которых только поступающий в слайс s запрос будет заблокирован.

Диаграмма интенсивностей переходов для состояния системы \mathbf{n} , $\mathbf{n} \in \Omega$, имеет вид, представленный на рис. 3.

Функция управления доступом к ресурсам системы определяется следующим образом:

$$f_s(\mathbf{n}) = \begin{cases} 1, & \text{если } \mathbf{n} \in (\Omega_s^{\text{dad}} \cup \Omega_s^{\text{vpad}}); \\ 0 & \text{в противном случае, } \mathbf{n} \in \Omega_s^{\text{block}}. \end{cases} \quad s \in \mathcal{S}.$$

Согласно диаграмме интенсивностей переходов рис. 3, рассматриваемый СП описывается следующей системой уравнений равновесия:

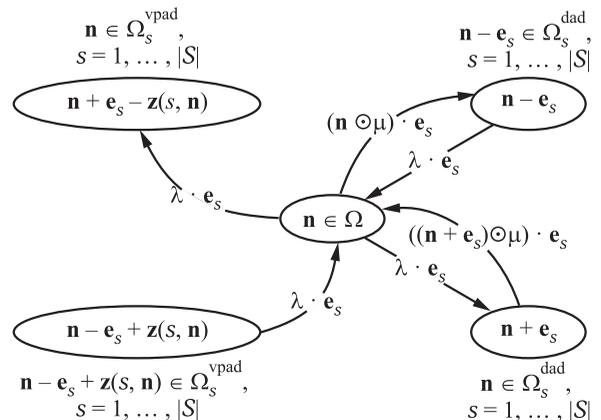


Рис. 3 Диаграмма интенсивностей переходов для состояния системы \mathbf{n} , $\mathbf{n} \in \Omega$

$$\begin{aligned}
 P(\mathbf{n}) & \left(\lambda \cdot \sum_{s \in \mathcal{S}} \left(I_{\Omega_s^{\text{dad}}}(\mathbf{n}) + I_{\Omega_s^{\text{vpad}}}(\mathbf{n}) \right) \mathbf{e}_s + \mathbf{n} \cdot \boldsymbol{\mu} \right) = \\
 & = \lambda \cdot \sum_{s \in \mathcal{S}} \left(P(\mathbf{n} - \mathbf{e}_s) I_{\Omega_s^{\text{dad}}}(\mathbf{n} - \mathbf{e}_s) + \right. \\
 & \left. + P(\mathbf{n} - \mathbf{e}_s + \mathbf{z}(s, \mathbf{n})) I_{\Omega_s^{\text{vpad}}}(\mathbf{n} - \mathbf{e}_s + \mathbf{z}(s, \mathbf{n})) \right) \mathbf{e}_s + \\
 & + \boldsymbol{\mu} \cdot \sum_{s \in \mathcal{S}} \left(P(\mathbf{n} + \mathbf{e}_s) I_{\Omega_s^{\text{dad}}}(\mathbf{n}) (\mathbf{n} + \mathbf{e}_s) \cdot \mathbf{e}_s \right) \mathbf{e}_s,
 \end{aligned}$$

где $P(\mathbf{n} \in \Omega)$ — стационарная вероятность того, что система находится в состоянии \mathbf{n} , а $I_o(\ast)$ — функция-индикатор¹.

В связи с реализацией механизма прерывания обслуживания запросов СП $\mathbf{X}(t)_{t>0}$, описывающий рассматриваемую систему, не является обратимым марковским процессом. В этом случае для вычисления стационарного распределения вероятностей состояний системы $\mathbf{P} = (P(\mathbf{n}))_{\mathbf{n} \in \Omega}$ — вектора-столбца размера $|\Omega|$ — может быть применен один из численных методов, например итерационный метод [11, 12]:

$$\mathbf{A}^T \mathbf{P} = \mathbf{0}, \quad \mathbf{P} \cdot \mathbf{j} = 1,$$

где \mathbf{A} — инфинитезимальная матрица размера $|\Omega|^2$, элементы $A(\mathbf{n}, \hat{\mathbf{n}})$, $\mathbf{n} \in \Omega$, $\hat{\mathbf{n}} \in \Omega$, которой определяются следующим образом:

при $\mathbf{n} \neq \hat{\mathbf{n}}$

$$A(\mathbf{n}, \hat{\mathbf{n}}) =$$

$$\begin{cases} \lambda \cdot \mathbf{e}_s, & \text{если } \hat{\mathbf{n}} = \mathbf{n} + \mathbf{e}_s, \mathbf{n} \in \Omega_s^{\text{dad}}, \\ & \text{или } \hat{\mathbf{n}} = \mathbf{n} + \mathbf{e}_s - \mathbf{z}(s, \mathbf{n}), \mathbf{n} \in \Omega_s^{\text{vpad}}; \\ (\mathbf{n} \odot \boldsymbol{\mu}) \cdot \mathbf{e}_s, & \text{если } \hat{\mathbf{n}} = \mathbf{n} - \mathbf{e}_s, \hat{\mathbf{n}} \in \Omega_s^{\text{dad}}; \\ 0 & \text{в противном случае, } \hat{\mathbf{n}} \in \Omega \setminus \{\mathbf{n}\}, \end{cases}$$

$$s = 1, \dots, |\mathcal{S}|;$$

при $\mathbf{n} = \hat{\mathbf{n}}$

$$A(\mathbf{n}, \mathbf{n}) = - \sum_{\hat{\mathbf{n}} \in \Omega \setminus \{\mathbf{n}\}} A(\mathbf{n}, \hat{\mathbf{n}}).$$

Рассчитав стационарное распределение вероятностей состояний системы, можно вычислить следующие КПЭ системы:

– среднее число запросов, обслуживаемых в системе

$$N = \sum_{\mathbf{n} \in \Omega} P(\mathbf{n}) \mathbf{n} \cdot \mathbf{j}; \quad (5)$$

– вероятность блокировки запроса, поступающего в систему

$$P^{\text{block}} = \sum_{\mathbf{n} \in \mathcal{B}} P(\mathbf{n}); \quad \mathcal{B} = \bigcap_{s=1}^{|\mathcal{S}|} \Omega_s^{\text{block}}; \quad (6)$$

– средний занятый ресурс,

$$K = \sum_{\mathbf{n} \in \Omega} P(\mathbf{n}) \mathbf{n} \cdot \mathbf{b}. \quad (7)$$

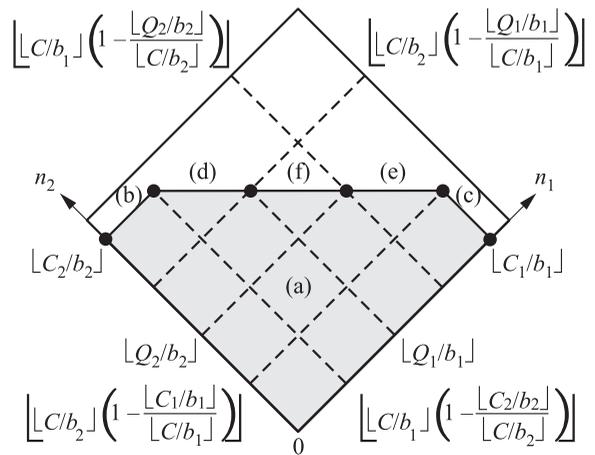
4 Примеры расчета вектор-функции прерывания обслуживания

Проиллюстрируем работу итерационного метода расчета числа и типа запросов-«нарушителей», которые должны быть прерваны для приема запроса, поступающего в гарантированную часть своего слайса, на примере моделей с двумя и тремя слайсами, для которых вычислим вектор-функцию прерывания обслуживания (1) запросов.

4.1 Модель сети с двумя слайсами

Для случая $\mathcal{S} = \{1, 2\}$ с учетом основных подмножеств системы (2)–(4) пространство состояний модели имеет вид, изображенный на рис. 4.

Приведем пример расчета вектор-функции прерывания обслуживания (1) для приема запроса



(a)	$\Omega_1^{\text{dad}} \cap \Omega_2^{\text{dad}}$	(d)	Ω_1^{vpad}
(b)	$\Omega_1^{\text{dad}} \setminus \Omega_2^{\text{dad}}$	(e)	Ω_2^{vpad}
(c)	$\Omega_2^{\text{dad}} \setminus \Omega_1^{\text{dad}}$	(f)	$\Omega_1^{\text{block}} \cap \Omega_2^{\text{block}}$

Рис. 4 Пространство состояний модели с двумя слайсами с учетом основных подмножеств системы

¹<https://mathworld.wolfram.com/CharacteristicFunction.html>.

Таблица 2 Исходные данные для сравнительного анализа [13]

СЛАЙС/УСЛУГА	СХЕМА С РЕЗЕРВИРОВАНИЕМ (РС)			СХЕМА С ПРЕРЫВАНИЕМ (ПС)		
	Параметр	Значение	Единица измерения	Параметр	Значение	Единица измерения
1/3K Cloud VR (Game)	C_1	1,0	Гбит/с	C_1	1,25	Гбит/с
	Q_1	1,0	Гбит/с	Q_1	1,0	Гбит/с
	b_1	0,1	Гбит/с	b_1	0,1	Гбит/с
	μ_1^{-1}	3600	с	μ_1^{-1}	3600	с
2/4K Live News Pushing (30 fps)	C_2	0,5	Гбит/с	C_2	0,65	Гбит/с
	Q_2	0,5	Гбит/с	Q_2	0,5	Гбит/с
	b_2	0,04	Гбит/с	b_2	0,04	Гбит/с
	μ_2^{-1}	1200	с	μ_2^{-1}	1200	с
3/4K On-Demand Video	C_3	0,5	Гбит/с	C_3	0,65	Гбит/с
	Q_3	0,5	Гбит/с	Q_3	0,5	Гбит/с
	b_3	0,03	Гбит/с	b_3	0,03	Гбит/с
	μ_3^{-1}	1800	с	μ_3^{-1}	1800	с
4/4K Live Broadcast Pushing (50fps)	C_4	0,5	Гбит/с	C_4	0,65	Гбит/с
	Q_4	0,5	Гбит/с	Q_4	0,5	Гбит/с
	b_4	0,063	Гбит/с	b_4	0,063	Гбит/с
	μ_4^{-1}	5400	с	μ_4^{-1}	5400	с

ρ	λ_s	СЦЕН. 1		СЦЕН. 2		СЦЕН. 3	
		СЛАЙСЫ	C	СЛАЙСЫ	C	СЛАЙСЫ	C
от 1 до 25	$\rho\mu_s$ запросов/с	1 и 2	1,5 Гбит/с	1, 2 и 3	2,0 Гбит/с	1, 2, 3 и 4	2,5 Гбит/с

в первый слайс $\mathbf{z}(1, \mathbf{n}) = (z_1(1, \mathbf{n}), z_2(1, \mathbf{n}))$. Рассмотрим состояние системы

$$\mathbf{n} = \left(\left\lfloor \frac{C}{b_1} \left(1 - \frac{\lfloor C_2/b_2 \rfloor}{\lfloor C/b_2 \rfloor} \right) \right\rfloor, \lfloor C_2/b_2 \rfloor \right),$$

которое согласно диаграмме на рис. 4 принадлежит множеству Ω_1^{vpad} , т.е. множеству состояний системы, в которых поступающий в первый слайс запрос будет принят на обслуживание за счет прерывания обслуживаемого(-ых) запроса(-ов) во втором слайсе.

Для исходных данных «СЦЕН. 1», представленных в табл. 2, получим $\mathbf{n} = (8, 16)$, $\mathbf{n} \in \Omega_1^{\text{vpad}}$.

Рассчитаем число $z_2(1, \mathbf{n})$ обслуживаемых запросов второго слайса, которое необходимо прервать для приема одного запроса в первый слайс. Воспользуемся начальным условием $\mathbf{z}(1, \mathbf{n}) = (0, 0)$, получим

$$z_2(1, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_2), R \left(\left\lceil \frac{(\mathbf{n} + \mathbf{e}_1) \cdot \mathbf{b} - C}{\mathbf{e}_2 \cdot \mathbf{b}} \right\rceil \right) \right\} = \min \{R(4), R(1)\} = \min \{4, 1\} = 1.$$

Перейдем к расчету числа $z_1(1, \mathbf{n})$ обслуживаемых запросов первого слайса, которое необходимо

прервать для приема одного запроса в первый слайс. Очевидно, что число $z_1(1, \mathbf{n})$ должно быть равно 0, так как запрос не может быть принят в слайс на обслуживание за счет прерывания обслуживаемого(-ых) запроса(-ов) в этом же слайсе. Воспользуемся текущим значением $\mathbf{z}(1, \mathbf{n}) = (0, z_2(1, \mathbf{n})) = (0, 1)$, получим

$$z_1(1, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_1), R \left(\left\lceil \frac{(\mathbf{n} + \mathbf{e}_1 - \mathbf{z}(1, \mathbf{n})) \cdot \mathbf{b} - C}{\mathbf{e}_1 \cdot \mathbf{b}} \right\rceil \right) \right\} = \min \{R(-2), R(0)\} = \min \{0, 0\} = 0.$$

Таким образом, в состоянии системы $\mathbf{n} = (8, 16)$, $\mathbf{n} \in \Omega_1^{\text{vpad}}$, вектор-функция прерывания $\mathbf{z}(1, \mathbf{n}) = (z_1(1, \mathbf{n}), z_2(1, \mathbf{n})) = (0, 1)$, т.е. поступающий в первый слайс запрос будет принят на обслуживание за счет прерывания одного запроса, обслуживаемого во втором слайсе.

4.2 Модель сети с тремя слайсами

Перейдем к трехмерному случаю $\mathcal{S} = \{1, 2, 3\}$. Рассмотрим пример расчета вектор-функции прерывания обслуживания (1) для приема запроса во второй слайс $\mathbf{z}(2, \mathbf{n}) = (z_1(2, \mathbf{n}), z_2(2, \mathbf{n}), z_3(2, \mathbf{n}))$, $\mathbf{n} \in \Omega_2^{\text{vpad}}$.

Сведем исходные данные для примера в табл. 3.

Таблица 3 Исходные данные для примера с тремя слайсами

C , Мбит/с	\mathbf{b} , Мбит/с	\mathbf{n}	\mathbf{N}^g
13	(1,3,1)	(6,1,4)	(2,2,2)

Рассчитаем число $z_3(2, \mathbf{n})$ обслуживаемых запросов третьего слайса — слайса с низшим приоритетом, которое необходимо прервать для приема одного запроса во второй слайс. Воспользуемся начальным условием $\mathbf{z}(2, \mathbf{n}) = (0, 0, 0)$, получим

$$z_3(2, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_3), R\left(\left\lceil \frac{(\mathbf{n} + \mathbf{e}_2) \cdot \mathbf{b} - C}{\mathbf{e}_3 \cdot \mathbf{b}} \right\rceil\right) \right\} = \min \{R(2), R(3)\} = \min \{2, 3\} = 2.$$

Покажем, что число $z_2(2, \mathbf{n})$ обслуживаемых запросов второго слайса, которое необходимо прервать для приема одного запроса во второй слайс, равно 0. Воспользуемся текущим значением функции $\mathbf{z}(2, \mathbf{n}) = (0, 0, z_3(2, \mathbf{n})) = (0, 0, 2)$, получим

$$z_2(2, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_2), R\left(\left\lceil \frac{(\mathbf{n} + \mathbf{e}_2 - \mathbf{z}(2, \mathbf{n})) \cdot \mathbf{b} - C}{\mathbf{e}_2 \cdot \mathbf{b}} \right\rceil\right) \right\} = \min \{R(-1), R(1)\} = \min \{0, 1\} = 0.$$

Согласно схеме управления доступом ПС при исчерпании возможности освободить ресурс за счет прерывания запросов-«нарушителей» более низкого приоритета для приема запроса в гарантированную часть второго слайса должно быть прервано обслуживание запросов-«нарушителей» более высокого приоритета. Так как число обслуживаемых запросов первого слайса превышает гарантированное значение, для приема одного запроса во второй слайс должно быть прервано обслуживание $z_1(2, \mathbf{n})$ запросов первого слайса. С учетом текущего значения вектор-функции прерывания

$$\mathbf{z}(2, \mathbf{n}) = (0, z_2(2, \mathbf{n}), z_3(2, \mathbf{n})) = (0, 0, 2)$$

получим

$$z_1(2, \mathbf{n}) = \min \left\{ R((\mathbf{n} - \mathbf{N}^g) \cdot \mathbf{e}_1), R\left(\left\lceil \frac{(\mathbf{n} + \mathbf{e}_2 - \mathbf{z}(2, \mathbf{n})) \cdot \mathbf{b} - C}{\mathbf{e}_1 \cdot \mathbf{b}} \right\rceil\right) \right\} = \min \{R(4), R(1)\} = \min \{4, 1\} = 1.$$

Таким образом, в состоянии системы $\mathbf{n} = (6, 1, 4)$, $\mathbf{n} \in \Omega_2^{\text{vpad}}$, вектор-функция прерывания обслуживания запросов имеет вид

$$\mathbf{z}(2, \mathbf{n}) = (z_1(2, \mathbf{n}), z_2(2, \mathbf{n}), z_3(2, \mathbf{n})) = (1, 0, 2),$$

т.е. поступающий во второй слайс запрос будет принят на обслуживание за счет прерывания двух запросов, обслуживаемых в третьем слайсе, и одного — в первом.

Далее перейдем к анализу основных КПЭ системы, описанных в разд. 3.

5 Численный анализ

Для анализа эффективности предложенной в работе схемы доступа ПС к радиоресурсам сети, основанной на реализации механизма прерывания обслуживания пользователей, проведем сравнительный анализ ее основных КПЭ с КПЭ известной схемы доступа РС, основанной на реализации механизма резервирования ресурсов. Сведем исходные данные для численного анализа в табл. 2. Рассмотрим зависимость среднего числа запросов, обслуживаемых в системе, (5), вероятности блокировки запроса любого типа, поступающего в систему, (6) (вероятности блокировки системы) и среднего числа занятых ресурсов в системе (7) от интенсивности предложенной нагрузки ρ , создаваемой в каждом слайсе. Положим $\rho_s = \rho, s \in \mathcal{S}$. Результаты сравнительного анализа схем ПС и РС представлены на рис. 5 и 6.

На рис. 5, *a* проиллюстрировано поведение среднего числа запросов N , обслуживаемых в системе, в зависимости от интенсивности предложенной нагрузки ρ для трех сценариев — с двумя, тремя и четырьмя слайсами. По графику видно, что чем больше слайсов в сети, тем выше среднее число обслуживаемых запросов и тем эффективнее схема ПС, в частности (см. рис. 6, *a*) для «СЦЕН. 1» эффективность схемы ПС может превышать эффективность схемы РС в 1,06 раза, а для «СЦЕН. 3» — в 1,12 раза.

С ростом интенсивности предложенной нагрузки ρ вероятность блокировки системы P^{block} увеличивается для всех рассматриваемых сценариев (см. рис. 5, *b*), причем она минимальна для сценария с наибольшим числом слайсов («СЦЕН. 3»). Однако на рис. 5, *b* и 6, *b* видно, что схема ПС эффективнее схемы РС только в диапазоне небольших нагрузок на систему, в частности (см. рис. 6, *b*) для «СЦЕН. 2» и «СЦЕН. 3» — диапазон от 1 до 4,43.

Анализ рис. 5, *в* и 6, *в* показывает, что система более эффективно использует ресурсы при применении схемы ПС. В частности (см. рис. 6, *в*), среднее число ресурсов, занятых в системе, при использовании схемы ПС для «СЦЕН. 1» может быть в 1,05 раза выше, чем при использовании схемы РС, а для «СЦЕН. 3» — в 1,12 раза.

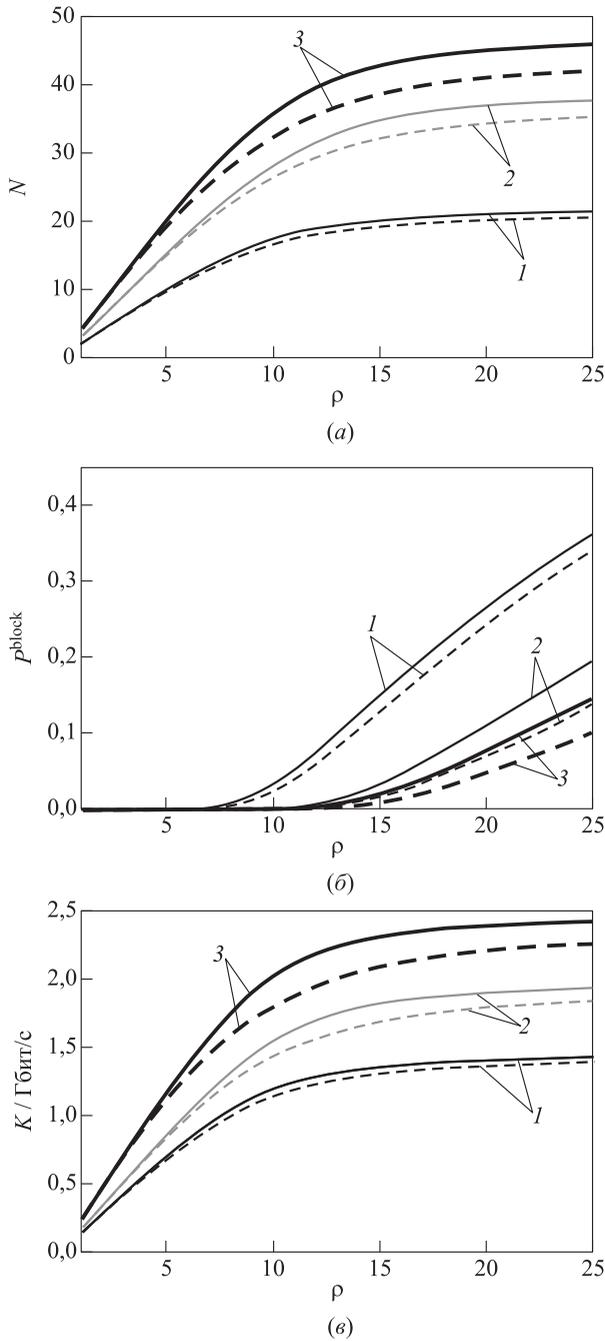


Рис. 5 Графики зависимостей среднего числа запросов, обслуживаемых в системе (а), вероятности блокировки системы (б) и среднего занятого ресурса (в) от интенсивности предложенной нагрузки: 1 — СЦЕН. 1; 2 — СЦЕН. 2; 3 — СЦЕН. 3; сплошные кривые — РС; штриховые кривые — РС

6 Заключение

В работе предложена схема доступа запросов пользователей к ресурсам беспроводной сети, основанная на реализации механизма прерывания об-

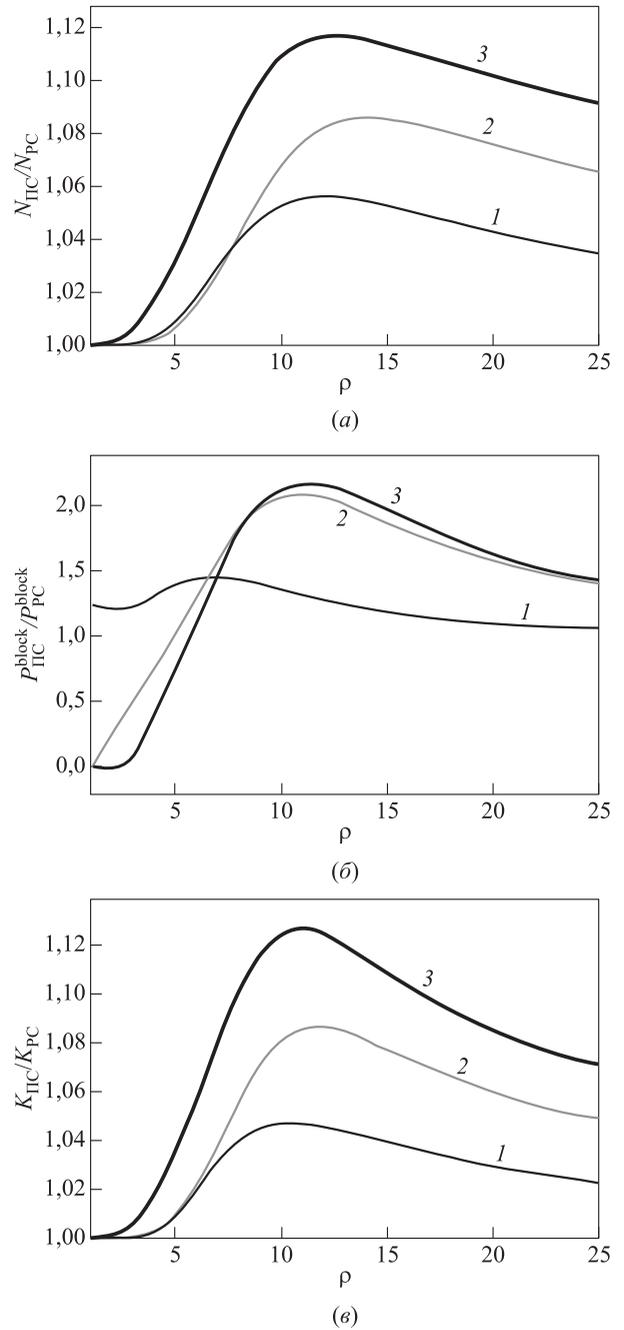


Рис. 6 Графики зависимостей соотношений средних чисел обслуживаемых в системе запросов (а), вероятности блокировки системы (б) и среднего занятого ресурса (в) для двух схем от интенсивности предложенной нагрузки: 1 — СЦЕН. 1; 2 — СЦЕН. 2; 3 — СЦЕН. 3

служивания пользователей в рамках технологии нарезки радиоресурсов сети Network Slicing.

Проведен сравнительный анализ, показывающий эффективность предложенной схемы по сравнению с известной схемой доступа, основан-

ной на механизме резервирования ресурсов. Результаты численного эксперимента показали, что в диапазоне небольших нагрузок на систему предложенная схема эффективнее схемы доступа с реализацией механизма резервирования. Результаты численного эксперимента выделяют следующие особенности применения предложенной схемы доступа по сравнению со схемой на основе реализации механизма резервирования:

- (1) эффективность в диапазоне небольших нагрузок в использовании физических ресурсов БС и емкостей слайсов;
- (2) значимое повышение эффективности при поддержке предоставления услуг для большего числа слайсов.

Литература

1. *Sultan A., Pope M.* Feasibility study on new services and markets technology enablers for network operation; Stage 1 (3GPP). Ver. 15.0.0, 2016. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3016>.
2. *Meredith J., Firmin F., Pope M.* Release 16 Description; Summary of Rel-16 Work Items (3GPP). Ver. 16.2.0, 2022. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3493>.
3. *Meredith J., Soveri M., Pope M.* Management and orchestration; 5G end to end Key Performance Indicators (KPI) (3GPP). Ver. 18.0.0, 2022. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3415>.
4. *Yarkina N., Correia L., Moltchanov D., Gaidamaka Y., Samouylov K.* Multi-tenant resource sharing with equitable-priority-based performance isolation of slices for 5G cellular systems // *Comput. Commun.*, 2022. Vol. 188. P. 39–51. doi: 10.1016/j.comcom.2022.02.019.
5. *Luu Q., Kerboeuf S., Kieffer M.* Admission control and resource reservation for prioritized slice requests with guaranteed SLA under uncertainties // *IEEE T. Netw. Serv. Man.*, 2022. Vol. 19. P. 3136–3153. doi: 10.1109/tnsm.2022.3160352.
6. *Rehman A., Mahmood I., Kamran M., Sanaullah M., Ijaz A., Ali J., Ali M.* Enhancement in quality-of-services using 5G cellular network using resource reservation protocol // *Phys. Commun. — Amst.*, 2022. Art. 101907. 10 p. doi: 10.1016/j.phycom.2022.101907.
7. *Kermani P., Kleinrock L.* Analysis of buffer allocation schemes in a multiplexing node // *Int. Conf. Comm.*, 1977. Vol. 2. P. 30–34.
8. *Katoun F., Kleinrock L.* Analysis of shared finite storage in a computer network node environment under general traffic conditions // *IEEE T. Commun.*, 1980. Vol. 28. P. 992–1003. doi: 10.1109/tcom.1980.1094756.
9. *Башарин Г. П., Самуйлов К. Е.* Об оптимальной структуре БП в сетях передачи данных с коммутацией пакетов. — М.: ВИНТИ, 1982. 70 с.
10. *Basharin G., Gaidamaka Y., Samouylov K.* Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks // *Autom. Control Comp. S.*, 2013. Vol. 47. P. 62–69. doi: 10.3103/s0146411613020028.
11. *Степанов С. Н.* Теория телетрафика: концепции, модели, приложения. — М.: Горячая линия-Телеком, 2015. 868 с.
12. *Zhou D., Chen Z., Pan E., Zhang Y.* Dynamic statistical responses of gear drive based on improved stochastic iteration method // *Appl. Math. Model.*, 2022. Vol. 108. P. 46–65. doi: 10.1016/j.apm.2022.03.020.
13. *Schoolar D., Lambert P., Nanbin W., Liang Z.* 5G service experience-based network planning criteria (Ovum Consulting) // Partnership with Huawei, 2019. <https://carrier.huawei.com/~media/CNBGV2/download/products/servies/5G-Planning-Criteria-White-Paper.pdf>.

Поступила в редакцию 15.01.23

PREEMPTION-BASED PRIORITIZATION SCHEME FOR NETWORK RESOURCES SLICING IN 5G SYSTEMS

K. Y. B. Adou¹, E. V. Markova¹, Yu. V. Gaidamaka^{1,2}, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

²Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The network slicing (NS) technology, which has been actively studied in recent years, is based on the representation of a common network infrastructure in the form of various customizable logical networks called slices and involves the division of mobile network operators into two groups — physical network infrastructure providers (InPs) and mobile virtual network operators (MVNOs). The MVNOs lease the physical resources of InPs

to create their own slices to provide services to their users with different quality of service requirements. In the present paper, for a network with NS technology, a scheme for accessing its radio resources is proposed that provides users with services with a guaranteed bit rate (GBR) and priority control based on the implementation of the user service interruption mechanism. The authors propose a scheme for accessing radio resources of a network under NS technology that provides users with services with GBR and priority control based on the implementation of the user service interruption mechanism. To evaluate the effectiveness of the proposed scheme, a comparative analysis of its characteristics with the characteristics of the access scheme based on the resource reservation mechanism was carried out.

Keywords: 5G; network slicing; quality of service; key performance indicators; priority management; service interruption; iterative method

DOI: 10.14357/19922264230113

Acknowledgments

The research was supported by the Russian Science Foundation grant No.22-79-10053, <https://rscf.ru/en/project/22-79-10053/>.

References

1. Sultan, A., and M. Pope. 2016. Feasibility study on new services and markets technology enablers for network operation; Stage 1 (3GPP). Ver. 15.0.0. Available at: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3016> (accessed January 30, 2023).
2. Meredith, J., F. Firmin, and M. Pope. 2022. Release 16 description; Summary of Rel-16 work items (3GPP). Ver. 16.2.0. Available at: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3493> (accessed January 30, 2023).
3. Meredith, J., M. Soveri, and M. Pope. 2022. Management and orchestration; 5G end to end Key Performance Indicators (KPI) (3GPP). Ver. 18.0.0. Available at: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3415> (accessed January 30, 2023).
4. Yarkina, N., L. M. Correia, D. Moltchanov, Y. Gaidamaka, and K. Samouylov. 2022. Multi-tenant resource sharing with equitable-priority-based performance isolation of slices for 5G cellular systems. *Comput. Commun.* 188:39–51. doi: 10.1016/j.comcom.2022.02.019.
5. Luu, Q., S. Kerboeuf, and M. Kieffer. 2022. Admission control and resource reservation for prioritized slice requests with guaranteed SLA under uncertainties. *IEEE T. Netw. Serv. Man.* 19:3136–3153. doi: 10.1109/tnsm.2022.3160352.
6. Rehman, A., I. Mahmood, M. Kamran, M. Sanaullah, A. Ijaz, J. Ali, and M. Ali. 2022. Enhancement in quality-of-services using 5G cellular network using resource reservation protocol. *Phys. Commun.* — *Amst.* 55:101907. 10 p. doi: 10.1016/j.phycom.2022.101907.
7. Kermani, P., and L. Kleinrock. 1977. Analysis of buffer allocation schemes in a multiplexing node. *Int. Conf. Comm.* 2:30–34.
8. Kamoun, F, and L. Kleinrock. 1980. Analysis of shared finite storage in a computer networks node environment under general traffic conditions. *IEEE T. Commun.* 28(7):992–1003. doi: 10.1109/tcom.1980.1094756.
9. Basharin, G. P., and K. E. Samouylov. 1982. *Ob optimal'noy strukture bufernoy pamyati v setyakh peredachi dannykh s kommutatsiyey paketov* [On the optimal structure of BP in data transmission networks with packet commutation]. Moscow: VINITI. 70 p.
10. Basharin, G. P., Yu. V. Gaidamaka, and K. E. Samouylov. 2013. Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks. *Autom. Control Comp. S.* 47(2):62–69. doi: 10.3103/s0146411613020028.
11. Stepanov, S. N. 2015. *Teoriya teletrafika: Kontseptsii, modeli, prilozheniya* [Theory of teletraffic: Concepts, models, and applications]. Moscow: Goryachaya liniya-Telekom. 868 p.
12. Zhou, D., Z. Chen, E. Pan, and Y. Zhang. 2022. Dynamic statistical responses of gear drive based on improved stochastic iteration method. *Appl. Math. Model.* 108:46–65. doi: 10.1016/j.apm.2022.03.020.
13. Scholar, D., P. Lambert, W. Nanbin, and Z. Liang. 2019. 5G service experience-based network planning criteria (Ovum Consulting). Partnership with Huawei. Available at: <https://carrier.huawei.com/~media/CNBGV2/download/products/services/5G-Planning-Criteria-White-Paper.pdf> (accessed January 30, 2023).

Received January 15, 2023

Contributors

Adou Kpangny Y. B. (b. 1993) — PhD student, research assistant, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; adu-k@rudn.ru

Markova Ekaterina V. (b. 1987) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; markova-ev@rudn.ru

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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; gaydamaka-yuv@rudn.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, 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 119133, Russian Federation; sshorgin@ipiran.ru

МНОГОМЕРНЫЕ БАТТЕРФЛЯИ В ЗАДАЧАХ ОПТИМИЗАЦИИ ПО CC-VaR

Г. А. Агасандян¹

Аннотация: Работа продолжает исследование технических проблем, связанных с применением континуального критерия VaR (CC-VaR) на многомерных рынках опционов. В предположении, что на рынке сценарными баттерфляями непосредственно не торгуют, разрабатывается методика получения их репликации из многомерных α -опционов — многомерного обобщения обычных одномерных опционов, таких как коллы и путы. Работа служит непосредственным расширением предложенного в предыдущей работе автора способа, позволяющего конструировать индикаторы базиса на многомерном сценарном рынке комбинациями многомерных бинарных опционов. Методика основывается на теоремах паритета для одномерного рынка традиционных опционов и пригодна для рынков произвольной размерности, но ее фактическая реализация проводится для двумерных рынков. Приводятся конструкции базисов из α -опционов — как однотипных, так и смешанных естественных с выделенным центром рынка. Теоретические представления оптимальных портфелей в этих базисах иллюстрируются на примере конкретного двумерного рынка.

Ключевые слова: базовые активы; многомерный рынок; функция рисков предпочтений инвестора; континуальный критерий VaR (CC-VaR); стоимостная и прогнозная плотности; опционы колл и пут; α -опционы; сценарные баттерфляи; базисы; центр рынка; портфели баттерфляев

DOI: 10.14357/19922264230114

1 Введение

Проблемы применения на рынках опционов введенного автором континуального критерия VaR (CC-VaR) рассматриваются в [1–5]. Настоящую работу можно рассматривать как продолжение исследования [6], в котором предлагались варианты репликации индикаторов базиса на многомерном сценарном рынке комбинациями так называемых ζ -опционов (многомерных бинарных опционов). Здесь подобная задача решается для более сложных инструментов — многомерных аналогов одномерных базисных баттерфляев, которые реплицируются комбинациями так называемых α -опционов — многомерных аналогов традиционных опционов типа колл и пут.

Основания для такого рассмотрения и его проблемы, связанные с применением континуального критерия VaR (CC-VaR), приведены в [6], там же вводятся многие обозначения, которые используются и здесь. Теоретической моделью при построении α -рынка служит также многомерный δ -рынок [5, 6].

В работе для многомерных рынков опционов решаются те же проблемы технического характера, что и для ζ -рынков — рынков многомерных бинарных опционов. Но на этот раз в отношении своего инструментария они в большей мере напоминают

проблемы традиционных рынков опционов, на которых в отсутствие баттерфляев в качестве объектов непосредственной торговли предлагается получать их в виде комбинаций коллов и путов.

2 Теоретический α -рынок и его свойства

Вновь рассматривается многомерный δ -рынок (однопериодный, теоретический и идеальный) с n (> 1) базовыми активами, векторы цен которых в конце периода $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $x_l \in X_l \subset \mathfrak{R}$, $l \in N = \{1, \dots, n\}$, образуют n -мерное множество $X = \prod_{l \in N} X_l$. На X заданы *прогнозная* $p(\mathbf{x})$ и *стоимостная* $c(\mathbf{x})$ плотности, порождающие вероятностные меры $P\{\cdot\}$ и $C\{\cdot\}$.

Платежная функция произвольного инструмента I обозначается $\pi(\mathbf{x}; I)$, его рыночная стоимость и средний, с точки зрения инвестора, доход, рассчитанные по плотностям $c(\mathbf{x})$ и $p(\mathbf{x})$ соответственно, определяются соотношениями:

$$|I| = \int_{\mathfrak{X}} \pi(\mathbf{x}; I) c(\mathbf{x}) d\mathbf{x}; \quad \|I\| = \int_{\mathfrak{X}} \pi(\mathbf{x}; I) p(\mathbf{x}) d\mathbf{x}.$$

Базис рынка составляют δ -инструменты $D(s)$, $s \in X$, с обобщенной n -мерной δ -функцией относительно s в качестве платежной:

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, agasand17@yandex.ru

$$D(s) = \prod_{l \in N} D_l(s_l),$$

$$\pi(x; D(s)) = \delta(x - s) = \prod_{l \in N} \delta(x_l - s_l). \quad (1)$$

Инструмент G с произвольной измеримой платежной функцией $g(x)$ и его стоимость имеют вид:

$$G = \int_X g(s) D(s) ds;$$

$$|G| = \int_X g(s) |D(s)| ds = \int_X g(s) c(s) ds.$$

Наряду с «полноправными» n -мерными инструментами на рынке присутствуют и их k -мерные версии, у которых $n - k$ координатных базовых активов представлены в форме одномерных единичных безрисковых инструментов.

Для индикаторов $H\{M\}$, $M \subset X$, безрискового актива $U = H\{X\}$ и их цен

$$H\{M\} = \int_M D(s) ds; \quad |H\{M\}| = \int_M c(s) ds;$$

$$|U| = C\{X\} = \int_X c(s) ds = \frac{1}{r},$$

где r — приравняваемый единице безрисковый доход за период.

На одномерном рынке опционы пут P_s и колл C_s со страйком s задаются своими платежными функциями:

$$\left. \begin{aligned} \pi(x; P_s) &= \max(0, s - x); \\ \pi(x; C_s) &= \max(0, x - s), \quad x, s \in X \subset \mathfrak{X}. \end{aligned} \right\} \quad (2)$$

Для них выполняется формула паритета (X — вектор базовых активов)

$$C_s - P_s = X - sU.$$

Нормированными спредами быка с парой страйков $s - h$, $s \in X$ ($h > 0$) и медведя с парой страйков s , $s + h \in X$ служат комбинации опционов соответственно

$$\left. \begin{aligned} S_{s;h}^{\text{bull}} &= \frac{1}{h} (C_{s-h} - C_s) = U + \frac{1}{h} (P_{s-h} - P_s); \\ S_{s;h}^{\text{bear}} &= \frac{1}{h} (P_{s+h} - P_s) = U + \frac{1}{h} (C_{s+h} - C_s) \end{aligned} \right\} \quad (3)$$

с платежными функциями

$$\left. \begin{aligned} \pi(x; S_{s;h}^{\text{bull}}) &= \min\left(1, \frac{1}{h} \max(0, x - (s - h))\right); \\ \pi(x; S_{s;h}^{\text{bear}}) &= \min\left(1, \frac{1}{h} \max(0, (s + h) - x)\right). \end{aligned} \right\} \quad (4)$$

Нормированные симметричные бафтерфляи с тройкой страйков $s - h$, s , $s + h \in X$ образуются комбинациями

$$B_{s;h} = \frac{1}{h} (C_{s-h} - 2C_s + C_{s+h}) =$$

$$= \frac{1}{h} (P_{s-h} - 2P_s + P_{s+h}) =$$

$$= U + \frac{1}{h} (P_{s-h} - P_s - C_s + C_{s+h}) \quad (5)$$

с платежными функциями

$$\pi(x; B_{s;h}) = \frac{1}{h} \max(0, h - |x - s|). \quad (6)$$

В комбинациях (3) и (5) безрисковый инструмент U выполняет функцию маржевого инструмента и применяется инвестором в соответствии с требованиями рынка. Формально верно еще одно представление:

$$B_{s;h} = \frac{1}{h} (C_{s-h} - C_s - P_s + P_{s+h}),$$

но оно не является *естественным* (страйки коллов в комбинации ниже страйков путов) и потому далее не используется.

Можно было бы рассматривать и не создающие принципиальных трудностей несимметричные бафтерфляи (с неравными по длине сценариями и неравномерной линейкой страйков), но они, как правило, не применяются на рынках и к тому же сильно загромождали бы изложение.

С целью алгоритмической автоматизации дальнейших построений для одномерных опционов P_s и C_s вводятся также обозначения O_s^- (и $O_{0;s}$) и O_s^+ (и $O_{1;s}$), которые могут обрастать дополнительными индексами координат $l \in N$:

$$O_{0;s} \equiv O_s^- \equiv P_s; \quad O_{1;s} \equiv O_s^+ \equiv C_s. \quad (7)$$

Многомерным обобщением одномерных опционов P_s и C_s служат n -мерные α -опционы $A_{\alpha;s}$ векторного типа α и с векторным страйком $s \in \mathfrak{X}^n$, задаваемые вместе с платежными функциями соотношениями

$$A_{\alpha;s} = \prod_{i \in N} O_{i\beta_i;s_i},$$

$$\pi(x, A_{\alpha;s}) = \prod_{l \in N} \pi(x_l; O_{l\beta_l;s_l}), \quad x \in \mathfrak{X}^n,$$

$$\pi(x_l; O_{l\beta_l;s_l}) = \omega_{l\beta_l;s_l}(x_l) =$$

$$= \max(0, \alpha_l(x_l - s_l)), \quad l \in N. \quad (8)$$

Как и в [6], вектор α с компонентами $\alpha_l = \pm 1$, $l \in N$, в индексах инструментов (или просто \pm) определяет векторный тип α -опционов $A_{\alpha;s}$. Вектор $\beta = (\alpha + 1)/2$, дублирующий α , вводится для удобства по техническим причинам и принимает для каждого $l \in N$ два значения:

$$\beta_l = \begin{cases} 0 & \text{для пута;} \\ 1 & \text{для колла.} \end{cases}$$

Для каждого векторного страйка s на n -мерном рынке могут котироваться 2^n типов α -опционов. Рынок n -мерных α -опционов с их k -мерными версиями, $k < n$, называется n -мерным α -рынком.

3 Двумерный дискретный α -рынок

В основе дискретного α -рынка лежит *сценарный* рынок — сценарная дискретизация двумерного теоретического δ -рынка. Как и для бинарного рынка, используется в большей мере адаптированная к двумерному случаю очевидная система обозначений, но учитывается и специфика требований α -рынка.

Цены двух базовых активов *теоретического* двумерного δ -рынка обозначаются x и y , страйки опционов — соответственно s и t , $x, s \in X = [a_1, b_1] \subset \mathfrak{R}$, $y, t \in Y = [a_2, b_2] \subset \mathfrak{R}$. Дискретизация осуществляется равномерным разбиением множества X на v_1 интервалов (сценариев), Y — на v_2 интервалов. Одномерные сценарии на X и Y даются формулами:

$$S_i = [x_{i-1}, x_i], \quad x_i = a_1 + ih_1, \quad h_1 = \frac{b_1 - a_1}{v_1}, \\ i \in I, \quad x_0 = a_1; \quad (9)$$

$$T_j = [y_{j-1}, y_j], \quad y_j = a_2 + jh_2, \quad h_2 = \frac{b_2 - a_2}{v_2}, \\ j \in J, \quad y_0 = a_2, \quad (10)$$

где $I = \{1, 2, \dots, v_1\}$, $J = \{1, 2, \dots, v_2\}$, а номер сценария совпадает с индексом его правой границы. Двумерными сценариями служат прямые произведения всех пар $S_i \times T_j$, $i \in I$, $j \in J$.

На сценарном рынке базис образуют индикаторы сценариев $D_{ij} = H\{S_i \times T_j\}$, но для α -рынка при той же структуре сценариев уместнее использовать иной базис — из баттерфляев B_{ij} , задаваемых с учетом определения (5), но для специально подобранных страйков. Страйками s_i и t_j одномерных опционов и упомянутых баттерфляев служат середины сценариев (9) и (10):

$$s_i = \frac{x_{i-1} + x_i}{2}, \quad i \in I; \quad t_j = \frac{y_{j-1} + y_j}{2}, \quad j \in J.$$

При этом параметр h для баттерфляев (5), равный длине сценариев, определяется в (9) и (10). Для удобства записи формул также доопределяются параметры $s_0 = a_1$, $s_{v_1+1} = b_1$, $t_0 = a_2$, $t_{v_2+1} = b_2$, но они страйками не являются.

Портфель с вектором g весов базисных баттерфляев в двумерном случае приобретает вид:

$$G = \sum_{i \in I, j \in J} g_{ij} B_{ij}. \quad (11)$$

Двумерным обобщением обычных опционов служат инструменты, характеризуемые парой страйков (s_i, t_j) , или просто (i, j) , с дополнительным указанием типа (лучше в терминах $\beta = (\beta_1, \beta_2)$):

$$A_{\beta_1\beta_2;ij} = O_{\beta_1;1,i} O_{\beta_2;2,j} = O_{\beta_1;i} O_{\beta_2;j}, \\ i \in I, \quad j \in J.$$

Также рассматриваются и их одномерные версии, обозначаемые A_i и A_j с маркером «точка» в позиции, отведенной координате безрискового актива.

Для представления произвольного инструмента G (11) в базисе из α -опционов (избыточным в сравнении с базисом из баттерфляев) все нормированные баттерфляи в (11) следует реплицировать в терминах α -опционов.

В соответствии с (8) для конструирования репликаций следует перемножать одномерные представления сценарных баттерфляев, выбирая подходящие сомножители из (3) и (5). Для одномерного рынка с v сценариями $i \in I$ справедливы такие репликации баттерфляев коллами и путами:

$$B_i = \begin{cases} U - \frac{O_1^+ - O_2^+}{h} = \frac{O_2^- - O_1^-}{h}, & i = 1; \\ \frac{O_{i-1}^+ - 2O_i^+ + O_{i+1}^+}{h} = \\ = \frac{O_{i-1}^- - 2O_i^- + O_{i+1}^-}{h} = \\ = U - \frac{O_i^- - O_{i-1}^-}{h} - \\ - \frac{O_i^+ - O_{i+1}^+}{h}, & 1 < i < v; \\ \frac{O_{v-1}^+ - O_v^+}{h} = U - \frac{O_v^- - O_{v-1}^-}{h}, & i = v. \end{cases} \quad (12)$$

Базисные инструменты для $i = \overline{1, v}$ являются спредами, но их для удобства также называем баттерфляями (*усеченными*). Инструмент U в выписанных соотношениях, как в (3) и (5), выполняет функцию маржевого инструмента.

Подобно сценарным базисам для ζ -рынка [6] построение двумерных базисов из α -опционов проводится на основе одномерных базисов, но их элементы на этот раз выбираются из (12). Строятся три варианта репликации базисов: два однотипных (один в путах, другой в коллах) и третий — смешанный естественный. Если в базисе v сценариев, а центральный страйк i_c , то

– однотипный базис при $\alpha = -1$ (в путах):

$$\left. \begin{aligned} B_1^- &= \frac{O_2^* - O_1^-}{h}; \\ B_i^- &= \frac{O_{i-1}^- - 2O_i^- + O_{i+1}^-}{h}; \\ B_v^- &= U - \frac{O_v^- - O_{v-1}^-}{h}; \end{aligned} \right\} \quad (13)$$

– однотипный базис при $\alpha = +1$ (в коллах):

$$\left. \begin{aligned} B_1^+ &\equiv U - \frac{O_1^+ - O_2^+}{h}; \\ B_i^+ &= \frac{O_{i-1}^+ - 2O_i^+ + O_{i+1}^+}{h}; \\ B_v^+ &\equiv \frac{O_{v-1}^+ - O_v^+}{h}; \end{aligned} \right\} \quad (14)$$

– смешанный естественный базис:

$$\left. \begin{aligned} B_1^m &\equiv \frac{O_2^- - O_1^-}{h}; \\ B_i^m &\equiv \frac{O_{i-1}^- - 2O_i^- + O_{i+1}^-}{h}, \quad 0 < i < i_c; \\ B_{i_c}^m &\equiv U - \frac{O_{i_c-1}^- - O_{i_c}^- - O_{i_c}^+ + O_{i_c+1}^+}{h}; \\ B_i^m &\equiv \frac{O_{i-1}^+ - 2O_i^+ + O_{i+1}^+}{h_i}, \quad i_c < i < v; \\ B_v^m &\equiv \frac{O_{v-1}^+ - O_v^+}{h}. \end{aligned} \right\} \quad (15)$$

4 Формирование базисов и платежных функций портфелей α -опционов

На основе соотношений (13)–(15) введенные в многомерном случае произвольной размерности конструкции здесь переписываются для двумерного α -рынка в однотипных и смешанных вариантах.

Поскольку каждый двумерный базисный баттерфляй определяется как произведение двух одномерных (что соответствует перемножению платежных функций), его репликации двумерными α -опционами находятся перемножением пары подходящих представлений из (13)–(15).

Однотипная репликация сценарных баттерфляев проводится α -опционами единого типа $\alpha = \{\alpha_1, \alpha_2\}$. Он фиксируется заранее, и потому используются более простые соотношения (13) и (14), а обозначение типа опциона опускается.

Каждое перемножение сумм одномерных опционов в (13) или (14) дает сумму парных произведений этих опционов, которые затем следует замещать согласно (8) эквивалентными двумерными α -опционами по правилам

$$\begin{aligned} 1 &\rightarrow U, \quad O_{1,i} O_{2,j} \rightarrow A_{ij}, \\ O_{1,i} U &\rightarrow A_{i\cdot}, \quad U O_{2,j} \rightarrow A_{\cdot j}. \end{aligned} \quad (16)$$

Смешанная репликация осуществляется аналогично, но указание типа в обозначениях необходимо, и потому правила трансформации приобретают вид:

$$\begin{aligned} 1 &\rightarrow U, \quad O_{1,i}^{\alpha_1} O_{2,j}^{\alpha_2} \rightarrow A_{ij}^\alpha = A_{\beta_1, \beta_2; ij}, \\ O_{1,i}^{\alpha_1} &\rightarrow A_{\beta_1; i\cdot} (= A_{i\cdot}^{\alpha_1}), \quad O_{2,j}^{\alpha_2} \rightarrow A_{\beta_2; \cdot j} (= A_{\cdot j}^{\alpha_2}), \\ \beta_1, \beta_2 &\in \{0, 1\}, \end{aligned} \quad (17)$$

На двумерном α -рынке в соответствии с числом возможных векторов α насчитываются четыре варианта однотипных базисов и один смешанный (естественный с заданным центром рынка).

Для каждого варианта с *однотипным* базисом и оптимальным портфелем фиксируется тип α , и он становится типом всех α -опционов варианта. В двумерном случае таких типов четыре: $\{-1, -1\}$; $\{-1, +1\}$; $\{+1, -1\}$; $\{+1, +1\}$. Последовательным применением правил (16) ко всем страйкам для каждого значения векторного параметра α находятся искомые четыре базиса. В однотипном случае для каждой компоненты рынка наличествуют три качественно различных представления по варианту страйка — двум крайним и общему внутреннему, и потому их $3^2 = 9$. В качестве примера приводится базис для $\alpha = \{-1, +1\}$, т.е. в терминах α -опционов A_{01} (остальные три *однотипных* базиса выписываются сходным образом), при этом в списке принимается $0 < i < v_1, 0 < j < v_2$:

$$\begin{aligned} B_{1,1} &= \frac{A_{1,1} - A_{1,2} - A_{2,1} + A_{2,2}}{h_1 h_2} + \frac{-A_{1\cdot} + A_{2\cdot}}{h_1}; \\ B_{1,j} &= \frac{-A_{1,j-1} + 2A_{1,j} - A_{1,j+1}}{h_1 h_2} + \frac{A_{2,j-1} - 2A_{2,j} + A_{2,j+1}}{h_1 h_2}; \\ B_{1,v_2} &= \frac{-A_{1,v_2-1} + A_{1,v_2} + A_{2,v_2-1} - A_{2,v_2}}{h_1 h_2}; \end{aligned}$$

$$\begin{aligned}
 B_{i_c, j_c} = & 1 + \frac{A_{00; i_c-1, j_c-1} - A_{00; i_c-1, j_c}}{h_1 h_2} + \\
 & + \frac{-A_{00; i_c, j_c-1} + A_{00; i_c, j_c} - A_{01; i_c-1, j_c}}{h_1 h_2} + \\
 & + \frac{A_{01; i_c-1, j_c+1} + A_{01; i_c, j_c} - A_{01; i_c, j_c+1}}{h_1 h_2} + \\
 & + \frac{-A_{10; i_c, j_c-1} + A_{10; i_c, j_c} + A_{10; i_c+1, j_c-1}}{h_1 h_2} + \\
 & + \frac{-A_{10; i_c+1, j_c} + A_{11; i_c, j_c} - A_{11; i_c, j_c+1}}{h_1 h_2} + \\
 & + \frac{-A_{11; i_c+1, j_c} + A_{11; i_c+1, j_c+1}}{h_1 h_2} + \\
 & + \frac{A_{0; i_c-1, \cdot} - A_{0; i_c, \cdot} - A_{1; i_c, \cdot} + A_{1; i_c+1, \cdot}}{h_1} + \\
 & + \frac{A_{0; \cdot, j_c-1} - A_{0; \cdot, j_c} - A_{1; \cdot, j_c} + A_{1; \cdot, j_c+1}}{h_2}.
 \end{aligned}$$

В этом списке присутствуют обозначения инструментов A с четырьмя и тремя индексами. В первой группе пара индексов до точки с запятой означает тип двумерного α -опциона (17), а после нее — его страйк. Во второй группе представлены одномерные версии двумерных α -опционов. Индекс до точки с запятой означает тип опциона, числовой индекс после нее — его страйк, а позиция маркера «точка» показывает координату безрискового актива.

Сценарные баттерфляи, полученные из α -опционов, позволяют произвольный инструмент на рынке представить в виде портфеля α -опционов. Для нахождения его доходов следует воспользоваться соотношениями (2) с учетом переопределения (7). Так, в однотипном случае платежная функция портфеля находится в соответствии с (16) по правилам:

$$\begin{aligned}
 U \rightarrow 1, \quad A_{ij} & \rightarrow \omega_{1;i}(x)\omega_{2;j}(y), \\
 A_{i \cdot} & \rightarrow \omega_{1;i}(x), \quad A_{\cdot j} \rightarrow \omega_{2;j}(y); \quad (18)
 \end{aligned}$$

$$\begin{aligned}
 \omega_{\beta_1; i}(x) & = \max(0, \alpha_1(x - s_i)), \quad i \in I; \\
 \omega_{2; i}(y) & = \max(0, \alpha_2(y - t_j)), \quad j \in J.
 \end{aligned}$$

Аналогично в соответствии с (17) записываются в смешанном случае правила формирования платежных функций:

$$\begin{aligned}
 U \rightarrow 1, \quad A_{\beta; ij} & \rightarrow \omega_{\beta_1, 1; i}(x)\omega_{\beta_2, 2; j}(y), \\
 A_{\beta_1; i \cdot} & \rightarrow \omega_{\beta_1, 1; i}(x), \quad A_{\beta_2; \cdot j} \rightarrow \omega_{\beta_2, 2; j}(y); \\
 \omega_{\beta_1, 1; i}(x) & = \max(0, \alpha_1(x - s_i)), \quad i \in I, \\
 \omega_{\beta_2, 2; j}(y) & = \max(0, \alpha_2(y - t_j)), \quad j \in J. \quad (19)
 \end{aligned}$$

5 Иллюстративный пример

Для построения мер $C\{\cdot\}$ и $P\{\cdot\}$ и их сравнительного анализа данные в примере заимствуются из [6]. Так, принимается $X = [0, 1)$, $Y = [0, 1)$, а для $F_{CX}(x)$ и $F_{CY}(y)$ выбираются бета-распределения с параметрами $\{3/2, 2\}$ и $\{3/2, 3\}$ соответственно, для $F_{PX}(x)$ и $F_{PY}(y) - \{2, 3\}$ и $\{2, 4\}$:

$$\begin{aligned}
 F_{CX}(x) & = \frac{x^{3/2}(5 - 3x)}{2}; \\
 F_{CY}(y) & = \frac{y^{3/2}(35 - 42y + 15y^2)}{8}; \\
 F_{PX}(x) & = x^2(6 - 8x + 3x^2); \\
 F_{PY}(y) & = y^2(10 - 20y + 15y^2 - 4y^3).
 \end{aligned}$$

Из них совместные функции распределения для обеих мер строятся как

$$\begin{aligned}
 F(x, y) & = \\
 & = F_X(x)F_Y(y)(1 + 3\kappa(1 - F_X(x))(1 - F_Y(y))). \quad (20)
 \end{aligned}$$

Искомые двумерные функции распределения $F_C(x, y)$ и $F_P(x, y)$ определяются подстановкой в (20) в качестве параметра, отвечающего за корреляционную связь компонент, соответственно $\kappa_c = 0$ и $\kappa_p = 0,2$. Из них простым смешанным дифференцированием по обоим переменным находятся плотности $c(x, y)$ и $p(x, y)$, но ввиду громоздкости записей они здесь не приводятся.

Двумерная дискретизация множества $X \times Y$ в примере проводится также при $v_1 = 6$ и $v_2 = 5$, а центральным выбирается страйк $i_c = 3, j_c = 3$.

В отличие от ζ -рынков [6], для которых в целях применения дискретного алгоритма находились стоимости сценарных индикаторов, для α -рынков следует вычислять стоимости сценарных баттерфляев. И потому для адекватного сравнения относительных доходов естественно вычислять и их средние доходы. И те и другие, а это векторы c^B и p^B , определяются интегрированием платежных функций (4) и (6) с плотностями $c(x, y)$ и $p(x, y)$ соответственно.

Применением к этим векторам дискретного алгоритма оптимизации [6], основанного на процедуре Неймана—Пирсона [7], определяется вектор весов базисных баттерфляев для оптимального двумерного портфеля. При этом в качестве функции рискованных предпочтений выбирается $\varphi(\varepsilon) = \varepsilon^2, \varepsilon \in [0, 1]$.

В результате получается вектор весов

$$\begin{aligned}
 g = \{ & 0,118; 0,159; 0,0113; 0,000219; 0,000008; 0,414; \\
 & 1,0; 0,228; 0,0151; 0,000989; 0,0739; 0,788; 0,602; \\
 & 0,0873; 0,00175; 0,0069; 0,309; 0,495; 0,176; 0,00191;
 \end{aligned}$$

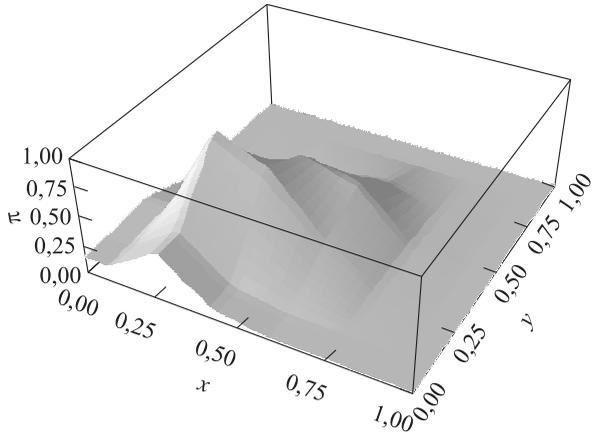


Рис. 1 Доходы оптимального опционного портфеля при дискретизации 6×5

$\{0,000907; 0,0254; 0,0405; 0,0291; 0,00159; 0,0000058;$
 $0,0000848; 0,00151; 0,00112; 0,000009\}.$

Он порождает оптимальный портфель (11) с инвестиционной суммой, средним доходом и средней доходностью соответственно $A = 0,28642$, $R = 0,364418$ и $y = 0,272317$. График его платежной функции изображен на рис. 1. Для сравнения на рис. 2 приведен аналогичный график для сценарного рынка при дискретизации 40×40 .

По понятным причинам графики платежных функций на рис. 1 и 2 демонстрируют большее взаимное сходство, чем аналогичная пара графиков из [6]. Но и различие между собой графиков на рис. 1 и 2 представляется естественным.

Остается определить оптимальные портфели в терминах α -опционов рассматриваемых типов. Для нахождения каждого такого представления в формулу (11) следует для всех пар (i, j) подставлять вместо индикаторов B_{ij} соответствующие им представления в α -опционах. В результате после упрощений получаются четыре однотипных портфеля и один смешанный. Так, оптимальный *однотипный* портфель для $\alpha = \{-1, +1\}$, т. е. образованный путами для первого актива и коллами — для второго:

$$G_{01} = 0,000006U + 16,369A_{11} - 35,107A_{12} + 12,681A_{13} + 5,641A_{14} + 0,416A_{15} + 1,774A_{16} - 12,549A_{21} + 48,875A_{22} - 39,318A_{23} + 1,263A_{24} + 1,729A_{25} - 3,812A_{26} - 16,162A_{31} + 9,717A_{32} + 21,364A_{33} - 15,431A_{34} + 0,512A_{35} + 1,636A_{36} + 4,007A_{41} - 20,268A_{42} + 19,612A_{43} + 3,707A_{44} - 7,058A_{45} + 0,366A_{46} + 7,603A_{51} - 2,899A_{52} - 13,593A_{53} + 5,279A_{54} + 3,609A_{55} + 0,031A_{56} + 0,731A_{61} - 0,318A_{62} - 0,745A_{63} - 0,458A_{64} +$$

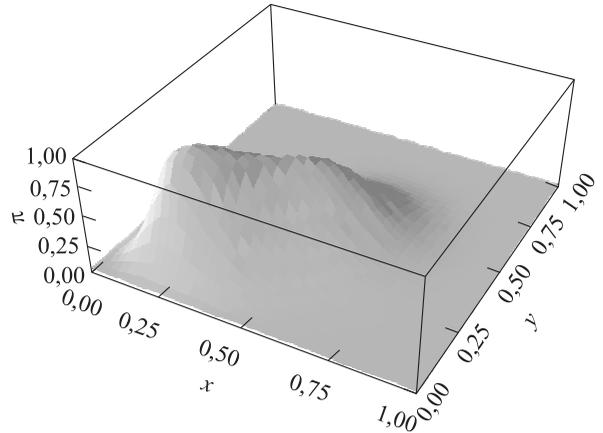


Рис. 2 Доходы сценарного портфеля при дискретизации 40×40

$+ 0,791A_{65} + 0,005A_{66} + 0,0004A_{67} + 0,007A_{68} - 0,009A_{69} - 0,004A_{70} + 0,006A_{71}.$

Платежные функции всех однотипных портфелей получаются по правилам (18). Проведенные расчеты подтверждают верность алгоритма. Все они, несмотря на внешнее различие их представлений, на идеальном рынке должны иметь единую платежную функцию с графиком, представленным на рис. 1.

Оптимальный *смешанный портфель* строится вновь по формуле (11), но в смешанном базисе естественного происхождения с выделенным центральным страйком (3, 3). В первом квадранте (относительно центра рынка) используются α -опционы A_{11} , во втором — A_{01} , в третьем — A_{00} , в четвертом — A_{10} .

Вычисления с применением (11) дают оптимальный *смешанный портфель*:

$$G_m = 0,602U + 16,369A_{00;11} - 35,107A_{00;12} + 18,737A_{00;13} - 12,549A_{00;21} + 48,876A_{00;22} - 36,326A_{00;23} - 3,82A_{00;31} - 13,769A_{00;32} + 17,589A_{00;33} - 6,056A_{01;13} + 5,641A_{01;14} + 0,416A_{01;15} - 2,992A_{01;23} + 1,263A_{01;24} + 1,729A_{01;25} + 9,049A_{01;33} - 6,903A_{01;34} - 2,145A_{01;35} - 12,342A_{10;31} + 23,486A_{10;32} - 11,144A_{10;33} + 4,007A_{10;41} - 20,268A_{10;42} + 16,261A_{10;43} + 7,603A_{10;51} - 2,899A_{10;52} - 4,704A_{10;53} + 0,731A_{10;61} - 0,318A_{10;62} - 0,413A_{10;63} + 5,871A_{11;33} - 8,528A_{11;34} + 2,657A_{11;35} + 3,351A_{11;43} + 3,707A_{11;44} - 7,058A_{11;45} - 8,889A_{11;53} + 5,279A_{11;54} + 3,609A_{11;55} - 0,333A_{11;63} - 0,458A_{11;64} +$$

$$\begin{aligned}
& + 0,791A_{11;65} + 1,3A_{0;1} + 0,943A_{0;2} - 2,243A_{0;3} + \\
& + 3,568A_{0;1} - 4,497A_{0;2} + 0,929A_{0;3} - 0,642A_{1;3} - \\
& - 2,085A_{1;4} + 2,492A_{1;5} + 0,234A_{1;6} - 2,573A_{1;3} + \\
& + 2,145A_{1;4} + 0,428A_{1;5}.
\end{aligned}$$

В этом портфеле 56 инструментов; среди них один безрисковый актив, 42 двумерных α -опциона A_{00} , A_{01} , A_{10} и A_{11} и 13 одномерных версий A_0 , A_1 в количествах 9, 9, 12, 12 и 6, 7 соответственно.

Двумерные α -опционы снабжены четырьмя индексами; первые два из них (до точки с запятой) показывают тип опциона по каждой координате в терминах β , другие два индекса — номера страйков.

Одномерные версии снабжены двумя индексами и маркером «точка». Один индекс до точки с запятой указывает тип опциона, индекс после нее — номер страйка, а маркер — координату подразумеваемого безрискового актива.

Платежная функция смешанного портфеля строится по правилам (19). Ее графиком служит все тот же график на рис. 1.

6 Заключение

В работе решена задача алгоритмического нахождения представлений многомерных баттерфляев при сценарной дискретизации рынка и построения из них базиса в терминах α -опционов — многомерного обобщения традиционных опционов колл и пут. На конкретном примере двумерного рынка продемонстрирована работа этого алгоритма и ее результат. Подобные расчеты могут быть без принципиальных трудностей реализованы и для рынков большей размерности. Поскольку

ку баттерфляя для одномерного рынка образуются из трех страйков, а индикаторы — из двух, то их многомерные репликации получаются еще более громоздкими. Фактические расчеты, проведенные для $n = 4$, показали, что уже объем перечня базисных инструментов однотипного базиса, например в терминах A_{0000} , аналогичного A_{00} из разд. 4, был соразмерен всему объему настоящей работы, а потому и на многомерных α -рынках лучше торговать непосредственно баттерфляями, а не α -опционами.

Литература

1. *Агасандян Г. А.* Применение непрерывного критерия VaR на финансовых рынках. — М.: ВЦ РАН, 2011. 299 с.
2. *Агасандян Г. А.* Непрерывный критерий VaR и оптимальный портфель инвестора // Управление большими системами, 2018. Вып. 73. С. 6–26.
3. *Агасандян Г. А.* Непрерывный критерий VaR на сценарных рынках // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 32–40.
4. *Агасандян Г. А.* Вычисление показателей оптимальных по CC-VaR портфелей на рынках опционов // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 75–84.
5. *Агасандян Г. А.* Многомерные рынки опционов и оптимизация по CC-VaR // Управление большими системами, 2020. Вып. 88. С. 5–25.
6. *Агасандян Г. А.* Многомерные бинарные рынки и CC-VaR // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 2–10.
7. *Кramer Г.* Математические методы статистики / Пер. с англ. — М.: Мир, 1975. 750 с. (*Cramer H.* Mathematical methods of statistics. — Princeton, NJ, USA: Princeton University Press, 1946. 575 p.)

Поступила в редакцию 09.03.22

MULTIDIMENSIONAL BUTTERFLIES IN PROBLEMS OF OPTIMIZATION ON CC-VaR

G. A. Agasandyan

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

Abstract: The work continues studying problems of using continuous VaR-criterion (CC-VaR) in financial markets. Again some technical problems are concerned. However, they emerge this time not in multidimensional relatively simple binary markets but in multidimensional markets that are an extension of one-dimensional traditional markets of options such as calls and puts. In assumption that scenario butterflies are not traded in markets directly, a method of receiving their replication from multidimensional options, i. e., α -options, is developed. It is based on options parity theorems and can be applied to markets of arbitrary dimension, but actual realization is conducted for two-dimensional markets. The bases constructions in terms of α -options both one-type and natural mixed with

selected market center are produced. Theoretical representations of optimal portfolios in these bases accompanied with the payoffs diagram are illustrated by the distinctive example of a two-dimensional market.

Keywords: underliers; multidimensional market; investor's risk preferences function; continuous VaR-criterion; cost and forecast densities; scenario indicators; bases; binary options; one-type portfolio; market center; mixed portfolio

DOI: 10.14357/19922264230114

References

1. Agasandyan, G. A. 2011. *Primenenie kontinual'nogo kriteriya VaR na finansovykh ryinkakh* [Application of continuous VaR-criterion in financial markets]. Moscow: CCRAS. 299 p.
2. Agasandyan, G. A. 2018. Kontinual'nyy kriteriy VaR i optimal'nyy portfel' investora [Continuous VaR-criterion and investor's optimal portfolio]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 73:6–26.
3. Agasandyan, G. A. 2018. Kontinual'nyy kriteriy VaR na stsenarnykh ryinkakh [Continuous VaR-criterion in scenario markets]. *Informatika i ee Primeneniya — Inform. Appl.* 12(1):32–40.
4. Agasandyan, G. A. 2019. Vychislenie pokazateley optimal'nykh po CC-VaR portfeley na ryinkakh opsionov [Performance estimations for optimal-on-CC-VaR portfolios in option markets]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):75–84.
5. Agasandyan, G. A. 2020. Mnogomernye rynki opsionov i optimizatsiya po CC-VaR [Multidimensional option markets and optimization on CC-VaR]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 88:5–25.
6. Agasandyan, G. A. 2022. Mnogomernye binarnye rynki i CC-VaR [Multidimensional binary markets and CC-VaR]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):2–10.
7. Cramer, H. 1946. *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press. 575 p.

Received March 9, 2022

Contributor

Agasandyan Gennady A. (b. 1941) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; agasand17@yandex.ru

ДАННЫЕ, ИНФОРМАЦИЯ И ЗНАНИЕ В НАУЧНОЙ ПАРАДИГМЕ ИНФОРМАТИКИ*

И. М. Зацман¹

Аннотация: Рассматриваются три базовых понятия информатики: данные, информация и знание. Предлагается вариант специфицирования этих понятий в рамках построения системы терминов научной парадигмы информатики как фундаментальной науки. С одной стороны, понятия «данные», «информация» и «знание» широко используются в научной литературе и учебниках по информатике, в частности при описании ее теоретических оснований. С другой стороны, до сих пор отсутствует консенсус по их смысловому содержанию. Сложившаяся ситуация, скорее всего, обусловлена распространенной пресуппозицией (имплицитным предположением), что рассматриваемые понятия выражают некоторые объективные сущности предметной области информатики. В статье предполагается, что они выражают интересующие сущности, которые по своей природе возникают как предметы мысли в результате договоренности, т. е. не являются объективно существующими. С точки зрения треугольника Фреге (предмет—понятие—слово, которое выражает понятие и обозначает предмет) для объективных сущностей первичной вершиной треугольника служит предмет, в результате изучения которого появляются понятие и слово. Для интересующих сущностей первичной вершиной служит понятие, варианты дефиниции которого необходимо обсудить в интересах достижения консенсуса. Если его удастся достичь, то именно в процессе обсуждения появляются предмет мысли, слово, его обозначающее и выражающее понятие, которые вместе и образуют треугольник Фреге. Цель статьи — специфицировать базовые понятия информатики как выражающие интересующие сущности и являющиеся первичными вершинами треугольника Фреге, распределить эти сущности по средам ее предметной области, выделяя границы между ними, и рассмотреть отношения между этими сущностями на выделенных границах.

Ключевые слова: научная парадигма; информатика как фундаментальная наука; данные; информация; знание; интересующие сущности информатики; треугольник Фреге

DOI: 10.14357/19922264230115

1 Введение

Томас Элиот строками поэмы «The Rock» [1] сформулировал следующие вопросы, используя слова «информация», «знание» и «мудрость»²:

‘Где та мудрость, которую мы потеряли в знании?
Где то знание, которое мы потеряли в информации?’

В 1989 г. Рассел Акофф добавил к ним четвертое слово — «данные», ввел понятие «иерархия³ DIKW» (data, information, knowledge, wisdom — данные, информация, знания, мудрость) и описал отношения между ними так: «Мудрость находится на вершине иерархии... По нисходящей от мудрости идут понимание, знание, информация и, в самом низу, данные. Каждое из перечисленных понятий [кроме данных] содержит в себе нижестоящие, например

не может быть мудрости без понимания и понимания без знания» [2].

Отметим, что при использовании иерархии DIKW рассматривают чаще только четыре ее уровня (данные, информация, знания, мудрость), т. е. для *понимания* отдельный уровень, как правило, не выделяется [3]. Иногда ограничиваются рассмотрением первых трех понятий — данные, информация, знания — и описанием отношений между ними [4].

Акофф приводит их толкование, что помогает понять смысл его фразы о том, что «*перечисленные понятия содержат в себе нижестоящие*». Данные определяются им как наборы символов, которые *характеризуют свойства объектов и событий, а также их окружение*. Эти наборы формируются в процессе наблюдения или зондирования.

* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, izatsman@yandex.ru

² Строки поэмы даны в переводе автора статьи. Поэма на английском языке включает следующие строки [1]:

Where is the wisdom that we have lost in knowledge?

Where is the knowledge that we have lost in information?

³ Иногда используют словосочетание «пирамида DIKW», но в статье Акоффа говорится именно об иерархии [2].

Информация определяется как *результат анализа данных*, выполняемый в том числе и с помощью компьютеров. Знание рассматривается им в контексте управления организационными системами в сфере экономики: это то, что делает возможным *преобразование информации в инструкции*, и это делает возможным управление такими системами. Данные Акофф сравнивает с металлической рудой, которая ценится меньше по сравнению с результатом ее преобразования (где металл выступает как метафора информации) [2].

Таким образом, когда Акофф говорит о том, что вышестоящие понятия включают нижестоящие, то речь идет о процессах их трансформации, а не о включенности их смыслового содержания, которая описывается родовидовыми или иными иерархическими отношениями [5]. На эту особенность описания иерархии DIKW обратила внимание Дженнифер Роули в обзоре работ, посвященных иерархии DIKW: «Обычно [в иерархии] информация определяется в терминах данных, знание в терминах информации и мудрость в терминах знаний, но существует меньше консенсуса в описании процессов, которые трансформируют сущности, расположенные ниже в иерархии, в те, которые находятся над ними, что приводит к отсутствию четкости определений этих терминов» [3].

Иерархия DIKW как упрощенная модель отношений между данными, информацией и знанием имеет как своих сторонников, так и критиков [6, 7]. Например, Давид Вайнбергер ее недостатки описал так: «Знание — это не просто результат фильтрации [сущностей нижних уровней иерархии] или применения некоторых алгоритмов. Оно — результат гораздо более сложного процесса [его генерации], который является *социальным, целенаправленным, контекстно и культурно обусловленным*. . . Наиболее важным в этом отношении является то, что там, где решения сложные и знание [необходимое для решения] получить трудно, *она не определяется информацией*, поскольку именно процесс познания в первую очередь определяет, какая информация необходима и как ее следует использовать. . . Представление о том, что знание (а тем более мудрость) является результатом применения фильтров на каждом уровне, рисует неверную картину. . . Знание — это результат *креативного* процесса, и оно может иметь *лакуны*» [6] (курсив мой. — И. З.).

Цель статьи — рассмотреть базовые понятия информатики как выражающие интересующие сущности и являющиеся первичными вершинами треугольника Фреге, распределить эти сущности по средам ее предметной области, выделяя грани-

цы между ними, и рассмотреть отношения между этими сущностями на выделенных границах в рамках предлагаемой научной парадигмы информатики [8].

Необходимость в более детальной спецификации базовых понятий информатики (по сравнению с иерархией DIKW) обусловлена тем, что при описании ее научных направлений отличные по своей природе сущности, а иногда имеющие *взаимоисключающие свойства*, часто называют одинаково, например словом «информация» [9, 10]. Таким образом, кардинальные понятийные различия сущностей предметной области информатики часто не выражаются лексически, что существенно усложняет описание отношений между этими сущностями. В статье различные по своей природе и принципиальным свойствам сущности предлагается специфицировать терминологически за счет использования отличающихся лексических средств их описания.

2 Контекст построения научной парадигмы

Сложность построения и сопоставления вариантов научной парадигмы информатики обусловлена отсутствием конвенциональных границ ее предметной области и консенсуса по смысловому содержанию ее базовых понятий [11–15]. Дискуссии о природе данных, информации, знания ведутся уже десятки лет. Согласно Ю. А. Шрейдеру, «споры о предмете [информатики] — это не столько споры об объективной истине, сколько стремление отстоять свой взгляд на предмет, сделать его фактом общественного сознания сообщества исследователей. Это не значит, что представление о предмете произвольно — оно должно выражать плодотворный и эвристичный взгляд на реально осуществляемую деятельность, возможности ее развития и перспективы использования» [16]. Однако «плодотворные взгляды» могут различаться концептуально, и в статье излагается только один из возможных вариантов.

Для построения предлагаемого варианта научной парадигмы информатики выбрано десять публикаций как источников его зарождения [16–25], из которых взяты следующие ключевые положения:

- согласно Ю. А. Шрейдеру, определение базовых понятий информатики — это результат договоренности в сообществе исследователей [16], а не итоги изучения объективно существующих предметов¹, и поэтому в рамках предлагаемой

¹ См. в [26] подход к определению информации, основанный на интерпретации объективных сущностей.

парадигмы они и позиционируются как *интерсубъективные сущности*¹ (курсив мой. — И. З.);

- согласно Р. С. Гиляревскому, «если *данные воспринимаются и интерпретируются человеком*, то они становятся для него *информацией*» (курсив мой. — И. З.) [19, с. 19];
- в отчете «Глубинное изменение — технологические переломные моменты и социальное воздействие» [22]² предложен перечень новых информационных технологий (ИТ), которые определяют кардинальный характер преобразования общества и экономики, получившего название «Четвертая промышленная революция» [23], и сказано, что прогнозируемый характер преобразования во многом будет обусловлен, в частности, теми ИТ, которые охватывают *сущности, принадлежащие средам разной природы* (курсив мой. — И. З.);
- в обзоре «Informatics Education in Europe» [24]³, опубликованном в 2017 г. и предваряющем разработку Европейской стратегии компьютерного образования «Informatics for All» [25], дана следующая характеристика современной информатики: «В то время как естественные науки определяются применительно к миру (world), в котором мы живем, информатику как научную дисциплину определить сложнее; у нее нет эмпирических основ, как у естественных наук; это нечто большее, чем мышление формальными символами, как в математике; и это далеко не просто компиляция инженерных принципов и технологий» (курсив мой. — И. З.), что также обуславливает интерсубъективность базовых понятий информатики;
- в статье [18] Нонака разделяет явное и неявное знание и дает схему процессов их преобразования, которую планируется использовать при определении одной из составляющих парадигмы информатики, а именно: ее базовых мето-

дов и моделей (эту составляющую планируется описать в отдельной статье);

- в работах [20, 21] Деннинг и Розенблум предложили сгруппировать научные дисциплины в четырех отраслях знания и включить исследование информационных трансформаций в технических, живых и социальных системах в четвертую отрасль знания;
- в докладе Кристена Ньюгора [17]⁴ дано следующее определение информатики: «это наука, область [исследований] которой охватывает информационные процессы и *связанные с ними феномены в артефактах, обществе и природе*» (курсив мой. — И. З.).

В работе Кристена Ньюгора дается ссылка на определение понятия «феномен» в словаре Webster⁵, а потом на примерах он дает расширенное толкование этого понятия, которое концептуально отличается от определения по словарю Webster 1960 г.: «Важными примерами феноменов являются: живые организмы, неодушевленные объекты. . . Мы также можем говорить о *когнитивных феноменах*, происходящих в сознании людей, в отличие от *явных [сенсорно воспринимаемых] феноменов*, находящихся вне сознания» (курсив мой — И. З.) [17].

Это расширенное толкование имплицитно вводит в предметную область информатики по Ньюгору разделение ее сущностей на категории разной природы: *ментальной* (например, результаты процессов генерации научного знания, происходящих в сознании исследователей) и *сенсорно воспринимаемой*, к которой относятся, в частности, знаковые формы представления результатов процессов генерации знания после его деления на концепты.

На основе определения информатики по Ньюгору в работе [29] было дано описание основания для построения верхнего уровня классификации сущностей ее предметной области как феноменов разной природы и формирования следующих 5 ее

¹ С точки зрения треугольника Фреге (*предмет—понятие—слово*) для объективных сущностей первичной вершиной треугольника служит *предмет*, в результате изучения которого появляются *понятие* и *слово*. Для интерсубъективных сущностей первичной вершиной служит *понятие*, варианты дефиниции которого необходимо обсудить в интересах достижения консенсуса. Если его удастся достичь, то именно в процессе обсуждения появляются *предмет* мысли, *слово*, его обозначающее и выражающее *понятие*, которые вместе и образуют треугольник Фреге в результате генезиса интерсубъективной сущности.

² В подготовке материалов для этого отчета, который был разработан под эгидой Всемирного экономического форума (Давос, Швейцария), принимали участие около 800 экспертов и руководителей отрасли информационных и коммуникационных технологий.

³ В обзор включены результаты двухлетнего мониторинга состояния систем преподавания информатики в регионах и странах Европы, включая РФ, а также в Израиле.

⁴ Доклад был представлен на Всемирном конгрессе IFIP (International Federation of Information Processing — Международная федерация по обработке информации) в Дублине в 1986 г.

⁵ Феномен — любой факт, обстоятельство или событие, которые сенсорно воспринимаются и которые могут быть научно описаны или оценены» [27]. Отметим, что толкование понятия «феномен», существенно расширенное по объему его значения, позже было включено в современный онлайн-словарь Merriam-Webster (в новом определении удалена сенсорная воспринимаемость для феноменов, поддающихся научному описанию) [28].

сред¹, включающих сущности одной и той же природы:

- (1) *ментальная среда* — это совокупность когнитивных феноменов, формируемых в процессах познания, происходящих в сознании людей (далее — концепты);
- (2) *информационная среда* — это совокупность сенсорно воспринимаемых феноменов, находящихся вне сознания;
- (3) *цифровая среда* — это совокупность компьютерных кодов;
- (4) *нейросреда* — это электрические потенциалы и магнитные поля, генерируемые мозгом, которые используются в ИТ управления роботизированной рукой [31] и в других ИТ, применяющих интерфейсы «мозг—компьютер»;
- (5) *ДНК-среда* — это совокупность естественных цепочек РНК и ДНК².

В соответствии с перечисленными средами верхний уровень классификации сущностей предметной области информатики включает как минимум 5 классов, каждый из которых содержит объекты одной среды: ментальной, информационной, цифровой, нейро- или ДНК-среды. При этом с ростом разнообразия природы сущностей верхний уровень классификации может пополняться новыми классами, природа сущностей которых отличается от природы сред, ранее включенных в этот уровень классификации [8].

Это может произойти, например, в том случае, когда при проектировании ИТ встретятся сущности, которые по своей природе не относятся ни к одной из ранее уже выделенных сред [32]. Таким образом, в предлагаемом варианте парадигмы информатики верхний уровень классификации ее предметной области предлагается сделать открытым, что обусловлено разнообразием природы ее сущностей и возможным включением в будущем в ее предметную область сущностей ранее не рассматривавшейся природы (открытость уровня классификации).

3 Составляющие научной парадигмы информатики

Согласно А. Соломонику, научная парадигма «зрелой» науки состоит из следующих четырех со-

ставляющих, которые могут разрабатываться отдельно, но объединяются в единую и цельную конструкцию [33, с. 23–24]:

- (1) философские основания;
- (2) аксиоматика;
- (3) классификация исследуемых объектов и процессов;
- (4) система терминов.

Сам термин «научная парадигма» трактуется им в соответствии с теорией Т. Куна, которая описывает процесс смены научных парадигм [34]. При этом А. Соломоник отмечает тот факт, что в книге Куна мы не находим ответа на вопрос: «Из чего должна состоять парадигма любой «зрелой» науки?» [33, с. 23].

Отметим, что в предлагаемом варианте парадигмы информатики планируется добавить пятую составляющую: базовые методы и модели. Благодаря этому появится возможность дать дефиниции ключевых понятий и описать отношения между ними в рамках системы терминов, тогда как процессы взаимного преобразования обозначаемых ими сущностей будут описаны в рамках базовых моделей и методов. Иными словами, описания понятий, терминов, которые их обозначают, и процессов преобразования обозначаемых ими сущностей будут распределены по разным составляющим парадигмы.

В этой статье спецификация значений трех интерсубъективных сущностей (данные, информация, знание), широко используемых, но не имеющих конвенциональных дефиниций в информатике, ограничена только тремя средами ее предметной области: ментальной, информационной и цифровой. Иначе говоря, в статье следующий уровень классификации сущностей предметной области информатики специфицирует только сущности ментальной, информационной и цифровой природы, а сущности остальных сред предметной области не рассматриваются.

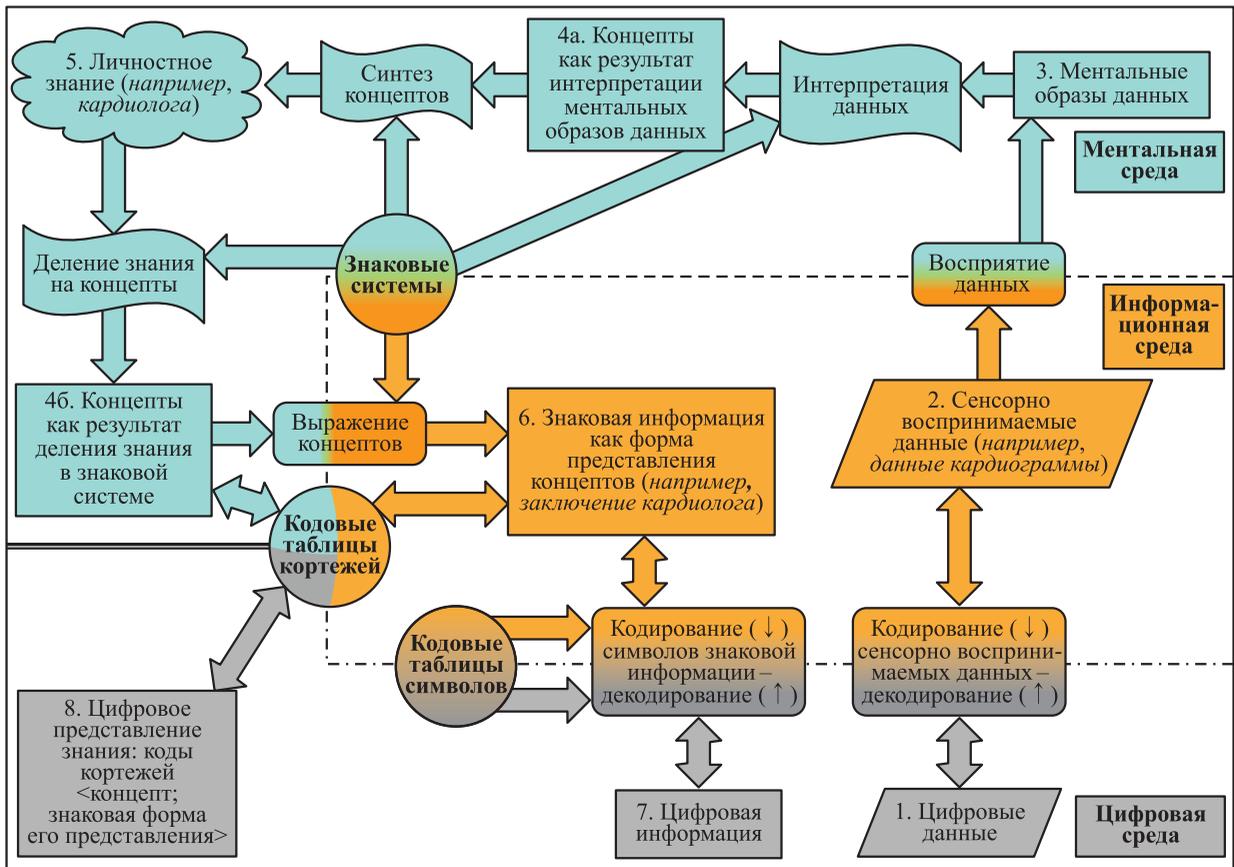
Важно отметить, что согласно предложенному определению информационной среды она содержит сенсорно воспринимаемые данные, например кардиограмму, сформированную в процессе регистрации электрических полей, образующихся при работе сердца. Информация из заключения кардиолога, созданная им как результат содержательного анализа данных кардиограммы, также является сенсорно воспринимаемой, знаковой и принадлежит

¹ Другое значение термина «среда» было определено К. К. Колиным для описания структуры научных исследований, относящихся к комплексу наук об информации [30], а не для определения классов сущностей как феноменов разной природы, относящихся к предметной области информатики.

² Например, модели трансляции естественных ДНК, созданные микробиологами, используются при разработке методов записи и хранения данных с использованием синтезированных цепочек ДНК.

к той же среде. Таким образом, информационная среда содержит как минимум *знаковую информацию* и *сенсорно воспринимаемые данные*. При их компьютерном кодировании получаем соответственно *цифровую информацию* и *цифровые данные*, т.е. две принципиально разные сущности цифровой природы. При этом в предлагаемом делении предметной области информатики цифровая среда включает также результаты кодирования кортежей вида <концепт; знаковая форма его представления>. Совокупность компьютерных кодов таких кортежей предлагается назвать *цифровым представлением знания*. Они включают и коды дефиниций концептов, и коды слов, которые их обозначают [35]. Использование кодов кортежей служит основой разрешения проблемы асимметрии в цифровой среде при кодировании синонимов и омонимов (например, коды кортежей <(заплетенные волосы); коса>, <(с.-х. орудие, используемое при скашивании); коса> и <(намывная полоса суши, причлененная к берегу [36]); коса> будут отличаться из-за разных кодов трех концептов, но коды слова «коса» будут одинаковыми в кодах кортежей).

Отметим, что процесс содержательного анализа данных кардиограммы и подготовки заключения состоит из нескольких этапов, в процессе описания которых дадим определения сущностям, выделяемым в трех средах. Сначала кардиолог сенсорно воспринимает *данные* кардиограммы, полученной в процессе *цифрового мониторинга* работы сердца и последующего *декодирования* его результатов. На следующем этапе появляются *ментальные образы данных* в сознании кардиолога как *результат сенсорного восприятия данных*. Затем генерируется *знание* кардиолога о наличии или отсутствии нарушений в работе сердца как *результат креативного процесса понимания ментальных образов данных*. Потом следует этап подготовки заключения на некотором естественном языке, включающий генерацию *концептов* и как результат *деление полученного знания* и их выражение словами этого языка. Если готовится двуязычное заключение, например параллельно на русском и английском языке, то деление знания на концепты выполняется по-разному в системах этих двух языков. Другими словами, деление одного и того же знания на концепты зависит от



Три среды предметной области информатики, восемь их сущностей и отношения между ними

используемого естественного языка как вербальной знаковой системы. Отметим, что виды знаковых систем, обуславливающих деление знания на концепты, относятся к третьему по счету уровню классификации, который планируется рассмотреть в отдельной статье.

Подведем итоги построения второго уровня классификации объектов предметной области информатики, на котором в этой статье определены восемь сущностей (см. рисунок), а также описаны и отношения между ними:

- ментальная среда содержит как минимум *ментальные образы данных* и *знание*, деление которого на *концепты* определяется используемым естественным языком (в общем случае деление на концепты определяется используемой знаковой системой того или иного вида);
- информационная среда содержит как минимум *сенсорно воспринимаемые данные* и *знаковую информацию*;
- цифровая среда содержит как минимум *цифровые данные*, *цифровую информацию* и *коды кортежей* вида (концепт; знаковая форма его представления).

Приведенный пример с кардиограммой иллюстрирует только отдельные виды отношений между восемью описанными сущностями. Обобщенный вариант отношений между ними приведен на рисунке, который включает и частный случай отношений между сущностями в примере с кардиограммой. Отметим, что кардиолог может использовать данные предыдущих кардиограмм и ранее сделанные заключения, компьютерные коды которых хранятся в информационной медицинской системе.

Пример с геофизическими данными и генерацией на их основе *невербальной знаковой информации* приведен в [37].

Опишем кратко отношения между восемью сущностями трех сред (см. рисунок):

- *концепты знания* человека — это либо результат деления этого знания на составляющие с использованием некоторой знаковой системы, либо результат восприятия и семантической интерпретации данных, либо результат понимания знаковой информации (на рисунке процесс понимания не показан);
- *ментальные образы данных* — это результат сенсорного восприятия данных человеком;

- *знаковая информация* — это результат выражения концептов знания с использованием некоторой знаковой системы или результат декодирования цифровой информации с использованием кодовых таблиц символов;
- *сенсорно воспринимаемые данные* — это результаты процессов наблюдения, мониторинга или зондирования, воспринимаемые органами чувств человека;
- *цифровая информация* — это результат компьютерного кодирования знаковой информации с использованием кодовых таблиц символов;
- *цифровые данные* — это результаты компьютерного кодирования сенсорно воспринимаемых данных или цифрового мониторинга (зондирования), который их порождает;
- *цифровое представление знания* — это результат компьютерного кодирования кортежей (с теоретической точки зрения их кодовые таблицы — это триединые по своей природе сущности¹).

4 Заключение

Верхний уровень классификации объектов информатики как одна из составляющих ее парадигмы, рассмотренный в работе [8] на примере пяти ее сред, дает возможность увидеть спектр теоретически возможных интерфейсов между объектами разной природы и при проектировании ИТ, и в процессе преподавания информатики. В данной статье начато описание следующего (второго) уровня классификации на примере трех сред (ментальной, информационной и цифровой). При этом второй уровень классификации предлагается сделать также открытым для расширения. Это даст возможность ввести еще один вид данных, формируемых при работе искусственных нейронных сетей, которые предлагается назвать ИИ-данными.

В статье рассмотрены два вида сущностей информационной среды и по три вида сущностей ментальной среды (знание, концепты и ментальные образы данных) и цифровой среды (цифровые данные, цифровая информация и коды кортежей). Дефиниции этих восьми сущностей предлагается включить не в классификацию, а в другую тесно связанную с ней составляющую научной парадигмы информатики — в систему терминов. Их распределение по средам дает возможность показать принципиальное теоретическое различие в средствах кодирования *символов* на границе между информационной и цифровой средами (например,

¹ Использование кодовых таблиц кортежей как триединых по своей природе сущностей, которые располагаются в точке соприкосновения трех сред (на рисунке это ментальная, информационная и цифровая среды), описано в работах [31, 35].

с помощью традиционных кодовых таблиц Unicode) и концептов в точке соприкосновения трех сред (например, с использованием кодов значений слов в тезаурусах [5]).

В заключение отметим, что необходимость перехода от схемы иерархии DIKW к более детально специфицированным сущностям предметной области информатики обусловлена, в частности, актуальным направлением в информатике, получившем название визуальной аналитики [4, 38].

Автор признателен А. А. Гончарову, С. Н. Гринченко и С. К. Дулину за обсуждение предлагаемого варианта научной парадигмы информатики и их предложения по редактированию ее положений.

Литература

1. *Eliot T. S.* The rock. — London: Faber and Faber Ltd., 1934. 86 p.
2. *Ackoff R.* From data to wisdom // *J. Appl. Syst. Anal.*, 1989. Vol. 16. P. 3–9.
3. *Rowley J.* The wisdom hierarchy: representations of the DIKW hierarchy // *J. Inf. Sci.*, 2007. Vol. 33. No. 2. P. 163–180.
4. *Chen M., Ebert D., Hagen H., Laramée R., Van Liere R., Ma K.-L., Ribarsky W., Scheuermann G., Silver D.* Data, information, and knowledge in visualization // *IEEE Comput. Graph.*, 2009. Vol. 29. No. 1. P. 12–19.
5. *Лукашевич Н. В.* Тезаурусы в задачах информационного поиска. — М.: Изд-во Московского ун-та, 2011. 512 с.
6. *Weinberger D.* The problem with the data–information–knowledge–wisdom hierarchy // *Harvard Bus. Rev.*, 2010. <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>.
7. *Frické M. H.* Data–Information–Knowledge–Wisdom (DIKW) pyramid, framework, continuum // *Encyclopedia of big data* / Eds. L. Schintler, C. McNeely. — Cham: Springer, 2018. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
8. *Зацман И. М.* О научной парадигме информатики: верхний уровень классификации объектов ее предметной области // *Информатика и её применения*, 2022. Т. 16. Вып. 4. С. 108–114.
9. *Newman J.* Some observations on the semantics of “Information” // *Inform. Syst. Front.*, 2001. Vol. 3. No. 2. P. 155–167.
10. *Díaz Nafria J. M.* What is information? A multidimensional concern // *Triple C*, 2010. Vol. 8. No. 1. P. 77–108.
11. *Denning P. J.* Opening statement // Ubiquity Symposium “What is Computation?” Proceedings, 2010. Vol. 10. <https://ubiquity.acm.org/article.cfm?id=1880067>.
12. *Conery J. S.* Computation is symbol manipulation // Ubiquity Symposium “What is Computation?” Proceedings, 2010. Vol. 10. <https://ubiquity.acm.org/article.cfm?id=1889839>.
13. *Frailey D. J.* Computation is process // Ubiquity Symposium “What is Computation?” Proceedings, 2010. Vol. 10. <https://ubiquity.acm.org/article.cfm?id=1891341>.
14. *Bajcsy R.* Computation and information // Ubiquity Symposium “What is Computation?” Proceedings, 2010. Vol. 10. <https://ubiquity.acm.org/article.cfm?id=1899473>.
15. *Rosenbloom P.* The computing sciences and STEM education // Ubiquity Symposium “The Science in Computer Science” Proceedings, 2014. Vol. 14. <http://ubiquity.acm.org/article.cfm?id=2590530>.
16. *Шрейдер Ю. А.* Информация и знание // Системная концепция информационных процессов. — М.: ВНИИСИ, 1988. С. 47–52.
17. *Nygaard K.* Program development as a social activity // 10th World Computer Congress Proceedings / Ed. H.-J. Kugler. — North Holland: Elsevier Science Publishers B. V., IFIP, 1986. P. 189–198.
18. *Nonaka I.* The knowledge-creating company // *Harvard Bus. Rev.*, 1991. Vol. 69. No. 6. P. 96–104.
19. Информатика как наука об информации / Под ред. Р. С. Гиляревского. — М.: ФАИР-ПРЕСС, 2006. 592 с.
20. *Denning P., Rosenbloom P.* Computing: The fourth great domain of science // *Commun. ACM*, 2009. Vol. 52. No. 9. P. 27–29.
21. *Rosenbloom P. S.* On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 307 p.
22. Deep shift — technology tipping points and societal impact // World Economic Forum. — Geneva, Switzerland, 2015. http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf.
23. *Шваб К.* Четвертая промышленная революция / Пер. с англ. — М.: Эксмо, 2018. 288 с. (*Schwab K.* The fourth industrial revolution. — Geneva, Switzerland: World Economic Forum, 2016. 172 p.)
24. Informatics Education in Europe: Are we all in the same boat? — New York, NY, USA: ACM, 2017. Technical Report of the Committee on European Computing Education. 251 p.
25. *Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E.* Informatics for all: The strategy. — New York, NY, USA: ACM, 2018. 16 p.
26. *Díaz Nafria J. M., Zimmermann R. E.* Emergence and evolution of meaning: The general definition of information (GDI) revisiting program — Part 2: The regressive perspective: Bottom-up // *Information*, 2013. Vol. 4. P. 240–261.
27. Webster’s New World dictionary of the American language / Eds. D. B. Guralnik, J. H. Friend. — New York, NY, USA: The World Publishing Company, 1960. 1760 p.
28. Definition of phenomenon (meaning 2c) // Merriam-Webster’s dictionary. <https://www.merriam-webster.com/dictionary/phenomenon>.
29. *Зацман И. М.* Теоретические основания компьютерного образования: среды предметной области информатики как основание классификации ее объектов // Системы и средства информатики, 2022. Т. 32. № 4. С. 77–89.

30. *Коллин К. К.* О структуре научных исследований по комплексной проблеме «Информатика» // Социальная информатика. — М.: ВКШ при ЦК ВЛКСМ, 1990. С. 19–33.
31. *Зацман И. М.* Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89.
32. *Зацман И. М.* Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2014. № 11. С. 1–15.
33. *Соломоник А.* Парадигма семиотики. — Минск: МЕТ, 2006. 335 с.
34. *Кун Т.* Структура научных революций / Пер. с англ. — М.: Прогресс, 1977. 302 с. (*Kuhn T.* The structure of scientific revolutions. — Chicago, IL, USA: University of Chicago Press, 1962. 264 p.)
35. *Зацман И. М.* Кодирование концептов в цифровой среде // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 97–106.
36. Географический энциклопедический словарь. Понятия и термины / Под ред. А. Ф. Трёшников. — М.: Сов. энциклопедия, 1988. 432 с.
37. *Зацман И. М.* Концептуальный поиск и качество информации. — М.: Наука, 2003. 272 с.
38. *Federico P., Wagner M., Rind A., Amor-Amoros A., Miksch S., Aigner W.* The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics // IEEE Conference on Visual Analytics Science and Technology Proceedings. — New York, NY, USA: IEEE, 2017. P. 92–103.

Поступила в редакцию 14.01.23

ON THE SCIENTIFIC PARADIGM OF INFORMATICS: DATA, INFORMATION, AND KNOWLEDGE

I. M. Zatsman

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

Abstract: Three basic notions of informatics — data, information, and knowledge — are considered. The variant of specification of these notions within the framework of constructing a system of terms of the scientific paradigm of informatics as a fundamental science is proposed. On the one hand, the notions of “data,” “information,” and “knowledge” are widely used in the scientific literature and textbooks on informatics, in particular, when describing its theoretical foundations. On the other hand, there is still no consensus on their semantic content. The current situation is most likely due to the widespread presupposition (implicit assumption) that the notions in question express some objective entities of the subject domain of informatics. The paper assumes that they express intersubjective entities that by their nature arise as objects of thought during an agreement process, that is, they are not objectively existing. From the point of view of the Frege’s triangle (subject – notion (concept) – word that expresses the notion and denotes the subject), for objective entities, the primary vertex of the triangle is the subject as a result of the study of which the notion and the word appear. For intersubjective entities, the primary vertex is the notion, the definitions of which must be discussed in the interests of reaching consensus. If it can be achieved, then it is during the process of discussion that the subject of thought appears, the word denoting it and expressing the notion which together form the Frege’s triangle. The aim of the paper is to specify the basic notions of informatics as expressing intersubjective entities and being the primary vertices of the Frege’s triangle, to distribute them among the media of the subject domain of informatics, highlighting the boundaries between the media, and to consider the relationship between the specified notions on these boundaries.

Keywords: scientific paradigm; informatics as fundamental science; data; information; knowledge; intersubjective entities of informatics; Frege’s triangle

DOI: 10.14357/19922264230115

Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Информатика”) of FRC CSC RAS (Moscow).

References

1. Eliot, T. S. 1934. *The rock*. London: Faber and Faber Ltd. 86 p.
2. Ackoff, R. 1989. From data to wisdom. *J. Appl. Syst. Anal.* 16(1):3–9.
3. Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* 33(2):163–180.

4. Chen, M., D. Ebert, H. Hagen, R. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. 2009. Data, information, and knowledge in visualization. *IEEE Comput. Graph.* 29(1):12–19.
5. Loukachevitch, N. V. 2011. *Tezaurusy v zadachakh informatsionnogo poiska* [Thesauri in information retrieval tasks]. Moscow: Izd-vo Moskovskogo un-ta. 512 p.
6. Weinberger, D. 2010. The problem with the data–information–knowledge–wisdom hierarchy. *Harvard Bus. Rev.* Available at: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> (accessed January 24, 2023).
7. Frické, M. H. 2018. Data–Information–Knowledge–Wisdom (DIKW) pyramid, framework, continuum. *Encyclopedia of big data*. Eds. L. Schintler and C. McNeely. Cham: Springer. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
8. Zatsman, I. 2022. O nauchnoy paradigme informatiki: verkhniy uroven' klassifikatsii ob'ektov ee predmetnoy oblasti [On the scientific paradigm of informatics: The classification high level of its objects]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):108–114.
9. Newman, J. 2001. Some observations on the semantics of “Information.” *Inform. Syst. Front.* 3(2):155–167.
10. Díaz Nafría, J. M. 2010. What is information? A multidimensional concern. *Triple C* 8(1):77–108.
11. Denning, P. J. 2010. Opening statement. *Ubiquity Symposium “The Science in Computer Science” Proceedings*. 10. Available at: <https://ubiquity.acm.org/article.cfm?id=1880067> (accessed January 24, 2023).
12. Conery, J. S. 2010. Computation is symbol manipulation. *Ubiquity Symposium “The Science in Computer Science” Proceedings*. 10. Available at: <https://ubiquity.acm.org/article.cfm?id=1889839> (accessed January 24, 2023).
13. Frailey, D. J. 2010. Computation is process. *Ubiquity Symposium “The Science in Computer Science” Proceedings*. 10. Available at: <https://ubiquity.acm.org/article.cfm?id=1891341> (accessed January 24, 2023).
14. Bajcsy, R. 2010. Computation and information. *Ubiquity Symposium “The Science in Computer Science” Proceedings*. 10. Available at: <https://ubiquity.acm.org/article.cfm?id=1899473> (accessed January 24, 2023).
15. Rosenbloom, P. S. 2014. The computing sciences and STEM education. *Ubiquity Symposium “The Science in Computer Science” Proceedings*. 14. Available at: <http://ubiquity.acm.org/article.cfm?id=2590530> (accessed January 24, 2023).
16. Shreyder, Yu. A. Informatsiya i znanie [Information and knowledge]. *Sistemnaya kontseptsiya informatsionnykh protsessov* [System conception of information processes]. Moscow: VNIISI Pubs. 47–52.
17. Nygaard, K. 1986. Program development as a social activity. *10th World Computer Congress Proceedings*. Ed. H.-J. Kugler. North Holland: Elsevier Science Pubs. B. V., IFIP. 189–198.
18. Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
19. Gilyarevskiy, R. S., ed. 2006. *Informatika kak nauka ob informatsii* [Informatics as information science]. Moscow: FAIR-PRESS. 592 p.
20. Denning, P., and P. Rosenbloom. 2009. Computing: The fourth great domain of science. *Commun. ACM* 52(9):27–29.
21. Rosenbloom, P. S. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 307 p.
22. Deep shift — technology tipping points and societal impact. 2015. *World Economic Forum*. Geneva, Switzerland. Available at: http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (accessed January 24, 2023).
23. Schwab, K. 2016. *The fourth industrial revolution*. Geneva, Switzerland: World Economic Forum. 172 p.
24. The Committee on European Computing Education. 2017. *Informatics education in Europe: Are we all in the same boat?* New York, NY: ACM. Technical Report. 251 p. Available at: <https://dl.acm.org/citation.cfm?id=3106077> (accessed January 24, 2023).
25. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2018. *Informatics for all: The strategy*. New York, NY: ACM. 16 p.
26. Díaz Nafría, J. M., and R. E. Zimmermann. 2013. Emergence and evolution of meaning: The general definition of information (GDI) revisiting program — Part 2: The regressive perspective: Bottom-up. *Information* 4:240–261.
27. Guralnik, D. B., and J. H. Friend, eds. 1960. *Webster's New World dictionary of the American language*. New York, NY: The World Publishing Co. 1760 p.
28. Definition of phenomenon (meaning 2c). *Merriam-Webster's dictionary*. Available at: <https://www.merriam-webster.com/dictionary/phenomenon> (accessed January 24, 2023).
29. Zatsman, I. 2022. Teoreticheskie osnovaniya komp'yuternogo obrazovaniya: sredi predmetnoy oblasti informatiki kak osnovanie klassifikatsii ee ob'ektov [Theoretical foundations of digital education: Subject domain media of informatics as the base of its objects' classification]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(4):77–89.
30. Kolin, K. K. 1990. O strukture nauchnykh issledovaniy po kompleksnoy probleme “Informatika” [On the structure of scientific research on the complex problem of “Informatics”]. *Sotsial'naya informatika* [Social informatics]. Moscow: VKSh. 19–33.
31. Zatsman, I. M. 2019. Interfeysy tret'ego poryadka v informatike [Third-order interfaces in informatics]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):82–89.
32. Zatsman, I. 2014. A table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):233–246.
33. Solomonick, A. 2006. *Paradigma semiotiki* [The paradigm of semiotics]. Minsk: MET Pubs. 335 p.

34. Kuhn, T. 1962. *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press. 264 p.
35. Zatsman, I. M. 2019. Kodirovanie kontseptov v tsifrovoy srede [Digital encoding of concepts]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):97–106.
36. Treshnikov, A. F., ed. 1988. *Geograficheskiy entsiklopedicheskiy slovar'. Ponyatiya i terminy* [Geographical encyclopedic dictionary. Concepts and terms]. Moscow: The Soviet Encyclopedia Pubs. 432 p.
37. Zatsman, I. 2003. *Kontseptual'nyy poisk i kachestvo informatsii* [Conceptual retrieval and quality of information]. Moscow: Nauka. 272 p.
38. Federico, P., M. Wagner, A. Rind, A. Amor-Amoros, S. Miksch, and W. Aigner. 2017. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. *IEEE Conference on Visual Analytics Science and Technology Proceedings*. New York, NY: IEEE. 92–103.

Received January 14, 2023

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, head of department, 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; izatsman@yandex.ru

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Агасандян Геннадий Аршавинович (р. 1941) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Аду Кпангни Ивь Бэрэнжэ (р. 1993) — аспирант, стажер-исследователь кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Архипов Павел Олегович (р. 1979) — кандидат технических наук, директор Орловского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального исследовательского университета)

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший 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 информатики 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

Дулин Сергей Константинович (р. 1950) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Научно-исследовательского и проектно-конструкторского института информатизации, автоматизации и связи на железнодорожном транспорте

Дулина Наталья Георгиевна (р. 1947) — кандидат технических наук, ведущий программист Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забжайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Игнатов Алексей Николаевич (р. 1991) — кандидат физико-математических наук, доцент кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального исследовательского университета)

Кабанов Юрий Михайлович (р. 1948) — доктор физико-математических наук, профессор Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кульченков Владислав Владимирович (р. 1989) — руководитель службы мониторинга рисков Банка ВТБ (ПАО)

Лери Марина Максумовна (р. 1969) — кандидат технических наук, научный сотрудник Института прикладных математических исследований Карельского научного центра Российской академии наук

Листопад Сергей Викторович (р. 1984) — кандидат технических наук, старший научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Малашенко Юрий Евгеньевич (р. 1946) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Маркова Екатерина Викторовна (р. 1987) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Назарова Ирина Александровна (р. 1966) — кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Нейчев Радослав Георгиев (р. 1994) — аспирант Московского физико-технического института

Розенберг Игорь Наумович (р. 1965) — доктор технических наук, профессор, член-корреспондент РАН, научный руководитель Научно-исследовательского и проектно-конструкторского института информатизации, автоматизации и связи на железнодорожном транспорте

Сидоренко Артур Павлович (р. 1998) — аспирант механико-математического факультета Московского государственного университета имени М. В. Ломоносова

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор кафедры «Мехатроника и теоретическая механика» Москов-

ского авиационного института (национального исследовательского университета)

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

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ушаков Владимир Георгиевич (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ушаков Николай Георгиевич (р. 1954) — доктор физико-математических наук, ведущий научный сотрудник Института проблем технологии микроэлектроники и особо чистых материалов Российской академии наук (Черноголовка); профессор Норвежского научно-технологического университета (г. Тронхейм)

Филиппских Сергей Леонидович (р. 1987) — младший научный сотрудник Орловского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Цуканов Максим Владимирович (р. 1995) — младший научный сотрудник Орловского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шibaев Иннокентий Андреевич (р. 1997) — аспирант Московского физико-технического института

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

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 ЛАТЭХ (.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 10 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/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. *Russ. 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. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

Описание неопубликованного документа:

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 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: iiep@frccsc.ru (Стригина Светлана Николаевна)

<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 10 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. *Russ. 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 neftegazooitdachi"* [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]. Moscow: IPI RAN. PhD Thesis. 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: iiep@frccsc.ru (to Svetlana Strigina)
<http://www.ipiran.ru/english/journal.asp>