

Информатика и её применения

Том 15 Выпуск 1 Год 2021

СОДЕРЖАНИЕ

Нормальные субоптимальные фильтры для дифференциальных стохастических систем, не разрешенных относительно производных	
<i>И. Н. Сеницын</i>	3
О некоторых частных случаях в задаче управления выходом стохастической дифференциальной системы по квадратичному критерию	
<i>А. В. Босов</i>	11
Связность конфигурационных графов в моделях сложных сетей	
<i>Ю. Л. Павлов</i>	18
Методы теории категорий в цифровом проектировании гетерогенных киберфизических систем	
<i>С. П. Ковалёв</i>	23
Методы обнаружения переводных заимствований в больших текстовых коллекциях	
<i>Р. В. Кузнецова, О. Ю. Бахтеев, Ю. В. Чехович</i>	30
Вариационная оптимизация модели глубокого обучения с контролем сложности	
<i>О. С. Гребенькова, О. Ю. Бахтеев, В. В. Стрижов</i>	42
Информационная модель весового облика летательных аппаратов	
<i>Л. Л. Вышинский, Ю. А. Флёров</i>	50
Оптимальное пороговое управление доступом в системе $M/M/s$ с неоднородными приборами и общим накопителем	
<i>Я. М. Агаларов</i>	57
Вероятностные характеристики индекса баланса факторов, имеющих обобщенные гамма-распределения	
<i>Е. Н. Арутюнов, А. А. Кудрявцев, Ю. Н. Недоливко</i>	65
Неасимптотический анализ статистики Бартлетта–Нанда–Пилая для данных большой размерности	
<i>А. А. Липатьев</i>	72
Архитектура распределенного решения задач анализа данных в области нейрофизиологии	
<i>Д. О. Брюхов, С. А. Ступников, Д. Ю. Ковалёв, И. А. Шанин</i>	78
Представление новых лексикографических знаний в динамических классификационных системах	
<i>А. А. Гончаров, И. М. Зацман, М. Г. Кружков</i>	86

Информатика и её применения

Том 15 Выпуск 1 Год 2021

СОДЕРЖАНИЕ

Проблемно-ориентированная актуализация словарных статей двуязычных словарей и медицинской терминологии: сопоставительный анализ <i>И. М. Зацман</i>	94
Моделирование стохастической динамики изменения состояний узлов и перколяционных переходов в социальных сетях с учетом самоорганизации и наличия памяти <i>Д. О. Жуков, Т. Ю. Хватова, А. Д. Зальцман</i>	102
О системной иерархии искусственного интеллекта <i>С. Н. Гринченко</i>	111
О точности нормальной аппроксимации при отсутствии нормальной сходимости <i>В. Ю. Королев, А. В. Дорофеева</i>	116
Об авторах	122
Правила подготовки рукописей	124
Requirements for manuscripts	127

НОРМАЛЬНЫЕ СУБОПТИМАЛЬНЫЕ ФИЛЬТРЫ ДЛЯ ДИФФЕРЕНЦИАЛЬНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ, НЕ РАЗРЕШЕННЫХ ОТНОСИТЕЛЬНО ПРОИЗВОДНЫХ

И. Н. Синицын¹

Аннотация: Рассмотрены вопросы синтеза среднеквадратичных (с.к.) нелинейных нормальных (гауссовских) субоптимальных фильтров (НСОФ) для дифференциальных стохастических систем (СтС), не разрешенных относительно производных. Представлены уравнения состояния и наблюдения нелинейных дифференциальных СтС, не разрешенных относительно производных. Один из разделов посвящен синтезу НСОФ при следующих условиях: (1) отсутствуют пуассоновские шумы в наблюдениях; (2) коэффициент при гауссовском шуме не зависит от состояния. Рассмотрен синтез НСОФ при аддитивных шумах в уравнениях состояния и наблюдения. Для иллюстрации методов синтеза НСОФ приведен пример. Рассмотрены вопросы качества НСОФ.

Ключевые слова: метод аналитического моделирования (МAM); метод нормальной аппроксимации (МНА); метод статистической линеаризации (МСЛ); нормальный субоптимальный фильтр (НСОФ); стохастическая система (СтС); стохастические системы, не разрешенные относительно производных; формирующий фильтр (ФФ)

DOI: 10.14357/19922264210101

1 Введение

Методы аналитического моделирования (МAM) широкополосных стохастических процессов (СтП) в нелинейных дифференциальных СтС, разрешенных относительно производных, представлены в [1–6].

В [7] представлено развитие методов параметрического аналитического моделирования [1, 2, 5] применительно к СтС, не разрешенным относительно производных. Рассмотрены уравнения нелинейных дифференциальных СтС, не разрешенных относительно производных с винеровскими и пуассоновскими белыми, а также автокоррелированными шумами. Дано обобщение МAM на основе метода нормальной аппроксимации (МНА) (метода статистической линеаризации (МСЛ)) на случай автокоррелированных шумов. Рассмотрены общие вопросы МAM на основе параметризации одно- и многомерных плотностей СтП нелинейных дифференциальных СтС, не разрешенных относительно производных, приводимых к СтС, разрешенным относительно производных. Особое внимание уделено МAM на основе метода ортогональных разложений (МОР) и метода квазиоментов (МКМ) для одно- и многомерных плотностей в случае эталонных гауссовских и пуассоновских плотностей.

В статье рассмотрены вопросы синтеза с.к. нелинейных НСОФ для дифференциальных СтС, не разрешенных относительно производных.

В разд. 2 представлены уравнения состояния и наблюдения нелинейных дифференциальных СтС, не разрешенных относительно производных.

Раздел 3 посвящен синтезу НСОФ при следующих условиях: (1) отсутствуют пуассоновские шумы в наблюдениях; (2) коэффициент при гауссовском шуме не зависит от состояния.

Синтез НСОФ при аддитивных шумах в уравнениях состояния и наблюдения рассмотрен в разд. 4.

Для иллюстрации методов синтеза НСОФ в разд. 5 приведен пример. Вопросы качества НСОФ рассмотрены в разд. 6.

Заключение и возможные обобщения содержатся в разд. 7.

2 Уравнения дифференциальных стохастических систем, не разрешенных относительно производных и наблюдений

Следуя [8], рассмотрим сначала векторную дифференциальную СтС с нелинейностями, описываемыми гладкими функциями:

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

$$\begin{aligned} \varphi &= \varphi \left(t, \Theta, X_t, \dot{X}_t, \dots, X_t^{(k)}, U_t \right) = 0, \\ X(t_0) &= X_0, \\ \dot{X}(t_0) &= \dot{X}_0, \dots, X^{(k)}(t_0) = X_0^{(k)}, \end{aligned} \quad (1)$$

при этом уравнение нелинейного формирующего фильтра (ФФ) возьмем в виде, разрешенном относительно возмущений U_t :

$$\begin{aligned} \dot{U}_t &= a^U(t, \Theta, U_t) + b^U(t, \Theta, U_t) V_t^U, \\ U(t_0) &= U_0. \end{aligned} \quad (2)$$

Здесь $a^U = a^U(t, \Theta, U_t)$ и $b^U(t, \Theta, U_t) — ($n^X \times 1$)- и ($n^X \times n^V$)-мерные функции; V_t^U — белый шум в строгом смысле [1, 2], допускающий представление в виде гауссовской и пуассоновской составляющих$

$$\begin{aligned} V_t^U &= \dot{W}_t^U, \\ W_t^U &= W_0^U(t, \Theta) + \int_{R_0^q} c^U(\Theta, \rho) P^0(t, \Theta, d\rho), \end{aligned}$$

где ν_t — его интенсивность:

$$\nu_t = \nu_t^W = \nu_t^{W_0} + \int_{R_0^q} c^U(\Theta, \rho) [c^U(\Theta, \rho)]^T \nu_P(t, \Theta, \rho) d\rho;$$

$c^U = c^U(\Theta, \rho)$ — известная векторная функция той же размерности, что и W_t^0 , а интеграл при любом $t \geq t_0$ представляет собой стохастический интеграл по централизованной пуассоновской мере $P^0(t, \Theta, \mathcal{A})$, независимой от W_0^U и имеющей независимые значения на попарно непересекающихся множествах; \mathcal{A} — борелевское множество пространства R_0^q с выколотым началом; $\nu_t^W, \nu_t^{W_0}$ и ν_P — интенсивности СтП W_t^U, W_0^U и P^0 . Уравнение (2) понимается в смысле Ито и имеет единственное решение в с.к. [1, 2].

Для гладких функций в (1), допускающих стохастические производные Ито до h -го порядка и статистическую линейризацию по Казакову [1–4], выполним следующие преобразования. Будем дифференцировать сполна по t левые части уравнений (1) по обобщенной формуле Ито [1, 2] до тех пор, пока не появятся производные белого шума. В результате получим следующую систему нелинейных дифференциальных уравнений:

$$\varphi = 0, \quad \dot{\varphi} = 0, \dots, \varphi^{(h)} = 0. \quad (3)$$

Затем заменим эту систему уравнений статистически линейризованной:

$$\begin{aligned} \varphi^{(i)} \left(t, \Theta, X_t^{(i)}, U_t \right) &= \\ &= \varphi_{0X'}^{(i)} \left(t, \Theta, m_t^{X^i}, K_t^{X^i}, m_t^U, K_t^U \right) + \end{aligned}$$

$$\begin{aligned} &+ k_{1X'}^{(i)} \left(t, \Theta, m_t^{X^i}, K_t^{X^i}, m_t^U, K_t^U \right) X_t^{i0} + \\ &+ k_{1U}^{(i)} \left(t, \Theta, m_t^{X^i}, K_t^{X^i}, m_t^U, K_t^U \right) U_t^0 = 0 \end{aligned} \quad (i = 1, 2, \dots, h). \quad (4)$$

Здесь $X_t' = [X_t^T \dot{X}_t^T \dots X_t^{(k-1)T}]^T$; $m_t^{X'}$ и $K_t^{X'}$ — вектор математического ожидания и ковариационная матрица; $k_{1X'}^{(i)} = k_{1X'}^{(i)}(t, \Theta, m_t^{X^i}, K_t^{X^i}, m_t^U, K_t^U)$ и $k_{1U}^{(i)} = k_{1U}^{(i)}(t, \Theta, m_t^{X^i}, K_t^{X^i}, m_t^U, K_t^U)$ — матричные коэффициенты статистической линейризации функций (4).

Далее введем вектор $\bar{X}_t = [X_t'^T X_t''^T]^T$, составленный из $X_t' = [X_t^T \dot{X}_t^T \dots X_t^{(k-1)T}]^T$ и вспомогательного вектора $X_t'' = [X_t^{(i)}]_{i=\overline{1, h}}$ на основе (4). В результате придем к уравнениям, разрешенным относительно дифференциалов, следующего вида:

$$d\bar{X}_t = a^{\bar{X}} dt + b^{\bar{X}} dW_0 + \int_{R_0^q} c^{\bar{X}} P^0(t, \Theta, du), \quad (5)$$

где $a^{\bar{X}} = a^{\bar{X}}(t, \Theta, \bar{X}_t)$; $b^{\bar{X}} = b^{\bar{X}}(t, \Theta, \bar{X}_t)$; $c^{\bar{X}} = c^{\bar{X}}(t, \Theta, \bar{X}_t, u)$.

Таким образом, имеем следующее утверждение [8].

Теорема 1. Пусть нелинейная негауссовская СтС (1), (2), не разрешенная относительно производных k -го порядка, удовлетворяет условиям:

- 1° функции (1) допускают обобщенные стохастические дифференциалы Ито вплоть до h -го порядка включительно и статистическую линейризацию по Казакову;
- 2° возмущения U_t негауссовские, причем уравнение ФФ (2) разрешено относительно возмущений U_t и имеет единственное с.к. решение;
- 3° СтП X_t' более гладкий, чем возмущение U_t .

Тогда система (1), (2) приводима к дифференциальной системе, разрешенной относительно производных (5).

Далее, следуя [9], предположим, что СтС удовлетворяет условиям теоремы 1, а также полностью наблюдаема. При этом уравнения наблюдения описываются уравнениями (5) при следующих условиях: во-первых, уравнения наблюдения не содержат пуассоновского шума ($c_1 \equiv 0$) и, во-вторых, коэффициент при винеровском шуме b_1^Y не зависит от состояния $\bar{X}_t(b_1^Y(\bar{X}_t, Y_t)) = b_1^Y(Y_t, t)$. Размерность вектора Y_t примем равной n^Y , а $\bar{X}_t = n^{\bar{X}}$.

В дальнейшем для краткости будем принимать $\hat{X}_t = X_t$, а уравнения «система плюс наблюдения» записывать в следующем виде:

$$dX_t = a(X_t, Y_t, t, \Theta) dt + b(X_t, Y_t, t, \Theta) dW_0 + \int_{R_0^q} c(X_t, Y_t, t, \Theta, v) P^0(dt, dv, \Theta); \quad (6)$$

$$dY_t = a_1(X_t, Y_t, t, \Theta) dt + b_1(Y_t, t, \Theta) dW_0. \quad (7)$$

Для СтС (6), (7) с аддитивными шумами имеют место следующие условия:

$$\left. \begin{aligned} b(X_t, Y_t, t, \Theta) &= b_0(t, \Theta); \\ c(X_t, Y_t, t, \Theta, v) &= 0, \\ b_1(Y_t, t, \Theta) &= b_{10}(t, \Theta). \end{aligned} \right\} \quad (8)$$

Условия (8) будут выполняться, если в последнем уравнении (3) с номером h появится белый шум, а коэффициент при нем будет зависеть только от t и Θ .

3 Нормальный субоптимальный фильтр для стохастических систем, не разрешенных относительно производных

Как известно из нелинейной теории фильтрации [9], для гауссовской СтС, поскольку гауссовское (нормальное) распределение, аппроксимирующее апостериорное распределение вектора X_t , полностью определяется апостериорными математическим ожиданием \hat{X}_t и ковариационной матрицей R_t вектора X_t , при аппроксимации апостериорного распределения вектора X_t нормальным распределением все математические ожидания для $d\hat{X}_t$ и dR_t будут определенными функциями \hat{X}_t , R_t и t , т. е. будут представлять собой стохастические дифференциальные уравнения, определяющие \hat{X}_t и R_t :

$$d\hat{X}_t = B(\hat{X}_t, Y_t, R_t, t, \Theta) = f(\hat{X}_t, Y_t, R_t, t, \Theta) dt + h(\hat{X}_t, Y_t, R_t, t, \Theta) dt \times \left[dY_t - f^{(1)}(\hat{X}_t, Y_t, R_t, t) dt \right]; \quad (9)$$

$$dR_t = \left\{ f^{(2)}(\hat{X}_t, Y_t, R_t, t, \Theta) - h(\hat{X}_t, Y_t, R_t, t, \Theta) (b_1 \nu b_1^T)(Y_t, t, \Theta) \times \right.$$

$$\left. \times h(\hat{X}_t, Y_t, R_t, t, \Theta)^T \right\} dt + \sum_{r=1}^{n_y} \rho_r(\hat{X}_t, Y_t, R_t, t, \Theta) \times \left[dY_r - f_r^{(1)}(\hat{X}_t, Y_t, R_t, t, \Theta) dt \right], \quad (10)$$

где $\nu = \nu(t)$ — интенсивность белого шума:

$$f(\hat{X}_t, Y_t, R_t, t, \Theta) = [(2\pi)^n |R_t|]^{-1/2} \int_{-\infty}^{\infty} a(Y_t, x, t, \Theta) \times \exp \left\{ - \left(x^T - \hat{X}_t^T \right) R_t^{-1} \frac{x - \hat{X}_t}{2} \right\} dx; \quad (11)$$

$$f^{(1)}(\hat{X}_t, Y_t, R_t, t, \Theta) = \left\{ f_r^{(1)}(\hat{X}_t, Y_t, R_t, t, \Theta) \right\} = [(2\pi)^{n_x} |R_t|]^{-1/2} \int_{-\infty}^{\infty} a_1(Y_t, x, t) \times \exp \left\{ - \left(x^T - \hat{X}_t^T \right) R_t^{-1} \frac{x - \hat{X}_t}{2} \right\} dx; \quad (12)$$

$$h(\hat{X}_t, Y_t, R_t, t, \Theta) = \left\{ [(2\pi)^{n_x} |R_t|]^{-1/2} \times \int_{-\infty}^{\infty} [x a_1(Y_t, x, t)^T + \nu \nu b_1^T(Y_t, x, t, \Theta)] \times \exp \left\{ - \left(x^T - \hat{X}_t^T \right) R_t^{-1} \frac{x - \hat{X}_t}{2} \right\} dx - \hat{X}_t f^{(1)}(\hat{X}_t, Y_t, R_t, t, \Theta)^T \right\} \times (b_1 \nu b_1^T)^{-1}(Y_t, t); \quad (13)$$

$$f^{(2)}(\hat{X}_t, Y_t, R_t, t, \Theta) = [(2\pi)^{n_x} |R_t|]^{-1/2} \int_{-\infty}^{\infty} \left\{ \left(x - \hat{X}_t \right) a(Y_t, x, t, \Theta)^T + a(Y_t, x, t, \Theta) \left(x^T - \hat{X}_t^T \right) + \nu \nu b_1^T(Y_t, x, t, \Theta) \right\} \times \exp \left\{ - \left(x^T - \hat{X}_t^T \right) R_t^{-1} \frac{x - \hat{X}_t}{2} \right\} dx; \quad (14)$$

$$\rho_r(\hat{X}_t, Y_t, R_t, t, \Theta) = [(2\pi)^{n_x} |R_t|]^{-1/2} \times \int_{-\infty}^{\infty} \left\{ \left(x - \hat{X}_t \right) \left(x^T - \hat{X}_t^T \right) \alpha_r(Y_t, x, t, \Theta) + \right.$$

$$\begin{aligned}
 & + \left(x - \hat{X}_t \right) \beta_r (Y_t, x, t, \Theta)^T \left(x^T - \hat{X}_t^T \right) + \\
 & \quad + \beta_r (Y_t, x, t, \Theta) \left(x^T - \hat{X}_t^T \right) \} \times \\
 & \times \exp \left\{ - \left(x^T - \hat{X}_t^T \right) R_t^{-1} \frac{x - \hat{X}_t}{2} \right\} dx \\
 & \quad (r = 1, \dots, n_y). \quad (15)
 \end{aligned}$$

Здесь α_r — r -й элемент матрицы-строки $(a_1 - \hat{a}_1^T)(b_1 \nu b_1^T)^{-1}$; β_{kr} — элемент k -й строки и r -го столбца матрицы $b \nu b^T (b_1 \nu b_1^T)^{-1}$, при этом $\beta_r = [\beta_{1r} \dots \beta_{pr}]^T$ ($r = 1, \dots, n_x$).

Таким образом, применительно к (6), (7) имеет место следующее утверждение.

Теорема 2. Пусть уравнения нелинейной гауссовской дифференциальной СтС (6) и (7) допускают применение МНА. Тогда в основе МНА одномерного апостериорного распределения лежат уравнения (9) и (10) при условиях (11)–(15) и начальных условиях.

Количество уравнений МНА одномерного апостериорного распределения определяется по формуле:

$$Q_{\text{МНА}} = n_x + \frac{n_x(n_x + 1)}{2} = \frac{n_x(n_x + 3)}{2}.$$

За начальные значения \hat{X}_t и R_t при интегрировании уравнений (9) и (10), естественно, следует принять условные математическое ожидание и ковариационную матрицу величины X_0 относительно Y_0 :

$$\begin{aligned}
 \hat{X}_0 &= M[X_0 | Y_0]; \\
 R_0 &= M \left[(X_0 - \hat{X}_0) (X_0^T - \hat{X}_0^T) | Y_0 \right].
 \end{aligned}$$

Если нет информации об условном распределении X_0 относительно Y_0 , то начальные условия можно взять в виде:

$$\hat{X}_0 = M X_0; \quad R_0 = M (X_0 - M X_0) (X_0^T - M X_0^T).$$

Если же и об этих величинах нет никакой информации, то начальные значения \hat{X}_t и R_t приходится задавать произвольно.

Теперь рассмотрим НСОФ для негауссовских СтС, не разрешенных относительно производных. В этом случае, как показано в [9], требуется ограниченность функций $f, f^{(1)}, h, \rho_r$ и функции

$$\bar{f}^{(2)} = f^{(20)} + M^N \left[\int_{R_0^q} c c^T \nu_P(t, \Theta) dv \right]$$

(теорема 3).

4 Нормальный субоптимальный фильтр для стохастических систем, не разрешенных относительно производных и с аддитивными шумами

Пусть функции a, b, a_1 и b_1 в (6), (7) удовлетворяют условиям:

$$\left. \begin{aligned}
 a &= a(X_t, Y_t, t, \Theta) = a(X_t, t, \Theta); \\
 a_1 &= a_1(X_t, Y_t, t, \Theta) = a_1(X_t, t, \Theta); \\
 b(X_t, Y_t, t, \Theta) dW &= b(t, \Theta) dW_1; \\
 b_1(X_t, t, \Theta) dW &= dW_2.
 \end{aligned} \right\} \quad (16)$$

Здесь W_1 и W_2 — независимые винеровские процессы размерности $n_{w_1} = n_x$ и $n_{w_2} = n_y$. Тогда после перехода от дифференциалов к производным будем иметь

$$\left. \begin{aligned}
 \dot{X}_t &= a(X_t, t, \Theta) + b(t, \Theta) V_1; \\
 Z_t = \dot{Y}_t &= a_1(X_t, t, \Theta) + V_2,
 \end{aligned} \right\} \quad (17)$$

где $V = [V_1 V_2]^T$ — нормальный белый шум интенсивности $\nu = \text{diag}(\nu_1, \nu_2)$.

Заменим (17) статистически линеаризованной системой, нелинейной относительно математических ожиданий m_t^x и m_t^z и линейной относительно центрированных составляющих $X_t^0 = X_t - m_t^x$ и $\hat{X}_t^0 = \hat{X}_t - \hat{m}_t^x$:

$$\dot{m}_t^x = a_{00}(m_t^x, K_t^x, t, \Theta); \quad (18)$$

$$m_t^z = a_{10}(m_t^x, K_t^x, t, \Theta); \quad (19)$$

$$\dot{X}_t^0 = a_{01}(m_t^x, K_t^x, t, \Theta) X_t^0 + b(t, \Theta) V_1; \quad (20)$$

$$Z_t^0 = a_{11}(m_t^x, K_t^x, t, \Theta) X_t^0 + V_2, \quad (21)$$

где $a_{00} = a_{00}(m_t^x, K_t^x, t, \Theta)$, $a_{10} = a_{10}(m_t^x, K_t^x, t, \Theta)$, $a_{01}(m_t^x, K_t^x, t, \Theta) = \partial a_0 / \partial m_t^x$ и $a_{11}(m_t^x, K_t^x, t, \Theta) = \partial a_{10} / \partial m_t^x$ — коэффициенты статистической линеаризации нелинейных функций a и a_1 , вычисляемые для нормального распределения $\mathcal{N}(m_t^x, K_t^x)$. При этом в силу (20) ковариационная матрица K_t^x будет определяться уравнением:

$$\begin{aligned}
 \dot{K}_t^x &= a_{11}(m_t^x, K_t^x, t, \Theta) K_t^x + \\
 & + K_t^x a_{11}(m_t^x, K_t^x, t, \Theta)^T + b(t, \Theta) \nu_1(t, \Theta) b(t, \Theta)^T. \quad (22)
 \end{aligned}$$

Применяя к модели (20), (21) уравнения линейного фильтра Калмана–Бьюси, получим искомые уравнения квазилинейного фильтра на основе МСЛ:

$$\begin{aligned}
 \dot{\hat{X}}_t &= a_{00}(m_t^x, K_t^x, t, \Theta) - \\
 & - a_{01}(m_t^x, K_t^x, t, \Theta) m_t^x + a_{01}(m_t^x, K_t^x, t, \Theta) \hat{X}_t +
 \end{aligned}$$

$$+ R_t a_{11} (m_t^x, K_t^x, t, \Theta)^T \nu_2(t, \Theta)^{-1} [Z_t - a_{11} (m_t^x, K_t^x, t, \Theta) \hat{X}_t - a_{10} (m_t^x, K_t^x, t, \Theta) + a_{11} (m_t^x, K_t^x, t, \Theta) m_t^x], \quad (23)$$

где $\hat{X}_0 = M X_0$;

$$\begin{aligned} \dot{R}_t = & a_{01} (m_t^x, K_t^x, t, \Theta) R_t + R_t a_{01} (m_t^x, K_t^x, t, \Theta)^T - \\ & - R_t a_{11} (m_t^x, K_t^x, t, \Theta)^T \nu_2(t, \Theta)^{-1} \times \\ & \times a_{11} (m_t^x, K_t^x, t, \Theta) R_t + b(t, \Theta) \nu_1(t, \Theta) b(t, \Theta)^T, \\ R_0 = & M \left[(X_0 - \hat{X}_0) (X_0 - \hat{X}_0)^T \right]. \quad (24) \end{aligned}$$

Входящие сюда m_t^x и K_t^x определяются из уравнений (18) и (19).

Полученные результаты можно сформулировать в виде теоремы.

Теорема 4. Пусть в условиях (16) и теоремы 1 уравнения нелинейной гауссовской дифференциальной СтС (17) допускают применение МСЛ, а линеаризованные уравнения удовлетворяют условиям стохастической наблюдаемости. Тогда НСОФ согласно МСЛ определяется уравнениями (23), (24) при условиях (18), (19) и (22) и соответствующих начальных условиях.

При чисто линейных наблюдениях, когда

$$a_1(X_t, t, \Theta) = b_1(t, \Theta) X_t + b_0(t, \Theta),$$

уравнения (23) и (24) упрощаются, поскольку

$$\begin{aligned} a_{10} (m_t^x, K_t^x, t, \Theta) &= b_0(t, \Theta); \\ a_{11} (m_t^x, K_t^x, t, \Theta) &= b_1(t, \Theta); \\ \beta_t &= R_t b_1(t, \Theta)^T \nu_2(t, \Theta)^{-1}, \end{aligned}$$

и принимают вид:

$$\begin{aligned} \dot{\hat{X}}_t &= a_{00} (m_t^x, K_t^x, t, \Theta) - a_{01} (m_t^x, K_t^x, t, \Theta) m_t^x + \\ &+ a_{01} (m_t^x, K_t^x, t, \Theta) \hat{X}_t + \\ &+ \beta_t \left[Z_t - b_1(t, \Theta) \hat{X}_t - b_0(t, \Theta) + b_1(t, \Theta) m_t^x \right], \\ \hat{X}_0 &= M X_0; \end{aligned}$$

$$\begin{aligned} \dot{R}_t &= a_{01} (m_t^x, K_t^x, t, \Theta) R_t + R_t a_{01} (m_t^x, K_t^x, t, \Theta)^T - \\ &- \beta_t b_1(t, \Theta) R_t + b(t, \Theta) \nu_1(t, \Theta) b(t, \Theta)^T, \\ R_0 &= M \left[(X_0 - \hat{X}_0) (X_0 - \hat{X}_0)^T \right]. \end{aligned}$$

Коэффициенты статистической линеаризации a_{00} , a_{01} , a_{10} и a_{11} и вспомогательная (инструментальная) матрица R_t размерности $n_z \times n_z$ не

содержат результаты наблюдений Z_t и могут быть определены отдельно (до получения результатов наблюдений). Тогда линеаризованные уравнения (18) и (20) в силу их простоты могут быть проинтегрированы в реальном масштабе времени в течение наблюдений системы. При этом возможна априорная оценка точности фильтра [9].

5 Пример

Рассмотрим скалярную нелинейную СтС вида

$$\varphi = -\dot{X}_t^3 - X_t^3 + X_t V(\Theta), \quad X_{t_0} = X_0. \quad (25)$$

Здесь $V(\Theta)$ — гауссовский белый шум интенсивности $\nu = \nu(\Theta)$. Выполним с (25) следующие преобразования. Проведем статистическую линеаризацию левой части (25), положив

$$\dot{X}_t^3 = \varphi_0 (m_t^{\dot{X}}, D_t^{\dot{X}}) + k_1 (m_t^{\dot{X}}, D_t^{\dot{X}}) \dot{X}_t^0,$$

где

$$\begin{aligned} \varphi_0 &= \varphi_0 (m_t^{\dot{X}}, D_t^{\dot{X}}) = D_t^{\dot{X}} \left[3 + \frac{(m_t^{\dot{X}})^2}{D_t^{\dot{X}}} \right]; \\ k_1 &= k_1 (m_t^{\dot{X}}, D_t^{\dot{X}}) = 3D_t^{\dot{X}} \left[1 + \frac{(m_t^{\dot{X}})^2}{D_t^{\dot{X}}} \right]. \end{aligned}$$

Учитывая, что $k_1 \neq 0$, заменим (25) приближенной СтС следующего вида:

$$\dot{X}_t = \alpha - \beta X_t^3 + \gamma X_t V(\Theta), \quad X(t_0) = X_0, \quad (26)$$

где

$$\begin{aligned} \alpha &= m_t^{\dot{X}} - \varphi_0 k_1^{-1} = m_t^{\dot{X}} - \frac{3 + (m_t^{\dot{X}})^2 / D_t^{\dot{X}}}{3 \left[1 + (m_t^{\dot{X}})^2 / D_t^{\dot{X}} \right]}; \\ \beta &= \gamma = k_1^{-1} = \frac{1}{3D_t^{\dot{X}} \left[1 + (m_t^{\dot{X}})^2 / D_t^{\dot{X}} \right]}. \end{aligned}$$

Система (26) при $\alpha = 0$ и $\beta = \gamma = 1$ подробно изучена в [8, 9]. Обобщая [8, 9] на случай $\beta = \gamma \neq 1$ с точностью до вероятностных моментов второго порядка, приходим к следующим взаимосвязанным уравнениям для $m_t^{\dot{X}}$, $D_t^{\dot{X}}$ [8]:

$$\begin{aligned} \dot{m}_t^{\dot{X}} &= \alpha - \beta m_t^{\dot{X}} \left[(m_t^{\dot{X}})^2 + 3D_t^{\dot{X}} \right], \quad m^{\dot{X}}(t_0) = m_0^{\dot{X}}; \\ \dot{D}_t^{\dot{X}} &= \beta^2 [\nu(\Theta) - 6D_t^{\dot{X}}] \left[(m_t^{\dot{X}})^2 + D_t^{\dot{X}} \right], \\ D^{\dot{X}}(t_0) &= D_0^{\dot{X}}. \end{aligned}$$

Пусть наблюдаемый процесс определяется уравнением

$$Z_t = \dot{Y}_t = X_t + V_2.$$

Процесс $W(t)$ состоит из двух независимых скалярных винеровских процессов W_1 и W_2 , слабыми с.к. производными которых служат белые шумы V_1 и V_2 соответственно. Соответствующую структуру имеют в этом случае матрицы

$$b(X_t, Y_t, t) = [X_t \ 0]; \quad b_1(Y_t, t) = [0 \ 1]; \quad \nu = \begin{bmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{bmatrix},$$

где ν_1 и ν_2 — интенсивности белых шумов V_1 и V_2 соответственно. Функции a и a_1 определяются формулами $a(X_t, Y_t, t) = -X_t^3$ и $a_1(X_t, Y_t, t) = X_t$. Уравнения (9) и (10) имеют вид:

$$\begin{aligned} \dot{\hat{X}}_t &= -\hat{X}_t (\hat{X}_t^2 + 3R_t) + \nu_2^{-1} R_t (Z_t - \hat{X}_t); \\ \dot{R}_t &= (\nu_1 - 6R_t) (\hat{X}_t^2 + R_t) - \nu_2^{-1} R_t^2. \end{aligned}$$

Эти уравнения приближенно определяют с.к. оптимальную оценку \hat{X}_t состояния системы и апостериорную дисперсию ошибки. За начальные значения \hat{X}_t и R_t следует взять априорные математическое ожидание и дисперсию величины X_0 , поскольку Y_0 не задано и может быть взято совершенно произвольно, независимо от X_0 .

Заметим, что полученный НСОФ упрощается, если дополнительно провести статистическую линеаризацию кубической функции в (25). В этом случае НСОФ будет обобщенным фильтром Калмана–Бьюси.

6 Качество нормального субоптимального фильтра

Алгоритмы синтеза НСОФ зависят от инструментальных параметров Θ . Они представляют собой случайные величины (СВ) или медленно меняющиеся функции времени. Применим методы теории чувствительности [10, 11] при гауссовских СВ Θ аналогично [8]. В результате получим для оценки качества НСОФ и условной функции потерь $\rho = R(\Theta, t)$, допускающей квадратическую аппроксимацию

$$\begin{aligned} \rho &= \rho(\Theta) = \rho(m^\Theta) + \\ &+ \sum_{i=1}^{p^\Theta} \rho'_i(m^\Theta) \Theta_i^0 + \sum_{i,j=1}^{p^\Theta} \rho''_{ij}(m^\Theta) \Theta_i^0 \Theta_j^0, \end{aligned}$$

показатель ε , равный

$$\varepsilon = \varepsilon_2^{1/4}, \quad \varepsilon_2 = M_{\mathcal{N}}[\rho(\Theta)]^2 - \rho(m^\Theta)^2,$$

где

$$\begin{aligned} M_{\mathcal{N}}[\rho(\Theta)^2] &= \rho(m^\Theta)^2 + \rho'(m^\Theta)^T K^\Theta \rho'(m^\Theta) + \\ &+ 2\rho(m^\Theta) \text{tr}[\rho''(m^\Theta) K^\Theta] + \\ &+ \{\text{tr}[\rho''(m^\Theta) K^\Theta]\}^2 + 2\text{tr}[\rho''(m^\Theta) K^\Theta]^2. \end{aligned}$$

Функции чувствительности ρ'_i и ρ''_{ij} находятся путем решения уравнений для функций чувствительности $\nabla^\Theta R_t$ и $(\nabla^\Theta)^2 R_t$ в силу (10):

$$\nabla^\Theta \dot{R}_t = \nabla^\Theta B; \quad (\nabla^\Theta)^2 \dot{R}_t = (\nabla^\Theta)^2 B$$

при нулевых начальных условиях.

7 Заключение

Для нормальных (гауссовских) нелинейный дифференциальных СтС, не разрешенных относительно производных, разработаны методы синтеза с.к. НСОФ (теоремы 2–4). На основе принципа эквивалентности гауссовских и негауссовских шумов получено обобщение на случай негауссовских дифференциальных СтС. Разработанные методы справедливы только в случае, когда уравнение наблюдения не содержит пуассоновских шумов, а коэффициент при гауссовском белом шуме не зависит от обобщенного вектора состояния. Особое внимание уделено синтезу НСОФ для СтС с аддитивными шумами. Приведен пример, иллюстрирующий методы. Разработан метод оценки качества НСОФ на основе теории чувствительности.

В настоящее время идет разработка учебно-экспериментального программного обеспечения по тематике [12–15] в случаях, когда в уравнениях состояния можно пренебречь постоянными времени высокого порядка.

В качестве обобщений могут рассматриваться следующие задачи:

- (1) синтез с.к. условно-оптимальных фильтров, экстраполяторов и интерполяторов на основе параметризации распределений (моментов, семиинвариантов, ортогональных разложений, эллипсоидальной аппроксимации и др.);
- (2) в том случае когда вектор возмущений и/или нелинейные функции заданы каноническими представлениями случайных функций, полученные результаты допускают обобщения, если воспользоваться [2, 3];
- (3) синтез с.к. условно-оптимальных фильтров для стохастических эредитарных интегродифференциальных систем, не разрешенных относительно производных.

Наконец, теоретический и практический интерес представляют задачи, когда функции (1) являются разрывными, а также стохастическими функциями отмеченных переменных.

Литература

1. Пугачёв В. С., Синицын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. (*Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering.* — Chichester — New York: John Wiley & Sons, 1987. 549 p.)
2. Пугачёв В. С., Синицын И., Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
3. Синицын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009. 768 с.
4. Синицын И. Н., Синицын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
5. Синицын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
6. Синицын И. Н., Синицын В. И. Аналитическое моделирование нормальных процессов в стохастических системах со сложными нелинейностями // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 2–4.
7. Синицын 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.
8. Синицын И. Н. Параметрическое аналитическое моделирование процессов в стохастических системах, не разрешенных относительно производных // Системы и средства информатики, 2017. Т. 27. № 1. С. 20–45.
9. Синицын И. Н. Фильтры Калмана и Пугачёва. — 2-е изд. — М.: Логос, 2007. 776 с.
10. Справочник по теории автоматического управления / Под ред. А. А. Красовского. — М.: Наука, 1987. 712 с.
11. Евланов А. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 568 с.
12. ГОСТ 23743–88. Изделия авиационной техники. Номенклатура показателей безопасности полета, надежности, контролепригодности, эксплуатационной и ремонтной технологичности.
13. Александровская Л. Н., Аронов И. З., Круглов В. И. и др. Безопасность и надежность технических систем. — М.: Университетская книга, Логос, 2008. 376 с.
14. Синицын И. Н. Лекции по теории систем интегрированной логистической поддержки. — М.: ТОРУС ПРЕСС, 2019. 1072 с.
15. Sinitsyn I. N., Shalamov A. S. Probabilistic modeling, estimation and control for CALS organization-technical-economic systems // Probability, combinatorics and control and control / Eds. A. Kostohryzov, V. Korolev. — Intechopen, 2019. P. 117–141. doi: 10.5772.79802.

Поступила в редакцию 29.06.2020

NORMAL SUBOPTIMAL FILTERING FOR DIFFERENTIAL STOCHASTIC SYSTEMS WITH UNSOLVED DERIVATIVES

I. N. Sinitsyn

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

Abstract: The article develops the series of papers dedicated to stochastic systems with unsolved derivatives. Methodological aspects of normal suboptimal filterings (NSOF) for stochastic systems with unsolved derivatives are presented. Nonlinear differential equations for state and observation are given at the following conditions: observation equations are Gaussian and do not depend on the state variable. One of sections is devoted to NSOF for Gaussian and non-Gaussian systems. Corresponding NSOF are given for additive noises. Also, an illustrative example is given. The NSOF quality analysis is considered.

Keywords: method of analytical modeling (MAM); method of normal approximation (MNA); method of statistical linearization (MSL); normal suboptimal filter; stochastic system (StS); stochastic systems with unsolved derivatives; shaping filter

DOI: 10.14357/19922264210101

References

1. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester – New York: John Wiley & Sons. 549 p.
2. Pugachev, V. S., and I. N. Sinitsyn. 2000, 2004. *Teoriya stokhasticheskikh sistem* [Stochastic systems. Theory and applications]. Moscow: Logos. 1000 p.
3. Sinitsyn, I. N. 2009. *Kanonicheskie predstavleniya sluchaynykh funktsiy i ikh primeneniye v zadachakh komp'yuternoy podderzhki nauchnykh issledovaniy* [Canonical expansions of random functions and their applications in computer-aided support]. Moscow: TORUS PRESS. 768 p.
4. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Leksii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation in stochastic systems]. Moscow: TORUS PRESS. 488 p.
5. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraziyakh [Parametric statistical and analytical modeling of distributions in stochastic systems on manifolds]. *Informatika i eePrimeneniya — Inform. Appl.* 7(2):4–16.
6. Sinitsyn, I. N., and V. I. Sinitsyn. 2014. Analiticheskoe modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi nelineynostyami [Analytical modeling of normal processes in stochastic systems with complex nonlinearities]. *Informatika i eePrimeneniya — Inform. Appl.* 8(3):2–4.
7. 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.
8. Sinitsyn, I. N. 2017. Parametricheskoe analiticheskoe modelirovanie protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel'no proizvodnykh [Parametric analytical modeling of wide band processes in stochastic systems with unsolved derivatives]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(1):20–45.
9. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd. ed. Moscow: Logos. 776 p.
10. Krasovskiy, A. A., ed. 1987. *Spravochnik po teorii avtomaticheskogo upravleniya* [Handbook for automatic control]. Moscow: Nauka. 712 p.
11. Evlanov, A. G., and V. M. Konstantinov. 1976. *Sistemy so slozhnymi parametrami* [Systems with random parameters]. Moscow: Nauka. 568 p.
12. GOST 23743-88. Izdeliya aviatsionnoy tekhniki. Nomenklatura pokazateley bezopasnosti poleta, nadezhnosti, kontroleprigodnosti, ekspluatatsionnoy i remontnoy tekhnologichnosti [Aircraft products. Nomenclature of flight safety indicators, reliability, controllability, operational and repair processability].
13. Aleksandrovskaia, L. N., I. Z. Aronov, V. I. Kruglov, et al. 2008. *Bezopasnost' i nadezhnost' tekhnicheskikh sistem* [Security and reliability of technical systems]. Moscow: Universitetskaya kniga, Logos. 376 p.
14. Sinitsyn, I. N. 2019. *Leksii po teorii sistem integrirrovannoy logisticheskoy podderzhki* [Lectures on theory of integrated logistic support systems]. Moscow: TORUS PRESS. 1072 p.
15. Sinitsyn, I. N., and A. S. Shalamov. 2019. Probabilistic modeling, estimation and control for CALS organization-technical-economic systems. *Probability, combinatorics and control and control*. Eds. A. Kostohryzov and V. Korablev. IntechOpen. 117–141. doi: 10.5772.79802.

Received June 29, 2020

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; sinitsin@dol.ru

О НЕКОТОРЫХ ЧАСТНЫХ СЛУЧАЯХ В ЗАДАЧЕ УПРАВЛЕНИЯ ВЫХОДОМ СТОХАСТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ СИСТЕМЫ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ*

А. В. Босов¹

Аннотация: Общее исследование задачи оптимального управления для диффузионного процесса Ито и линейного управляемого выхода с квадратичным критерием качества выполнено в предыдущих работах автора (совместно с А. И. Стефановичем). Анализ имеющихся результатов позволяет выделить некоторые модели, носящие частный характер в отношении общей постановки, но имеющие особую практическую значимость. В данной статье рассмотрены две такие частные модели. Первая модель определяется предположением о линейном сносе в уравнении состояния при сохранении нелинейной диффузии. Показано, что такая модель обеспечивает линейность оптимальному управлению и отсутствие необходимости для его реализации решать параболическое уравнение. При этом квадратичной функции Беллмана в задаче не возникает, соответствующее выражение описывается, как и в общем случае, решением параболического уравнения и сохраняет содержательную стохастическую интерпретацию, выражаемую формулой Фейнмана–Каца. Вторая модель реализует предположение о зависимости возмущений в уравнениях состояния и выхода. Модифицированное уравнение динамического программирования решается тем же способом, что и в общем случае, рассмотренном в предыдущих работах, в том числе в рамках объединенной модели, включающей оба представленных случая. Данная модель окажется особо востребованной в задачах с неполной информацией, когда предположение о наличии полной информации о состоянии и выходе будет заменяться описанием системы наблюдения, в которой выход интерпретируется как косвенные наблюдения за состоянием. Кратко обсуждается численный пример, детально исследованный в предыдущих работах автора (совместно с А. И. Стефановичем), так как оказывается, что он удовлетворяет предположению о линейном сносе в уравнении состояния и, соответственно, полученные ранее приближенные решения удается уточнить.

Ключевые слова: стохастическое дифференциальное уравнение; оптимальное управление; управление выходом; дифференциальные системы с мультипликативными и зависимыми возмущениями

DOI: 10.14357/19922264210102

1 Введение

Общая постановка задачи управления линейным выходом стохастической дифференциальной системы по квадратичному критерию качества, предложенная в [1], использует в качестве модели для состояния y_t управляемой по выходу системы стохастическое дифференциальное уравнение Ито с винеровским процессом v_t (как и в цитируемых работах, в качестве непринципиального упрощения далее рассматривается скалярный случай):

$$dy_t = A_t(y_t) dt + \Sigma_t(y_t) dv_t, \quad y_0 = Y. \quad (1)$$

Предположения в отношении нелинейных функций $A_t(y)$ и $\Sigma_t(y)$ ограничены лишь требованием существования единственного решения.

Моделью связанного с состоянием y_t линейного выхода z_t также является уравнение Ито с независимым винеровским процессом w_t , линейное по y_t, z_t и управлению u_t :

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z. \quad (2)$$

Функции a_t, b_t, c_t и σ_t предполагаются ограниченными, процесс управления u_t — допустимый с полной обратной связью $u_t = u_t(y, z)$ [2], что обеспечивает существование и единственность решения уравнения (2).

Для оптимизации $U_0^T = \{u_t, 0 \leq t \leq T\}$ используется квадратичный целевой функционал:

$$J(U_0^T) = E \left\{ \int_0^T \left(S_t (s_t y_t - g_t z_t - h_t u_t)^2 + G_t z_t^2 + H_t u_t^2 \right) dt + S_T (s_T y_T - g_T z_T)^2 + G_T z_T^2 \right\}, \quad (3)$$

где S_t, G_t и H_t — неотрицательные ограниченные функции.

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

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

Реализация полученных в [1] соотношений для оптимального управления $(U^*)_0^T = \{u_t^*, 0 \leq t \leq T\}$, доставляющего минимум $J(U_0^T)$, требует решения дифференциальных уравнений в частных производных параболического типа. В [3] с этой целью используется традиционный метод сеток, в [4] к свойствам оптимального управления в части, связанной с параболическим уравнением, добавляется формула Фейнмана–Каца, на основании которой в [5] реализуется приближенное решение на основании имитационного моделирования и метода Монте-Карло.

В данной работе по отношению к модели (1), (2) сделаны два частных предположения, имеющих особое прикладное значение. Во-первых, уравнение (1) упрощается предположением о линейности сноса $A_t(y)$ при сохранении нелинейной диффузии $\Sigma_t(y)$. Такие модели широко известны как линейные модели с мультипликативными по состоянию возмущениями, их исследованию посвящено множество работ, обзор которых может представлять самостоятельный исследовательский интерес. В качестве близкой к тематике данной статьи можно указать, например, работу [6], где задача управления ставится в терминах H_∞ -робастности или, точнее, H_2/H_∞ -управления. Таких работ очень много, что связано как с практической востребованностью, так и с наличием успешно применимых к этим моделям исследовательских инструментов, в отличие от моделей полностью нелинейных.

Второе уточнение в исходную модель вносит отказ от предположения о независимости возмущений v_t и w_t . Такое уточнение уместно с учетом формирования линейной модели управляемой системы с мультипликативными возмущениями. В условиях полностью наблюдаемого состояния и выхода присутствие в уравнениях для y_t и z_t одного и того же возмущения вполне возможно. Второе полезное приложение зависимые возмущения дадут при решении задачи с неполной информацией, когда состояние y_t неизвестно, а выход z_t интерпретируется как косвенные наблюдения. В таких приложениях на первый план выходят задачи стохастической фильтрации. Общие постановки такого характера детально исследованы в монографии [7].

2 Частная модель

«состояние—выход» и решение задачи оптимизации

Для формальной постановки задачи воспользуемся теми же уравнениями (1), (2), внося в их

описание следующие дополнения. Во-первых, будем предполагать, что функция сноса в (1) имеет вид:

$$A_t = A_t(y) = A_t^a y + A_t^b, \quad (4)$$

т. е. является линейной по аргументу y . Неслучайные функции A_t^a и A_t^b предполагаются ограниченными, что обеспечивает выполнение условий Ито для существования решения (1).

Второе уточнение состоит в отказе от независимости винеровских процессов v_t и w_t в (1), (2). Для описания этих возмущений введем два независимых стандартных винеровских процесса ξ_t и η_t и будем считать, что v_t и w_t порождаются следующим линейным преобразованием:

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} = \begin{pmatrix} 1 - \Lambda_t & \Lambda_t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix}, \quad (5)$$

где функция Λ_t предполагается ограниченной: $|\Lambda_t| \leq 1, 0 \leq t \leq T$. Такое преобразование обеспечивает коррелированность возмущений: $E\{v_t w_t\} = \Lambda_t t$, в том числе допускается возможность не только зависимости, но и равенства возмущений $v_t = w_t$ при $\Lambda_t = 1$.

Решение задачи, как и в общем случае [1], дает метод динамического программирования. Обозначив функцию Беллмана

$$\begin{aligned} V_t &= V_t(y, z) = \\ &= \inf_{U_t^T} E \left\{ \int_0^T (S_\tau (s_\tau y_\tau - g_\tau z_\tau - h_\tau u_\tau)^2 + G_\tau z_\tau^2 + \right. \\ &\quad \left. + H_\tau u_\tau^2) d\tau + S_T (s_T y_T - g_T z_T)^2 + G_T z_T^2 | \mathcal{F}_t^{y,z} \right\}, \end{aligned}$$

где $\mathcal{F}_t^{y,z}$ — σ -алгебра, порожденная $y_\tau, z_\tau, 0 \leq \tau \leq t, E\{\cdot | \mathcal{F}\}$ — оператор условного математического ожидания относительно \mathcal{F} , и опуская аргументы у функций $A_t(y)$ и $\Sigma_t(y)$, записываем уравнение динамического программирования

$$\begin{aligned} \frac{\partial V_t}{\partial t} + \frac{1}{2} \left((1 - \Lambda_t)^2 + \Lambda_t^2 \right) \Sigma_t^2 \frac{\partial^2 V_t}{\partial y^2} + \sigma_t^2 \frac{\partial^2 V_t}{\partial z^2} + \\ + \Lambda_t \Sigma_t \sigma_t \frac{\partial^2 V_t}{\partial y \partial z} + \\ + \min_u \left[A_t \frac{\partial V_t}{\partial y} + (a_t y + b_t z + c_t u) \frac{\partial V_t}{\partial z} + \right. \\ \left. + S_t (s_t y - g_t z - h_t u)^2 + G_t z^2 + H_t u^2 \right] = 0, \\ V_T(y, z) = S_T (s_T y - g_T z)^2 + G_T z^2. \quad (6) \end{aligned}$$

Отличие этого уравнения от полученного для случая независимых возмущений v_t и w_t состоит в до-

полнительном слагаемом $\Lambda_t \Sigma_t \sigma_t \partial^2 V_t / (\partial y \partial z)$. При этом сохраняется возможность реализовать главное предположение, используемое для решения, а именно: представление функции Беллмана в виде

$$V_t = \alpha_t z^2 + \beta_t(y)z + \gamma_t(y), \quad (7)$$

так как $\partial^2 V_t / (\partial y \partial z) = \partial \beta_t / \partial y$, что дает ту же структуру решаемого уравнения, что и в [1], и сводит задачу оптимизации к поиску достаточно гладких функций α_t , $\beta_t(y)$ и $\gamma_t(y)$, обеспечивающих решение уравнения (6). При этом оптимальное управление, доставляющее минимум (3), имеет тот же вид:

$$\begin{aligned} u_t^* &= u_t^*(y_t, z_t) = \\ &= -\frac{1}{2} (S_t h_t^2 + H_t)^{-1} (c_t (2\alpha_t z_t + \beta_t(y_t)) + \\ &\quad + 2S_t (s_t y_t - g_t z_t) h_t) \end{aligned} \quad (8)$$

с учетом предположения $S_t h_t^2 + H_t > 0$ и равенства $\partial V_t / \partial z = 2\alpha_t z_t + \beta_t(y)$.

Уточнение уравнений, определяющих искомые коэффициенты α_t , $\beta_t(y)$ и $\gamma_t(y)$, в связи с предположением (5) о коррелированности возмущений дает

$$\begin{aligned} \frac{d\alpha_t}{dt} + 2\alpha_t (b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t) + \\ + (S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2) g_t^2 + G_t - \\ - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t^2 = 0, \quad \alpha_T = S_T g_t^2 + G_T; \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial \beta_t}{\partial t} + A_t \frac{\partial \beta_t}{\partial y} + \frac{1}{2} ((1 - \Lambda_t)^2 + \Lambda_t^2) \Sigma_t^2 \frac{\partial^2 \beta_t}{\partial y^2} + \\ + M_t y + N_t \beta_t = 0, \quad \beta_T = -2S_T s_T g_T y; \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{\partial \gamma_t}{\partial t} + A_t \frac{\partial \gamma_t}{\partial y} + \frac{1}{2} ((1 - \Lambda_t)^2 + \Lambda_t^2) \Sigma_t^2 \frac{\partial^2 \gamma_t}{\partial y^2} + L_t = 0, \\ \gamma_T = S_T s_T^2 y^2, \end{aligned} \quad (11)$$

где

$$\begin{aligned} M_t &= 2 \left(\alpha_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) - \right. \\ &\quad \left. - \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t g_t \right); \\ N_t &= b_t - (S_t h_t^2 + H_t)^{-1} c_t S_t h_t g_t - \\ &\quad - (S_t h_t^2 + H_t)^{-1} c_t^2 \alpha_t; \\ L_t &= L_t(y) = \Lambda_t \Sigma_t \sigma_t \frac{\partial \beta_t}{\partial y} + \sigma_t^2 \alpha_t + \\ &\quad + \beta_t \left(a_t + (S_t h_t^2 + H_t)^{-1} c_t S_t h_t s_t \right) y + \\ &\quad + \left(S_t - (S_t h_t^2 + H_t)^{-1} S_t^2 h_t^2 \right) s_t^2 y^2 - \\ &\quad - \frac{1}{4} (S_t h_t^2 + H_t)^{-1} c_t^2 \beta_t^2. \end{aligned}$$

Остается учесть частное условие на модель (4). На уравнение Риккати (9) это предположение никак не влияет. Также не претерпит изменений параболическое уравнение (11). Принципиально изменится только уравнение (10). Действительно, поскольку в этом уравнении коэффициенты M_t и N_t не зависят от y , а в силу (4) $A_t(y)$ линейно по y , то решение (10) можно искать в виде

$$\beta_t = \beta_t(y) = \beta_t^a y + \beta_t^b,$$

т. е. в виде линейной функции. Для уравнения (11) линейность $A_t(y)$ дает единственное небольшое изменение — это явное представление $\partial \beta_t / \partial y = \beta_t^a$ в соотношении, определяющем свободный член L_t . В уравнении (10), прежде всего, устраняется лапласиан и оно фактически перестает быть уравнением в частных производных, преобразовываясь в систему обыкновенных дифференциальных уравнений:

$$\begin{aligned} \frac{d\beta_t^a}{dt} y + \frac{d\beta_t^b}{dt} + \\ + (A_t^a y + A_t^b) \beta_t^a + M_t y + N_t (\beta_t^a y + \beta_t^b) = 0, \end{aligned}$$

откуда для β_t^a и β_t^b получается

$$\left. \begin{aligned} \frac{d\beta_t^a}{dt} + (A_t^a + N_t) \beta_t^a + M_t = 0, \\ \beta_T^a = -2S_T s_T g_T; \\ \frac{d\beta_t^b}{dt} + A_t^b \beta_t^a + N_t \beta_t^b = 0, \quad \beta_T^b = 0, \end{aligned} \right\} \quad (12)$$

т. е. система из двух линейных уравнений.

Оптимальное управление (8) принимает вид:

$$\begin{aligned} u_t^* &= -\frac{1}{2} (S_t h_t^2 + H_t)^{-1} (2(c_t \alpha_t - S_t g_t h_t) z_t + \\ &\quad + (c_t \beta_t^a + 2S_t s_t h_t) y_t + c_t \beta_t^b). \end{aligned} \quad (13)$$

Полученное решение, т. е. соотношения (12), (13), характерно тем, что оптимальное управление в рассмотренном частном случае может быть найдено аналитически — без численного решения параболического уравнения (11). Очевидно, что к такому же результату приводит и другой частный случай — классическая линейно-гауссовская постановка. В ней функция Беллмана (7) является квадратичной, что означает и линейность коэффициента $\beta_t(y)$, и квадратичность коэффициента $\gamma_t(y)$ по y , т. е. кроме того, что решение (10) может быть найдено в виде $\beta_t = \beta_t^a y + \beta_t^b$, еще и решение (11) представляется в виде

$$\gamma_t = \gamma_t^a y^2 + \gamma_t^b y + \gamma_t^c. \quad (14)$$

Рассмотренная постановка представляет более интересный частный вариант. Оказалось, что для линейности коэффициента $\beta_t(y)$ достаточно линейного сноса $A_t(y)$ в (1), а линейность диффузии $\Sigma_t(y)$ не требуется. В отличие от линейно-гауссовского случая функция Беллмана при этом не является квадратичной, решение уравнения (11) в виде (14) не представляется именно из-за нелинейности $\Sigma_t(y)$, но простое линейное представление оптимального управления (13) в этом практически важном частном случае получается.

3 Численный пример

Для иллюстрации рассмотренного частного случая линейного сноса в модели состояния (1) оказались применимы выполненные в работах [3, 5] эксперименты с математической моделью для показателя RTT (Round-Trip Time) сетевого протокола TCP (Transmission Control Protocol), предложенной в [8], следующего вида:

$$dy_t = (1 - 0,1y_t) dt + 0,5\sqrt{y_t} dv_t, \\ y_0 = Y \sim N(15,9). \quad (15)$$

Здесь $N(M, D)$ — нормальное распределение со средним M и дисперсией D .

Выход для (15) задается уравнением

$$dz_t = 0,1y_t dt - z_t dt + u_t dt + dw_t, \\ z_0 = Z \sim N(9,9), \quad (16)$$

целевой функционал

$$J(U_0^T) = E \left\{ \int_0^T ((y_t - z_t)^2 + z_t^2 + u_t^2) dt + \right. \\ \left. + (y_T - z_T)^2 + z_T^2 \right\}. \quad (17)$$

Поскольку в работе [3] эксперименты с данным примером (приближенное вычисление коэффициента $\beta_t(y)$, оптимального управления (8) и его качества $J(U_0^t)$) проведены традиционным методом сеток, а в [5] — методом Монте-Карло на основе формулы Фейнмана–Каца, то здесь можно получить эталонное решение, вычислив $\beta_t(y)$ по формулам (12), а оптимальное управление — по формуле (13). Для этого вначале решается уравнение Риккати (9) для α_t . С учетом того, что в модели выхода (16) и целевом функционале (17) все коэффициенты постоянны, это уравнение решается аналитически, а именно:

$$\alpha_t = \frac{C_\alpha e^{2\sqrt{3}x}(1 + \sqrt{3}) - 1 + \sqrt{3}}{1 - C_\alpha e^{2\sqrt{3}x}},$$

где

$$C_\alpha = \frac{3 - \sqrt{3}}{3 + \sqrt{3}} e^{-10\sqrt{3}}.$$

Далее решалась система (12) линейных уравнений для β_t^a и β_t^b :

$$\frac{d\beta_t^a}{dt} - (1,1 - \alpha_t) \beta_t^a + 2\alpha_t + 2 = 0; \\ \frac{d\beta_t^b}{dt} - (1 + \alpha_t) \beta_t^b + \beta_t^a = 0.$$

Для этого был использован неявный метод Эйлера с шагом $\delta_t = 0,001$. Это тот же шаг, что использовался в сеточном методе для расчета $\beta_t(y)$ в [3] и для дискретизации и моделирования решений (15), (16) в [5].

Выполненный расчет иллюстрируется рис. 1, на котором изображен фрагмент поверхности $\beta_t(y)$, и рис. 2, на котором представлены отдельные сечения этой поверхности в точках $t = 0,0; 1,0; 2,0; 3,0; 4,0; 4,1; 4,2; 4,3; 4,4; 4,5; 4,6; 4,7; 4,8; 4,9; 4,95; 5,0$, т. е. образующие поверхность прямые $\beta_t^a y + \beta_t^b$.

Визуальное сравнение поверхности на рис. 2 с аппроксимациями, полученными в [3, 5], пока-

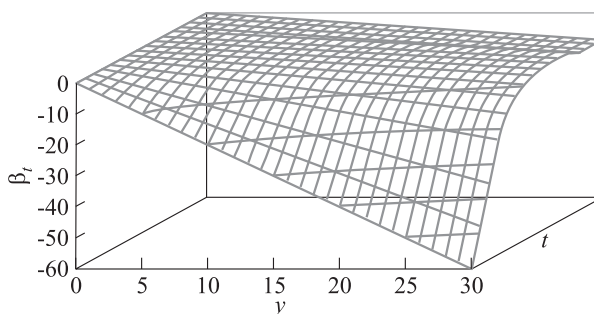


Рис. 1 Поверхность $\beta_t(y)$

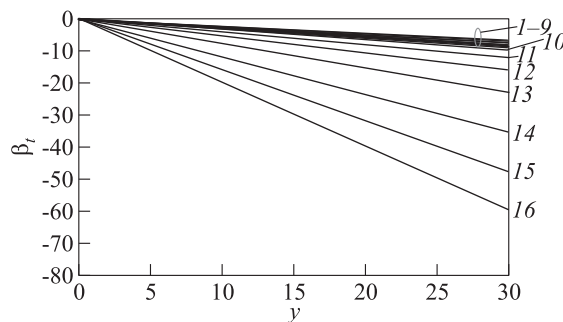


Рис. 2 Сечения поверхности $\beta_t(y)$: 1 — $t = 0,0$; 2 — $1,0$; 3 — $2,0$; 4 — $3,0$; 5 — $4,0$; 6 — $4,1$; 7 — $4,2$; 8 — $4,3$; 9 — $4,4$; 10 — $4,5$; 11 — $4,6$; 12 — $4,7$; 13 — $4,8$; 14 — $4,9$; 15 — $4,95$; 16 — $t = 5,0$

зывает их полное сходство, относительные ошибки в большинстве точек (t, y) составляют 1%–2%, в отдельных, особенно близких к границе, достигают 8%–12%. Это свидетельствует о достаточно высокой точности методов, представленных в [3, 5], так как в данном расчете точность ограничена только шагом решения линейного дифференциального уравнения методом Эйлера, поэтому результат можно обоснованно считать эталонным и использовать для сравнения с действительно приближенными решениями из [3, 5]. Наконец, окончательный вывод о качестве аппроксимаций в сравнении с точным решением дает обеспечиваемое методом расчета оптимального управления значение целевого функционала (17). Заметим, что в [5] показано преимущество метода Монте-Карло в сравнении с сеточной аппроксимацией, обеспечившего меньшее значение $J(U_0^T)$ в терминальной точке на величину порядка 5%. Вычисленное здесь точное оптимальное управление дает по критерию преимущество еще порядка 3%.

Заключение

В рамках исследования задачи оптимизации линейного выхода нелинейной дифференциальной системы по квадратичному критерию, выполненному в работах [1, 3–5], получено решение задачи в общем случае нелинейного уравнения Ито, описывающего состояние оптимизируемой системы. Для практической реализации оптимального управления приходится решать дифференциальное уравнение в частных производных параболического типа, для чего имеются вполне эффективные численные методы, причем технически разные, что полезно, например, для контроля правильности выполненного расчета. Но открытым при этом остается вопрос точности аппроксимаций, т. е. близости приближенного решения к точному. Для ответа на него требуется эталонная модель, пример, в котором решение можно было бы вычислить точно. Именно эта цель и достигнута в данной работе. Частный случай модели состояния, интерпретируемый как модель с мультипликативными возмущениями, дал возможность решить действительно нелинейную задачу фактически аналитически и не апеллировать к классической линейно-гауссовской постановке. Таким образом, хотя бы на

модельном примере удалось продемонстрировать эффективность предложенных ранее аппроксимаций рассматриваемой задачи.

Также полезным для дальнейших исследований представляется и возможность использования в модели «состояние–выход» зависимых возмущений. Уточнение уравнений для оптимального решения станет в дальнейшем основанием для рассмотрения постановок, предполагающих отказ от предположения о наблюдаемости состояния, понимания выхода как косвенных наблюдений и поиска управления как функции этих наблюдений.

Литература

1. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. I. Оптимальное решение методом динамического программирования // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 99–106.
2. Флеминг У., Ришел Р. Оптимальное управление детерминированными и стохастическими системами / Пер. с англ. — М.: Мир, 1978. 316 с. (Fleming W. H., Rishel R. W. Deterministic and stochastic optimal control. — New York, NY, USA: Springer-Verlag, 1975. 222 p.)
3. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. II. Численное решение уравнений динамического программирования // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 9–15.
4. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. III. Анализ свойств оптимального управления // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 41–49.
5. Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. IV. Альтернативное численное решение // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 24–30.
6. Chen B. S., Zhang W. Stochastic H_2/H_1 control with state-dependent noise // IEEE T. Automat. Contr., 2004. Vol. 49. Iss. 1. P. 45–56.
7. Луицер Р. Ш., Ширяев А. Н. Статистика случайных процессов (нелинейная фильтрация и смежные вопросы). — М.: Наука, 1974. 696 с.
8. Bohacek S., Rozovskii B. A diffusion model of roundtrip time // Comput. Stat. Data An., 2004. Vol. 45. Iss. 1. P. 25–50.

Поступила в редакцию 12.12.2020

ON SOME SPECIAL CASES IN THE PROBLEM OF STOCHASTIC DIFFERENTIAL SYSTEM OUTPUT CONTROL BY THE QUADRATIC CRITERION

A. V. Bosov

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

Abstract: A general study of the optimal control problem for the Ito diffusion process and linear controlled output with the quadratic quality criterion was carried out in the author’s previous works (coauthored by A. I. Stefanovich). An analysis of the available results allows to single out some models that are of a particular nature in relation to the general setting but have special practical significance. This article examines two such particular models. The first model is determined by the assumption of linear drift in the equation of state while maintaining nonlinear diffusion. It is shown that such a model provides linearity to the optimal control and the absence of the need to solve a parabolic equation for its implementation. But in this case, the quadratic Bellman function does not appear in the problem, the corresponding expression, as in the general case, is described by the solution of a parabolic equation and retains a meaningful stochastic interpretation expressed by the Feynman–Katz formula. The second model implements the assumption about the dependence of disturbances in the equations of state and output. The modified dynamic programming equation is solved in the same way as in the general case considered in previous works, including and within the framework of a combined model involving both cases presented. This model will be especially useful in the problems with incomplete information, when the assumption of the presence of complete information about the state and output will be replaced by a description of the observation system, in which the output is interpreted as indirect observations of the state. A numerical example, studied in detail in the author’s previous works (coauthored by A. I. Stefanovich), is briefly discussed, since it turns out that it satisfies the assumption of linear drift in the equation of state and, accordingly, the previously obtained approximate solutions can be refined.

Keywords: stochastic differential equation; optimal control; system output control; stochastic differential systems with multiplicative and dependent disturbances

DOI: 10.14357/19922264210102

Acknowledgments

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

References

1. Bosov, A. V., and A. I. Stefanovich. 2018. Upravlenie vykhodom stokhasticheskoy differentsial’noy sistemy po kvadraticnomu kriteriyu. I. Optimal’noe reshenie metodom dinamicheskogo programmirovaniya [Stochastic differential system output control by the quadratic criterion. I. Dynamic programming optimal solution]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):99–106.
2. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p.
3. Bosov, A. V., and A. I. Stefanovich. 2019. Upravlenie vykhodom stokhasticheskoy differentsial’noy sistemy po kvadraticnomu kriteriyu. II. Chislennoe reshenie uravneniy dinamicheskogo programmirovaniya [Stochastic differential system output control by the quadratic criterion. II. Dynamic programming equations numerical solution]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):9–15.
4. Bosov, A. V., and A. I. Stefanovich. 2019. Upravlenie vykhodom stokhasticheskoy differentsial’noy sistemy po kvadraticnomu kriteriyu. III. Analiz svoystv optimal’nogo upravleniya [Stochastic differential system output control by the quadratic criterion. III. Optimal control properties analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):41–49.
5. Bosov, A. V., and A. I. Stefanovich. 2020. Upravlenie vykhodom stokhasticheskoy differentsial’noy sistemy po kvadraticnomu kriteriyu. IV. Al’ternativnoe chislennoe reshenie [Stochastic differential system output control by the quadratic criterion. IV. Alternative numerical decision]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):24–30.
6. Chen, B. S., and W. Zhang. 2004. Stochastic H_2/H_1 control with state-dependent noise. *IEEE T. Automat. Contr.* 49(1):45–56.

7. Liptser, R. S., and A. N. Shiryaev. 2001. *Statistics of random processes II. Applications*. Berlin: Springer-Verlag. 402 p.
8. Bohacek, S., and B. Rozovskii. 2004. A diffusion model of roundtrip time. *Comput. Stat. Data An.* 45(1):25–50.

Received December 12, 2020

Contributor

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

СВЯЗНОСТЬ КОНФИГУРАЦИОННЫХ ГРАФОВ В МОДЕЛЯХ СЛОЖНЫХ СЕТЕЙ*

Ю. Л. Павлов¹

Аннотация: Рассматриваются конфигурационные графы, степени вершин которых являются независимыми случайными величинами, одинаково распределенными по обобщенному дискретному степенному закону. Связи между вершинами формируются равновероятно в соответствии со степенями вершин. Эти случайные графы часто используются для моделирования сложных сетей коммуникаций, таких как интернет и социальные сети. В статье предполагается, что распределение степеней вершин неизвестно, поскольку зависит от медленно меняющейся функции с неизвестными свойствами. При стремлении числа вершин к бесконечности найдены условия, при выполнении которых граф становится асимптотически достоверно связным. При этих условиях получены оценки скорости сходимости к нулю вероятности того, что граф не связан. Для доказательства результатов статьи использовались свойства устойчивых распределений и медленно меняющихся функций.

Ключевые слова: случайные графы; конфигурационные графы; случайные степени вершин; связность графа

DOI: 10.14357/19922264210103

1 Введение

Случайные графы широко используются при моделировании сложных сетей коммуникаций, таких как интернет, транспортные, телефонные, социальные сети и т.д. [1]. Такие модели обычно соответствуют известным свойствам реальных сетей, обнаруженным в ходе многочисленных эмпирических исследований и подробно описанным в различных публикациях (см., например, [2]). Наблюдения показали, что в графах, описывающих топологию сетей, степени вершин можно считать независимыми одинаково распределенными случайными величинами. Одно из важнейших свойств, присущих сетям коммуникаций различной природы, состоит в том, что число вершин степени, не меньшей чем k , при больших k пропорционально $k^{-\tau}$, при этом значения положительного параметра τ могут отличаться для разных сетей. Поэтому во многих работах предполагается, что распределение случайной величины ξ , равной степени любой вершины графа, можно задать следующим образом:

$$P\{\xi \geq k\} = \frac{h(k)}{k^\tau}, \quad k = 1, 2, \dots, \quad (1)$$

где $h(k)$ — медленно меняющаяся функция.

В настоящее время одной из наиболее популярных моделей сложных сетей коммуникаций служит

так называемый конфигурационный граф, определение которого было дано в статье [3].

Обозначим через N число вершин графа. Степени вершин задаются N независимыми реализациями случайной величины ξ , и они равны числу инцидентных каждой вершине полуребер, т.е. ребер, для которых смежные вершины еще не определены. Все полуребра различимы (занумерованы). Граф строится путем попарного равновероятного соединения полуребер друг с другом для образования ребер. Разумеется, сумма степеней всех вершин любого графа должна быть четной, в случае нечетной суммы в граф вводится вспомогательная вершина единичной степени. В статье [4] было замечено, что эта вспомогательная вершина не влияет на основные асимптотические свойства графа при $N \rightarrow \infty$. Поэтому ниже не будем учитывать эту дополнительную вершину, что фактически соответствует предположению о четности суммы степеней. Если же это предположение неверно, то, как легко убедиться, следуя доказательству полученных в данной статье результатов, дополнительная вершина не влияет на эти результаты. Заметим еще, что описанная конструкция конфигурационных графов допускает появление петель и кратных ребер.

Асимптотические свойства случайных конфигурационных графов при $N \rightarrow \infty$ изучались многи-

* 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 КарНЦ РАН, ФИЦ «Карельский научный центр РАН», pavlov@krc.karelia.ru

ми авторами, наиболее полные обзоры таких работ можно найти в [1, 5]. Известно, что если $\tau \in (1, 2)$, то асимптотически достоверно, т. е. с вероятностью, стремящейся к единице, граф содержит единственную гигантскую компоненту связности, число вершин в которой пропорционально N , в то время как объемы других компонент имеют порядок $o(N)$.

В статье [4] рассматривался простейший вариант распределения (1) в предположении, что $h(k) \equiv 1$. Для таких графов даны оценки объема и диаметра гигантской компоненты связности и подробно исследована ее структура.

В [6] впервые рассматривались условные конфигурационные графы с распределением (1) степеней вершин в случае $h(k) \equiv 1$ при условии, что число ребер известно. В [7] такие графы исследовались уже при условии, что число ребер ограничено сверху. Далее, в [8] в распределении (1) степеней вершин условных графов медленно меняющаяся функция $h(k)$ уже не предполагалась известной, но задавалась асимптотика вероятностей стремящихся к бесконечности значений степеней вершин:

$$P\{\xi = k\} \sim \frac{d}{k^g (\ln k)^\eta}, \quad (2)$$

где $k \rightarrow \infty$, $g > 1$, $\eta \geq 0$. Нетрудно проверить, что распределение (1) удовлетворяет условию (2), если $d = \tau$, $g = \tau + 1$ и $h(k) = (\ln k)^{-\eta}$. Хотя значения $h(k)$ в (1) известны не для всех k , знание асимптотического поведения вероятностей (2) существенно облегчило доказательство полученных в [8] результатов.

Обозначим A_N событие, состоящее в том, что граф не связан. В теореме 4.15 книги [9] показано, что если $P\{\xi = 1\} = P\{\xi = 2\} = 0$, то при $N \rightarrow \infty$ асимптотически достоверно граф состоит из единственной компоненты связности, содержащей все N вершин. В [9] установлено также, что в этом случае $P\{A_N\} = O(1/N)$. В статье [10] были найдены условия асимптотически достоверной связности конфигурационных графов, степени вершин которых обладают свойством (2), в том числе и в случаях $P\{\xi = 1\} > 0$, $P\{\xi = 2\} > 0$.

Таким образом, осталась нерешенной задача нахождения условий, при выполнении которых асимптотически достоверно связан случайный конфигурационный граф, степени вершин которого заданы распределением (1) с неизвестной медленно меняющейся функцией $h(k)$, необязательно обеспечивающей выполнение соотношения (2). Основным результатом настоящей статьи является доказанная ниже теорема, в которой найдены такие условия и даны оценки скорости сходимости к нулю вероятности того, что граф не связан. Согласно теореме 4.4 [9], если $M\xi(\xi - 1)/M\xi > 1$, то граф

асимптотически достоверно содержит больше одной компоненты связности. Нетрудно видеть, что это имеет место в случае $\tau > 1$, поэтому в теореме рассматривается только случай $0 < \tau \leq 1$.

Доказательство теоремы основано на идеях, изложенных в доказательстве теоремы 4.15 [9], но с существенными изменениями, связанными с использованием, как и в [10], методов исследования локальной сходимости распределений сумм независимых случайных величин к устойчивым законам [11]. Кроме того, впервые при решении подобных задач используются общие свойства медленно меняющихся функций, изложение соответствующей теории можно найти, например, в [12].

В следующем разделе в виде теоремы сформулирован основной результат статьи, доказательство этой теоремы приводится в разд. 3.

2 Основной результат

Пусть конфигурационный граф содержит N вершин, степени которых независимы и одинаково распределены в соответствии с (1), где $h(k)$, $k \geq 1$, — измеримая медленно меняющаяся функция. Предположим для простоты, что максимальный шаг распределения случайной величины ξ равен единице. Введем последовательность B_N , $N = 1, 2, \dots$, при $N \rightarrow \infty$ удовлетворяющую условию

$$B_N \sim (Nh([B_N]))^{1/\tau}, \quad (3)$$

где $[x]$ означает целую часть числа x . Отсюда и из свойств медленно меняющихся функций очевидным образом вытекает, что $B_N \rightarrow \infty$. Примером построения такой последовательности для распределения, имеющего свойство (2), могут служить величины вида $B_N = N^{1/\tau}$, если $\eta = 0$, а если $0 < \eta < 1$, то

$$B_N = \left(N \left(\frac{\tau}{\ln N} \right)^\eta \right)^{1/\tau}. \quad (4)$$

Теперь можно сформулировать основной результат статьи.

Теорема. Пусть $N \rightarrow \infty$ и $P\{\xi = 2\} > 0$. Тогда с вероятностью, сколь угодно близкой к единице, справедливы следующие утверждения.

1. Если $P\{\xi = 1\} = 0$ и $\tau = 1$, то $P\{A_N\} = O(1/\ln N)$.
2. Если $P\{\xi = 1\} = 0$ и $0 < \tau < 1$, то $P\{A_N\} = O(N/B_N)$.
3. Если $P\{\xi = 1\} > 0$ и $0 < \tau < 1/2$, то $P\{A_N\} = O(N^2/B_N)$.

3 Доказательство теоремы

Пусть произошло событие A_N . Тогда множество V всех вершин графа можно представить в виде $V = V_1 \cup V_2$, где V_1 и V_2 — непересекающиеся множества вершин такие, что не существует ребер, соединяющих вершины из V_1 с вершинами из V_2 . Можем считать, что $|V_1| \leq |V_2|$, где $|V_1|$ и $|V_2|$ — мощности множеств V_1 и V_2 соответственно. Тогда $|V_1| \leq N/2$. Обозначим Ω множество всех возможных разбиений V на V_1 и V_2 . Пусть случайные величины ξ_1, \dots, ξ_N равны степеням вершин $1, \dots, N$ соответственно. Понятно, что распределения этих независимых случайных величин совпадают с (1). Представляет интерес предельное поведение их суммы

$$\zeta_N = \xi_1 + \dots + \xi_N. \quad (5)$$

Обозначим также

$$\zeta_N^{(1)} = \sum_{i \in V_1} \xi_i, \quad \zeta_N^{(2)} = \sum_{i \in V_2} \xi_i. \quad (6)$$

Ясно, что общее число различных графов с суммой степеней вершин (5) равно $(\zeta_N - 1)!!$. Учитывая равновероятность соединения полуребер при образовании ребер, находим, что

$$P\{A_N\} \leq \sum_{V_1, V_2 \in \Omega} \frac{(\zeta_N^{(1)} - 1)!! (\zeta_N^{(2)} - 1)!!}{(\zeta_N - 1)!!}.$$

Отсюда следует оценка вероятности того, что граф не связан:

$$P\{A_N\} \leq \sum_{V_1 \in \Omega} \prod_{j=1}^{\zeta_N^{(1)}/2} \frac{\zeta_N^{(1)} - 2j + 1}{\zeta_N - 2j + 1}.$$

Верхний предел произведения в этом выражении равен $\zeta_N^{(1)}/2$, поскольку, очевидно, $\zeta_N^{(1)}$, как и $\zeta_N^{(2)}$, принимает четные значения.

Пусть $N \rightarrow \infty$. Обозначим $F(x)$ функцию распределения случайной величины ξ . Из (1) следует, что

$$F(x) = \begin{cases} 0, & \text{если } x < 0; \\ 1 - \frac{h(x)}{x^\tau} (1 + o(1)) & \text{при } x \rightarrow \infty. \end{cases} \quad (7)$$

В силу (7) функция $F(x)$ удовлетворяет условиям теоремы 2.6.1 книги [11], следовательно, она принадлежит области притяжения устойчивого закона $G(x)$ с показателем τ . Применяв теорему 2.2.2 [11], находим, что если $0 < \tau < 1$, то логарифм характеристической функции $\varphi_G(t)$ устойчивого закона $G(x)$ имеет вид:

$$\ln \varphi_G(t) = i\gamma t - c|t|^\tau \left(1 + i\beta \frac{t}{|t|} \tan \frac{\pi\tau}{2} \right), \quad (8)$$

где γ — некоторая постоянная, а значения параметров β и c с помощью (7) определены в ходе доказательства теоремы 2.2.2 [11]: $\beta = -1$;

$$c = \Gamma(1 - \tau) \cos \frac{\pi\tau}{2}, \quad (9)$$

где $\Gamma(x)$ — гамма-функция.

Вид логарифма характеристической функции $\varphi_G(t)$ при $\tau = 1$ определяется аналогичным образом:

$$\ln \varphi_G(t) = i\gamma t - \frac{\pi}{2} |t| \left(1 + i \frac{t}{|t|} \frac{2}{\pi} \ln |t| \right). \quad (10)$$

В выражениях (8) и (10) константу γ можно сделать равной нулю, подобрав нужным образом нормирующие множители в определении области притяжения закона $G(x)$ [11]. Обозначим $\varphi(t)$ характеристическую функцию случайной величины ξ . Из (8), (10) и теоремы 2.6.5 [11] следует, что в окрестности нуля

$$\ln \varphi(t) = -c|t|^\tau l(t) \left(1 + i\beta \frac{t}{|t|} \omega(t, \tau) \right), \quad (11)$$

где $\beta = -1$, а c определено в (9) в случае $0 < \tau < 1$, а если $\tau = 1$, то $\beta = 1$, $c = \pi/2$, функция $l(t)$ является медленно меняющейся и

$$\omega(t, \tau) = \begin{cases} \tan \frac{\pi\tau}{2}, & 0 < \tau < 1; \\ \frac{2}{\pi} \ln |t|, & \tau = 1. \end{cases} \quad (12)$$

В доказательстве теоремы 2.6.8 [11, с. 104–108] показано, что для распределения (1) при $t \rightarrow 0$

$$l(t) = h \left(\frac{1}{|t|} \right) (1 + o(1)). \quad (13)$$

Очевидно, что $\varphi(0) = 1$. Рассмотрим $\varphi(t)$ при фиксированных $t \neq 0$. Пусть $0 < \tau < 1$. Используя (3), (9), (11)–(13) и проводя простые вычисления, находим, что если $N \rightarrow \infty$, то

$$\varphi^N \left(\frac{t}{B_N} \right) \rightarrow \exp \left\{ -c|t|^\tau \left(1 - i \frac{t}{|t|} \tan \frac{\pi\tau}{2} \right) \right\}.$$

Это соотношение показывает, что распределения сумм ζ_N слабо сходятся к устойчивому закону с показателем τ . Заметим, что, согласно теореме 4.2.1 [11], на самом деле имеет место и локальная сходимость. Это значит, что при достаточно

больших N и любом сколь угодно малом $\varepsilon > 0$ существует положительная константа L такая, что

$$P \left\{ \frac{B_N}{L} \leq \zeta_N \leq LB_N \right\} > 1 - \varepsilon. \quad (14)$$

Пусть $\tau = 1$. С помощью (3), (11)–(13) и известных свойств медленно меняющихся функций [12] нетрудно вывести, что при $N \rightarrow \infty$ существует стремящаяся к нулю последовательность $q(N)$ такая, что при любом фиксированном $t \neq 0$

$$\begin{aligned} \varphi^N \left(\frac{t}{B_N} \right) \exp\{-itN(\ln N)(1 + q(N))\} \rightarrow \\ \rightarrow \exp \left\{ -\frac{\pi}{2} |t| \left(1 + i \frac{t}{|t|} \frac{2}{\pi} \ln |t| \right) \right\}. \end{aligned} \quad (15)$$

Из (15) и локальной предельной теоремы 4.2.1 [11] теперь вытекает, что асимптотически достоверно

$$\zeta_N \sim N \ln N. \quad (16)$$

Рассмотрим предельное поведение $\zeta_N^{(1)}$. Если объем множества V_1 конечен, то из (1) и (6) нетрудно получить, что сумма $\zeta_N^{(1)}$ асимптотически достоверно конечна. Пусть $|V_1| \rightarrow \infty$. Если $0 < \tau < 1$, то по аналогии с (3) и (14) находим, что при достаточно большом $|V_1|$ и достаточно малом $\varepsilon > 0$ существует положительная константа L_1 такая, что

$$P \left\{ \frac{B_N^{(1)}}{L_1} \leq \zeta_N^{(1)} \leq L_1 B_N^{(1)} \right\} > 1 - \varepsilon, \quad (17)$$

где

$$B_N^{(1)} = (|V_1| h(|V_1|))^{1/\tau}.$$

Если же $\tau = 1$, то замечаем, что, подобно (16),

$$\zeta_N^{(1)} \sim |V_1| \ln |V_1|. \quad (18)$$

Теперь для того, чтобы получить утверждения 1–3, достаточно повторить доказательство теоремы статьи [10], в котором (4) заменить на (3) и для оценки предельного поведения сумм ζ_N и $\zeta_N^{(1)}$ использовать соотношения (14), (16)–(18).

Литература

1. Hofstad R. Random graphs and complex networks. — Cambridge: Cambridge University Press, 2017. Vol. 1. 337 p.
2. Faloutsos C., Faloutsos P., Faloutsos M. On power-law relationships of the Internet topology // Comput. Commun. Rev., 1999. Vol. 29. P. 251–262.
3. Bollobas B. A. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // Eur. J. Combin., 1980. Vol. 1. P. 311–316.
4. Reittu H., Norros I. On the power-law random graph model of massive data networks // Perform. Evaluation, 2004. Vol. 55. P. 3–23.
5. Durrett R. Random graph dynamics. — Cambridge: Cambridge University Press, 2007. 212 p.
6. Павлов Ю. Л., Челюкова И. А. Случайные графы Интернет-типа и обобщенная схема размещения // Дискретная математика, 2008. Т. 20. Вып. 3. С. 3–18.
7. Павлов Ю. Л., Хворостянская Е. В. О предельных распределениях степеней вершин конфигурационных графов с ограниченным числом ребер // Математический сборник, 2016. Т. 207. Вып. 3. С. 93–110.
8. Павлов Ю. Л. Условные конфигурационные графы со случайным параметром степенного распределения степеней // Математический сборник, 2018. Т. 209. Вып. 2. С. 120–137.
9. Hofstad R. Random graphs and complex networks. Vol. 2 // Notes RGCNII, November 16, 2020. 341 p. <https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>.
10. Павлов Ю. Л. О связности конфигурационных графов // Дискретная математика, 2019. Т. 31. Вып. 2. С. 115–123.
11. Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. — М: Наука, 1965. 524 с.
12. Bigham N. H., Goldie C. M., Teugels J. L. Regular variations. Encyclopedia of mathematics and its applications. — Cambridge: Cambridge University Press, 1987. Vol. 27. 513 p.

Поступила в редакцию 15.04.2020

CONNECTIVITY OF CONFIGURATION GRAPHS IN COMPLEX NETWORK MODELS

Yu. L. Pavlov

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

Abstract: The author considers configuration graphs whose degrees of vertices are independent and identically distributed according to the generalized power-law distribution. Connections between vertices are equiprobably

formed in compliance with their degrees. Such random graphs are often used for modeling complex communication networks like the Internet and social networks. It is assumed that the distribution of vertex degrees is unknown because it depends on a slowly varying function with unknown properties. The conditions are found under which a graph is asymptotically almost surely connected as the number of vertices tends to infinity. Under these conditions, estimates of the convergence rate to zero of the probability that the graph is not connected are obtained. The results in the present paper are proved using the properties of stable distributions and slowly varying functions.

Keywords: random graphs; configuration graphs; random vertex degrees; graph connectivity

DOI: 10.14357/19922264210103

Acknowledgments

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

References

- Hofstad, R. 2017. *Random graphs and complex networks*. Cambridge: Cambridge University Press. Vol. 1. 337 p.
- Faloutsos, C., P. Faloutsos, and M. Faloutsos. 1999. On power-law relationships of the Internet topology. *Comput. Commun. Rev.* 29:251–262.
- Bollobas, B. A. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* 1:311–316.
- Reittu, H., and I. Norros. 2004. On the power-law random graph model of massive data networks. *Perform. Evaluation* 55:3–23.
- Durrett, R. 2007. *Random graph dynamics*. Cambridge: Cambridge University Press. 212 p.
- Pavlov, Yu. L., and I. A. Cheplyukova. 2008. Random graphs of Internet type and the generalised allocation scheme. *Discrete Mathematics Applications* 18(5):447–463.
- Pavlov, Yu. L., and E. V. Khvorostyanskaya. 2016. On the limit distributions of the degrees of vertices in configuration graphs with a bounded number of edges. *Sb. Math.* 207(3):400–417.
- Pavlov, Yu. L. 2018. Conditional configuration graphs with discrete power-law distribution of vertex degrees. *Sb. Math.* 209(2):258–275.
- Hofstad, R. 2020. Random graphs and complex Networks. Vol. 2. *Notes RGCNII* Available at: <https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf> (accessed January 11, 2021).
- Pavlov, Yu. L. 2019. O svyaznosti konfiguratsionnykh grafov [On connectivity of configuration graphs]. *Discrete Mathematics Applications* 31(2):115–123.
- Ibragimov, I. A., and Yu. V. Linnik. 1965. *Nezavisimye i stacionarno svyazannye velichiny* [Independent and stationary sequences of random variables]. Moscow: Nauka. 524 p.
- Bigham, N. H., C. M. Goldie, and J. L. Teugels. 1987. *Regular variations. Encyclopedia of mathematics and its applications*. Cambridge: Cambridge University Press. Vol. 27. 513 p.

Received April 15, 2020

Contributor

Pavlov Yuri L. (b. 1949) — Doctor of Science in physics and mathematics, principal scientist, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation; pavlov@krc.karelia.ru

МЕТОДЫ ТЕОРИИ КАТЕГОРИЙ В ЦИФРОВОМ ПРОЕКТИРОВАНИИ ГЕТЕРОГЕННЫХ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

С. П. Ковалёв¹

Аннотация: Развивается предложенный ранее математический аппарат на базе теории категорий, предназначенный для формального описания и строгого исследования процедур инженерной деятельности на базе математического и компьютерного моделирования. При помощи аппарата описаны и исследованы высокоавтоматизированные процедуры проектирования гетерогенных киберфизических систем на основе цифровых двойников, востребованные грядущей четвертой промышленной революцией. Для этого впервые введена конструкция категории мультизапятой, объектами которой служат архитектурные модели некоторой гетерогенной киберфизической системы с заданной схемой структурной иерархии, представленные с некоторой фиксированной точки зрения, а морфизмы отвечают действиям по подбору составных частей (СЧ) для сборки системы из них. Рассмотрено применение категории мультизапятой в решении прямых и обратных задач проектирования отдельных систем и состоящих из них так называемых систем систем (СС).

Ключевые слова: киберфизическая система; цифровой двойник; порождающее проектирование; система систем; теория категорий; категория мультизапятой

DOI: 10.14357/19922264210104

1 Введение

Концепция киберфизической системы (cyber-physical system) состоит в развитии традиционных автоматизированных систем в направлении максимально тесной интеграции мира физических объектов с виртуальным миром управления в целях повышения качества и оперативности управления объектами. Такая система включает разнородные физические компоненты, оснащенные большим числом цифровых датчиков и исполнительных механизмов, позволяющих скомпоновать и поддерживать в системе цифровой двойник (digital twin) — виртуальную модель, воспроизводящую и задающую состояние и поведение оригинала в реальном времени [1]. Структура и начальное информационное наполнение двойника формируются в цикле высокоавтоматизированного цифрового проектирования системы.

Примерами киберфизических систем служат «умные» здания, города, энергетические системы, сетевые воинские формирования и т. д.

В рамках парадигмы грядущей четвертой промышленной революции (Industrie 4.0) степень автоматизации проектирования и производства повышается вплоть до полной замены человека искусственным интеллектом в цикле порождающе-

го проектирования (generative design) [2]. Однако традиционные информационные технологии, в том числе на основе машинного обучения, не способны обеспечить искомый уровень автоматизации по крайней мере в двух важных аспектах:

- (1) интеграция разноплановых видов и языков моделирования;
- (2) минимизация потребности в вычислительных ресурсах для решения оптимизационных задач проектирования.

Актуальна разработка новых подходов и методов цифрового проектирования, в том числе основанных на перспективном математическом аппарате.

В настоящей работе, следуя современным тенденциям [3–5], в качестве основы используется теория категорий. Модели компонентов и систем рассматриваются как объекты подходящих категорий, морфизмы в которых описывают действия, связанные со сборкой сложных изделий. Строятся и исследуются теоретико-категорные конструкции, описывающие методы цифрового проектирования киберфизических систем на абстрактном концептуальном уровне и открывающие новые пути к автоматизации. В частности, впервые введена конструкция категории мультизапятой.

¹Институт проблем управления им. В. А. Трапезникова Российской академии наук, kovalyov@sibnet.ru

2 Принципы цифрового проектирования гетерогенных киберфизических систем

Задача проектирования состоит в формировании архитектуры системы и ее описании в различных моделях. Согласно основополагающему стандарту ISO/IEC/IEEE 42010:2011 [6], описание архитектуры включает представления, отражающие точки зрения различных заинтересованных сторон: пользователей, производителей, проектировщиков, эксплуатирующего персонала и т. д. В результате «обезличивания» (устранения субъективных факторов заинтересованных сторон) и «гранулирования» (расщепления в целях уменьшения пересечений) выделяются типовые точки зрения, такие как пространственное расположение (геометрическая форма), поведение, технология производства, надежность и т. п. Соответствующие модели в совокупности образуют электронно-цифровой макет системы [7] — основу цифрового двойника.

По структуре архитектурная модель системы представляет собой ориентированный граф, узлы которого помечены моделями составляющих системных единиц, а ребра — описаниями действий по иерархической сборке СЧ от деталей до системы в целом. Модели задаются в различных формах, таких как геометрические фигуры и тела, системы дифференциальных уравнений, множества с операциями и отношениями, базы данных и т. д. Все модели некоторого заданного вида и описания всех действий предоставляются проектировщику в виде

(виртуального) каталога, из которого он выбирает строительные блоки для архитектуры.

Построение архитектуры усложняется, когда система включает множество существенно разнородных СЧ. В этом случае для каждой СЧ имеется свой самостоятельный каталог моделей и априори не ясно, в каких терминах описывать действия по сборке систем из таких СЧ. Можно зафиксировать лишь общую схему структурной иерархии СЧ, обусловленную их природой. Такая «схема деления» [8] формируется в самом начале жизненного цикла изделия, когда определяется его концептуальный облик, и считается неизменной в процессе проектирования. Киберфизические системы могут включать в себя такие СЧ (подсистемы), как механическая (несущая), гидравлическая, электрическая, тепловая, защитная, контрольно-измерительная, информационно-коммуникационная (которая в свою очередь разделяется на программное и аппаратное обеспечение). Ярким примером служит «умное» здание, крупноблочная структура которого упрощенно показана на рис. 1 [9].

Чтобы превратить схему деления в полноценную архитектурную модель, необходимо отобразить ее в рамках некоторой точки зрения на изделие. Такое отображение возможно, поскольку каждая точка зрения определяет как некоторый аспект каждой СЧ, так и действия, связанные с составлением в этом аспекте сложных единиц из более простых. Например, с точки зрения пространственного расположения сборка сводится к взаимному расположению и полной либо частичной относительной фиксации геометрических форм СЧ. С точки

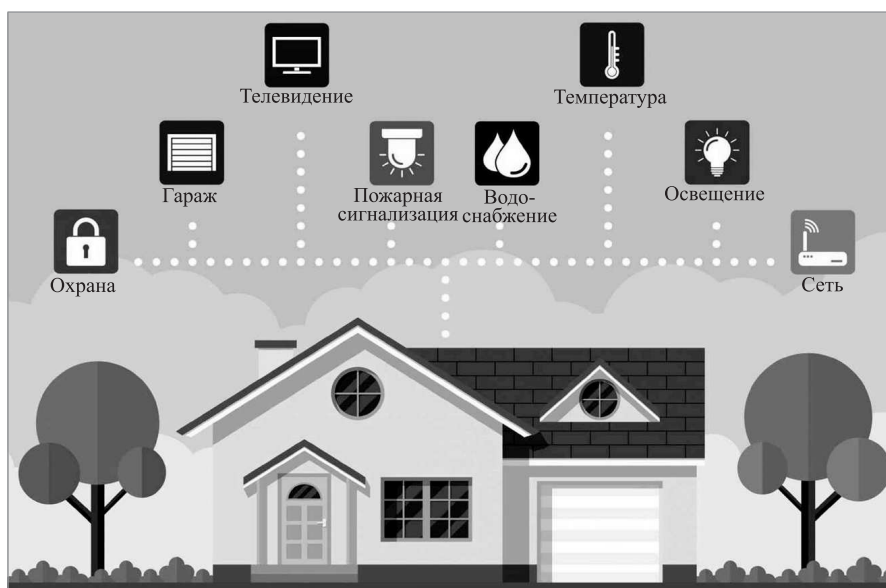


Рис. 1 Структурная схема «умного» здания [9]

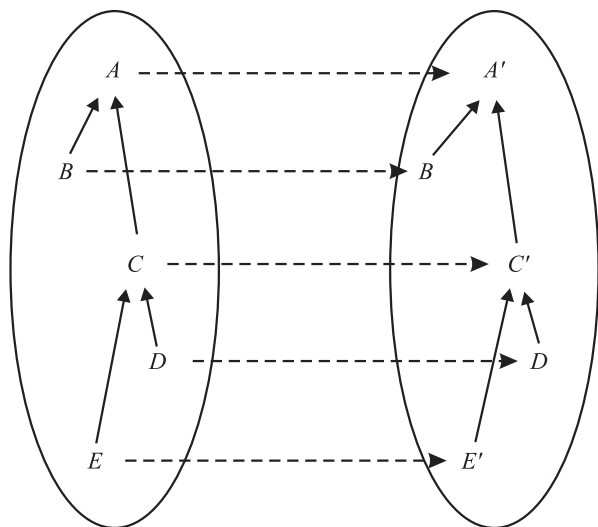


Рис. 2 Структурно целостный переход между вариантами ЭСИ

зрения поведения СЧ описываются сценариями — фрагментами предполагаемой истории их поведения, представленного потоком дискретных событий, так что описания действий по сборке сценариев поведения сложных систем отображают вклад сценариев поведения СЧ.

Представление архитектуры системы с некоторой точки зрения, заданное в цифровом виде, называется электронной структурой изделия (ЭСИ) [10]. На разных стадиях жизненного цикла системы формируются и прорабатываются разные виды ЭСИ, соответствующие точкам зрения различных заинтересованных сторон: конструктивная, производственно-технологическая, эксплуатационная и т. д. Компьютерные инструменты проектирования помогают проектировщику подбирать СЧ и действия в ЭСИ так, чтобы результирующая система удовлетворяла требованиям заинтересованных сторон. Классические системы автоматического проектирования способны автоматически решать прямые задачи проектирования — строить виртуальную модель изделия (вершину иерархии ЭСИ), исходя из вручную выбранных проектировщиком СЧ и действий по сборке, рассчитывать ее характеристики и сопоставлять с требованиями. Однако для технологий типа порождающего проектирования этого недостаточно: компьютер должен самостоятельно автоматически подбирать из каталога СЧ и способы их сборки так, чтобы в наибольшей степени удовлетворить требованиям.

Ключевую роль в порождающем проектировании играет удобное для компьютерной навигации пространство проектирования (design space), состоящее из всех допустимых вариантов ЭСИ. В рам-

ках некоторой фиксированной точки зрения удобство перехода между вариантами обеспечивается, в частности, когда каждая СЧ исходного варианта переходит в заменяющую ее СЧ целевого варианта посредством некоторого действия. Более того, «хороший» переход является структурно целостным в том смысле, что образующие его действия однозначно комбинируются с действиями, составляющими ЭСИ с выбранной точки зрения [11], как показано на рис. 2. Пространства проектирования гетерогенных систем с такими переходами будут далее описаны на строгом алгебраическом языке, что открывает перспективу для применения эффективных алгоритмов поисковой оптимизации с привлечением средств компьютерной алгебры.

Сложность проектирования киберфизических систем усугубляется тем обстоятельством, что такие системы редко проектируются по отдельности. Напротив, на практике обычно возникает потребность в целой системе таких систем (System of Systems, SoS), в качестве СЧ которой выступают полноценные и вполне самостоятельные киберфизические системы. В рамках СС они обмениваются информацией и вступают в коллаборацию в целях оптимизации управления многошаговыми многосубъектными процессами с помощью средств типа платформы «умных» сервисов [1]. Часто встречается такой класс СС, как группа однотипных киберфизических объектов, организованная для достижения общей цели, например ударная группировка беспилотных летательных аппаратов. Синтез таких групп относится к прямым задачам проектирования СС. А обратные задачи требуют проведения иерархически организованных процедур поисковой оптимизации в пространствах проектирования СЧ (имеющих, вообще говоря, разные схемы деления). Примером такой обратной задачи служит порождающее проектирование «умной» городской агломерации, состоящей из зданий и инфраструктурных систем различного назначения. В качестве подхода к решению таких задач далее рассмотрим приведение к плоскому поиску путем «отрисовки» (подстановки в развернутом виде) ЭСИ СЧ в ЭСИ СС.

3 Категория мультизапятой

Будем пользоваться теоретико-категорными конструкциями и обозначениями, введенными в работах [5, 12]. Каталоги моделей системных единиц описываются подходящими категориями. В свою очередь, все возможные модели любых системных единиц в представлении, выражающем

некоторую фиксированную точку зрения на систему, также описываются категорией, которую будем обозначать через C . Например, при представлении систем с точки зрения пространственного расположения в качестве C выступает категория твердотельных геометрических моделей **MBS**, содержащаяся в категории множеств **Set**, а при представлении поведения — категория дискретно-событийных имитационных моделей **Pomset**, которая является конкретной категорией над **Set**.

Представление архитектуры системы с точки зрения C (C -ЭСИ) сводится к C -диаграмме, в вершинах которой находятся представления СЧ, а стрелки представляют действия по сборке. Пусть I — форма (схема) этой диаграммы. Напомним, что I — это малая категория (и через **Cat** обозначается полная подкатегория в «категории всех категорий» **CAT**, состоящая из всех малых категорий). Каталоги моделей СЧ задаются семейством категорий $D_i, i \in |I|$ (напомним, что через $|I|$ обозначается множество вершин схемы I и его можно рассматривать как дискретную подкатегорию в I). Для каждой СЧ имеется правило представления с точки зрения C , которое по соображениям корректности задается функтором вида $F_i : D_i \rightarrow C$ для i -й СЧ (будем обозначать через F любое такое семейство функторов, индексированное множеством вершин схемы). Таким образом, архитектурная модель некоторой конкретной системы получается, если выбрать по одному объекту $A_i \in D_i, i \in |I|$, и некоторую диаграмму $\Delta : I \rightarrow C$, удовлетворяющую условию $\Delta(i) = F_i(A_i), i \in |I|$. Заметим, что точно такой же формальный вид имеют модели, выражающие комплексное представление архитектуры с нескольких точек зрения одновременно: если отдельные точки зрения образуют множество Q и им отвечают категории $C_q, q \in Q$, то комплексное представление описывается произведением категорий $\prod_{q \in Q} C_q$, которое и выступает в качестве C .

Процедуры подбора и замены СЧ в ходе проектирования формально описываются преобразованиями архитектурных моделей, не изменяющими ни схему деления, ни правила представления СЧ. Такими преобразованиями очевидным образом служат естественные преобразования диаграмм, индуцированные действиями из каталогов СЧ, а именно: преобразованием модели $((A_i, i \in |I|), \Delta)$ в модель $((A'_i, i \in |I|), \Delta')$ является любое семейство морфизмов $f_i : A_i \rightarrow A'_i, i \in |I|$ (где каждый морфизм f_i принадлежит категории D_i) такое, что для любых вершин схемы $i, k \in |I|$ и стрелки $h : i \rightarrow k$ выполняется условие

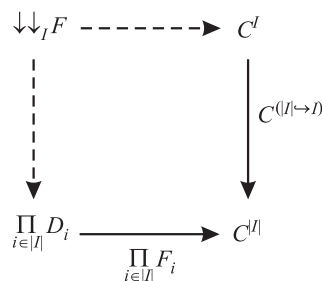
$$F_k(f_k) \circ \Delta(h) = \Delta'(h) \circ F_i(f_i).$$

Именно это условие выражает на языке теории категорий структурную целостность перехода между моделями, соответствующего замене СЧ [11].

Легко проверить, что для любых фиксированных C, I и F совокупность всех архитектурных моделей и всех их преобразований образует категорию. В теории категорий давно известен один частный случай этой конструкции, где в качестве I выбрана схема вида $0 \rightarrow 1$. Категория архитектурных моделей для этого случая встречается во многих задачах, называется категорией запятой (comma category) [12, § II.6] и обозначается через $F_0 \downarrow F_1$. Поэтому будем называть произвольную категорию архитектурных моделей вышеописанного вида *категорией мультизапятой* (multicomma) и обозначать через $\Downarrow_I F$. Пара $\langle I, F \rangle$ называется формой (shape) категории мультизапятой, а категория C — представлением (representation) категории мультизапятой.

Примечательно, что категорию мультизапятой можно получить при помощи универсальных конструкций в «категории всех категорий» **CAT**, а именно: произведения, декартова квадрата и экспоненты.

Теорема 1. *Категория мультизапятой $\Downarrow_I F$ изоморфна вершине (объекту, находящемуся в левом верхнем углу) следующего декартова квадрата в **CAT**:*



Доказательство. Проверяется непосредственно по правилам вычисления пределов и экспонент в **CAT**. \square

Штриховые стрелки декартова квадрата из теоремы 1 задают два канонических «забывающих» функтора, определенных на любой категории мультизапятой. Первый функтор, заданный левой вертикальной стрелкой, извлекает из архитектурной модели набор всех СЧ. Он унивалентен (faithful), так что можно трактовать его как функтор выделения «носителя» модели, по аналогии с функторами, выделяющими носитель у алгебраических систем, топологических пространств и т. п. Второй функтор, заданный верхней горизонтальной стрелкой, извлекает из модели представление структуры системы с точки зрения C и не имеет прямого аналога

в универсальной алгебре. Будем называть его *функтором структуры* и обозначать через ∇_I^F , так что

$$\nabla_I^F : \Downarrow_I F \rightarrow C^I : ((A_i, i \in |I|), \Delta : I \rightarrow C) \mapsto \Delta.$$

В свою очередь, экспонента C^I допускает каноническое вложение в категорию диаграмм \mathbf{DC} , представляющую каталог всех формально возможных структур систем [13], с точки зрения C . При помощи этого вложения и функтора структуры можно представить многие процедуры проектирования в виде функторов, определенных на категории мультизапятой. Рассмотрим в качестве примера построение копредела диаграммы, формально описывающее решение классической прямой задачи цифрового проектирования — формирование модели цельной системы известной структуры. Предположим, что любая C -диаграмма со схемой I имеет копредел. Тогда имеется вложение $lc : C^I \hookrightarrow LC$, где через LC обозначена полная подкатегория в \mathbf{DC} , состоящая из всех диаграмм, имеющих копредел. Вычисление копредела задается функтором $\text{colim} : LC \rightarrow C$. Таким образом, получается функтор представления процедуры сборки систем формы I с точки зрения C :

$$\text{colim} \circ lc \circ \nabla_I^F : \Downarrow_I F \rightarrow C.$$

А для решения обратных задач проектирования киберфизических систем категория $\Downarrow_I F$ служит естественным «строительным материалом» для пространства проектирования, поскольку целевые функции, определяющие степень соответствия вариантов архитектуры системы требованиям, можно задавать функторами на таком пространстве. Действительно, областью значения целевой функции всегда является линейно упорядоченное множество, а его, как хорошо известно, можно представить категорией [12, §1.2]: объектами такой категории служат все элементы множества, а морфизмами — все пары (x, y) такие, что $x \leq y$ (так что между любыми двумя объектами имеется не более одного морфизма). Интересна ситуация, когда целевая функция выступает функцией объектов функтора, действующего в такую категорию из нетривиальной (достаточно богатой морфизмами) подкатегории в $\Downarrow_I F$ или в двойственной категории $(\Downarrow_I F)^{\text{op}}$. В этом случае можно применить оптимизационные алгоритмы типа градиентного спуска, выполняющие навигацию вдоль морфизмов этой подкатегории, с расчетом пути методами компьютерной алгебры.

Обратимся к задачам проектирования СС. Прямые задачи, предполагающие «подъем» конструкций в моделях на уровень структур, могут быть решены в категории мультизапятой «почленно».

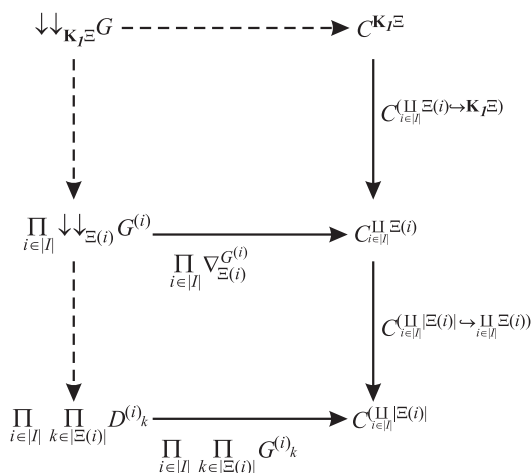
Например, рассмотрим I как схему, вершины которой представляют членов некоторой группы однотипных киберфизических объектов, архитектура которых характеризуется формой $\langle K, G_k : D_k \rightarrow C, k \in |K| \rangle$, а стрелки представляют внутригрупповые коллаборационные связи. Предположим, что во всех категориях $D_k, k \in |K|$, все диаграммы формы I имеют копределы и все функторы G_k сохраняют их. Тогда произвольная $(\Downarrow_K G)$ -диаграмма формы I , представляющая возможную структуру группы, также имеет копредел: компоненты его носителя строятся по отдельности на объектах и морфизмах каждой СЧ D_k , затем переносятся в C посредством функторов G_k , после чего дополняются универсальными стрелками копределов до C -диаграммы формы K , образуя $(\Downarrow_K G)$ -объект, представляющий проектируемую группу как единый цельный объект.

Конструкция мультизапятой хорошо подходит и для решения обратных задач проектирования СС, поскольку ведет себя естественно относительно процедуры отрисовки диаграмм, состоящих из диаграмм. Действительно, формально структура СС описывается диаграммой, в вершинах которой находятся системы, структуры которых в свою очередь описываются подходящими диаграммами (вообще говоря, имеющими различные схемы). Тем самым схема СС задается диаграммой вида $\Xi : I \rightarrow \mathbf{Cat}$. Отрисовка такой диаграммы — это преобразование в малую категорию, порожденную заменой каждой вершины $i \in |I|$ схемой $\Xi(i)$, а каждой стрелки $h : i \rightarrow l$ — совокупностью стрелок, по одной для каждой вершины s схемы $\Xi(i)$, направленной из s в вершину $\Xi(h)(s)$ схемы $\Xi(l)$, с наложением подходящих условий естественности. Вместе с \mathbf{Cat} -диаграммами можно отрисовывать и их морфизмы, так что имеется функтор отрисовки $\mathbf{K}_I : \mathbf{D}(\mathbf{Cat}) \rightarrow \mathbf{Cat}$ (это частный случай общей конструкции отрисовки — умножения в монаде диаграмм \mathbf{D} [13]).

Теорема 2. Пусть заданы произвольные схема I , диаграмма $\Xi : I \rightarrow \mathbf{Cat}$ и семейство функторов $G_k^{(i)} : D_k^{(i)} \rightarrow C, k \in |\Xi(i)|, i \in |I|$. Категория мультизапятой $\Downarrow_{\mathbf{K}_I \Xi} G$ изоморфна вершине следующего декартова квадрата в \mathbf{CAT} :

$$\begin{array}{ccc} \Downarrow_{\mathbf{K}_I \Xi} G & \dashrightarrow & C^{\mathbf{K}_I \Xi} \\ \downarrow & & \downarrow C^{(\prod_{i \in |I|} \Xi(i) \rightarrow \mathbf{K}_I \Xi)} \\ \prod_{i \in |I|} \Downarrow_{\Xi(i)} G^{(i)} & \xrightarrow{\prod_{i \in |I|} \nabla_{\Xi(i)}^{G^{(i)}}} & C^{\prod_{i \in |I|} \Xi(i)} \end{array}$$

Доказательство. Сформируем «почленное» произведение декартовых квадратов, задающих согласно теореме 1 все категории мультизапятой $\coprod_{i \in I} G^{(i)}$, $i \in I$. Если заменить в нем правое вертикальное ребро экспонентой суммы, то можно пристроить его снизу к декартову квадрату, указанному в условии теоремы:



Здесь наружный прямоугольник является декартовым [14, предложение 11.10]. А поскольку $|\mathbf{K}_I \Xi| \cong \prod_{i \in I} |\Xi(i)|$ для любой **Cat**-диаграммы Ξ со схемой I (отрисовка не добавляет и не удаляет объекты), по теореме 1 объект в левом верхнем углу действительно изоморфен категории мультизапятой $\coprod_{\mathbf{K}_I \Xi} G$. □

4 Заключение

Теория категорий обладает большим потенциалом применения в технологиях четвертой промышленной революции, в том числе в цифровом проектировании гетерогенных киберфизических систем. В настоящее время возможности применения предложенных теоретико-категорных методов исследуются на макете программного инструмента разработки цифровых двойников энергетических систем [15]. В ходе развития инструмента до промышленного уровня готовности возникнет много новых задач для дальнейших исследований.

Литература

1. Tao F., Qi Q., Wang L., Nee A. Y. C. Digital twins and cyber-physical systems toward smart manufacturing and

Industry 4.0: Correlation and comparison // Engineering, 2019. Vol. 5. P. 653–661.

2. Kowalski J. CAD is a lie: Generative design to the rescue. — San Rafael, CA, USA: Autodesk, 2016. <https://www.autodesk.com/redshift/generative-design>.

3. Baez J. C., Erbele J. Categories in control // Theor. Appl. Categ., 2015. Vol. 30. Iss. 24. P. 836–881.

4. Wisnesky R., Breiner S., Jones A., Spivak D. I., Subrahmanian E. Using category theory to facilitate multiple manufacturing service database integration // J. Comput. Inf. Sci. Eng., 2017. Vol. 17. Iss. 2. Art. ID: 021011.

5. Ковалёв С. П. Методы теории категорий в модельно-ориентированной системной инженерии // Информатика и её применения, 2017. Т. 11. Вып. 3. С. 42–50.

6. ГОСТ Р 57100-2016/ISO/IEC/IEEE 42010:2011. Системная и программная инженерия. Описание архитектуры. — М.: Стандартинформ, 2016. 32 с.

7. Gherghina G., Tutunea D., Popa D. About digital mock-up for mechanical products // J. Industrial Design Engineering Graphics, 2015. Vol. 10. No. 2. P. 19–22.

8. Рафальский В. В., Рафальская Л. Г., Старостина А. В. Информационная модель схемы деления // Автоматизация процессов управления, 2009. № 3. С. 22–28.

9. What is a smart building and how can it benefit you? — Milford, MA, USA: Comark, 2016. <https://comarkcorp.com/smart-building-can-benefit>.

10. ГОСТ 2.053-2013. Единая система конструкторской документации. Электронная структура изделия. Общие положения. — М.: Стандартинформ, 2014. 10 с.

11. Ковалёв С. П. Алгебраическое моделирование жизненного цикла крупномасштабных гетерогенных систем в аспектах // Управление развитием крупномасштабных систем: Мат-лы X Междунар. конф. — М.: ИПУ РАН, 2017. Т. II. С. 266–268.

12. Маклейн С. Категории для работающего математика / Пер. с англ. — М.: Физматлит, 2004. 352 с. (Mac Lane S. Categories for the working mathematician. — New York, NY, USA: Springer, 1978. 317 p.)

13. Ковалёв С. П. Теория категорий как математическая прагматика модельно-ориентированной системной инженерии // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 95–104.

14. Adámek J., Herrlich H., Strecker G. E. Abstract and concrete categories. — New York, NY, USA: John Wiley, 1990. 507 p.

15. Ковалёв С. П. Проектирование информационного обеспечения цифровых двойников энергетических систем // Системы и средства информатики, 2020. Т. 30. № 1. С. 66–81.

Поступила в редакцию 12.10.2019

METHODS OF THE CATEGORY THEORY IN DIGITAL DESIGN OF HETEROGENEOUS CYBER-PHYSICAL SYSTEMS

S. P. Kovalyov

V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation

Abstract: A mathematical device built upon the category theory is developed which was previously proposed to formally describe and rigorously explore engineering procedures based on mathematical and computer modeling. With the help of the device, highly automated procedures for designing heterogeneous cyber-physical systems on top of digital twins, demanded by the upcoming fourth industrial revolution, are described and explored. For this purpose, the novel construction of the multicomma category is introduced, whose objects are the architectural models of a heterogeneous cyber-physical system with a certain fixed structural hierarchy scheme represented from a certain architecture viewpoint, and morphisms describe actions associated with selection of constituents for assembling a system from them. The application of the multicomma category in solving direct and inverse problems of designing individual systems and complex systems of systems is considered.

Keywords: cyber-physical system; digital twin; generative design; system of systems; category theory; multicomma category

DOI: 10.14357/19922264210104

References

1. Tao, F., Q. Qi, L. Wang, and A. Y. C. Nee. 2019. Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison. *Engineering* 5:653–661.
2. CAD is a lie: Generative design to the rescue. Available at: <https://www.autodesk.com/redshift/generative-design/> (accessed December 9, 2020).
3. Baez, J. C., and J. Erbele. 2015. Categories in control. *Theor. Appl. Categ.* 30(24):836–881.
4. Wisnesky, R., S. Breiner, A. Jones, D. I. Spivak, and E. Subrahmanian. 2017. Using category theory to facilitate multiple manufacturing service database integration. *J. Comput. Inf. Sci. Eng.* 17(2):021011.
5. Kovalyov, S. P. 2017. Metody teorii kategoriy v model'no-orientirovannoy sistemnoy inzhenerii [Methods of category theory in model-based systems engineering]. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):42–50.
6. GOST R ISO/IEC/IEEE 42010:2011. 2016. Sistemnaya i programmaya inzheneriya. Opisanie arkhitektury [System and software engineering. Description of architecture]. Moscow: Standardinform Pubs. 32 p.
7. Gherghina, G., D. Tutunea, and D. Popa. 2015. About digital mock-up for mechanical products. *J. Industrial Design Engineering Graphics* 10(2):19–22.
8. Rafal'skiy, V. V., L. G. Rafal'skaya, and A. V. Starostina. 2009. Informatsionnaya model' skhemy deleniya [Information model of decomposition scheme]. *Avtomatizatsiya protsessov upravleniya* [Automation of Control Processes] 3:22–28.
9. What is a smart building and how can it benefit you? 2016. Milford, MA: Comark. Available at: <https://comarkcorp.com/smart-building-can-benefit/> (accessed December 9, 2020).
10. GOST 2.053-2013. 2014. Edinaya sistema konstruktorskoy dokumentatsii. Elektronnyaya struktura izdeliya. Obshchie polozheniya [Unified system for design documentation. Electronic structure of the product. General Provisions]. Moscow: Standardinform Pubs. 10 p.
11. Kovalyov, S. P. 2017. Algebraicheskoe modelirovanie zhiznennogo tsikla krupnomasshtabnykh geterogennykh sistem v aspektakh [Aspectwise algebraic modeling of large-scale heterogeneous systems life cycle]. *10th Conference (International) "Management of Large-Scale Systems Development" Proceedings*. Moscow. 2:266–268.
12. Mac Lane, S. 1978. *Categories for the working mathematician*. New York, NY: Springer. 317 p.
13. Kovalyov, S. P. 2018. Teoriya kategoriy kak matematicheskaya pragmatika model'no-orientirovannoy sistemnoy inzhenerii [Category theory as a mathematical pragmatics of model-based systems engineering]. *Informatika i ee Primeneniya — Inform. Appl.* 12(1):95–104.
14. Adámek, J., H. Herrlich, and G. E. Strecker. 1990. *Abstract and concrete categories*. New York, NY: John Wiley. 507 p.
15. Kovalyov, S. P. 2020. Proektirovanie informatsionnogo obespecheniya tsifrovyykh dvoynikov energeticheskikh sistem [Information architecture of the power system digital twin]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1):66–81.

Received October 22, 2019

Contributor

Kovalyov Sergey P. (b. 1972) — Doctor of Science in physics and mathematics, leading scientist, V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; kovalyov@sibnet.ru

МЕТОДЫ ОБНАРУЖЕНИЯ ПЕРЕВОДНЫХ ЗАИМСТВОВАНИЙ В БОЛЬШИХ ТЕКСТОВЫХ КОЛЛЕКЦИЯХ*

Р. В. Кузнецова¹, О. Ю. Бахтеев², Ю. В. Чехович³

Аннотация: Рассматривается задача обнаружения переводных заимствований. Для решения предлагается использовать моноязыковой подход — свести задачу обнаружения заимствований к одному языку, используя машинный перевод. В связи со спецификой рассматриваемой задачи предлагаемый алгоритм обнаружения должен быть устойчив к неоднозначностям перевода. Предлагается декомпозировать задачу на несколько этапов. Сначала отбираются документы-кандидаты, устойчивость к неоднозначности перевода достигается за счет замены слов на метки кластеров, полученных с помощью дистрибутивной модели. Затем происходит сравнение найденных кандидатов и рассматриваемого документа, для этого используется отображение текстовых фрагментов документов в векторное пространство высокой размерности. Вычислительный эксперимент проводится для языковой пары «русский–английский» на двух выборках — синтетическом корпусе и на статьях из журналов, входящих в Российский индекс научного цитирования (РИНЦ).

Ключевые слова: автоматическая обработка текстов; машинный перевод; глубокое обучение; переводные заимствования; обнаружение переводных заимствований; дистрибутивная семантика

DOI: 10.14357/19922264210105

1 Введение

Проблема некорректных текстовых заимствований актуальна для сферы образования и научных исследований [1]. По материалам исследования [2], проведенного в 2013 г., более 1500 диссертаций по историческим наукам, защищенных в России после 2000 г., содержат значительные заимствования из других диссертаций.

Для задачи обнаружения заимствований в рамках одного языка высокую полноту поиска показывают промышленные инструменты [1], работа которых основана на представлении документов в виде набора перекрывающихся друг друга послонных n -грамм (шинглов) [3]. Такой подход позволяет эффективно проводить поиск точных текстовых заимствований, но не позволяет обнаруживать заимствования с большой долей перефразированного текста или со вставками текста, переведенного с другого языка.

Существуют несколько подходов, описывающих проблему поиска переводных заимствований для некоторых пар языков [4, 5], например для пары испанский–английский. Настоящая работа посвящена обнаружению переводных заимствований для пары языков русский–английский. Данная

пара нечасто встречается в литературе и не является родственной. Выбор пары языков русский–английский обусловлен преобладанием англоязычных публикаций в интернете и лучшим знанием этого языка по сравнению с другими. Аналогично работам [6, 7] в данной статье предлагается описание алгоритма полного цикла поиска заимствований — сначала ведется поиск документов-кандидатов по внешней коллекции, затем происходит их детальное сравнение с проверяемым документом. Предлагается алгоритм, основанный на моноязыковом анализе документов, схожем с проведенным в работах [8, 9] — проверяемый документ переводится на английский язык с использованием системы машинного перевода с дальнейшим сравнением текстовых фрагментов внутри документов.

В ряде работ, посвященных поиску переводных заимствований, используются дополнительные ресурсы, такие как тезаурусы и онтологии. В работах [4, 5] авторы предлагают использовать базы знаний для извлечения информации о близости между текстами. В работе [5] предлагается алгоритм, основанный на комбинации нейронных сетей и графов знаний. Основной недостаток этого подхода — ресурсоемкость: использование мультиязычных онтологий и баз знаний требует больших вычисли-

* Работа выполнена при поддержке РФФИ (проект 18-07-01441) и Фонда содействия развитию малых форм предприятий в научно-технической сфере (проект 44116).

¹ Московский физико-технический институт, rita.kuznetsova@phystech.edu

² Компания Антиплагиат; Московский физико-технический институт, bakhteev@ap-team.ru

³ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, chehovich@ap-team.ru

тельных мощностей для построения семантических графов для каждого текстового фрагмента, а также сравнения полученных семантических графов.

В данной работе предлагается декомпозиция алгоритма обнаружения переводных заимствований для поиска по большим текстовым коллекциям. Общая схема алгоритма включает следующие шаги.

1. *Машинный перевод* — перевод проверяемого документа на английский язык. Для этого используется система статистического машинного перевода [10].
2. *Поиск документов-кандидатов* — для проверяемого документа находятся наиболее релевантные документы-кандидаты, для этого используется модификация алгоритма шинглов.
3. *Сравнение документов* — текст разбивается на фрагменты, строится отображение каждой фразы в векторное пространство. Для каждого вектора проверяемого документа находятся ближайшие векторы из документов-кандидатов, после чего проводится классификация пар данных векторов на схожие и несхожие пары текстовых фрагментов.

Так как в предлагаемом алгоритме используется моноязыковой анализ заимствований, то задача близка к задаче обнаружения перефразированного текста. Ряд подходов [11–14] к решению этой задачи используют векторные представления фраз, полученные с помощью нейронных сетей глубокого обучения. В работе [13] предлагается нейронный мешок слов (*англ.* Neural Bag-of-Words) и глубокие усредняющие сети (*англ.* Deep averaging networks). В данной статье предлагается использовать выходы нейронной сети как векторные представления текстовых фрагментов для дальнейшего приближенного алгоритма поиска ближайшего соседа [15].

В работе исследуются свойства предлагаемого метода обнаружения переводных заимствований. Проводится анализ моделей глубокого обучения, используемых на этапе сравнения документов, а также составной оптимизируемой функции. Проверка качества предложенного метода проводится как на синтетической выборке, так и на статьях из журналов, входящих в РИНЦ. Проводится анализ ошибок. Предложенный метод поиска заимствований сравнивается с базовым алгоритмом поиска заимствований, основанным на использовании машинного перевода и алгоритме шинглов.

2 Постановка задачи

Пусть заданы коллекции документов на английском языке

$$D_e = \{d_e^j\}_{j=1}^N$$

и русском языке

$$D_r = \{d_r^i\}_{i=1}^M.$$

Документы на русском и английском языке представимы в виде конкатенации текстовых фрагментов:

$$d_e^j = [s_{e_1}^j \sqcup \dots \sqcup s_{e_h}^j]; \quad d_r^i = [s_{r_1}^i \sqcup \dots \sqcup s_{r_k}^i].$$

Пусть задана выборка

$$\mathcal{D} = \left\{ (d_e^l, d_r^l), \text{RL}^l \right\}_{l=1}^L,$$

где каждой паре документов $(d_e^l, d_r^l)_{d_e^l \in D_e, d_r^l \in D_r}$ сопоставлен список пар фрагментов

$$\text{RL} = \left[(s_{e_1}^l, s_{r_1}^l), \dots, (s_{e_{k(l)}}^l, s_{r_{k(l)}}^l) \right].$$

Для каждой пары $(s_{e_k}^l, s_{r_k}^l)$ известно, что фрагмент $s_{r_k}^l$ является переводом фрагмента $s_{e_k}^l$.

Модель f задается как последовательное выполнение функций *filter* и *comparison*, где

$$\text{filter}: (d_r^i, D_e)_{d_r^i \in D_r} \rightarrow D_e^{\text{retrieved}_i} \subset D_e,$$

$$\text{comparison}: (d_r^i, D_e^{\text{retrieved}_i})_{d_r^i \in D_r} \rightarrow \text{RL}^i.$$

Здесь RL^i — список пар фрагментов. Функция *filter* отвечает за сужение числа документов коллекции, сравниваемых с проверяемым документом, и позволяет проводить дальнейшее более детальное сравнение *comparison* с использованием ресурсоемких вычислительных алгоритмов, основанных на моделях глубокого обучения.

Качество модели f оценивается с помощью функций *Precision* и *Recall*:

$$\text{Precision} = \frac{|\cup_{l=1}^L \text{RL}^l \cap (\cup_{i=1}^M \text{RL}^i)|}{|\cup_{i=1}^M \text{RL}^i|},$$

$$\text{Recall} = \frac{|\cup_{l=1}^L \text{RL}^l \cap (\cup_{i=1}^M \text{RL}^i)|}{|\cup_{l=1}^L \text{RL}^l|}.$$

Требуется найти функцию f , максимизирующую F1, среднее гармоническое показателей *Precision* и *Recall*:

$$\hat{f} = \arg \max_{f \in \mathcal{F}} \text{F1}(f, \mathcal{D}),$$

$$\text{F1} = \frac{2 \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

где \mathcal{F} — заданное семейство моделей.

3 Поиск документов-кандидатов

Одним из алгоритмов поиска документов-кандидатов в задачах обнаружения дословных заимствований и поиска *почти-дубликатов* текста служит алгоритм, основанный на построении инвертированного индекса, в котором каждый документ коллекции представляется набором *шинглов* [3], т. е. набором перекрывающихся n -грамм. Проверяемый документ также разбивается на шинглы, после чего проводится поиск документов по инвертированному индексу с наибольшим совпадением шинглов. В данной работе предлагается обобщение алгоритма шинглов, позволяющее улучшить качество поиска кандидатов в случае обнаружения переводных заимствований.

Предлагается функция *filter* следующего вида:

$$\text{filter}(d_r^i, D_e) = \arg \max_{D'_e \subset D_e, |D'_e|=k} \sum_{d_e^j \in D'_e} \sum_{h \in \mathcal{H}(d_e^j)} \mathbf{I}[h \in \mathcal{H}(d_r^i)] / \left(|d_e^j \in D_e : h \in \mathcal{H}(d_e^j)|^\alpha + \text{const} \right).$$

Здесь \mathcal{H} — множество n -грамм документа, упорядоченная последовательность n меток кластеров, где процедура формирования кластеров описана ниже; $\alpha \in \mathbb{R}$; k — оптимизируемый гиперпараметр.

Для уменьшения влияния неоднозначности перевода на поиск документов-кандидатов предлагается заменять слова на соответствующие им метки кластеров:

$$\{x_1, \dots, x_n\} \rightarrow \{\text{class}(x_1), \dots, \text{class}(x_n)\} = h,$$

где x_1, \dots, x_n — слова. Кластеры предварительно выделены из текстового корпуса и содержат семантически близкие слова. Для уменьшения неоднозначности перевода перед разбиением на n -граммы предлагается удалять из текста стоп-слова и проводить лемматизацию. Для учета возможных перестановок слов, возникающих после перевода текста, слова внутри каждой n -граммы сортируются в лексикографическом порядке.

В данной работе для получения кластеров используется модель векторного представления слов, основанная на дистрибутивной гипотезе. Кластеризация проводится с использованием косинусной функции расстояния

$$\cos(\mathbf{c}_1, \mathbf{c}_2) = \frac{\langle \mathbf{c}_1, \mathbf{c}_2 \rangle}{\|\mathbf{c}_1\|_2 \|\mathbf{c}_2\|_2}, \quad (1)$$

где \mathbf{c}_1 и \mathbf{c}_2 — векторы из одного векторного пространства.

Ниже приведены примеры полученных кластеров:

- [beer, beers, brewing, ale, brew, brewery, pint, stout, guinness, ipa, brewed, lager, ales, brews, pints, cask];
- [brilliant, excellent, exceptional, finest, outstanding, super, terrific].

4 Сравнение документов

Для сравнения найденных документов-кандидатов $D_e^{\text{retrieved}_i}$ и проверяемого документа (d_r^i) используется модель векторного представления фразы — тексты разбиваются на фрагменты и сравниваются соответствующие им векторы. Ниже представлены детали алгоритма сравнения, а также анализ предлагаемой оптимизационной задачи.

4.1 Модель векторного представления фразы

Рассмотрим подробнее этап построения отображения фрагмента в вектор. Пусть каждому слову документа на языке коллекции поставлен в соответствие вектор $\mathbf{v} \in \mathbb{R}^u$ размерности u . Для простоты будем полагать, что все фрагменты на языке коллекции имеют ограниченную длину n_{col} . Тогда моделью векторизации фрагмента будем называть отображение

$$\mathbf{h} : \mathbb{W} \times \mathbb{R}^{u \times n_{\text{col}}} \rightarrow \mathbb{R}^u,$$

где \mathbb{W} — пространство параметров модели. Объекты из множества $\mathbb{R}^{u \times n_{\text{col}}}$ являются последовательной конкатенацией векторов векторных представлений слов для фрагментов выборки:

$$\mathbf{x} \in [\mathbf{v}_1, \dots, \mathbf{v}_{n_{\text{col}}}]^T, \mathbf{x} \in \mathbb{R}^{u \times n_{\text{col}}}.$$

Для работы с фрагментами длиной меньше n_{col} определим некоторый вектор, обозначающий пустое слово.

Модель оптимизируется в режиме частичного обучения с учителем. В качестве оптимизируемой функции используется составная функция ошибки, представляющая собой сумму ошибки реконструкции и ошибки отступа:

$$\alpha E_{\text{rec}}(\mathbf{X}_{\text{rec}}, \mathbf{w}) + (1 - \alpha) E_{\text{me}}(\mathbf{X}_{\text{me}}, \mathbf{w}) \rightarrow \min_{\mathbf{w} \in \mathbb{W}}, \quad (2)$$

где E_{rec} — ошибка реконструкции; E_{me} — ошибка отступа; \mathbf{X}_{rec} и \mathbf{X}_{me} — обучающие выборки; \mathbf{w} — параметры модели; α — настраиваемый гиперпараметр. Рассмотрим подробнее каждое слагаемое функции ошибки.

Первое слагаемое функции ошибки соответствует модели автокодировщика. Пусть задана выборка $\mathbf{X}_{\text{rec}} \subset \mathbb{R}^{u \times n_{\text{col}}}$. Модель \mathbf{h} выступает в качестве функции кодирования информации о выборке \mathbf{X}_{rec} . Пусть также задана вспомогательная функция декодирования \mathbf{g} , восстанавливающая исходное векторное представление \mathbf{x} по выходам модели \mathbf{h} :

$$\mathbf{r}(\mathbf{x}, \mathbf{w}) = \mathbf{g}(\cdot, \mathbf{w}) \circ \mathbf{h}(\mathbf{x}, \mathbf{w}) \approx \mathbf{x}, \mathbf{x} \in \mathbb{R}^{u \times n_{\text{col}}}.$$

Минимизируемая ошибка реконструкции выглядит следующим образом:

$$E_{\text{rec}}(\mathbf{X}_{\text{rec}}, \mathbf{w}) = \frac{1}{|\mathbf{X}_{\text{rec}}|} \sum_{\mathbf{x} \in \mathbf{X}_{\text{rec}}} \|\mathbf{x} - \mathbf{r}(\mathbf{x}, \mathbf{w})\|_2^2. \quad (3)$$

Выбор ошибки реконструкции в качестве оптимизируемой функции можно обосновать, используя результаты статьи [16]. Будем пользоваться результатами, доказанными в работе [16], где было показано, что автокодировщики с регуляризацией специального вида позволяют оценить распределение $p(\mathbf{X})$ объектов, принадлежащих генеральной совокупности.

Теорема 1 [16]. Пусть p — дифференцируемая плотность вероятности и $\forall \mathbf{x}_i \in \mathbb{R}^{u \times n_{\text{col}}} p(\mathbf{x}_i) \neq 0$. Пусть \mathcal{L}_{σ^2} — функция потерь вида

$$\begin{aligned} \mathcal{L}_{\sigma^2} &= \\ &= \int_{\mathbb{R}^{u \times n_{\text{col}}}} p(\mathbf{x}) \left[\|\mathbf{x} - \mathbf{r}(\mathbf{x}, \mathbf{w})\|_2^2 + \sigma^2 \left\| \frac{\partial \mathbf{r}(\mathbf{x}, \mathbf{w})}{\partial \mathbf{x}} \right\|_F^2 \right] d\mathbf{x}, \end{aligned}$$

где \mathbf{r} дважды дифференцируема; $0 \leq \sigma \in \mathbb{R}$. Пусть $\hat{\mathbf{w}}$ — оптимум функции \mathcal{L}_{σ^2} по параметрам моделей кодирования и декодирования, доставляющий минимум \mathcal{L}_{σ^2} . Тогда

$$\hat{\mathbf{r}}_{\sigma^2}(\mathbf{x}, \mathbf{w}) = \mathbf{x} + \sigma^2 \frac{\partial \log p(\mathbf{x})}{\partial \mathbf{x}} + o(\sigma^2), \quad \sigma^2 \rightarrow 0.$$

Используя результаты теоремы 1, можно сделать следующее утверждение.

Теорема 2. Плотность вероятности представима в виде:

$$\frac{\hat{\mathbf{r}}_{\sigma^2}(\mathbf{x}, \mathbf{w}) - \mathbf{x}}{\sigma^2} \approx -\frac{\partial}{\partial \mathbf{x}} E(\mathbf{x}),$$

где $\mathbf{x} = (1/Z) \exp(-E(\mathbf{x}))$, Z — нормировочная константа.

Доказательство.

$$\mathbf{r}_{\sigma^2}(\mathbf{x}, \hat{\mathbf{w}}) = \mathbf{x} + \sigma^2 \frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{x}) + o(\sigma^2);$$

$$\frac{\mathbf{r}_{\sigma^2}(\mathbf{x}, \hat{\mathbf{w}}) - \mathbf{x}}{\sigma^2} = \frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{x}) + o(1);$$

$$\frac{\mathbf{r}_{\sigma^2}(\mathbf{x}, \hat{\mathbf{w}}) - \mathbf{x}}{\sigma^2} \approx \frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{x}).$$

Представляя $\log p(\mathbf{x})$ в форме $-E(\mathbf{x}) - \log Z$, получим искомое выражение.

Таким образом, при устремлении регуляризатора σ к нулю получается языковая модель, т. е. распределение вероятностей на множестве \mathbf{X} — множестве текстовых последовательностей.

Второе слагаемое составной функции ошибки — ошибка отступа [12]. Для оптимизации этой функции ошибки используется выборка $\mathbf{X}_{\text{me}} = \{(\mathbf{x}_i, \mathbf{x}_j)\}$, состоящая из пар объектов:

$$\mathbf{X}_{\text{me}} = [\mathbf{X}_{\text{me}}^A; \mathbf{X}_{\text{me}}^B] \subset \mathbb{R}^{u \times n_{\text{col}}} \times \mathbb{R}^{u \times n_{\text{col}}};$$

$$E_{\text{me}} = \frac{1}{|\mathbf{X}_{\text{me}}|} \left(\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{X}_{\text{me}}} \max(0, \delta - c_-) + \max(0, \delta - c_+) \right), \quad (4)$$

где

$$c_- = \cos(\mathbf{h}(\mathbf{x}_i, \mathbf{w}), \mathbf{h}(\mathbf{x}_j, \mathbf{w})) - \cos(\mathbf{h}(\mathbf{x}_i, \mathbf{w}), \mathbf{h}(\mathbf{x}_{i'}, \mathbf{w}));$$

$$c_+ = \cos(\mathbf{h}(\mathbf{x}_i, \mathbf{w}), \mathbf{h}(\mathbf{x}_j, \mathbf{w})) - \cos(\mathbf{h}(\mathbf{x}_j, \mathbf{w}), \mathbf{h}(\mathbf{x}_{j'}, \mathbf{w}));$$

δ — отступ; \cos — функция расстояния (1),

$$\mathbf{x}_{i'} = \arg \max_{\mathbf{x}_{i'} \in \mathbf{X}^B, \mathbf{x}_{i'} \neq \mathbf{x}_i} \cos(\mathbf{x}_i, \mathbf{x}_{i'});$$

$$\mathbf{x}_{j'} = \arg \max_{\mathbf{x}_{j'} \in \mathbf{X}^A, \mathbf{x}_{j'} \neq \mathbf{x}_j} \cos(\mathbf{x}_j, \mathbf{x}_{j'}).$$

Следующая теорема объясняет поведение данного слагаемого при проводимой оптимизации параметров \mathbf{w} модели \mathbf{h} .

Теорема 3. Пусть выполнены следующие условия.

1. Задан гиперпараметр $\delta \in (0, 2)$.
2. Мощность выборки $|\mathbf{X}_{\text{me}}|$ ограничена следующей величиной:

$$\begin{aligned} |\mathbf{X}_{\text{me}}|(|\mathbf{X}_{\text{me}}| - 1) &\leq \\ &\leq \sqrt{\pi} \frac{\Gamma((u-1)/2)}{\Gamma(u/2)} \left(\int_0^{\arccos(1-\delta)} \sin^{u-2} x dx \right)^{-1}. \quad (5) \end{aligned}$$

3. Подвыборки \mathbf{X}_{me}^A и \mathbf{X}_{me}^B содержат все элементы в единственном числе, ни один элемент не встречается в обеих выборках.

Тогда существует непрерывное отображение $\hat{\mathbf{h}}$ из множества векторных представлений слов $\mathbb{R}^{u \times n_{\text{col}}}$ в векторное пространство \mathbb{R}^u , доставляющее глобальный минимум функции $E_{\text{me}} = 0$.

Доказательство. Построим отображение $\hat{\mathbf{h}}$ явно. Положим для каждой пары $(\mathbf{x}_1, \mathbf{x}_2)$: $\hat{\mathbf{h}}(\mathbf{x}_1) = \hat{\mathbf{h}}(\mathbf{x}_2)$.

Тогда функция E_{me} выглядит следующим образом с точностью до множителя:

$$E_{\text{me}} = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{X}_{\text{me}}} \max \left(0, \delta - 1 + \cos \left(\hat{\mathbf{h}}(\mathbf{x}_i), \hat{\mathbf{h}}(\mathbf{x}_j) \right) \right) + \max \left(0, \delta - 1 + \cos \left(\hat{\mathbf{h}}(\mathbf{x}_j), \hat{\mathbf{h}}(\mathbf{x}_i) \right) \right).$$

Область значений функции ограничена снизу нулем, который достигается при выполнении условий:

$$1 - \delta \geq \cos \left(\hat{\mathbf{h}}(\mathbf{x}), \hat{\mathbf{h}}(\mathbf{x}') \right)$$

для любой пары $\mathbf{x} \in \mathbf{X}_{\text{me}}^A$, $\mathbf{x}' \in \mathbf{X}_{\text{me}}^B$, $(\mathbf{x}, \mathbf{x}') \notin \mathbf{X}_{\text{me}}$, $(\mathbf{x}', \mathbf{x}) \notin \mathbf{X}_{\text{me}}$. Число пар, описанных выше, в множестве \mathbf{X}_{me} при выполнении третьего условия теоремы равно $|\mathbf{X}_{\text{me}}|(|\mathbf{X}_{\text{me}}| - 1)$. Назначим значение отображения $\hat{\mathbf{h}}$ для каждой такой пары так, чтобы $\cos(\hat{\mathbf{h}}(\mathbf{x}), \hat{\mathbf{h}}(\mathbf{x}')) \leq 1 - \delta$.

Существование такого отображения следует из задачи о нахождении сферического кода максимального размера для сферы в пространстве размерности u и углом $\arccos(1 - \delta)$. В работе [17] представлена нижняя оценка для размерности выборки, удовлетворяющей заданным условиям. Оценка соответствует правой части неравенства (5). Так как выборка \mathbf{X}_{me} конечна, то для построения непрерывной функции, заданной условиями, описанными выше, можно использовать интерполяционные полиномы, что и требовалось доказать.

Заметим, что предложенное в теореме отображение является непрерывным, поэтому для приближения данного отображения можно использовать нейросетевые модели. По теореме Цыбенко отображения из класса нейросетевых моделей будут приближать непрерывные модели сколь угодно хорошо [18].

Таким образом, составная оптимизируемая функция (2) позволяет получить модель, которая, с одной стороны, обладает обобщающими свойствами, за которые отвечает языковая модель (3), с другой стороны, эффективно разделяет схожие и несхожие фразы из обучающей выборки (4). Гиперпараметр α отвечает за вклад каждого из оптимизируемых слагаемых в данную функцию.

4.2 Классификатор

Для каждого вектора фразы $\mathbf{h}(\mathbf{x}_{r_a}^i)$ из проверяемого документа d_r^i находится v ближайших векторов по косинусной функции расстояния (1) для фрагментов из документов-кандидатов $D_e^{\text{retrieved}_i}$,

используя метод приближенного поиска ближайшего соседа. Основная цель данной процедуры — сократить число пар фрагментов для классификации для снижения ресурсоемкости этапа сравнения документа.

Для векторного представления пары фрагментов $(\mathbf{h}(\mathbf{x}_{e_b}^j), \mathbf{h}(\mathbf{x}_{r_a}^i))$ рассматривается следующее решающее правило:

$$f_{\text{frag}}((\mathbf{h}(\mathbf{x}_{e_b}^j), \mathbf{h}(\mathbf{x}_{r_a}^i))) = \begin{cases} 1, & \text{если } \cos(\mathbf{h}(\mathbf{x}_{e_b}^j), \mathbf{h}(\mathbf{x}_{r_a}^i)) > t_1 \\ & \text{и } p(\mathbf{h}(\mathbf{x}_{e_b}^j), \mathbf{h}(\mathbf{x}_{r_a}^i)) > t_2; \\ 0 & \text{иначе,} \end{cases} \quad (6)$$

где p — вероятность классификатора; t_1 — порог косинусной функции расстояния (1); t_2 — минимальный порог вероятности классификатора.

В качестве признаков используется конкатенация разницы по модулю и покомпонентное произведение компонент вектора $[\|\mathbf{h}(\mathbf{x}_{e_b}^j) - \mathbf{h}(\mathbf{x}_{r_a}^i)\|, \mathbf{h}(\mathbf{x}_{e_b}^j) \odot \mathbf{h}(\mathbf{x}_{r_a}^i)]$. В качестве классификатора выступает модель случайного леса.

5 Вычислительный эксперимент

Для анализа качества предложенного алгоритма был проведен ряд вычислительных экспериментов как на синтетической выборке [19], так и на реальных коллекциях документов. В данном разделе приводятся детали порождения синтетических выборок и эксперименты, проведенные на них.

5.1 Синтетическая коллекция переводных заимствований

Для порождения переводных заимствований были использованы документы из английской и русской версии сайта Wikipedia. В качестве коллекции документов D_e были использованы 100 тыс. статей из английской версии Wikipedia. В качестве коллекции проверяемых документов D_r использовалась случайная подвыборка документов из русской версии Wikipedia. Для порождения заимствований для каждого документа $d_r^i \in D_r$ применялся следующий алгоритм.

1. Выбрать документы-кандидаты $\{d_e^j\}$ из коллекции D_e . Для уменьшения разброса лексики в документах-кандидатах и проверяемом документе выбор документов-кандидатов проводился из подвыборки 500 наиболее релевантных документов для проверяемого документа d_r^i . Для определения релевантности использовалась $\text{tf} \cdot \text{idf}$ -мера. Число документов-кандидатов выбиралось случайно от 1 до 10.

2. Выбрать предложения из документов-кандидатов $\{d_e^j\}$ случайным образом и перевести их на русский язык.
3. Заменить случайные предложения из проверяемого документа d_r^i на переведенные предложения из документов-кандидатов. Доля замененных предложений из проверяемого документа d_r^i выбиралась случайно от 20% до 80%.

5.2 Оптимизация параметров рассматриваемых моделей

В качестве модели векторного представления слов использовалась библиотека fastText [20], оптимизация параметров которой проводилась на английской версии Wikipedia. Размерность векторного пространства для векторного представления слов и фрагментов была установлена как 100. Для оптимизации модели векторного представления текстовых фрагментов использовался алгоритм AdaDelta с параметрами $\varepsilon = 10^{-6}$, $\mu = 0,95$ и L2-регуляризация $\lambda_2 = 10^{-6}$. Для итоговой функции потерь (2) были установлены следующие значения гиперпараметров: $\delta = 0,3$; $\alpha = 0,1$. Пороги классификатора (6) были подобраны на основе процедуры кросс-валидации: $t_1 = 0,6$; $t_2 = 0,5$. Для построения кластеров была использована агломеративная кластеризация на векторах слов. В качестве меры близости слов рассматривалась косинусная функция расстояния (1) между соответствующими векторными представлениями. Итоговая модель содержала 30 тыс. кластеров для 777 тыс. слов. В качестве моделей кодирования \mathbf{h} и декодирования \mathbf{g} использовалась рекуррентная модель GRU (gated recurrent unit) [21]. В качестве системы машинного перевода использовался Moses [10], модель которого была обучена на 18,5 млн параллельных предложений из корпусов Opus [22]. В качестве выборки для минимизации ошибки реконструкции E_{rec} (3) использовались 10 млн предложений из английской версии Wikipedia. Второе слагаемое функции потерь (4) использует информацию о похожих предложениях $\mathbf{X}_{\text{me}} = \{(\mathbf{x}_i, \mathbf{x}_j)\}$. В качестве выборки таких предложений использовались пары параллельных предложений из корпуса OpenSubtitles [22].

5.3 Детали вычислительного эксперимента

Было проведено три эксперимента на синтетических данных.

1. Поиск кандидатов. В данном эксперименте анализировалось качество полученной модели

кластеров слов. В качестве базового эксперимента для сравнения рассматривался алгоритм, основанный на шинглах без приведения слов к меткам кластеров.

2. Сравнение фрагментов текста. В данном эксперименте рассматривался случай, когда отбор кандидатов был проведен полностью корректно: $\text{Recall@10} = 1,0$. В качестве базового алгоритма также выступал алгоритм, основанный на шинглах: проверяемый документ d_r^i переводился на английский язык. После этого полученный текст проходил лемматизацию и разбивался на множество перекрывающихся 4-грамм. Для учета возможных перестановок слов при переводе слова внутри каждой 4-граммы сортировались. Результатом сравнения двух документов выступало множество совпавших отсортированных 4-грамм.
3. Эксперимент, оценивающий качество всего алгоритма (поиск кандидатов и сравнение фрагментов текста). Данный эксперимент позволял оценить качество представленного алгоритма в целом.

Результаты эксперимента по поиску кандидатов представлены в табл. 1. Представленный алгоритм, основанный на построении кластеров, дает лучшее качество, чем базовый алгоритм, основанный на шинглах.

Результаты экспериментов по сравнению фрагментов текста представлены в табл. 2. Представленный алгоритм показывает точность, сравнимую с точностью базового алгоритма, и полноту, значительно превосходящую полноту базового алгоритма. Точность базового алгоритма объясняется тем, что данный алгоритм учитывает схожесть только почти-дубликатов текста.

В третьем эксперименте, учитывавшем качество представленного алгоритма в целом, были получены следующие показатели: $\text{Precision} = 0,83$; $\text{Recall} = 0,79$; $\text{F1} = 0,80$.

Таблица 1 Результаты эксперимента по поиску кандидатов

Алгоритм	Recall@10
Базовый	0,93
Представленный	0,95

Таблица 2 Результаты экспериментов по поиску схожих фрагментов текста

Алгоритм	Precision	Recall	F1
Базовый	0,99	0,15	0,26
Представленный	0,93	0,80	0,85

6 Результаты экспериментов на реальной коллекции научных документов

Для апробации представленного алгоритма был проведен эксперимент по поиску переводных заимствований на коллекции документов из электронной библиотеки eLibrary.ru. Данная библиотека содержит научные документы, входящие в РИНЦ. Данный ресурс также содержит дополнительные метаданные для каждого документа: заголовок, авторов документа, язык документа и принадлежность к тематике, соответствующей Государственному рубрикатору научно-технической информации (ГРНТИ). Для апробации алгоритма в качестве проверяемых документов D_r были подготовлены 2,5 млн документов на русском языке.

В качестве коллекции документов D_e использовались документы из английской версии Wikipedia, документы на английском языке из eLibrary.ru и статьи ресурса arXiv.org. Суммарное число полученных документов составило 7,6 млн.

В силу большого числа проверяемых документов D_r для дальнейшего анализа рассматривались документы, содержащие значительное число найденных заимствований. Была получена 21 тыс. документов со значительным числом заимствований. Из них были проанализированы 7,6 тыс. документов, выбранных случайно. Основной целью эксперимента было обнаружение переводных заимствований, когда заимствование произошло из англоязычного документа в русскоязычный документ. В то же время при анализе полученных результатов был выявлен ряд других срабатываний представленного алгоритма, которые были в дальнейшем разделены на несколько типов:

- переводные заимствования — документ содержит заимствования, переведенные с английского языка, выданные за оригинальный текст;
- другие заимствования — заимствования из русскоязычных ресурсов или заимствования, направление которых нельзя определить по датам документов;
- двуязычные статьи — работы одного и того же автора на двух языках;
- самоцитирование — цитирование автором его англоязычной работы;
- цитирование законов — использование формулировок нормативных актов;
- ошибочные срабатывания — ложно-положительные срабатывания представленного алгоритма;

Таблица 3 Результаты экспериментов для коллекции документов eLibrary.ru

Тип	Количество
Переводные заимствования	921
Другие заимствования	2548
Двуязычные статьи	788
Самоцитирование	669
Цитирование законов	1567
Ошибочные срабатывания	507
Другое	698
Всего	7689

– другое — срабатывания, которые сложно отнести к какой-либо категории из-за нехватки метаданных или плохого качества текстов.

Результаты экспериментов представлены в табл. 3. Заметим, что были проанализированы только 36% всех срабатываний алгоритма, поэтому можно предварительно оценить число документов с переводными заимствованиями по всей коллекции в 2,5 тыс., что составляет 0,1% всех документов. Заметим, что результаты были получены в автоматическом режиме и требуют дальнейшей экспертной верификации.

Распределение доли заимствований в проанализированных документах представлено на рис. 1. Средняя доля заимствований составляет 20%.

Для анализа научных тематик, в которых переводные заимствования происходят наиболее часто, были проанализированы документы, отнесенные к типу *переводные заимствования*. Около 70% проанализированных документов были классифицированы по 10 научным рубрикам. Наибольшая часть документов оказалась распределена между рубриками «Экономика. Народное хозяйство. Экономические науки» и «Право. Юридические науки». За-

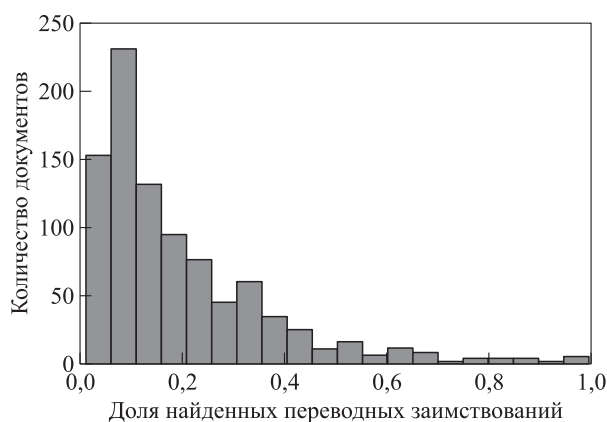


Рис. 1 Гистограмма распределения доли заимствования в тексте

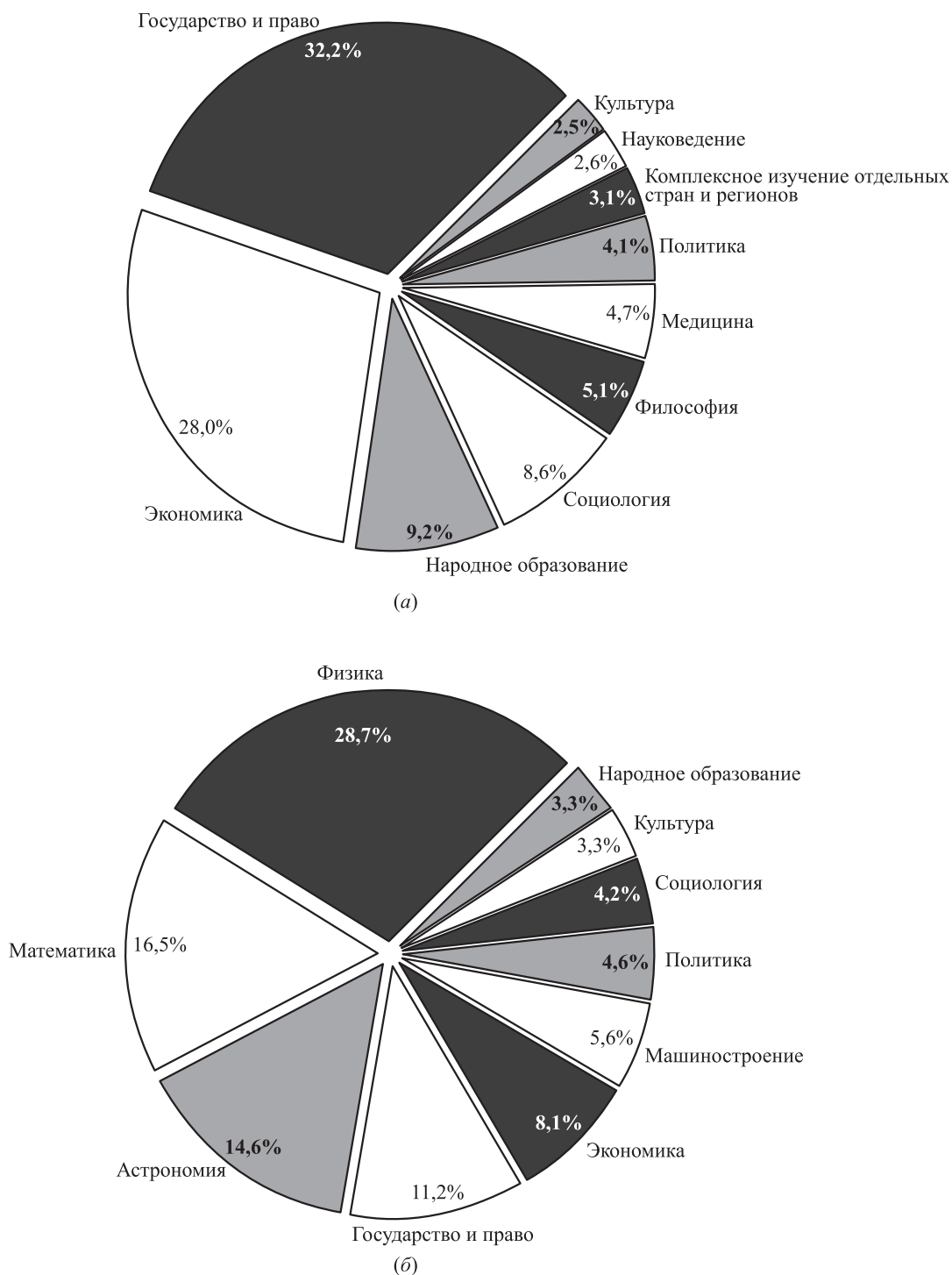


Рис. 2 Распределение заимствований по рубрикам ГРНТИ для типов *переводные заимствования* (а) и *двуязычные статьи* (б)

метим, что распределение по рубрикам заимствований, отнесенных к типу *двуязычные статьи*, значительно отличается от данного распределения. Диаграммы десяти наиболее представительных рубрик для данных типов срабатываний показаны на рис. 2.

Анализ ложно-отрицательных срабатываний. Для анализа ложно-отрицательных срабатываний представленного алгоритма была проанализирована полнота нахождения двуязычных документов. Оценка полноты была проведена с помощью мета-

данных, полученных из eLibrary.ru. Анализ срабатываний алгоритма показал, что только 85% документов были найдены алгоритмом корректно. Заметим, что представленная оценка полноты является грубой, так как учитывает только полные переводы текстов.

Основная причина ложно-отрицательных срабатываний — низкое качество машинного перевода. Другой проблемой, значительно повлиявшей на качество нахождения двуязычных статей, является используемый алгоритм поиска кандидатов, позволяющий находить только близкие по структуре заимствования. Кроме того, значительная часть проанализированных документов имела некорректную кодировку, что также повлияло на полноту поиска документов.

Анализ ложно-положительных срабатываний. Для анализа ложно-положительных срабатываний были проанализированы вручную 90 документов, отнесенных к типу *ошибочные срабатывания*. Основная проблема ложно-положительных срабатываний состояла в некорректном векторном представлении предложений, содержащих именованные сущности, не встречаемые в обучающей выборке, а также содержащих слова, незнакомые модели машинного перевода. Также было замечено, что алгоритм сравнения документов часто находил общие фразы вида «Работа посвящена следующей проблеме. . .» и т. п. Несмотря на корректность данных срабатываний, общие фразы представленного вида встречаются в большом числе документов и потому не должны рассматриваться как переводные заимствования. Общий процент документов с ложно-положительными срабатываниями составил 7%.

7 Заключение

В работе предложен алгоритм обнаружения переводных заимствований. Предложена декомпозиция алгоритма обнаружения переводных заимствований, позволяющая проводить эффективный поиск заимствований на больших текстовых коллекциях. Проведен анализ предложенного метода обнаружения заимствований, а также составной функции ошибки, используемой для оптимизации модели глубокого обучения. Для анализа качества представленного алгоритма были проведены эксперименты на синтетических данных для пары языков русский–английский. Качество алгоритма было также продемонстрировано на коллекции русскоязычных документов, входящих в РИНЦ. В дальнейшем планируется развитие предложенного алгоритма: использование модели векторного

представления предложений для задачи поиска кандидатов и улучшение качества отображения, ставящего в соответствие фразе вектор.

Авторы выражают свою благодарность Г. О. Еременко, ООО «Научная электронная библиотека», за предоставленные материалы.

Литература

1. *Никитов А. В., Орчаков О. А., Чехович Ю. В.* Плагиат в работах студентов и аспирантов: проблема и методы противодействия // Университетское управление: практика и анализ, 2012. Т. 5. С. 61–68.
2. *Khritankov A., Botov P., Surovenko N., Tsarkov S., Viuchnov D., Chekhovich Y.* Discovering text reuse in large collections of documents: A study of theses in history sciences // Artificial Intelligence and Natural Language & Information Extraction, Social Media and Web Search FRUCT Conference. — IEEE, 2015. P. 26–32.
3. *Зеленков И. В., Сегалович И. В.* Сравнительный анализ методов определения нечетких дубликатов для Web-документов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. 9-й Всеросс. научн. конф. RCDL. — Переславль-Залесский: Университет г. Переславля, 2007. С. 166–174.
4. *Franco-Salvador M., Gupta P., Rosso P.* Cross-language plagiarism detection using a multilingual semantic network // European Conference on Information Retrieval / Eds. P. Serdyukov, P. Braslavski, S. O. Kuznetsov, et al. — Lecture notes in computer science ser. — Berlin–Heidelberg: Springer, 2013. Vol. 7814. P. 710–713.
5. *Franco-Salvador M., Gupta P., Rosso P., Banchs R.* Cross-language plagiarism detection over continuous-space-and knowledge graph-based representations of language // Knowl.-Based Syst., 2016. Vol. 111. P. 87–99.
6. *Grman J., Ravas R.* Improved implementation for finding text similarities in large collections of data // Notebook papers of CLEF 2011 Labs and Workshops / Eds. V. Petras, P. Forner, P. D. Clough. — Amsterdam, The Netherlands, 2011. Vol. 1177. 6 p. <http://ceur-ws.org/Vol-1177/CLEF2011wn-PAN-GrmanEt2011.pdf>.
7. *Grozea C., Popescu M.* The encoplot similarity measure for automatic detection of plagiarism // Notebook papers of CLEF 2011 Labs and Workshops / Eds. V. Petras, P. Forner, P. D. Clough. — Amsterdam, The Netherlands, 2011. Vol. 1177. <http://ceur-ws.org/Vol-1177/CLEF2011wn-PAN-GrozeaEt2011.pdf>.
8. *Muhr M., Kern R., Zechner M., Granitzer M.* External and intrinsic plagiarism detection using a cross-lingual retrieval and segmentation system // Notebook papers of CLEF 2010 Labs and Workshops / Eds. M. Braschler, D. Harman, E. Pianta. — Padua, Italy, 2010. Vol. 1176. <http://ceur-ws.org/Vol-1176/CLEF2010wn-PAN-MuhrEt2010.pdf>.

9. *Bakhteev O., Kuznetsova R., Romanov A., Khritankov A.* A monolingual approach to detection of text reuse in Russian–English collection // Artificial Intelligence and Natural Language & Information Extraction, Social Media and Web Search FRUCT Conference. — IEEE, 2015. P. 3–10.
10. *Koehn P., Hoang Hien, Birch A., et al.* Moses: Open source toolkit for statistical machine translation // 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions Proceedings. — ACL, 2007. P. 177–180.
11. *Tai K., Socher R., Manning C.* Improved semantic representations from tree-structured long short-term memory networks // 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings. — ACL, 2015. Vol. 1. P. 1556–1566.
12. *Wieting J., Bansal M., Gimpel K., Livescu K.* Towards universal paraphrastic sentence embeddings // arXiv.org, 2015. arXiv:1511.08198 [cs.CL].
13. *Iyyer M., Manjunatha V., Boyd-Graber J. Daume H.* Deep unordered composition rivals syntactic methods for text classification // 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings. — ACL, 2015. Vol. 1. P. 1681–1691.
14. *Kuznetsova R., Bakhteev O., Ogal'tsov A.* Variational learning across domains with triplet information // 3rd Workshop on Bayesian Deep Learning. — Montreal, Canada. <http://bayesiandeeplearning.org/2018/papers/65.pdf>.
15. *Wang J., Shen H., Song J., Ji J.* Hashing for similarity search: A survey // arXiv.org, 2014. 29 p. arXiv:1408.2927 [cs.DS].
16. *Alain G., Bengio Y.* What regularized auto-encoders learn from the data-generating distribution // J. Mach. Learn. Res., 2014. Vol. 15. No. 1. P. 3563–3593.
17. *Jenssen M., Joos F., Perkins W.* On kissing numbers and spherical codes in high dimensions // Adv. Math., 2018. Vol. 335. P. 307–321.
18. *Cybenko G.* Approximation by superpositions of a sigmoidal function // Math. Control Signal., 1989. Vol. 2. No. 4. P. 303–314.
19. Синтетическая выборка для задачи обнаружения переводных заимствований. https://tiny.cc/cl_ru_en.
20. *Bojanowski P., Grave E., Joulin A., Mikolov T.* Enriching word vectors with subword information // Transactions Association for Computational Linguistics, 2017. Vol. 5. P. 135–146.
21. *Chung J., Gulcehre C., Cho K., Bengio Y.* Empirical evaluation of gated recurrent neural networks on sequence modeling // arXiv.org, 2014. 9 p. arXiv:1412.3555 [cs.NE].
22. *Tiedemann J.* News from OPUS — a collection of multilingual parallel corpora with tools and interfaces // Advances in natural language processing. — Amsterdam/Philadelphia: John Benjamins, 2009. Vol. 5. P. 237–248.

Поступила в редакцию 19.03.2020

METHODS OF CROSS-LINGUAL TEXT REUSE DETECTION IN LARGE TEXTUAL COLLECTIONS

R. V. Kuznetsova¹, O. Yu. Bakhteev^{1,2}, and Yu. V. Chekhovich³

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

²Antiplagiat Co., 42-1 Bolshoy Blvd., Moscow 121205, Russian Federation

³A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper investigates the cross-lingual text reuse detection problem. The paper proposes a monolingual approach to this problem: to translate the suspicious document into the language of the collection for the further monolingual analysis. One of the major requirements for the proposed method is robustness to the machine translation ambiguity. The further document analysis is divided into two steps. At the first step, the authors retrieve documents-candidates which are likely to be the source of the text reuse. For the robustness, the authors propose to retrieve the documents using word clusters that are constructed using distributional semantics. At the second step, the authors compare the suspicious document with candidates using sentence embeddings that are obtained by deep learning neural networks. The experiment was conducted for the “English–Russian” language pair both on the synthetic data and on the articles included in the Russian Science Citation Index.

Keywords: natural language processing; machine translation; deep learning; cross-lingual text reuse detection; distributional semantics

DOI: 10.14357/19922264210105

Acknowledgments

This research was supported by RFBR (project 18-07-01441) and Foundation for Assistance to Small Innovative Enterprises in Science and Technology (project 44116).

References

1. Nikitov, A. V., O. A. Orchakov, and Y. V. Chekhovich. 2012. Plagiat v rabotakh studentov i aspirantov: problema i metody protivodeystviya [Plagiarism in papers of students and graduate students: The problem and methods of counteraction]. *Universitetskoe upravlenie: praktika i analiz* [University Management: Practice and Analysis] 5:61–68.
2. Khritankov, A. S., P. V. Botov, N. S. Surovenko, S. V. Tsarkov, D. V. Viuchnov, and Y. V. Chekhovich. 2015. Discovering text reuse in large collections of documents: A study of theses in history sciences. *Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference Proceedings*. IEEE. 26–32.
3. Zelenkov, I. V., and I. V. Segalovich. 2007. Sravnitel'nyy analiz metodov opredeleniya nechetkikh dublikatov dlya Web-dokumentov [Comparative analysis of methods for determining fuzzy duplicates for Web-documents]. *Tr. 9-y Vseross. nauchn. konf. "Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kolekcii"* [9th All-Russian Scientific Conference "Digital libraries: Advanced Methods and Technologies, Electronic Collections" Proceedings]. Pereslavl-Zalessky: Pereslavl-Zalessky University. 166–174.
4. Franco-Salvador, M., P. Gupta, and P. Rosso. 2013. Cross-language plagiarism detection using a multilingual semantic network. *European Conference on Information Retrieval*. Eds. P. Serdyukov, P. Braslavski, S. O. Kuznetsov, et al. Lecture notes in computer science ser. Berlin–Heidelberg: Springer. 7814:710–713.
5. Franco-Salvador, M., P. Gupta, P. Rosso, and R. E. Banchs. 2016. Cross-language plagiarism detection over continuous-space-and knowledge graph-based representations of language. *Knowl.-Based Syst.* 111:87–99.
6. Grman, J., and R. Ravas. 2011. Improved implementation for finding text similarities in large collections of data. *Notebook papers of CLEF 2011 Labs and Workshops*. Eds. V. Petras, P. Forner, and P. D. Clough. 1177. 6 p. <http://ceur-ws.org/Vol-1177/CLEF2011wn-PAN-GrmanEt2011.pdf> (accessed January 18, 2021).
7. Grozea, C., and M. Popescu. 2011. The encoplot similarity measure for automatic detection of plagiarism. *Notebook papers of CLEF 2011 Labs and Workshops*. Eds. V. Petras, P. Forner, and P. D. Clough. Amsterdam, The Netherlands. 1177. Available at: <http://ceur-ws.org/Vol-1177/CLEF2011wn-PAN-GrozeaEt2011.pdf> (accessed January 18, 2021).
8. Muhr, M., R. Kern, M. Zechner, and M. Granitzer. 2010. External and intrinsic plagiarism detection using a cross-lingual retrieval and segmentation system. *Notebook paper of CLEF 2010 Labs and Workshops*. Eds. M. Braschler, D. Harman, and E. Pianta. Padua, Italy. 1176. Available at: <http://ceur-ws.org/Vol-1176/CLEF2010wn-PAN-MuhrEt2010.pdf> (accessed January 18, 2021).
9. Bakhteev, O., R. Kuznetsova, A. Romanov, and A. Khritankov. 2015. A monolingual approach to detection of text reuse in Russian–English collection. *Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference Proceedings*. IEEE. 3–10.
10. Koehn, P., Hien Hoang, A. Birch, et al. 2007. Moses: Open source toolkit for statistical machine translation. *45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions Proceedings*. ACL. 177–180.
11. Tai, K. S., R. Socher, and C. D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings*. ACL. 1:1556–1566.
12. Wieting, J., M. Bansal, K. Gimpel, and K. Livescu. 2015. Towards universal paraphrastic sentence embeddings. 19 p. Available at: <https://arxiv.org/abs/1511.08198> (accessed January 18, 2021).
13. Iyyer, M., V. Manjunatha, J. Boyd-Graber, and H. Daumé. 2015. Deep unordered composition rivals syntactic methods for text classification. *53rd Annual Meeting of the Association for Computational Linguistics and the 7th Joint Conference (International) on Natural Language Processing Proceedings*. ACL. 1:1681–1691.
14. Kuznetsova, R., O. Bakhteev, and A. Ogaltsov. 2018. Variational learning across domains with triplet information. *3rd Workshop on Bayesian Deep Learning Proceedings*. Available at: <http://bayesiandeeplearning.org/2018/papers/65.pdf> (accessed January 18, 2021).
15. Wang, J., H. T. Shen, J. Song, and J. Ji. 2014. Hashing for similarity search: A survey. 29 p. Available at: <https://arxiv.org/abs/1408.2927> (accessed January 18, 2021).
16. Alain, G., and Y. Bengio. 2014. What regularized auto-encoders learn from the data-generating distribution. *J. Mach. Learn. Res.* 15(1):3563–3593.
17. Jenssen, M., F. Joos, and W. Perkins. 2018. On kissing numbers and spherical codes in high dimensions. *Adv. Math.* 335:307–321.
18. Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signal.* 2(4):303–314.
19. Sinteticheskaya vyborka dlya zadachi obnaruzheniya perevodnykh zaimstvovaniy [Synthetic dataset for the cross-lingual text reuse detection problem]. Available at: https://tiny.cc/cl_ru_en (accessed January 18, 2021).

20. Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov. 2017. Enriching word vectors with subword information. *Transactions Association for Computational Linguistics* 5:135–146.
21. Chung, J., C. Gulcehre, K. Cho, and Y. Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. 9 p. Available at: <https://arxiv.org/abs/1412.3555> (accessed January 18, 2021).
22. Tiedemann, J. 2009. News from OPUS – a collection of multilingual parallel corpora with tools and interfaces. *Advances in natural language processing*. Amsterdam/Philadelphia: John Benjamins. 5:237–248.

Received March 19, 2020

Contributors

Kuznetsova Rita V. (b. 1990) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; rita.kuznetsova@phystech.edu

Bakhteev Oleg Yu. (b. 1993) — assistant professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; Head of Research Department, Antiplagiat Co., 42-1 Bolshoy Blvd., Moscow 121205, Russian Federation; bakhteev@ap-team.ru

Chekhovich Yury V. (b. 1976) — Candidate of Science (PhD) in physics and mathematics, Head of Department, 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; chehovich@ap-team.ru

ВАРИАЦИОННАЯ ОПТИМИЗАЦИЯ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ С КОНТРОЛЕМ СЛОЖНОСТИ*

О. С. Гребенькова¹, О. Ю. Бахтеев², В. В. Стрижов³

Аннотация: Исследуется задача построения модели глубокого обучения. Предлагается способ контроля ее сложности. Под сложностью модели понимается минимальная длина описания, минимальный объем информации, который требуется для передачи информации о модели и о выборке. Предлагается метод оптимизации параметров модели, основанный на представлении модели глубокого обучения в виде гиперсети с использованием байесовского подхода. Под гиперсетью понимается модель, которая порождает параметры оптимальной модели. Вводятся вероятностные предположения о распределении параметров модели глубокого обучения. Предлагается алгоритм, максимизирующий нижнюю вариационную оценку байесовской обоснованности модели. Вариационная оценка рассматривается как условная величина, зависящая от требуемой сложности модели. Для анализа качества предлагаемого алгоритма проведены эксперименты на выборке MNIST.

Ключевые слова: вариационная оптимизация модели; гиперсети; глубокое обучение; нейронные сети; байесовский вывод; заданная сложность модели

DOI: 10.14357/19922264210106

1 Введение

В работе рассматривается задача оптимизации модели глубокого обучения. Под моделью глубокого обучения понимается суперпозиция дифференцируемых по параметрам функций. Построение модели заданной сложности — одна из фундаментальных проблем глубокого обучения, так как по построению данное семейство моделей имеет избыточное число параметров [1]. В работе предполагается, что сложность модели задана заранее. Под сложностью модели понимается ее обоснованность.

Предлагаемый метод заключается в представлении параметров модели глубокого обучения в виде гиперсети. Гиперсеть — модель, которая задает параметры модели. На вход такой модели подается информация о структуре модели, а в результате работы порождается вектор параметров входной модели. Другими словами, гиперсеть — это отображение из множества переменных, отвечающих за сложность модели, во множество параметров модели. В работе [2] рассмотрены статистические и динамические гиперсети для порождения

весов сверточных и рекуррентных моделей соответственно.

В данной работе используется байесовский подход. Вводятся вероятностные предположения о параметрах моделей глубокого обучения. В работах [1, 3] предлагается использовать в качестве функции ошибки для оптимизации параметров модели глубокого обучения минимальную длину описания. Минимальная длина описания — минимальный объем информации, который требуется для передачи информации о модели и о выборке. Также в работах [1, 4] получены виды аппроксимаций для обоснованности модели для различных классов априорных распределений параметров.

Альтернативным подходом к построению модели заданной сложности выступает задача порождения и выбора оптимальной структуры моделей глубокого обучения. В работе [5] рассматривается возможность порождения широкого класса сверточных моделей как частей обобщенной модели, которая называется «фабрикой» (*англ.* *fab-ric*). Данная структура позволяет обойти процесс оптимизации параметров и проверки качества одиночных моделей. В работах [6, 7] представлены

*Настоящая статья содержит результаты проекта «Математические методы интеллектуального анализа больших данных», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М. В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 11.12.2018 № 13/1251/2018. Работа выполнена при поддержке РФФИ (проекты 19-07-01155 и 19-07-00875).

¹Московский физико-технический институт, grebenkova.os@phystech.edu

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

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

подходы к решению задачи выбора структуры нейросети с использованием дифференцируемых алгоритмов — стохастического (*англ.* Stochastic Neural Architecture Search — SNAS) и дифференцируемого (*англ.* Differentiable Neural Architecture Search — DNAS) методов поиска нейронных архитектур. Особенность работы [7] заключается в решении задачи выбора архитектуры модели, удовлетворяющей эксплуатационным требованиям: быстродействию на различных типах процессоров.

В данной работе исследуется поведение обобщенной функции обоснованности модели. Исследуется влияние априорного распределения на сложность модели. Для контроля сложности предлагается рассматривать задачу оптимизации параметров гиперсети. Данная модель порождает модели наперед заданной сложности с меньшими вычислительными затратами, чем в случае оптимизации модели, получаемой напрямую. Работа схожа с работой [8], где также исследовалась возможность получения гиперсети для предсказания наилучших гиперпараметров оптимизации модели. Вычислительный эксперимент проводился на выборке рукописных цифр MNIST [9].

2 Постановка задачи

Задана выборка:

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

Здесь $\mathbf{x}_i \in \mathbb{R}^m$; $y_i \in \{1, \dots, Y\}$, где Y — число классов. Рассмотрим модель

$$\mathbf{f}(\mathbf{x}, \mathbf{w}) : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^Y,$$

где $\mathbf{w} \in \mathbb{R}^n$ — пространство параметров модели. Пусть задано априорное распределение вектора параметров в пространстве \mathbb{R}^n :

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

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

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}$$

является апостериорным распределением вектора параметров \mathbf{w} при заданной выборке \mathcal{D} . Пусть также задано вариационное распределение

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

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

Для модели \mathbf{f} и соответствующего ей вектора параметров \mathbf{w} определим логарифмическую функцию правдоподобия выборки:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}|\mathbf{w}) = \log p(\mathcal{D}|\mathbf{w}).$$

Оптимальные значения \mathbf{w} находятся из максимизации $\mathcal{L}(\mathcal{D})$ — логарифма обоснованности модели:

$$\mathcal{L}(\mathcal{D}) = \log p(\mathcal{D}) = \log \int_{\mathbf{w} \in \mathbb{R}^n} p(\mathcal{D}|\mathbf{w})p(\mathbf{w}) d\mathbf{w}. \quad (1)$$

Так как вычисление интеграла (1) относится к вычислительно сложным задачам, рассмотрим вариационный подход к решению задачи. Оценим интеграл (1):

$$\begin{aligned} \mathcal{L}(\mathcal{D}) &= \log p(\mathcal{D}) = \\ &= \int_{\mathbf{w} \in \mathbb{R}^n} q(\mathbf{w}) \log \frac{p(\mathcal{D}, \mathbf{w})}{q(\mathbf{w})} d\mathbf{w} - \int_{\mathbf{w} \in \mathbb{R}^n} q(\mathbf{w}) \log \frac{p(\mathbf{w}|\mathcal{D})}{q(\mathbf{w})} d\mathbf{w} \geq \\ &\geq \int_{\mathbf{w} \in \mathbb{R}^n} q(\mathbf{w}) \log \frac{p(\mathcal{D}, \mathbf{w})}{q(\mathbf{w})} d\mathbf{w} = \int_{\mathbf{w} \in \mathbb{R}^n} q(\mathbf{w}) \log \frac{p(\mathbf{w})}{q(\mathbf{w})} d\mathbf{w} + \\ &+ \int_{\mathbf{w} \in \mathbb{R}^n} q(\mathbf{w}) \log p(\mathcal{D}|\mathbf{w}) d\mathbf{w} = \mathcal{L}_{\mathbf{w}}(\mathcal{D}, \mathbf{w}) + \mathcal{L}_E(\mathcal{D}). \quad (2) \end{aligned}$$

Первое слагаемое формулы (2) — это различие между апостериорным и априорным распределением параметров, задающее сложность распределением параметров относительно априорных предположений. Оно определяется расстоянием Кульбака—Лейблера, т.е. расстоянием между вариационным распределением $q(\mathbf{w})$ и априорным распределением $p(\mathbf{w})$:

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

Второе слагаемое формулы (2) представляет собой математическое ожидание правдоподобия выборки $\mathcal{L}_{\mathcal{D}}(\mathcal{D}|\mathbf{w})$:

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

Обобщенная обоснованность модели — это один из показателей сложности модели [1]. Рассматривается задача оптимизации параметров модели по обобщенной функции обоснованности \mathcal{L} :

$$\mathcal{L}(\lambda) = \lambda \mathcal{L}_{\mathbf{w}}(\mathcal{D}, \mathbf{w}) - \mathcal{L}_E(\mathcal{D}), \quad (3)$$

где параметр сложности λ контролирует влияние априорного распределения на выбор итоговой модели.

Введем множество допустимых значений параметра сложности $\lambda \in \Lambda \subset \mathbb{R}^+$. Требуется найти такое отображение $\mathfrak{G} : \Lambda \longrightarrow \mathbb{R}^n$, при котором для

произвольного значения параметра сложности $\lambda \in \Lambda$ параметры доставляли бы максимум следующей функции:

$$\mathfrak{G}(\lambda) = \arg \max_{\mathbf{w} \in \mathbb{R}^n} (\log p(\mathcal{D}|\mathbf{w}) - \lambda D_{KL}(q(\mathbf{w})||p(\mathbf{w}))). \quad (4)$$

3 Построение гиперсети для контроля сложности модели

Решение задачи оптимизации (4) для произвольного значения $\lambda \in \Lambda$ относится к вычислительно сложным задачам. В данной работе для ее решения предлагается использовать гиперсеть.

Пусть задано множество параметров Λ , контролирующих сложность модели. Гиперсеть — это параметрическое отображение из множества Λ во множество параметров модели:

$$\mathbf{G} : \Lambda \times \mathbb{R}^u \rightarrow \mathbb{R}^n,$$

где \mathbb{R}^u — множество допустимых параметров гиперсети. Рассмотрены два вида гиперсетей: (1) с отображением во множество матриц низкого ранга и (2) с линейной аппроксимацией. Пусть \mathbf{f} — функция, переводящаяся λ в скрытый слой; \mathbf{U}_1 и \mathbf{U}_2 — матрицы, переводящие из скрытого слоя в нужную размерность, их конкатенация принадлежит пространству параметров гиперсети: $[\mathbf{U}_1, \mathbf{U}_2] = \mathbf{U} \in \mathbb{R}^u$; \mathbf{B}_1 — матрица, не зависящая от λ . Тогда первая реализация гиперсети имеет вид:

$$\mathbf{G}_{\text{lowrank}}(\lambda) = (\mathbf{f}(\lambda)\mathbf{U}_1)^\top (\mathbf{f}(\lambda)\mathbf{U}_2) + \mathbf{B}_1, \quad (5)$$

где параметр λ — случайное число, реализуемое для каждой подвыборки при оптимизации параметров. Вторая реализация имеет вид:

$$\mathbf{G}_{\text{linear}}(\lambda) = \lambda \mathbf{b}_2 + \mathbf{b}_3, \quad (6)$$

где \mathbf{b}_2 и \mathbf{b}_3 — константы, не зависящие от λ .

Для аппроксимации оптимизационной задачи (4) предлагается оптимизировать параметры гиперсети $\mathbf{U} \in \mathbb{R}^u$ по случайно порожденным значениям параметра сложности $\lambda \in \Lambda$:

$$\begin{aligned} E_{\lambda \sim P(\lambda)} (\log p(\mathcal{D}|\mathbf{w}) - \lambda D_{KL}(q(\mathbf{w})||p(\mathbf{w}))) \rightarrow \\ \rightarrow \max_{\mathbf{U} \in \mathbb{R}^u}, \quad (7) \end{aligned}$$

где $P(\lambda)$ — некоторое распределение на множестве Λ . В данной работе в качестве распределения использовалось следующее: $\log \lambda \sim \mathcal{U}(-1, 2)$, где \mathcal{U} — равномерное распределение.

4 Вычислительный эксперимент

Для анализа свойств обобщенной задачи оптимизации (7) и предложенного метода построения гиперсети был проведен вычислительный эксперимент на выборке рукописных цифр MNIST [9]. Проведено сравнение следующих методов:

- (а) построения модели напрямую без использования гиперсети (4);
- (б) построения модели напрямую без использования гиперсети (4) с оптимизацией за одну эпоху;
- (в) построение с использованием гиперсети (5);
- (г) построение с использованием гиперсети (5) с дообучением итоговой модели за одну эпоху;
- (д) построение с использованием гиперсети (6);
- (е) построение с использованием гиперсети (6) с дообучением итоговой модели за одну эпоху.

Второй метод позволяет определить, насколько использование гиперсетей (5), (6) с дообучением за одну эпоху эффективно для оптимизации параметров модели в сравнении с оптимизацией модели напрямую. Для каждой из моделей проводились пять запусков, результаты которых усреднялись.

Для каждой модели проводилось прореживание параметров с применением подхода, описанного в [1]. Критерием удаления параметров выступала относительная плотность модели:

$$\rho(w_i) \propto \exp\left(-\frac{\mu_i^2}{2\sigma_i^2}\right).$$

Рассматривались следующие критерии качества модели.

1. Точность классификации

$$\text{Accuracy} = 1 - \frac{1}{m} \sum_{i=1}^m [f(\mathbf{x}_i, \mathbf{w}) \neq y_i],$$

где m — длина тестовой выборки.

2. Стабильность модели S : отношение точности исходной модели к точности модели с прореживанием 90% параметров. Данная величина показывает, насколько качество модели стабильно относительно удаления значительного числа параметров.
3. Число обновлений параметров модели. Этот показатель определяется как число всех обновлений значений параметров модели за все эпохи и характеризует сложность итоговой оптимизации.
4. Обобщенная обоснованность модели \mathfrak{L} (3).

Была рассмотрена нейросеть состоящая из двух слоев: 100 и 10 нейронов соответственно, где второй слой отвечает за функцию softmax:

$$f(\mathbf{x}, \mathbf{w}) = \text{softmax}(\mathbf{w}_2^\top \text{ReLU}(\mathbf{w}_1^\top \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2).$$

Здесь \mathbf{w}_1 и \mathbf{b}_1 — параметры первого слоя нейросети; \mathbf{w}_2 и \mathbf{b}_2 — параметры второго слоя нейросети:

$$\text{softmax}(\mathbf{x})_i = \frac{\exp(x_i)}{\sum_{j=1}^k \exp(x_j)}, \quad i = 1, \dots, k;$$

$$\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x}). \quad (8)$$

Нейросеть запускалась для разных значений параметра сложности $\lambda \in \Lambda$.

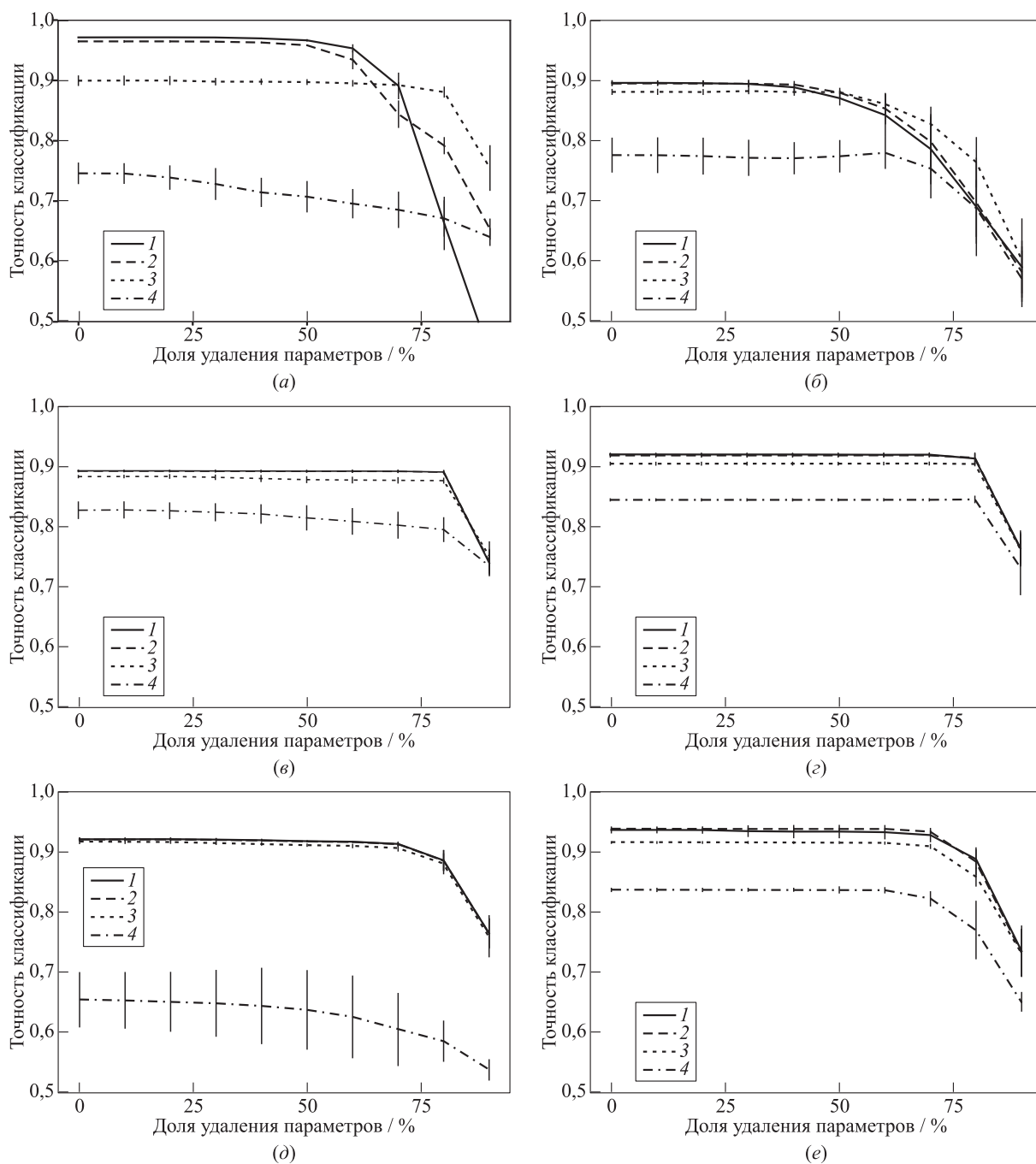


Рис. 1 График зависимости точности классификации от процента удаленных параметров для моделей без использования гиперсети (4) (a), без использования гиперсети (4) с оптимизацией за одну эпоху (б), с использованием гиперсети (5) (в), с использованием гиперсети (5) с дообучением итоговой модели за одну эпоху (г), с использованием гиперсети (6) (д) и с использованием гиперсети (6) с дообучением итоговой модели за одну эпоху (e): 1 — $\lambda = 10^{-1}$; 2 — 10^0 ; 3 — 10^1 ; 4 — $\lambda = 10^2$

На рис. 1, а показано, как меняется точность классификации при удалении параметров указанным методом. Из графика видно, что вариационный метод позволяет удалить $\approx 60\%$ параметров при $\lambda = 10^{-1}$ и 10^0 и $\approx 80\%$ параметров при $\lambda = 10^1$ и 10^2 без значительной потери точности классификации. При дальнейшем удалении качество для всех значений снижается. При больших значениях $\lambda > 100$ получается избыточно упрощенная модель, которая содержит малое число параметров. Таким образом, удаление параметров нейросети при данном значении λ слабо влияет на точность классификации. Однако изначальная точность невысока.

Для обеих моделей с использованием гиперсетей использовался оптимизатор SGD (stochastic gradient descent). Обучение проводилось на протяжении 25 эпох. В качестве A_{pr}^{-1} используется $\text{diag}(0, 1)$. Для первой модели (5) был рассмотрен случай с 50 нейронами в скрытом слое и с функцией активации ReLU (8). При обучении второй модели (6) каждая подвыборка проходила процесс оптимизации с пятью разными значениями сэмплируемой λ . Прореживание нейросетей запускалось для разных значений параметра сложности $\lambda \in \Lambda$.

На рис. 1, в показано, как меняется точность классификации при удалении параметров указанным методом для модели с низкоранговой аппроксимацией. Как видно из графика, средняя точность классификации относительно базового эксперимента для малых значений $\lambda \in [10^{-1}; 10^1]$ понизилась. Также сильно увеличилось отклонение от среднего. При этом для всех значений $\lambda \in \Lambda$ получили более стабильную модель: точность класси-

фикации меньше зависит от удаления параметров. Однако наблюдаем большую потерю точности при удалении более 80%.

На рис. 1, д показано, как меняется точность классификации при удалении параметров указанным методом для модели с линейной аппроксимацией. Линейная модель показала еще более стабильные результаты относительно предыдущих экспериментов. При этом точность классификации для небольших значений λ , $\lambda \in [10^{-1}; 10^1]$ остается постоянной при удалении до 60% процентов параметров и равной $\approx 90\%$. Отклонения от среднего незначительные для небольших значений λ , $\lambda \in [10^{-1}; 10^1]$. Далее данные модели были дообучены независимо от гиперсети в течение одной эпохи и эксперимент с прореживанием был запущен еще раз.

На рис. 1, з показано, как меняется точность у дообученной модели с низкоранговой аппроксимацией при удалении параметров. Как видно из графика, после обучения точность классификации увеличилась, уменьшилось отклонение от среднего. Стабильность модели осталась прежней, и для всех значений λ точность значительно падает при удалении более 80% параметров. Для значения $\lambda = 10^2$ модель показала улучшение в точности классификации и большую стабильность относительно предыдущей версии модели.

На рис. 1, е показано, как меняется точность у дообученной модели с линейной аппроксимацией при удалении параметров. Как видно из графика, после обучения точность классификации увеличилась для всех значений λ , значительно увеличилось

Таблица 1 Точность и число обновлений параметров моделей

Модель	Точность				Число обновлений параметров
	$\lambda = 10^{-1}$	$\lambda = 10^0$	$\lambda = 10^1$	$\lambda = 10^2$	
(а) Без гиперсети	0,87686	0,901570	0,8674	0,73031	7 458 976 000
(б) Без гиперсети с дообучением	0,81682	0,81986	0,83477	0,78924	298 359 040
(в) Гиперсеть (5)	0,87719	0,87666	0,87466	0,83124	4 165 376 600
(г) Гиперсеть (5) с дообучением	0,90262	0,90167	0,88876	0,83902	4 239 966 360
(д) Гиперсеть (6)	0,900679	0,90021	0,89218	0,53857	3 729 488 000
(е) Гиперсеть (6) с дообучением	0,90104	0,91456	0,89538	0,80627	3 804 077 760

Таблица 2 Стабильность моделей

Модель	Стабильность S			
	$\lambda = 10^{-1}$	$\lambda = 10^0$	$\lambda = 10^1$	$\lambda = 10^2$
(а) Без гиперсети	2,209571	1,476841	1,194326	1,165526
(б) Без гиперсети с дообучением	1,518516	1,537249	1,473905	1,362817
(в) Гиперсеть (5)	1,208003	1,203938	1,177115	1,125796
(г) Гиперсеть (5) с дообучением	1,20265	1,201760	1,186034	1,156385
(д) Гиперсеть (6)	1,206345	1,205434	1,208482	1,216405
(е) Гиперсеть (6) с дообучением	1,281112	1,287463	1,255834	1,289508

Таблица 3 Обобщенная обоснованность модели

Модель	\mathcal{L}			
	$\lambda = 10^{-1}$	$\lambda = 10^0$	$\lambda = 10^1$	$\lambda = 10^2$
(а) Без гиперсети	-9 035,229	-24 338,234	-56 679,427	-128 928,671
(б) Без гиперсети с дообучением	-32 788,838	-51 832,864	-186 124,637	-696 240,551
(в) Гиперсеть (5)	-24 566,315	-30 949,930	-56 720,932	-166 657,021
(г) Гиперсеть (5) с дообучением	-19 994,677	-27 220,746	-55 508,397	-132 758,414
(д) Гиперсеть (6)	-24 603,767	-28 189,602	-58 147,425	-177 139,477
(е) Гиперсеть (6) с дообучением	-20 776,473	-26 262,996	-57 948,826	-134 340,962

отклонение от среднего при удалении более чем 60% параметров. Понижилась стабильность модели, но она по-прежнему выше, чем в экспериментах с полноранговой моделью. Для значения $\lambda = 10^2$ модель показала улучшения в точности классификации.

На рис. 1, б показано, как меняется точность у дообученной в течение одной эпохи модели без гиперсети при удалении параметров. Из графика видно, что одной эпохи недостаточно для значительного улучшения модели, построенной напрямую.

Общие результаты экспериментов представлены в табл. 1–3. Несмотря на незначительную потерю в качестве, гиперсеть позволяет получить схожие результаты в сравнении с обычными моделями, но при существенно меньших вычислительных затратах. Более того, по графикам видно, что модель сохраняет схожие свойства при прореживании.

5 Заключение

В работе рассматривалась задача оптимизации модели глубокого обучения с наперед заданной сложностью. Итоговый метод заключался в представлении модели глубокого обучения в виде гиперсети. Использовался байесовский подход. Были введены вероятностные предположения о параметрах моделей глубокого обучения. По результатам вычислительного эксперимента можно сделать вывод о том, что модели на основе гиперсети имеют меньшую точность классификации, чем обычные модели. Однако при использовании гиперсети снижаются вычислительные затраты и сохраняются свойства моделей при прореживании.

В дальнейшем планируется исследовать теоретические свойства гиперсетей, а также улучшить

предложенные модели для построения сетей глубокого обучения с контролем сложности.

Литература

1. Graves A. Practical variational inference for neural networks // Advances in neural information processing systems / Eds. J. Shawe-Taylor, R. Zemel, P. Barlett, et al. — ACM, 2011. Vol. 24. P. 2348–2356.
2. Ha D., Dai A. M., Le Q. V. HyperNetworks // arXiv.org, 2016. arXiv:1609.09106 [cs.LG]. 29 p.
3. Kuznetsov M. P., Tokmakova A. A., Strijov V. V. Analytic and stochastic methods of structure parameter estimation // Informatica, 2016. Vol. 27. P. 607–624.
4. Стрижов В. В., Бахтеев О. Ю. Выбор модели глубокого обучения субоптимальной сложности // Автоматика и телемеханика, 2018. № 8. С. 129–147.
5. Saxena S., Verbeek J. Convolutional neural fabrics // Advances in neural information processing systems / Eds. D. Lee, M. Sugiyama, U. Luxburg, et al. — ACM, 2016. Vol. 29. P. 4053–4061.
6. Xie S., Zheng H., Liu C., Lin L. SNAS: Stochastic neural architecture search // arXiv.org, 2019. arXiv:1812.09926 [cs.LG]. 17 p.
7. Wu B., Dai X., Zhang P., Wang Y., Sun F., Wu Y., Tian Y., Vajda P., Jia Y., Keutzer K. FBNet: Hardware-aware efficient convnet design via differentiable neural architecture search // IEEE/CVF Conference on Computer Vision and Pattern Recognition. — IEEE, 2019. Vol. 1. P. 10726–10734.
8. Lorraine J., Duvenaud D. Stochastic hyperparameter optimization through hypernetworks // arXiv.org, 2018. arXiv:1802.09419 [cs.LG]. 9 p.
9. LeCun Y., Cortes C., Burges C. The MNIST dataset of handwritten digits, 1998. <http://yann.lecun.com/exdb/mnist/index.html>.

Поступила в редакцию 14.08.2020

VARIATIONAL DEEP LEARNING MODEL OPTIMIZATION WITH COMPLEXITY CONTROL

O. S. Grebenkova¹, O. Yu. Bakhteev^{1,2}, and V. V. Strijov^{1,3}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation

²Antiplagiat Co., 42-1 Bolshoy Blvd., Moscow 121205, Russian Federation

³A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: This paper investigates the problem of deep learning model optimization. The authors propose a method to control model complexity. The minimum description length is interpreted as the complexity of the model. It acts as the minimal amount of information that is required to transfer information about the model and the dataset. The proposed method is based on representation of a deep learning model. The authors propose a form of a hypernet using the Bayesian inference. A hypernet is a model that generates parameters of an optimal model. The authors introduce probabilistic assumptions about the distribution of parameters of the deep learning model. The paper suggests maximizing the evidence lower bound of the Bayesian model validity. The authors consider the evidence bound as a conditional value that depends on the required model complexity. The authors analyze this method in computational experiments on the MNIST dataset.

Keywords: model variational optimization; hypernets; deep learning; neural networks; Bayesian inference; model complexity control

DOI: 10.14357/19922264210106

Acknowledgments

This paper contains results of the project “Statistical methods of machine learning” which is carried out within the framework of the program “Center of Big Data Storage and Analysis” of the National Technology Initiative Competence Center. It is supported by the Ministry of Science and Higher Education of the Russian Federation according to the agreement between the M. V. Lomonosov Moscow State University and the Foundation of Project Support of the National Technology Initiative from 11.12.2018, No. 13/1251/2018. This research was supported by RFBR (projects 19-07-01155 and 19-07-00875).

References

1. Graves, A. 2011. Practical variational inference for neural networks. *Advances in neural information processing systems*. Eds. J. Shawe-Taylor, R. Zemel, P. Barlett, *et al.* ACM. 24:2348–2356.
2. Ha, D., A. M. Dai, and Q. V. Le. 2017. HyperNetworks. 29 p. Available at: <https://arxiv.org/pdf/1609.09106.pdf> (accessed January 25, 2021).
3. Kuznetsov, M. P., A. A. Tokmakova, and V. V. Strijov. 2016. Analytic and stochastic methods of structure parameter estimation. *Informatika* 27(3):607–624.
4. Strijov, V. V., and O. Yu. Bakhteev. 2018. Deep learning model selection of suboptimal complexity. *Automat. Rem. Contr.* 79(8):1474–1488.
5. Saxena, S., and J. Verbeek. 2016. Convolutional neural fabrics. *Advances in neural information processing systems*. Eds. D. Lee, M. Sugiyama, U. Luxburg, *et al.* — ACM. 29:4053–4061.
6. Xie, S., H. Zheng, C. Liu, and L. Lin. 2019. SNAS: Stochastic neural architecture search. 17 p. Available at: <https://arxiv.org/abs/1812.09926> (accessed January 25, 2021).
7. Wu, B., X. Dai, P. Zhang, Y. Wang, F. Sun, Y. Wu, Y. Tian, P. Vajda, Y. Jia., and K. Keutzer. 2019. FBNet: Hardware-aware efficient convnet design via differentiable neural architecture search. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings*. IEEE. 1:10726–10734
8. Lorraine, J., and D. Duvenaud. 2018. Stochastic hyperparameter optimization through hypernetworks. 9 p. Available at: <https://arxiv.org/pdf/1802.09419.pdf> (accessed January 25, 2021).
9. LeCun, Y., C. Cortes, and C. Burges. 1998. The MNIST dataset of handwritten digits. Available at: <http://yann.lecun.com/exdb/mnist/> (accessed January 25, 2021).

Received August 14, 2020

Contributors

Grebenkova Olga S. (b. 1999) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; grebenkova.os@phystech.edu

Bakhteev Oleg Yu. (b. 1993) — assistant professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; Head of Research Department, Antiplagiat Co., 42-1 Bolshoy Blvd., Moscow 121205, Russian Federation; bakhteev@ap-team.ru

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; professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; strijov@phystech.edu

ИНФОРМАЦИОННАЯ МОДЕЛЬ ВЕСОВОГО ОБЛИКА ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Л. Л. Вышинский¹, Ю. А. Флёров²

Аннотация: Статья посвящена описанию информационной модели весового облика летательных аппаратов (ЛА). Под весовым обликом ЛА здесь понимается совокупность взаимосвязанных между собой информационных объектов, содержащих описание конструкции, параметров и характеристик ЛА, достаточное для проведения весовых расчетов, весового анализа и весового контроля на всех этапах жизненного цикла изделия. Описанная информационная модель может служить основой логической схемы базы данных (БД) при разработке автоматизированных систем весового проектирования (АСВП). В статье модель весового облика ЛА описана в терминах сетевых структур данных.

Ключевые слова: автоматизация проектирования; летательный аппарат; весовое проектирование; весовая модель; дерево конструкции; генератор проектов

DOI: 10.14357/19922264210107

1 Информационная структура весовой модели летательного аппарата

В работе [1] была представлена АСВП ЛА, которая разработана как многопользовательская информационная система клиент-серверной архитектуры с централизованной БД сетевого типа. Настоящая статья посвящена описанию структурно-параметрической весовой модели ЛА, которая была положена в основу логической схемы БД этой системы. Выбор сетевой модели данных был обусловлен спецификой предметной области, основными объектами которой являются конструкции, состоящие из различных агрегатов, узлов и деталей, объединенные многочисленными связями и отношениями. Сетевая модель данных в наибольшей степени приспособлена для описания сложно организованных систем.

В основе сетевой модели данных лежат два основных понятия — структурированная запись данных (Record) и набор записей (Set), отражающий бинарные отношения типа один ко многим. Записи данных представляют собой поименованные кортежи фиксированной длины:

$$r = (x_1, x_2, \dots, x_n),$$

где $x_i \in X_i, i = 1, \dots, n; r \in R$. Множества X_i определяют типы данных атрибутов x_i кортежа, а множество R определяет тип записи r .

Типы данных атрибутов кортежей могут быть примитивными (целыми и действительными числами, строками, булевыми значениями, перечислимыми типами) или агрегатными типами, являющимися векторами, структурами (в смысле языка С) или более сложными конструкциями. В описании произвольной предметной области экземпляры записей представляют конкретные объекты этой прикладной области, а наборы задают связи между объектами — бинарные отношения один ко многим.

Тип набора определяется как упорядоченная поименованная пара

$$S = \langle R_i \rightarrow R_j \rangle. \quad (1)$$

Экземпляр s набора типа S определяется как пара, состоящая из экземпляра записи типа R_i , который принято называть владельцем набора, и подмножества экземпляров записей типа R_j , которые называют членами набора. Множество подчиненных членов набора может быть упорядочено по ключу, который задается при описании набора. Такие наборы называют ключевыми наборами. Возможно создание так называемых сингулярных наборов типа $S = \langle \emptyset \rightarrow R \rangle$, где формально владелец набора не задается, а членами такого набора являются записи из упорядоченного подмножества записей типа R . Возможно создание набора с одинаковыми типами владельца и членов набора, так называемые рекурсивные наборы $S = \langle R \rightarrow R \rangle$, которые

¹Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, wyshinsky@mail.ru

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, fler@ccas.ru

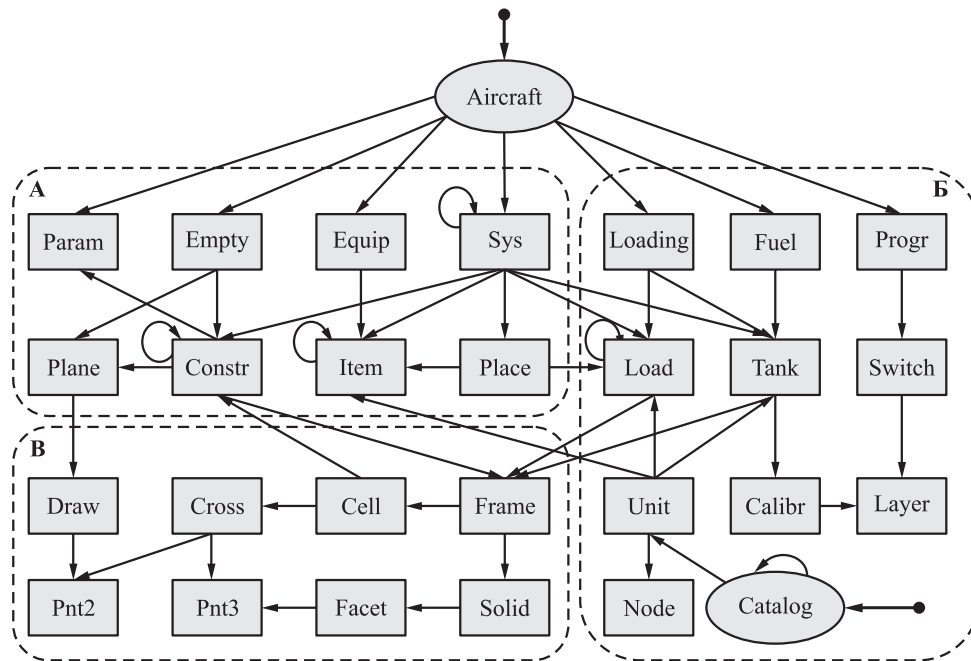


Рис. 1 Сетевая структура весовой модели ЛА: А — пустой снаряженный ЛА; Б — полезная нагрузка; В — геометрические модели

устанавливают отношения между записями одного типа, например отношения наследования или отношения вхождения. Такие наборы используются для описания строго иерархических структур. Заметим, что в таких наборах множества членов разных экземпляров набора не должны пересекаться.

Описание модели предметной области в парадигме сетевой модели данных задается множеством типов записей и множеством типов наборов. Графически такое описание может быть представлено в виде ориентированного графа, у которого узлам соответствуют типы записей, а дуге (ребру), идущей от узла R_i к узлу R_j , соответствует тип набора (1). Сингулярным наборам соответствуют дуги, входящие в узлы «извне», а рекурсивным наборам — дуги, выходящие из некоторого узла графа и возвращающиеся в тот же узел.

На рис. 1 приведено графическое изображение сетевой структуры весовой модели, положенной в основу АСВП ЛА. Узлам этой структуры соответствуют типы записей весовой модели, а стрелкам — связи между записями — наборы.

Основная функция весовой модели состоит в информационном обеспечении решения задач весового проектирования на всех этапах жизненного цикла ЛА. Вся текущая информация о структуре объекта, о массово-инерционных характеристиках (МИХ), о геометрических и других параметрах, не-

обходимых для весовых расчетов, для весового анализа и для весового контроля в процессе создания и эксплуатации ЛА, может быть размещена в структурах весовой модели. Вход в эту структуру осуществляется через сингулярный набор $\langle \emptyset \rightarrow \text{Aircraft} \rangle$, задающий множество представленных в БД АСВП ЛА.

Обязательными атрибутами всех записей в весовой модели являются их идентификационные данные — это номера, идентификаторы, шифры, наименования в зависимости от роли записи в весовой модели. Для записей типа **Aircraft** идентификационными атрибутами служат паспортные данные ЛА (обозначение, производитель, тип ЛА, его назначение, год выпуска).

Центральной задачей весового проектирования является вычисление МИХ ЛА и его компонент — массы, положения центра масс, статических моментов, собственных и полных моментов инерции.

Другая важная задача весовой модели состоит в обеспечении весового контроля в процессе проектирования, изготовления и сборки изделий в процессе опытного и серийного производства. Для этого весовая модель должна содержать полную информацию о конструкции ЛА и о вариантах целевого использования будущего изделия. Эти два аспекта в весовой модели на рис. 1 выделены в блоках А и Б. Блок А представляет собой структурно-параметрическую модель пустого снаряженного

изделия, а блок **В** содержит информацию о возможных вариантах загрузки ЛА, о параметрах и характеристиках топливных баков, а также о программах заправки и выработки топлива. Структуры, объединенные в блоке **В**, предназначены для описания геометрических моделей, использующихся при расчетах МИХ и для визуализации результатов расчетов.

2 Структурно-параметрическая модель пустого снаряженного летательного аппарата

Структурно-параметрическая модель пустого снаряженного ЛА содержит описание всех элементов конструкции изделия и его снаряжения, а также полный состав конструктивных параметров ЛА и его агрегатов, влияющих на их МИХ. Параметры в весовой модели ЛА задаются записями типа **Param**. Атрибуты этих записей включают идентификатор параметра, наименование, тип (*real*, *int*, *string*, *boolean*), способ задания (интерактивный ввод, вычисление по формулам, сами формулы, зависящие от других параметров), область допустимых значений. Набор типа определяет полное множество параметров ЛА. Это множество содержит весовые, геометрические, конструктивные и другие данные, от которых зависят масса, конструкция и другие МИХ ЛА. Множество весовых параметров ЛА представляет весовой облик изделия.

Структура пустого снаряженного ЛА состоит из двух ветвей: **Empty** — структуры пустого ЛА и **Equip** — состава снаряжения. Ветвь пустого ЛА отражает конструкцию планера и размещаемого внутри планера стационарного комплекса бортового оборудования. Обычно в весовой модели представлено несколько альтернативных вариантов конструкции, включая возможные варианты состава комплектующих систем. Варианты пустого изделия задаются набором $\langle \text{Aircraft} \rightarrow \text{Empty} \rangle$. Идентифицирующими атрибутами записей типа **Empty** служат обозначение или номер варианта. Основными расчетными параметрами являются МИХ пустого изделия. Наборы $\langle \text{Empty} \rightarrow \text{Constr} \rangle$ и $\langle \text{Constr} \rightarrow \text{Constr} \rangle$ образуют иерархическую структуру пустого ЛА, которую называют деревом конструкции изделия. Записи типа **Constr** являются узлами дерева конструкции и в совокупности представляют детальное описание агрегатов ЛА.

Описание каждого элемента конструкции может быть дополнено множеством конструктивных параметров, которое задается набором типа

$\langle \text{Constr} \rightarrow \text{Param} \rangle$. Визуализация общих видов ЛА и элементов конструкций обеспечивается наличием в весовой модели записей типа **Plan** и наборами типа $\langle \text{Empty} \rightarrow \text{Plan} \rangle$ и $\langle \text{Constr} \rightarrow \text{Plan} \rangle$. Записи типа **Plan** содержат идентификатор чертежа и тип общего вида (план, вид сбоку или спереди).

Локализация элементов конструкции в системе координат (СК) в весовой модели задается набором типа $\langle \text{Sys} \rightarrow \text{Constr} \rangle$, владельцем которого выступает запись типа **Sys**. Множество используемых в процессе проектирования ЛА систем координат в весовой модели описывается набором $\langle \text{Aircraft} \rightarrow \text{Sys} \rangle$. Это множество должно образовывать строго иерархическую структуру. Иерархия СК задается рекурсивным набором типа $\langle \text{Sys} \rightarrow \text{Sys} \rangle$, а атрибутами записей типа **Sys** служат идентификатор СК, название или комментарий к СК, а также положение локальной СК относительно вышестоящей СК, которое задается кортежем $(x, y, z, \alpha, \beta, \gamma)$. Пересчет центра масс и расчет полных моментов инерции осуществляется от локальной СК конструкции и далее по иерархии в главную СК. На рис. 2 приведено описание структуры дерева конструкции ЛА. Там же дан текст процедуры расчета МИХ, использующей эту структуру.

Дерево конструкции пустого ЛА является центральным объектом весовой модели как по значимости его в процессе проектирования ЛА, так и по объему создаваемой и хранящейся в БД информации. Число записей в дереве конструкции исчисляется десятками, а то и сотнями тысяч. Однако, благодаря использованию сетевых структур в весовой модели ЛА, расчет МИХ представляет собой хотя и длительную, но достаточно просто организованную рекурсивную процедуру. Приведенная процедура позволяет рассчитывать МИХ любой конструкции, вплоть до всего дерева пустого ЛА, состоящего из сотен тысяч деталей.

Расчет МИХ снаряжения ЛА ведется аналогично расчету пустого изделия. Снаряжение — это объекты, которые размещаются на борту ЛА в процессе его подготовки перед полетом и предназначены для обеспечения выполнения полетного задания. Снаряжение ЛА обычно подразделяют на базовое и специальное. К базовому снаряжению относят экипаж, средства и системы жизнеобеспечения экипажа и пассажиров, системы технологического обеспечения полета, а также заправляемые жидкости и газы, включая невырабатываемый остаток топлива. К специальному снаряжению относят технологические устройства и системы, предназначенные для выполнения целевых задач полетного задания, в том числе средства закрепления перевозимых грузов в отсеках ЛА и на внешних узлах подвески.

```

// ОПИСАНИЕ ДЕРЕВА КОНСТРУКЦИИ:
type t_Constr: struct // тип конструкции
    // атрибуты записи:
    (text cod, // обозначение
    text name, // наименование
    text type, // тип конструкции
    real m_thr, // теоретическая масса
    real m_lime, // лимитная масса
    real m_plot, // чертежная масса
    real m_fact, // фактическая масса
    real m_fast, // масса крепежа
    t_gab gab, // габариты конструкции
    t_pos pops, // положение в локальной СК
    text mat, // материал конструкции
    real dens, // плотность материала
    t_mich mich // МИХ — M, X, Y, Z, Jxx, Jyy, ...
);

// дерево конструкции:
type Tree_Constr: document
(record Constr: t_Constr; // узлы дерева

set Constr_Constr // набор, задающий
owner Constr // отношение
member Constr; // подчиненности
);

// ПРОЦЕДУРА РАСЧЕТА МИХ КОНСТРУКЦИИ:
proc eval_constr (Tree_Constr.Constr RC0, t_Constr C0)
// RC0 — ссылка на запись вершины дерева
// C0 — запись вершины дерева
{ var Tree_Constr.Constr RCN, t_Constr CN,
// RCN ссылка на подчиненные C0 записи
// CN подчиненные C0 записи
real PMx, PMy, PMz; // статические моменты C0
// построение списка подчиненных вершин:
Tree_Constr.Constr_Constr_mem (RC0, 0, RCN, CN);
// если список пуст (RCN=null), то выход из процедуры
while (isnotnull(RCN)) // цикл по подчиненным вершинам
{ // рекурсивный спуск по дереву:
call eval_constr_constr(RCN, CN);
if (CN.mich.M>0) // при M<=0 запись пропускается
{ PMx:=C0.mich.M*C0.mich.X; PMy:=C0.mich.M*C0.mich.Y;
PMz:= C0.mich.M* C0.mich.Z; C0.mich.M += CN.mich.M;
C0.mich.X := (PMx + CN.mich.M * CN.mich.X)/C0.mich.M;
C0.mich.Y := (PMy + CN.mich.M * CN.mich.Y)/C0.mich.M;
C0.mich.Z := (PMz + CN.mich.M * CN.mich.Z)/C0.mich.M;
C0.mich.Jxx += CN.mich.Jxx; C0.mich.Jyy += CN.mich.Jyy;
C0.mich.Jzz += CN.mich.Jzz; C0.mich.Jxy += CN.mich.Jxy;
C0.mich.Jxz += CN.mich.Jxz; C0.mich.Jyz += CN.mich.Jyz;
} Tree_Constr.Constr_Constr_next (RCN, CN);
} // выход из цикла
} // конец процедуры

```

Рис. 2 Фрагмент описания структуры весовой модели ЛА

Варианты снаряжения в весовой модели задаются записями типа **Equip**, а все множество вариантов задается набором $\langle \text{Aircraft} \rightarrow \text{Equip} \rangle$. Каждый вариант снаряжения состоит из множества элементов снаряжения (наборы типа $\langle \text{Equip} \rightarrow \text{Item} \rangle$). Элементы снаряжения могут быть организованы в иерархию средствами рекурсивных наборов типа $\langle \text{Item} \rightarrow \text{Item} \rangle$. Как правило, существуют несколько типовых вариантов комплектации экипажа и базового снаряжения. Весовая модель должна содержать перечень используемых вариантов снаряжения и их характеристики. Размещение снаряжения на борту ЛА задается аналогично конструкция пустого ЛА с помощью наборов типа $\langle \text{Sys} \rightarrow \text{Item} \rangle$. В дополнение к такому способу размещения отдельные элементы снаряжения могут быть привязаны к определенным точкам конструкции ЛА, местам размещения, которые задаются записями типа **Place**. Привязка элемента снаряжения к месту размещения обеспечивается наборами типа $\langle \text{Place} \rightarrow \text{Item} \rangle$. Места размещения снаряжения локализованы в одной из СК изделия $\langle \text{Sys} \rightarrow \text{Place} \rangle$.

Перечень элементов снаряжения может быть каталогизирован, что позволит упростить компоновку взлетной конфигурации ЛА и выполнение расчетов МИХ пустого снаряженного ЛА. Связь элементов снаряжения с соответствующим ката-

логом в весовой модели реализуется с помощью набора типа $\langle \text{Unit} \rightarrow \text{Place} \rangle$, владельцем которого выступают записи типа **Unit**, представляющие каталогизированные элементы снаряжения.

3 Модель размещения полезной нагрузки летательного аппарата

Целевую нагрузку ЛА вместе с заправляемым топливом обычно называют полезной нагрузкой. На рис. 1 структуры модели размещения на борту ЛА полезной нагрузки сосредоточены в блоке **Б**. Модели полезной нагрузки задаются наборами типа $\langle \text{Aircraft} \rightarrow \text{Loading} \rangle$ — множеством вариантов целевой загрузки ЛА, наборами типа $\langle \text{Aircraft} \rightarrow \text{Fuel} \rangle$ — множеством вариантов стартовой заправки ЛА топливом и $\langle \text{Aircraft} \rightarrow \text{Progr} \rangle$ — множество вариантов программ выработки топлива.

Каждый из вариантов целевой нагрузки, задаваемых записями типа **Loading**, состоит из множества различных элементов нагрузки (наборы типа $\langle \text{Loading} \rightarrow \text{Loading} \rangle$). Элементы нагрузки (записи типа **Load**) — это внешние объекты по отношению к ЛА. Разумеется, специфика размещаемой нагрузки связана с типом ЛА. Для пассажирских ЛА

целевой нагрузкой считаются пассажиры и их багаж, для транспортных самолетов — разнообразные перевозимые коммерческие грузы, для военных — вооружение.

Разнообразие элементов нагрузки требует их классификации и каталогизации. Для этой цели в рамках весовой модели предназначена специальная структура каталогов нагрузки. Каталоги элементов нагрузки в весовой модели задаются сингулярными наборами типа $\langle \emptyset \rightarrow \text{Catalog} \rangle$. Как правило, каталоги состоят из нескольких разделов и подразделов. Их структура задается рекурсивными наборами $\langle \text{Catalog} \rightarrow \text{Catalog} \rangle$. Множество элементов нагрузки каждого из разделов классификатора представляется наборами $\langle \text{Catalog} \rightarrow \text{Unit} \rangle$. В записях типа **Unit** дается исчерпывающая информация об элементе нагрузки, включая его обозначение и полное название, а главное, все необходимые для размещения и расчетов данные — тип нагрузки, МИХ, габариты, форма и пр.

В дополнение к описанию технических характеристик в каталоге может быть задана дополнительная информация об узлах крепления (наборы типа $\langle \text{Unit} \rightarrow \text{Node} \rangle$), конструктивно выполняемых на элементах нагрузки для пристыковки к ним других объектов. Такие узлы характерны для временно устанавливаемых на военных ЛА переходных балок, предназначенных для подвески вооружения.

Записи типа **Load** связываются с каталогом элементов нагрузки с помощью наборов типа $\langle \text{Unit} \rightarrow \text{Load} \rangle$. Элементы нагрузки могут быть многоярусными (переходная балка — узел крепления — подвеска). Это отражается в структуре весовой модели рекурсивным набором типа $\langle \text{Load} \rightarrow \text{Load} \rangle$. Структуры размещения полезной нагрузки непосредственно связаны со структурами пустого снаряженного ЛА. Положение нагрузки, размещаемой на борту или на подвесках, связывается с СК пустого ЛА (наборы типа $\langle \text{Sys} \rightarrow \text{Load} \rangle$) или непосредственно с местом размещения, например с местом в пассажирском салоне, местом на стеллажах багажных отсеков, с точкой подвески под крылом или под фюзеляжем (наборы типа $\langle \text{Place} \rightarrow \text{Load} \rangle$). В состав варианта загрузки ЛА могут входить кроме элементов целевой нагрузки временно размещаемые подвесные или вкладные топливные баки (наборы типа $\langle \text{Loading} \rightarrow \text{Tank} \rangle$). Характеристики пустых подвесных баков, которые задаются в записях типа **Tank**, берутся из каталога (наборы типа $\langle \text{Unit} \rightarrow \text{Tank} \rangle$), а размещение задается наборами типа $\langle \text{Sys} \rightarrow \text{Tank} \rangle$ и $\langle \text{Place} \rightarrow \text{Tank} \rangle$.

Важнейшей составляющей полезной нагрузки является топливо. В весовой модели ЛА варианты стартовой заправки топливом задаются записями

типа **Fuel**. Общий стартовый объем топлива зависит от конфигурации топливной системы, от наличия временно размещаемых подвесных или вкладных топливных баков. Конфигурация топливных емкостей задается наборами типа $\langle \text{Fuel} \rightarrow \text{Tank} \rangle$. Это множество включает все множество топливных емкостей, включая и временно устанавливаемые баки, и стационарные, встроенные в конструкцию ЛА. Располагаемый объем топливных емкостей, МИХ заправленного топлива определяются геометрией баков. От геометрии зависит и характер изменения МИХ топлива по мере его выработки в полете ЛА. Основными характеристиками баков, которые позволяют анализировать изменения общей массы ЛА и положение его центра масс в процессе выработки топлива, служат тарифовочные характеристики баков — зависимости положения центра тяжести и моментов инерции топлива в баке от массы невыработанного остатка, а также от углов тангажа и крена ЛА. В весовой модели могут быть представлены несколько вариантов тарифовочных характеристик (наборы типа $\langle \text{Tank} \rightarrow \text{Calibr} \rangle$). Построение тарифовочных характеристик основано на разбиении всего объема топлива в баке на тонкие слои параллельными плоскостями. Каждый слой топлива (записи типа **Layer**) характеризуется своей массой и другими МИХ. Варианты тарифовочных характеристик представляются упорядоченными множествами слоев (наборы типа $\langle \text{Calibr} \rightarrow \text{Layer} \rangle$).

Процесс выработки топлива из топливных баков управляется специальными программами топливной системы, которые обеспечивают определенный порядок перекачки топлива в расходные баки силовой установки. Эти программы должны обеспечивать равномерность и бесперебойность подачи топлива при любых эволюциях ЛА, а также сохранение центровки ЛА в заданных пределах. Под управлением программы выработки происходит изменение состояния различных устройств топливной системы — клапанов, вентилях, перекачивающих насосов. В весовой модели ЛА возможно задание нескольких программ (записями типа **Progr**) для различных конфигураций и состава топливных баков. Каждая программа характеризуется последовательностью переключений управляющих устройств — $\langle \text{Progr} \rightarrow \text{Switch} \rangle$. Между переключениями происходит выработка топлива из определенных баков. Эти периоды можно связать с выработкой очередных слоев топлива из разных баков. Множество вырабатываемых слоев между переключениями определяется наборами $\langle \text{Switch} \rightarrow \text{Layer} \rangle$. Таким образом, каждой программе выработки топлива соответствует упорядоченная последовательность вырабатываемых объемов

топлива из разных баков. Это позволяет строить таблицы и графики изменения положения центра масс и моментов инерции текущего остатка топлива в полете.

Массово-инерционные характеристики топлива на старте и их изменение в процессе выработки всецело зависят от геометрических форм топливных баков. Поэтому одной из предпосылок эффективности проведения расчетов является наличие в рамках весовой модели ЛА адекватных геометрических моделей трехмерных (3D) конструкций.

4 Геометрические расчетные модели

В проектировании ЛА геометрическое моделирование занимает центральное место. И хотя основная нагрузка в части решения задач весового проектирования — это численные расчеты, без геометрического, пространственного представления рассчитываемых изделий расчеты невозможны. Выше, на схеме информационной весовой модели (см. рис. 1), структуры геометрических моделей отображены в блоке **В**. На рис. 3 приведены три вида геометрических объектов, представимых в структурах весовой модели ЛА.

Двумерные модели уже упоминались при описании пустого ЛА. Это чертежи общих видов ЛА (рис. 3, *а*) и чертежи элементов конструкции, которые в структуре весовой модели представлены записями типа **Plan**. Любой чертеж состоит из множества элементов чертежа, которое представлено наборами типа $\langle \mathbf{Plan} \rightarrow \mathbf{Draw} \rangle$. Под элементами чертежа (записи типа **Draw**) понимаются такие объекты, как ломаные линии, произвольные многоугольники и параметрически задаваемые плоские фигуры. Атрибутами записей типа **Draw** кроме позиционного номера служат общепринятые при отображении плоских рисунков параметры — цвет, толщина и тип линий, цвет и характер заливки плоских фигур, их положение на поле чертежа. Ломаные линии и плоские фигуры задаются наборами

типа $\langle \mathbf{Draw} \rightarrow \mathbf{Pnt2} \rangle$, которые представляют множества нумерованных двумерных точек — записей типа $\mathbf{Pnt2} = (n, x, y)$.

Структуры 3D представления объектов в весовой модели ЛА формируются записями типа **Frame**. **Frame** — это головная запись, содержащая идентификационное обозначение геометрической модели и комментариев. Трехмерное геометрическое представление в весовой модели ЛА могут иметь три типа объектов — конструкции пустого ЛА (наборы типа $\langle \mathbf{Constr} \rightarrow \mathbf{Frame} \rangle$), топливные баки (наборы типа $\langle \mathbf{Tank} \rightarrow \mathbf{Frame} \rangle$) и элементы целевой нагрузки и снаряжения (наборы типа $\langle \mathbf{Unit} \rightarrow \mathbf{Frame} \rangle$). Трехмерное геометрическое представление может быть реализовано в двух формах — с помощью 3D каркасной модели поверхности объекта (рис. 3, *б*) и в виде 3D-модели твердого тела (рис. 3, *в*). Каркасная модель задается набором отсеков (набор типа $\langle \mathbf{Frame} \rightarrow \mathbf{Cell} \rangle$), которые в свою очередь состоят из упорядоченного множества плоских пространственных сечений (наборы типа $\langle \mathbf{Cell} \rightarrow \mathbf{Cross} \rangle$ и $\langle \mathbf{Cross} \rightarrow \mathbf{Pnt3} \rangle$), где $\mathbf{Pnt3} = (n, x, y, z)$ — записи, представляющие трехмерные нумерованные точки).

Геометрическое описание отсеков может использоваться для представления конструктивных отсеков пустого ЛА и размещения внутри этих отсеков других конструкций (наборы $\langle \mathbf{Cell} \rightarrow \mathbf{Constr} \rangle$). Трехмерная модель твердого тела может состоять из 3D-отсеков (записи типа **Solid**, которые задаются наборами плоских сечений (наборы $\langle \mathbf{Frame} \rightarrow \mathbf{Solid} \rangle$ и $\langle \mathbf{Solid} \rightarrow \mathbf{Facet} \rangle$). Плоские пространственные сечения в 3D-моделях (записи типа **Facet**) представляют грани 3D-отсеков (наборы типа $\langle \mathbf{Facet} \rightarrow \mathbf{Pnt3} \rangle$). В расчетах МИХ топлива и элементов конструкции пустого ЛА используют только 3D-модели. На стадиях рабочего проектирования ЛА в современных условиях 3D-модели конструкций ЛА строятся в специализированных системах геометрического моделирования и при необходимости могут передаваться в весовую модель ЛА в стандартных форматах представления твердых тел. Каркасные модели могут применяться при

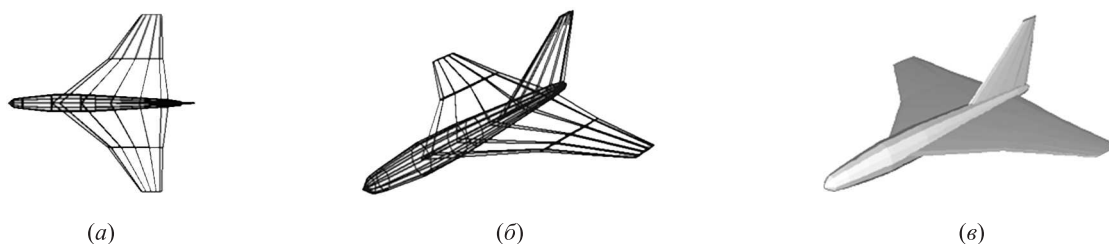


Рис. 3 Геометрические объекты, представимые в весовой модели ЛА

первичном задании форм трехмерных объектов на начальных стадиях проектирования, при этом кар-касные модели легко преобразуются в 3D-представление геометрии. Визуализация геометрических моделей возможна при всех формах задания геометрии.

5 Заключение

Представленная в статье информационная весовая модель ЛА, как уже говорилось, была разработана в процессе создания АСВП самолетов. Модель была апробирована при реализации ряда программных модулей этой системы. В работе [2]

был предложен подход к созданию цифрового весового паспорта самолета, который также основан на представленной в статье информационной модели весового облика ЛА.

Литература

1. Вышинский Л. Л., Флёров Ю. А., Широков Н. И. Автоматизированная система весового проектирования самолетов // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 18–30. doi: 10.14357/19922264180103.
2. Вышинский Л. Л., Курьянский М. К., Флеров Ю. А. Цифровая модель весового паспорта летательного аппарата // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 3–10. doi: 10.14357/19922264190401.

Поступила в редакцию 20.01.2020

INFORMATION MODEL OF AIRCRAFT WEIGHT PROFILE

L. L. Vyshinsky and Yu. A. Flerov

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

Abstract: The article is devoted to the description of the information model of the weight profile of an aircraft. The weight profile of an aircraft is understood as a set of interconnected information objects containing a description of the structure, parameters, and characteristics of the aircraft sufficient for weight calculations, weight analysis, and weight control at all stages of the product life cycle. The described information model can serve as the basis for a scheme of a database in the development of automated weight design systems. In the article, the weight model of an aircraft is described in terms of network data structures.

Keywords: design automation; aircraft; weight design; weighting model; design tree; project generator

DOI: 10.14357/19922264210107

References

1. Vyshinsky, L. L., Yu. A. Flerov, and N. I. Shirokov. 2018. Avtomatizirovannaya sistema vesovogo proektirovaniya samoletov [Computer-aided system of aircraft weight design]. *Informatika i ee Primeneniya* — *Inform Appl.* 12(1):18–30. doi: 10.14357/19922264180103.
2. Vyshinsky, L. L., M. K. Kuryansky, and Yu. A. Flerov. 2018. Tsifrovaya model' vesovogo pasporta letatel'nogo apparata [Digital model of the aircraft's weight passport]. *Informatika i ee Primeneniya* — *Inform Appl.* 13(4):3–10. doi: 10.14357/19922264190401.

Received January 20, 2020

Contributors

Vyshinsky Leonid L. (b. 1941) — Candidate of Science (PhD) 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; wysh@ccas.ru

Flerov Yuri A. (b. 1942) — Doctor of Science in physics and mathematics, professor, Corresponding Member of the Russian Academy of Sciences, 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; fler@ccas.ru

ОПТИМАЛЬНОЕ ПОРОГОВОЕ УПРАВЛЕНИЕ ДОСТУПОМ В СИСТЕМЕ $M/M/s$ С НЕОДНОРОДНЫМИ ПРИБОРАМИ И ОБЩИМ НАКОПИТЕЛЕМ*

Я. М. Агаларов¹

Аннотация: Рассматривается система $M/M/s$ с неоднородными приборами и общим накопителем с возможностью управления длиной очереди с целью максимизации среднего предельного дохода. Функция дохода включает плату за успешно обслуженную заявку, штраф за каждую отклоненную заявку, штрафы за единицу времени простоя каждого прибора, штраф за единицу времени ожидания заявки (или за превышение допустимого времени ожидания заявки), затраты, связанные с техническим обслуживанием мест в накопителе. Ставится задача максимизации предельного дохода на множестве простых пороговых стратегий управления длиной очереди. Доказано свойство выпуклости функции дохода и получены условия существования конечного оптимального порога длины очереди.

Ключевые слова: система массового обслуживания; оптимизация; пороговая стратегия; длина очереди

DOI: 10.14357/19922264210108

1 Введение

Задача ограничения доступа к ресурсам (управления потоками) и распределения допущенной нагрузки по ресурсам с целью оптимальной их загрузки — одна из наиболее важных подзадач общей задачи управления потоковыми системами, в частности информационными системами [1]. Для разработки и исследования процедур управления потоками и распределением нагрузки в таких системах успешно используется метод математического моделирования, при этом в качестве математических моделей потоковых систем в основном используются системы массового обслуживания (СМО) различного типа.

Для СМО с неоднородными приборами подзадача оптимальной загрузки ресурсов рассматривалась во многих работах, и в некоторых из них получены точные аналитические результаты, позволяющие найти оптимальный план диспетчеризации, минимизирующий среднее число заявок в системе (см., например, работы [2–6]). Опишем пример решения указанной подзадачи, приведенный в работе [2]. Рассматривается СМО типа $M/M/K/N$ ($0 \leq N \leq \infty$) с K неоднородными приборами, N местами для ожидания, пуассоновским входящим потоком интенсивности λ и экспоненциально распределенными временами обслуживания с параметрами μ_i , $i = 1, \dots, K$. Ставится задача оптимального выбора свободного прибора, на

который в момент поступления будет направлена вновь поступившая заявка, если очередь пуста, или в момент освобождения прибора будет направлена первая заявка в очереди, если она не пуста. Если в момент поступления очередь не пуста, то новая заявка становится в очередь, если есть свободные места ожидания, а иначе теряется. Пересылка заявок с одного прибора на другой не допускается. Для данной СМО получен следующий оптимальный план диспетчеризации, минимизирующий среднее число заявок в системе в стационарном режиме ее работы: если необходимо направить заявку на прибор, то нужно выбрать свободный прибор с наибольшей производительностью.

Аналогичное решение получено в работе [6] для СМО с двумя гетерогенными приборами, которые различаются по скорости обслуживания и характеристикам надежности (например, более медленный прибор абсолютно надежен, в то время как более быстрый прибор подвержен случайным сбоям). С помощью марковской модели процесса принятия решений доказано, что оптимально всегда направлять клиентов на более быстрый прибор, когда он доступен, независимо от частоты его отказов и скорости ремонта, если система стабильна. Для более медленного прибора существует оптимальная пороговая политика, которая зависит от длины очереди и состояния более быстрого прибора.

Ниже в рассматриваемой СМО используется такой же план распределения заявок по приборам,

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00692, 19-07-00739 и 20-07-00804).

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

как и в представленных выше работах, за исключением того, что при отсутствии очереди разрешается пересылка заявки с прибора меньшей производительности на освободившийся прибор большей производительности.

Одним из наиболее часто применяемых на практике способов ограничения заявок доступа в систему стала так называемая пороговая стратегия, суть которой заключается в том, что решение о допуске (план, политика допуска) заявки в систему меняется в момент достижения наблюдаемой величиной определенного порогового значения.

Из множества работ, посвященных пороговым стратегиям, отметим работы [7–11], в которых исследованы задачи, близкие по постановке к рассматриваемой в данной статье. В [7] рассматривается $M/M/1$ с управляемой нагрузкой, в которой можно отклонять заявки по мере прибытия и в каждый момент принятия решения корректировать скорость обслуживания. Цель состоит в минимизации предельных (долгосрочных) средних затрат, которые включают затраты на эксплуатацию мест хранения, затраты за единицу времени обслуживания заявки при заданной скорости обслуживания, штраф за отклонение задания. Разработан алгоритм вычисления оптимальной политики, который позволяет на каждом шаге вычислять точность приближения политики. В работе [8] в качестве оптимальной политики ограничения длины очереди, минимизирующей затраты СМО типа $GI/GI/1$ с нетерпеливыми заявками в условиях интенсивного трафика, предлагается использовать решение диффузионной аппроксимирующей задачи. Заявка после поступления в очередь через экспоненциально распределенный интервал времени покидает систему, если она за это время не обслужилась. Система при поступлении заявки в очередь получает платеж, при отклонении заявки и уходе необслуженной заявки из очереди платит соответствующие штрафы. В [9] сформулирована оптимизационная задача допуска для СМО типа $M/M/K/N$ с однородными приборами, когда есть затраты, связанные с отклонением заявок, простоем приборов и отказом допущенных в систему заявок от обслуживания. Предполагается, что процесс прибытия является пуассоновским, время обслуживания экспоненциально и заявка отказывается от обслуживания после экспоненциального времени пребывания в очереди. Заданы стоимость одной отклоненной заявки, стоимость отказа одной заявки, стоимость единицы времени простоя одного прибора. Построен соответствующий марковский процесс принятия решений и показано, что оптимальная политика имеет пороговую форму. Решена аппроксимирующая задача управ-

ления диффузией. Задача минимизации предельных средних затрат при интенсивном трафике для последовательности СМО типа $M/M/1$ с возможностью управления длиной очереди и скоростью обслуживания рассмотрена в работе [10]. Функция затрат включает штраф за каждую отвергнутую заявку, затраты, связанные с корректировкой скорости обслуживания, и штраф за каждую отказавшуюся заявку. Для построения оптимального управления СМО в условиях интенсивного трафика (выбора оптимального размера буфера и оптимальной скорости обслуживания) предлагается использовать оптимальную стратегию для аппроксимирующей задачи управления диффузией.

В работах [11, 12] для СМО типа $M/M/K$ с однородными приборами доказана унимодальность стоимостной целевой функции от порогового значения числа заявок в системе: в работе [11] в случае, когда целевая функция учитывает плату за одну обслуженную заявку, штраф за отклонение заявки и стоимость единицы среднего времени ожидания заявки, а в работе [12] в случае, когда целевая функция учитывает плату за одну своевременно обслуженную заявку, штраф за отклонение заявки и штраф за несвоевременно обслуженную заявку.

Ниже в работе обобщаются результаты работ [11, 12] для СМО типа $M/M/K$ на случай неоднородных приборов и наличия штрафа за единицу среднего времени простоя каждого прибора.

2 Формулирование задачи

Рассматривается СМО с s неоднородными приборами обслуживания, на которую поступает пуассоновский поток заявок с интенсивностью λ . Предполагается, что приборы имеют условные номера $1, \dots, N$ и закон распределения времени обслуживания на i -м приборе является показательным с параметром μ_i , $\mu_1 \geq \dots \geq \mu_s$. Заявки допускаются в систему согласно простой пороговой стратегии следующим образом: поступившая заявка допускается в систему (занимает любое свободное место в накопителе), если в момент ее поступления число заявок в системе меньше порогового значения k , $k \geq s$, в противном случае она отклоняется и повторно не возвращается. При распределении заявок по приборам предпочтение отдается наиболее производительным: если в момент поступления в систему есть свободный прибор, то новая заявка занимает свободный прибор с наименьшим номером и начинает сразу обслуживаться, иначе становится в конец очереди на обслуживание. Заявка после окончания обслуживания сразу покидает систему, освободив прибор и место в накопителе. Если в оче-

реди нет заявок и производительность освободившегося прибора строго больше, чем интенсивность обслуживания одного из занятых приборов, то заявка с прибора наименьшей среди занятых приборов производительности снимается и пересылается на освободившийся прибор. Параллельное обслуживание заявки на нескольких приборах не допускается.

Отметим, что процесс обслуживания в рассматриваемой СМО описывается марковским процессом гибели и размножения с непрерывным временем с множеством состояний $\{i : i = 0, \dots, k\}$. Обозначим через λ интенсивность входного потока, $\tilde{\mu}_i = \mu_1 + \dots + \mu_i$, β_{ij} — интенсивность перехода процесса из состояния i в состояние j , $i = 0, \dots, k$, $j = 0, \dots, k$. Для β_{ii+1} и β_{ii-1} в данном случае имеют место равенства:

$$\beta_{ii+1} = \lambda, \quad i = 0, \dots, k-1;$$

$$\beta_{ii-1} = \begin{cases} \tilde{\mu}_s, & i = s+1, \dots, k; \\ \tilde{\mu}_i, & i = 1, \dots, s. \end{cases}$$

Согласно формулам (см. [13]) для стационарных вероятностей состояний процесса гибели и размножения π_j^k в случае рассматриваемой СМО справедливы формулы:

$$\pi_j^k = \begin{cases} \pi_0^k \frac{\lambda^j}{\prod_{i=1}^j \tilde{\mu}_i}, & 1 \leq j \leq s; \\ \pi_0^k \frac{\lambda^s}{\prod_{i=1}^s \tilde{\mu}_i} \left(\frac{\lambda}{\tilde{\mu}_s}\right)^{j-s}, & s < j \leq k, \end{cases} \quad (1)$$

где

$$\pi_0^k = \left[1 + \sum_{j=1}^s \frac{\lambda^j}{\prod_{i=1}^j \tilde{\mu}_i} + \frac{\lambda^s}{\prod_{i=1}^s \tilde{\mu}_i} \sum_{j=1}^{k-s} \left(\frac{\lambda}{\tilde{\mu}_s}\right)^j \right]^{-1}.$$

В качестве показателя эффективности системы используется $D(k)$ — «доход» системы в единицу времени в стационарном режиме работы при стратегии k :

$$D(k) = C_0 \Lambda_{\text{ВЫХ}}(k) - C_1 \Lambda_{\text{ОТК}}(k) - C_2 W_{\text{ОЖ}}(k) - \sum_{i=1}^s C_{3i} Q_{\text{ПР},i}(k) - C_4 k, \quad (2)$$

где $\Lambda_{\text{ВЫХ}}(k)$ — интенсивность обслуженных заявок; $\Lambda_{\text{ОТК}}(k)$ — интенсивность отклоняемых заявок; $W_{\text{ОЖ}}(k)$ — суммарное среднее время ожидания в очереди обслуженных в единицу времени заявок; $Q_{\text{ПР},i}(k)$ — вероятность простоя i -го прибора; $C_0 \geq 0$ — плата, получаемая системой, если поступившая заявка будет обслужена системой; $C_1 \geq 0$ —

штраф за отклонение поступившей заявки; $C_2 \geq 0$ — штраф за единицу времени ожидания заявки в очереди; C_{3i} — штраф за единицу времени простоя i -го прибора; $C_4 \geq 0$ — стоимость технического обслуживания одного места в накопителе в единицу времени.

Задача оптимизации пороговой стратегии сформулирована в виде математической задачи

$$k^* = \arg \max_{k \geq s} D(k). \quad (3)$$

3 Решение задачи

Рассмотрим вложенную цепь Маркова, где переходы цепи определяются моментами поступления заявок и состояние системы есть число заявок, находящихся в системе в момент поступления [13]. Отметим, что стационарные вероятности состояний вложенной цепи также вычисляются по формулам (1).

Обозначим через q_i^k среднее значение дохода системы, получаемого в состоянии i цепи при стратегии k без учета простоя приборов. Тогда функцию $D(k)$ можно переписать в виде:

$$D(k) = \lambda d(k) + d_{\text{ПР}}(k), \quad (4)$$

где

$$d(k) = \sum_{i=0}^k \pi_i^k q_i^k; \quad d_{\text{ПР}}(k) = \sum_{i=1}^s C_{3i} Q_{\text{ПР},i}(k).$$

Здесь

$$q_i^k = \begin{cases} C_0 - C_2 W_i(k) - C_4 k \bar{v}, & s \leq i < k; \\ q_{k-1}^k - C_1 - C_0, & i = k; \\ C_0 - C_4 k \bar{v}, & 0 \leq i \leq s-1, \end{cases} \quad (5)$$

где $W_i(k)$ — суммарное среднее время ожидания заявок в очереди в состоянии i ; $\bar{v} = 1/\lambda$.

Для величины $W_i(k)$ имеют место равенства (см. [11]):

$$W_i(k) = \begin{cases} \frac{1}{\tilde{\mu}_s} \left[\frac{1}{2} \sum_{m=1}^{i+2-s} (m-1) m r_m - (i+1-s) \sum_{m=1}^{i+2-s} m r_m - \frac{1}{2} (i+1-s)(i+2-s) \sum_{m=i+3-s}^{\infty} r_m \right], & i = s-1, \dots, k-1; \\ 0, & i = 0, \dots, s-1. \end{cases}$$

Отсюда и из формул (5) следуют равенства:

$$q_{i+1}^k = \begin{cases} q_i^k - \frac{C_2}{\tilde{\mu}_s} \sum_{m=1}^{i+2-s} mr_m - \\ - \frac{C_2(i+2-s)}{\tilde{\mu}_s} \sum_{m=i+3-s}^{\infty} r_m, & s-1 \leq i \leq k-2; \\ q_i^k = C_0 - C_4k\bar{v}, & i = 0, \dots, s-2; \end{cases} \quad (6)$$

$$q_k^k = q_{k-1}^k - C_0 - C_1.$$

Для стационарных вероятностей состояний справедлива формула:

$$\pi_{i+1}^{k+1} = A_{k+1}\pi_i^k, \quad i = s-1, \dots, k. \quad (7)$$

Здесь

$$A_{k+1} = \frac{1 - Q_{s-1}^{k+1}}{1 - Q_{s-2}^k},$$

где

$$Q_{s-1}^{k+1} = \sum_{i=0}^{s-1} \pi_i^{k+1}; \quad Q_{s-2}^k = \begin{cases} 0, & s = 1; \\ \sum_{i=0}^{s-2} \pi_i^k, & s \geq 2. \end{cases}$$

Рассмотрим разность $d(k) - d(k+1)$. Используя (6) и (7), находим:

$$\begin{aligned} d(k) - d(k+1) &= \sum_{i=0}^k \pi_i^k q_i^k - \sum_{i=0}^{k+1} \pi_i^{k+1} q_i^{k+1} = \\ &= (C_0 - C_4k)\bar{v} \sum_{i=0}^{s-2} \pi_i^k + \sum_{i=s-1}^k \pi_i^k q_i^k - \\ &- [C_0 - C_4(k+1)\bar{v}] \sum_{i=0}^{s-1} \pi_i^{k+1} - \sum_{i=s}^{k+1} \pi_i^{k+1} q_i^{k+1} = \\ &= (C_0 - C_4k)\bar{v} (Q_{s-2}^k - Q_{s-1}^{k+1}) + C_4Q_{s-1}^{k+1}\bar{v} + \\ &+ \sum_{i=s-1}^k \pi_i^k q_i^k - A_{k+1} \sum_{i=s-1}^k \pi_i^k q_{i+1}^{k+1}. \end{aligned}$$

Введем для краткости изложения следующее обозначение:

$$\Delta_{i+1}^{k+1} = q_{i+1}^{k+1} - q_i^{k+1}, \quad s-1 \leq i \leq k.$$

Заметим, что

$$\begin{aligned} q_i^{k+1} &= q_i^k - C_4\bar{v}, \quad s-1 \leq i \leq k-1; \\ q_{k+1}^{k+1} &= q_k^{k+1} - C_1 - C_0 = \\ &= q_{k-1}^{k+1} + \Delta_k^{k+1} - C_1 - C_0 = \\ &= q_{k-1}^k + \Delta_k^{k+1} - C_1 - C_0 - C_4\bar{v} = \\ &= q_k^k + \Delta_k^{k+1} - C_4\bar{v}. \end{aligned}$$

Находим

$$\begin{aligned} d(k) - d(k+1) &= (C_0 - C_4k)\bar{v} (Q_{s-2}^k - Q_{s-1}^{k+1}) + \\ &+ C_4Q_{s-1}^{k+1}\bar{v} + \sum_{i=s-1}^k \pi_i^k q_i^k - A_{k+1} \sum_{i=s-1}^k \pi_i^k q_{i+1}^{k+1} - \\ &- A_{k+1} \left(\sum_{i=s-1}^{k-1} \pi_i^k \Delta_{i+1}^{k+1} + \pi_k^k \Delta_k^{k+1} \right) = \\ &= -(C_0 - C_4k)\bar{v} \frac{Q_{s-1}^{k+1} - Q_{s-2}^k}{1 - Q_{s-2}^k} + \\ &+ \frac{Q_{s-1}^{k+1} - Q_{s-2}^k}{1 - Q_{s-2}^k} \left(C_0 - C_4k \right) Q_{s-2}^k \bar{v} + \sum_{i=s-1}^k \pi_i^k q_i^k + \\ &+ C_4\bar{v} - A_{k+1} \left(\sum_{i=s-1}^{k-1} \pi_i^k \Delta_{i+1}^{k+1} + \pi_k^k \Delta_k^{k+1} \right) = \\ &= (1 - A_{k+1}) \left[d(k) - (C_0 - C_4k)\bar{v} + C_4 \frac{\bar{v}}{1 - A_{k+1}} - \right. \\ &\left. - \frac{A_{k+1}}{1 - A_{k+1}} \left(\sum_{i=s-1}^{k-1} \pi_i^k \Delta_{i+1}^{k+1} + \pi_k^k \Delta_k^{k+1} \right) \right]. \quad (8) \end{aligned}$$

Из (6) имеем:

$$\begin{aligned} \Delta_{i+1}^{k+1} &= \\ &= -\frac{C_2}{\tilde{\mu}_s} \left[\sum_{m=1}^{i+2-s} mr_m - (i+2-s) \sum_{m=i+3-s}^{\infty} r_m \right], \\ & \quad i = s-1, \dots, k-1; \quad (9) \end{aligned}$$

$$\begin{aligned} d(k) - d(k+1) &= (1 - A_{k+1}) \left[d(k) - (C_0 - C_4k)\bar{v} + \right. \\ &+ C_4 \frac{\bar{v}}{1 - A_{k+1}} + \frac{1}{1 - A_{k+1}} \frac{C_2}{\tilde{\mu}_s} \left[\sum_{i=s}^k \pi_i^{k+1} \times \right. \\ &\left. \left[\sum_{m=1}^{i+1-s} mr_m - (i+1-s) \sum_{m=i+2-s}^{\infty} r_m \right] + \right. \\ &\left. \left. + \pi_{k+1}^{k+1} \left[\sum_{m=1}^{k+1-s} mr_m - (k+1-s) \sum_{m=k+2-s}^{\infty} r_m \right] \right] \right]. \end{aligned}$$

Обозначив

$$\begin{aligned} F(k) &= \frac{1}{1 - A_{k+1}} \left[\sum_{i=s}^k \pi_i^{k+1} \times \right. \\ &\times \left[\sum_{m=1}^{i+1-s} mr_m - (i+1-s) \sum_{m=i+2-s}^{\infty} r_m \right] + \\ &\left. + \pi_{k+1}^{k+1} \left[\sum_{m=1}^{k+1-s} mr_m - (k+1-s) \sum_{m=k+2-s}^{\infty} r_m \right] \right]; \quad (10) \end{aligned}$$

$$G(k) = (C_0 - C_4 k) \bar{v} - C_4 \frac{\bar{v}}{1 - A_{k+1}} - \frac{C_2}{\tilde{\mu}_s} F(k),$$

из (8) и (9) получим

$$d(k) - d(k+1) = (1 - A_{k+1}) [d(k) - G(k)]. \quad (11)$$

Для $(1 - A_{k+1})F(k)$ из (10) после простых преобразований получим:

$$\begin{aligned} \frac{1}{\tilde{\mu}_s} (1 - A_{k+1}) F(k) &= \frac{1}{\tilde{\mu}_s} \sum_{i=s}^k \pi_i^{k+1} \sum_{m=1}^{i+1-s} m r_m + \\ &+ \frac{1}{\tilde{\mu}_s} \pi_{k+1}^{k+1} \sum_{m=1}^{k+1-s} m r_m + \\ &+ \frac{1}{\tilde{\mu}_s} \sum_{i=s}^k \pi_i^{k+1} (i+1-s) \sum_{m=i+2-s}^{\infty} r_m + \\ &+ \frac{1}{\tilde{\mu}_s} (k+1-s) \pi_{k+1}^{k+1} \sum_{m=k+2-s}^{\infty} r_m. \end{aligned}$$

Легко показать, что правая часть последнего равенства выражает среднее время ожидания поступившей заявки за период пребывания вложенной цепи в произвольно взятом состоянии. Тогда отсюда очевидно следует, что в рассматриваемой СМО с увеличением $k \geq s$ величина $(1 - A_{k+1})F(k)$ возрастает и, следовательно, $G(k)$ убывает, так как $0 < A_{k+1} < 1$ возрастает по k (см. [11]). Поскольку $G(k)$ — невозрастающая функция, то из (11) (см. [14]) следует, что $d(k)$ — унимодальная функция, причем при $C_2 = 0$ она всюду возрастает и ее точка максимума k^0 равна ∞ ($k^0 = \infty$), при $d(s) \geq G(s)$ всюду убывает и $k^0 = s$, в остальных случаях существует глобальный оптимум $s < k^0 < \infty$. Более того, из формулы (11), свойства монотонности функций $G(k)$ и A_{k+1} , а также неравенства $0 < A_{k+1} < 1$ следует, что $d(k)$ — выпуклая вверх функция.

Находим

$$\begin{aligned} d_{\text{пр}}(k) - d_{\text{пр}}(k+1) &= \\ &= \sum_{i=1}^s C_{3i} \sum_{j=0}^{i-1} \pi_j^k - \sum_{i=1}^s C_{3i} \sum_{j=0}^{i-1} \pi_j^{k+1} = \\ &= \sum_{i=1}^s C_{3i} \sum_{j=0}^{i-1} \frac{\lambda^j}{\prod_{i=1}^j \tilde{\mu}_s} (\pi_0^k - \pi_0^{k+1}) = \sum_{i=1}^s C_{3i} \times \\ &\times \sum_{j=0}^{i-1} \pi_0^k \frac{\lambda^j}{\prod_{i=1}^j \tilde{\mu}_s} \pi_0^{k+1} \frac{\lambda^s}{\prod_{i=1}^s \tilde{\mu}_i} \left(\frac{\lambda}{\tilde{\mu}_s} \right)^{k+1-s} = \\ &= \pi_{k+1}^{k+1} \sum_{i=1}^s C_{3i} \sum_{j=0}^{i-1} \pi_j^k = \pi_{k+1}^{k+1} \sum_{i=1}^s C_{3i} Q_{\text{пр},i}(k). \quad (12) \end{aligned}$$

Как следует из (12), функция $d_{\text{пр}}(k)$ — монотонно убывающая и выпуклая вниз функция, т.е. $D(k)$ — сумма двух выпуклых вверх функций (см. (4)).

Таким образом, доказано следующее утверждение.

Утверждение 1. *Функция $D(k)$ — выпуклая вверх функция, и решение задачи (3) удовлетворяет условиям:*

$$k^* = \begin{cases} \infty, & \text{если } C_2 = 0; \\ s, & \text{если } d_{\text{пр}}(s+1) - d_{\text{пр}}(s) < \\ & < \lambda[d(s) - d(s+1)], \end{cases}$$

иначе $0 < k^* < \infty$.

Отметим, что аналогичный результат имеет место и тогда, когда C_2 — штраф за превышение временем пребывания заявки в очереди допустимого предельного значения (дедлайна) $t_d > 0$, т.е. при замене в формуле (2) функции $W_{\text{ож}}(k)$ на $\Lambda_d(k)$ — интенсивность обслуженных заявок, время ожидания которых в очереди превышает t_d .

Обозначим показатель эффективности через $\tilde{D}(k)$:

$$\begin{aligned} \tilde{D}(k) &= C_0 \Lambda_{\text{вых}}(k) - C_1 \Lambda_{\text{отк}}(k) - C_2 \Lambda_d(k) - \\ &- \sum_{i=1}^s C_{3i} Q_{\text{пр},i}(k) - C_4 k. \end{aligned}$$

Рассмотрим вложенную цепь Маркова. Для краткости изложения положим $C_4 = 0$. Тогда имеет место равенство:

$$\tilde{D}(k) = \lambda \tilde{d}(k) + \sum_{i=1}^s C_{3i} Q_{\text{пр},i}(k). \quad (13)$$

Здесь

$$\tilde{d}(k) = \sum_{i=0}^k \pi_i^k \tilde{q}_i^k,$$

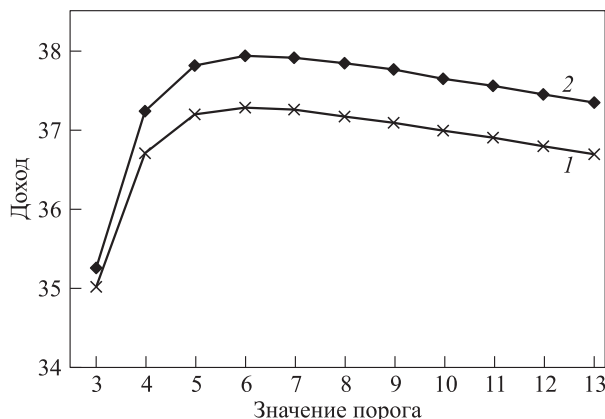
где

$$\tilde{q}_i^k = \begin{cases} C_0, & \text{если } 0 \leq i < s; \\ (C_0 + C_2) \Gamma_{\tilde{\mu}_s, i-s+1}(t_d) - C_2, & \text{если } s \leq i < k; \\ -C_1, & \text{если } i = k; \end{cases}$$

$\Gamma_{a,b}(\cdot)$ — функция гамма-распределения с параметрами a и b . Отсюда при $0 \leq i \leq k-2$ получим:

$$\tilde{q}_{i+1}^k = \begin{cases} \tilde{q}_i^k, & 0 \leq i \leq s-2; \\ \tilde{q}_i^k - (C_0 + C_2) \frac{(\tilde{\mu}_s t_d)^{i+1}}{(i+1)!} e^{-\tilde{\mu}_s t_d}, & s-1 \leq i \leq k-2. \end{cases}$$

Проведя такие же преобразования, как и выше для разности $d(k) - d(k+1)$ (см. также [12]), получим



Зависимость дохода системы от порогового значения длины очереди

$$\tilde{d}(k) - \tilde{d}(k + 1) = (1 - A_{k+1}) [\tilde{d}(k) - \tilde{G}(k)] ,$$

где

$$\tilde{G}(k) = -(C_0 + C_2) \frac{\pi_0^k}{1 - A_{k+1}} e^{-\tilde{\mu}_s t_n} \frac{\lambda^s}{\prod_{i=1}^s \tilde{\mu}_i} \times \sum_{j=s}^k \left(\frac{\lambda}{\tilde{\mu}_s} \right)^{j-s} \frac{(\tilde{\mu}_s t_n)^{j-s}}{(j-s)!} .$$

Рассуждая так же, как и в [12], можно доказать, что $\pi_0^k / (1 - A_{k+1})$ не зависит от k и $\tilde{G}(k)$ убывает по $k \geq s$. Таким образом, для $\tilde{d}(k)$ выполняются те же условия, что и для $d(k)$, из которых следовала выпуклость $d(k)$ по $k \geq s$, т.е. $\tilde{d}(k)$ — тоже выпуклая функция. Тогда из (13) получаем, что $\tilde{D}(k)$ — выпуклая функция по $k \geq s$ и для нее также справедливо утверждение 1.

Утверждение 1 справедливо также в случае однородных приборов с интенсивностью обслуживания $\mu_i = \mu, i = 1, \dots, s$, если $\Lambda_d(k)$ — интенсивность обслуженных заявок, время пребывания которых в системе превысило дедлайн. В этом случае имеем:

$$\tilde{q}_i^k = \begin{cases} (C_0 + C_2)\Gamma_{\mu,1}(t_n) - C_2, & \text{если } 0 \leq i < s; \\ (C_0 + C_2)\Gamma_{\tilde{\mu},i+1}(t_n) * \Gamma_{\mu,1}(t_n) - C_2, & \text{если } s \leq i < k; \\ -C_1, & \text{если } i = k; \end{cases}$$

$$\tilde{q}_{i+1}^k = \begin{cases} \tilde{q}_i^k, & 0 \leq i \leq s - 2; \\ \tilde{q}_i^k - (C_0 + C_2) [\Gamma_{\tilde{\mu}_s, i-s+1}(t_n) - \Gamma_{\tilde{\mu}_s, i-s+2}(t_n)] * \Gamma_{\mu,1}(t_n), & s - 1 \leq i \leq k - 2, \end{cases}$$

где * — знак свертки функций.

На рисунке проиллюстрированы примеры зависимости предельного дохода системы в единицу

времени от порогового значения длины очереди при параметрах $C_0 = 20, C_1 = 10, C_2 = 0,1, C_{31} = 1,5, C_{32} = 1, C_{33} = 0,5, C_4 = 0,05, \lambda = 2, \mu_1 = 1,5, \mu_2 = 1,0$ и $\mu_3 = 0,5$ (1) и $C_0 = 20, C_1 = 10, C_2 = 0,1, C_{31} = 1,5, C_{32} = 1, C_{33} = 0,5, C_4 = 0,05, \lambda = 6, \mu_1 = 3, \mu_2 = 2$ и $\mu_3 = 1$ (2, график сжат по оси ординат с коэффициентом 1/3).

4 Заключение

Полученный в данной работе результат обобщает результаты исследования задачи оптимизации стоимостной целевой функции на множестве простых пороговых стратегий в СМО с управляемой очередью, приведенных в работах [11, 12]. В настоящей работе использован тот же подход, что и в [11, 12], и доказано, что целевая функция удовлетворяет условиям теоремы из работы [14], где приведены достаточные условия унимодальности целевой функции. Отметим также, что приведенное в работе утверждение доказано с применением этого подхода и в рамках задачи, рассмотренной в работе [9] для СМО с нетерпеливыми заявками (заявка после поступления в очередь через экспоненциально распределенный интервал времени покидает систему, если она за это время не обслужилась, и за это система платит штраф), т.е. для этой задачи выводятся необходимые и достаточные условия конечности оптимального порога, из которых следует точное алгоритмическое ее решение.

В дальнейшем представляет интерес применение данного подхода к исследованию свойства унимодальности целевой функции для СМО с более общими параметрами и пороговой стратегией управления очередью.

Литература

1. *Клейнрок Л.* Вычислительные системы с очередями / Пер. с англ. — М.: Мир, 1979. 600 с. (*Kleinrock L.* Queueing systems. Vol. II: Computer applications. — New York, NY, USA: Wiley, 1976. 576 p.)
2. *Stidham Sh., Jr., Weber R.* A survey of Markov decision models for control of networks of queues // Queueing Syst., 1993. Vol. 13. P. 291–314.
3. *Рыков В. В.* Об условиях монотонности оптимальных политик управления системами массового обслуживания // Автоматика и телемеханика, 1999. № 9. С. 92–106.
4. *Rykov V., Efrosinin D.* Optimal control of queueing systems with heterogeneous servers // Queueing Syst., 2004. Vol. 46. P. 389–407.

5. *Efrosinin D.* Queueing model of a hybrid channel with faster link subject to partial and complete failures // *Ann. Oper. Res.*, 2013. Vol. 202. Iss. 1. P. 75–102.
6. *Özkan E., Kharoufeh J. P.* Optimal control of a two-server queueing system with failures // *Probab. Eng. Inform. Sc.*, 2014. Vol. 28. Iss. 4. P. 489–527.
7. *Ata B., Shneorson S.* Dynamic control of an $M/M/1$ service system with adjustable arrival and service rates // *Manage. Sci.*, 2006. Vol. 52. Iss. 11. P. 1778–1791.
8. *Ward A., Kumar S.* Asymptotically optimal admission control of a queue with impatient customers // *Math. Oper. Res.*, 2008. Vol. 33. Iss. 1. P. 167–202.
9. *Kocaga Y. L., Ward A. R.* Admission control for a multi-server queue with abandonment // *Queueing Syst.*, 2010. Vol. 65. Iss. 3. P. 275–323.
10. *Ghosh A. P., Weerasinghe A. P.* Optimal buffer size and dynamic rate control for a queueing system with impatient customers in heavy traffic // *Stoch. Proc. Appl.*, 2010. Vol. 120. Iss. 11. P. 2103–2141.
11. *Агаларов Я. М., Ушаков В. Г.* Об унимодальности функции дохода СМО типа $G/M/s$ с управляемой очередью // *Информатика и её применения*, 2019. Т. 13. Вып. 1. С. 55–61.
12. *Агаларов Я. М., Коновалов М. Г.* Доказательство унимодальности целевой функции в задаче порогового управления нагрузкой на сервер // *Информатика и её применения*, 2019. Т. 13. Вып. 2. С. 2–6.
13. *Бочаров П. П., Печинкин А. В.* Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
14. *Агаларов Я. М.* Признак унимодальности целочисленной функции одной переменной // *Обозрение прикладной и промышленной математики*, 2019. Т. 26. Вып. 1. С. 65–66.

Поступила в редакцию 01.08.2020

OPTIMAL THRESHOLD-BASED ADMISSION CONTROL IN THE $M/M/s$ SYSTEM WITH HETEROGENEOUS SERVERS AND A COMMON QUEUE

Ya. M. Agalarov

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

Abstract: The article discusses the $M/M/s$ system with heterogeneous servers and a common queue equipped with the mechanism to control the queue length in order to maximize the average marginal profit. The profit function includes a fee for successfully serviced customers, a fine for each rejected customer, a fine for idle period for each server, a fine for waiting (or for exceeding the allowable waiting time), and costs associated with queue maintenance. The problem is to maximize the marginal profit on a set of simple threshold-based queue length control policies. The property of convexity of the profit function is proved and conditions for existence of a finite optimal threshold of the queue length are obtained.

Keywords: queueing system; optimization; threshold strategy; queue length

DOI: 10.14357/19922264210108

Acknowledgments

The reported study was partly funded by RFBR, projects Nos. 18-07-00692, 19-07-00739, and 20-07-00804.

References

1. Kleinrock, L. 1976. *Queueing systems. Vol. II: Computer applications*. New York, NY: Wiley. 576 p.
2. Stidham, Sh., Jr., and R. Weber. 1993. A survey of Markov decision models for control of networks of queues. *Queueing Syst.* 13:291–314.
3. Rykov, V.V. 1999. On monotonicity conditions for optimal policies for the control of queueing systems. *Automat. Rem. Contr.* 60(9):1290–1301.
4. Rykov, V., and D. Efrosinin. 2004. Optimal control of queueing systems with heterogeneous servers. *Queueing Syst.* 46:389–407.
5. Efrosinin, D. 2013. Queueing model of a hybrid channel with faster link subject to partial and complete failures. *Ann. Oper. Res.* 202(1):75–102.
6. Özkan, E., and J. P. Kharoufeh. 2014. Optimal control of a two-server queueing system with failures. *Probab. Eng. Inform. Sc.* 28(4):489–527.

7. Ata, B., and S. Shneerson. 2006. Dynamic control of an $M/M/1$ service system with adjustable arrival and service rates. *Manage. Sci.* 52(11):1778–1791.
8. Ward, A., and S. Kumar. 2008. Asymptotically optimal admission control of a queue with impatient customers. *Math. Oper. Res.* 33(1):167–202.
9. Kocaga, Y. L., and A. R. Ward. 2010. Admission control for a multi-server queue with abandonment. *Queueing Syst.* 65(3):275–323.
10. Ghosh, A. P., and A. P. Weerasinghe. 2010. Optimal buffer size and dynamic rate control for a queueing system with impatient customers in heavy traffic. *Stoch. Proc. Appl.* 120(11):2103–2141.
11. Agalarov, Ya. M., and V. G. Ushakov. 2019. Ob unimodal'nosti funktsii dokhoda SMO tipa $G/M/s$ s upravlyaemoy ochered'yu [On the unimodality of the income function of a type $G/M/s$ queueing system with controlled queue]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):55–61.
12. Agalarov, Ya. M., and M. G. Konovalov. 2019. Dokazatel'stvo unimodal'nosti tselevoy funktsii v zadache porogovogo upravleniya nagruzkoy na server [Proof of the unimodality of the objective function in $M/M/N$ queue with threshold-based congestion control]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):2–6.
13. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
14. 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 Applied and Industrial Mathematics] 26(1):65–66.

Received August 1, 2020

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

ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ИНДЕКСА БАЛАНСА ФАКТОРОВ, ИМЕЮЩИХ ОБОБЩЕННЫЕ ГАММА-РАСПРЕДЕЛЕНИЯ*

Е. Н. Арутюнов¹, А. А. Кудрявцев², Ю. Н. Недоливко³

Аннотация: Приводятся основные вероятностные характеристики индекса баланса в байесовской постановке в предположении, что негативный и позитивный факторы имеют априорные обобщенные гамма-распределения. Формулировка задачи сводится к изучению характеристик масштабной смеси обобщенных гамма-законов. Особое внимание уделяется случаю, в котором распределения факторов имеют параметры формы противоположных знаков. Приводятся моментные характеристики и различные представления для плотности в терминах гамма-экспоненциальной функции, функций Фокса и Макдональда, а также обобщенной гипергеометрической функции. Метод анализа основан на применении преобразования Меллина и его обращении. Приводятся новые свойства гамма-экспоненциальной функции. Полученные результаты могут найти широкое применение в естественно-научных моделях, использующих для описания процессов и явлений распределения с положительным неограниченным носителем.

Ключевые слова: байесовский подход; обобщенное гамма-распределение; гамма-экспоненциальная функция; модели баланса; смешанные распределения; преобразование Меллина; функция Фокса; гипергеометрическая функция

DOI: 10.14357/19922264210109

1 Введение

В работе [1] подробно описывалась байесовская модель баланса, в которой один из основных показателей, так называемый индекс баланса ρ , представляет собой отношение двух случайных величин с заданными априорными распределениями: негативно влияющего на функционирование системы n -фактора λ к позитивному p -фактору μ .

В ряде предыдущих работ авторов рассматривались различные априорные распределения из гамма-класса. Основным требованием к изучаемым распределениям было ограничение на знаки параметров формы. Общий вид плотности индекса баланса факторов, имеющих распределения с параметрами формы одного знака, был сформулирован в [2] в терминах гамма-экспоненциальной функции [3]:

$$Ge_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta),$$

$$x \in \mathbb{R}, \quad 0 \leq \alpha < 1, \quad \beta > 0.$$

В предположении, что независимые n -фактор λ , имеющий обобщенное гамма-распреде-

ние, или распределение Крицкого–Менкеля [4, 5], $GG(v, q, \theta)$ с плотностью

$$f_{\lambda}(x) = \frac{|v|x^{vq-1}e^{-(x/\theta)^v}}{\theta^{vq}\Gamma(q)},$$

$$v \neq 0, \quad q > 0, \quad \theta > 0, \quad x > 0,$$

и p -фактор μ , имеющий распределение $GG(u, p, \alpha)$, удовлетворяют ограничению $uv > 0$ на параметры формы, справедливо следующее утверждение [2].

Теорема 1. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$, причем $uv > 0$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность

$$f_{\rho}(x) = \begin{cases} \frac{|v|\alpha^{vq}x^{vq-1}}{\theta^{vq}\Gamma(p)\Gamma(q)} Ge_{v/u, vq/u+p} \left(- \left(\frac{\alpha x}{\theta} \right)^v \right), & |u| > |v|; \\ \frac{|u|\theta^{up}x^{-up-1}}{\alpha^{up}\Gamma(p)\Gamma(q)} Ge_{u/v, up/v+q} \left(- \left(\frac{\alpha x}{\theta} \right)^{-u} \right), & |v| > |u|; \\ \frac{|v|(\alpha/\theta)^{vq}x^{vq-1}}{B(p, q)(1 + (\alpha x/\theta)^v)^{p+q}}, & u = v. \end{cases}$$

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 20-07-00655).

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

² Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, nubigena@mail.ru

³ Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, mouse98@mail.ru

Утверждение теоремы 1 может быть легко переформулировано для масштабной смеси обобщенных гамма-распределений с параметрами формы противоположных знаков, поскольку если случайная величина ξ имеет распределение $GG(v, q, \theta)$, то случайная величина $1/\xi$ имеет распределение $GG(-v, q, 1/\theta)$.

Заметим, что при $uv < 0$ нахождение плотности ρ напрямую сопряжено с принципиальными трудностями вычисления интеграла $\int_0^\infty y^{r-1} e^{-(y/\alpha)^u - (y/\theta)^v} dy$.

Далее описывается метод получения вероятностных характеристик индекса баланса ρ , основанный на обращении преобразования Меллина.

2 Основные результаты

Пусть ξ — неотрицательная случайная величина с функцией распределения $F_\xi(x)$. Преобразование Меллина случайной величины ξ определяется [6, 7] следующим образом:

$$M_\xi(s) = \int_0^\infty x^s dF_\xi(x), \quad s \in \mathbb{C}; \quad (1)$$

$M_\xi(s)$ предполагается конечным; величина 0^s предполагается равной нулю для всех s .

Поскольку далее рассматривается преобразование Меллина смесей обобщенных гамма-распределений, имеющих неотрицательный носитель, ограничимся определением (1), допускающим обобщение на случай произвольных случайных величин.

Приведем ряд известных свойств преобразования Меллина неотрицательных случайных величин [6, 7].

Лемма 1. 1. Интеграл (1) существует для всех значений параметра s из некоторой полосы $\mathcal{D}_\xi = \{s : \sigma_{\xi,1} \leq \text{Re}(s) \leq \sigma_{\xi,2}\}$, содержащей мнимую ось i , возможно, вырождающейся в эту ось.

2. Преобразование Меллина взаимно однозначно определяет распределение случайной величины.

3. Пусть ξ, η — независимые случайные величины. Тогда

$$M_{\xi\eta}(s) = M_\xi(s)M_\eta(s), \quad s \in \mathcal{D}_\xi \cap \mathcal{D}_\eta.$$

4. Для точек непрерывности плотности распределения случайной величины ξ справедлива следующая формула:

$$f_\xi(x) = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{c-iT}^{c+iT} x^{-s-1} M_\xi(s) ds, \quad x > 0,$$

где линия $(c - i\infty, c + i\infty)$ лежит в области аналитичности $M_\xi(s)$.

5. Для точек непрерывности $a, b, x > 0$ функции распределения случайной величины ξ справедливы формулы:

$$\begin{aligned} F_\xi(b) - F_\xi(a) &= \\ &= \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{-U}^U \frac{e^{-it \ln a} - e^{-it \ln b}}{it} M_\xi(t) dt; \\ F_\xi(x) &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \text{Im} \{ M_\xi(it) x^{-it} \} \frac{dt}{t}. \end{aligned}$$

Непосредственно из леммы 1 вытекают свойства преобразования Меллина обобщенного гамма-распределения и его мультипликативных смесей.

Лемма 2. 1. Если случайная величина λ имеет обобщенное гамма-распределение $GG(v, q, \theta)$, то

$$M_\lambda(s) = \frac{\theta^s}{\Gamma(q)} \Gamma\left(q + \frac{s}{v}\right), \quad q + \frac{\text{Re}(s)}{v} > 0.$$

2. Если независимые случайные величины λ и μ имеют распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$ соответственно, то

$$\begin{aligned} M_\rho(s) &= \frac{(\theta/\alpha)^s}{\Gamma(q)\Gamma(p)} \Gamma\left(q + \frac{s}{v}\right) \Gamma\left(p - \frac{s}{u}\right), \\ & \quad q + \frac{\text{Re}(s)}{v} > 0, \quad p - \frac{\text{Re}(s)}{u} > 0, \end{aligned}$$

причем $M_\rho(s)$ — аналитическая функция в области \mathcal{D}_ρ .

Непосредственно из леммы 2 следуют утверждения о моментных характеристиках и функции распределения отношения двух случайных величин, имеющих обобщенное гамма-распределение.

Теорема 2. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$. Тогда для случайной величины $\rho = \lambda/\mu$

$$\begin{aligned} E\rho^k &= \frac{(\theta/\alpha)^k}{\Gamma(q)\Gamma(p)} \Gamma\left(q + \frac{k}{v}\right) \Gamma\left(p - \frac{k}{u}\right), \\ & \quad q + \frac{k}{v} > 0, \quad p - \frac{k}{u} > 0; \end{aligned}$$

$$E e^{it \ln \rho} = \frac{(\theta/\alpha)^{it}}{\Gamma(q)\Gamma(p)} \Gamma\left(q + \frac{it}{v}\right) \Gamma\left(p - \frac{it}{u}\right), \quad t \in \mathbb{R}.$$

Теорема 3. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет функцию распределения

$$F_\rho(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \operatorname{Im} \left\{ \frac{(\theta/\alpha)^{it}}{\Gamma(q)\Gamma(p)} \Gamma\left(q + \frac{it}{v}\right) \Gamma\left(p - \frac{it}{u}\right) x^{-it} \right\} \frac{dt}{t}.$$

Найдем представление плотности случайной величины ρ . Для этого рассмотрим H -функцию Фокса [8]

$$H_{kl}^{mn} \left[z \left| \begin{matrix} (a_1, A_1), \dots, (a_k, A_k) \\ (b_1, B_1), \dots, (b_l, B_l) \end{matrix} \right. \right] = \frac{1}{2\pi i} \times \int_L \frac{\prod_{j=1}^m \Gamma(b_j + B_j s) \prod_{j=1}^n \Gamma(a_j - A_j s)}{\prod_{j=n+1}^k \Gamma(a_j + A_j s) \prod_{j=m+1}^l \Gamma(b_j - B_j s)} z^{-s} ds,$$

где $0 \leq m \leq l, 0 \leq n \leq k, A_{j_1}, B_{j_2} > 0, j_1 = 1, \dots, k, j_2 = 1, \dots, l$. Везде далее в качестве контура L рассматривается область интегрирования $(c - i\infty, c + i\infty) \subset \mathcal{D}_\rho$.

Теорема 4. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность

$$f_\rho(x) = \frac{x^{-1}}{\Gamma(q)\Gamma(p)} \times \begin{cases} H_{11}^{11} \left[\frac{\alpha x}{\theta} \left| \begin{matrix} \left(p, \frac{1}{u}\right) \\ \left(q, \frac{1}{v}\right) \end{matrix} \right. \right], & u > 0, v > 0; \\ H_{02}^{20} \left[\frac{\alpha x}{\theta} \left| \begin{matrix} \left(q, \frac{1}{v}\right), \left(p, -\frac{1}{u}\right) \\ - \end{matrix} \right. \right], & u < 0, v > 0; \\ H_{20}^{02} \left[\frac{\alpha x}{\theta} \left| \begin{matrix} \left(q, -\frac{1}{v}\right), \left(p, \frac{1}{u}\right) \\ - \end{matrix} \right. \right], & u > 0, v < 0; \\ H_{11}^{11} \left[\frac{\alpha x}{\theta} \left| \begin{matrix} \left(q, -\frac{1}{v}\right) \\ \left(p, -\frac{1}{u}\right) \end{matrix} \right. \right], & u < 0, v < 0. \end{cases}$$

Доказательство. Согласно п. 4 леммы 1 и п. 2 леммы 2

$$f_\rho(x) = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{c-iT}^{c+iT} x^{-s-1} \mathcal{M}_\rho(s) ds = \frac{x^{-1}}{2\pi i \Gamma(q)\Gamma(p)} \times \lim_{T \rightarrow \infty} \int_{c-iT}^{c+iT} \Gamma\left(q + \frac{s}{v}\right) \Gamma\left(p - \frac{s}{u}\right) \left(\frac{\alpha x}{\theta}\right)^{-s} ds,$$

откуда следует утверждение теоремы.

Теоремы 1 и 4 дают возможность представить гамма-экспоненциальную функцию через H -функцию Фокса.

Следствие 1. Для всех $0 < \alpha < 1$ и $\beta > 0$ имеет место равенство

$$Ge_{\alpha, \beta}(-x) = H_{11}^{11} \left[x \left| \begin{matrix} (\beta, \alpha) \\ (0, 1) \end{matrix} \right. \right], \quad x > 0.$$

Вид плотности случайной величины ρ допускает альтернативное представление в случае, когда $u = -v$. Рассмотрим функцию Макдональда (модифицированную функцию Бесселя третьего рода) [9], которую можно представить в виде:

$$K_\nu(z) = \int_0^\infty e^{-z \operatorname{ch} t} \operatorname{ch}(\nu t) dt, \quad |\arg(z)| < \frac{\pi}{2},$$

и функцию Мейера [8]

$$G_{kl}^{mn} \left[z \left| \begin{matrix} a_1, \dots, a_k \\ b_1, \dots, b_l \end{matrix} \right. \right] = \frac{1}{2\pi i} \times \int_L \frac{\prod_{j=1}^m \Gamma(b_j + s) \prod_{j=1}^n \Gamma(a_j - s)}{\prod_{j=n+1}^k \Gamma(a_j + s) \prod_{j=m+1}^l \Gamma(b_j - s)} z^{-s} ds,$$

где $0 \leq m \leq l, 0 \leq n \leq k$.

Следствие 2. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(-v, p, \alpha)$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность:

$$f_\rho(x) = \frac{2|v|x^{(vp+vq)/2-1}}{(\theta/\alpha)^{(vp+vq)/2} \Gamma(p)\Gamma(q)} K_{p-q} \left(2\sqrt{\left(\frac{\alpha x}{\theta}\right)^v} \right).$$

Доказательство. Утверждение следствия 2 вытекает из соотношений [8]:

$$H_{kl}^{mn} \left[z \left| \begin{matrix} (a_1, 1), \dots, (a_k, 1) \\ (b_1, 1), \dots, (b_l, 1) \end{matrix} \right. \right] = G_{kl}^{mn} \left[z \left| \begin{matrix} a_1, \dots, a_k \\ b_1, \dots, b_l \end{matrix} \right. \right];$$

$$G_{kl}^{mn} \left[z \left| \begin{matrix} a_1 - \alpha, \dots, a_n - \alpha, a_{n+1} + \alpha, \dots, a_k + \alpha \\ b_1 + \alpha, \dots, b_m + \alpha, b_{m+1} - \alpha, \dots, b_l - \alpha \end{matrix} \right. \right] = z^\alpha G_{kl}^{mn} \left[z \left| \begin{matrix} a_1, \dots, a_k \\ b_1, \dots, b_l \end{matrix} \right. \right];$$

$$K_\nu(2\sqrt{x}) = \frac{1}{2} G_{02}^{20} \left[x \left| \begin{matrix} - \\ \frac{\nu}{2}, -\frac{\nu}{2} \end{matrix} \right. \right];$$

$$K_\nu\left(\frac{2}{\sqrt{x}}\right) = \frac{1}{2} G_{20}^{02} \left[x \left| \begin{matrix} \frac{\nu}{2}, -\frac{\nu}{2} \\ - \end{matrix} \right. \right]$$

и свойства [9] $K_\nu(x) = K_{-\nu}(x)$.

Также утверждение следствия 2 может быть получено напрямую, как показано в [10] для $v > 0$.

Для некоторых значений параметров формы априорных распределений факторов λ и μ может быть получен альтернативный вид плотности индекса баланса ρ при помощи гипергеометрической функции.

Через

$$(\alpha)_i = \alpha(\alpha + 1) \cdots (\alpha + i - 1), (\alpha)_0 = 1, \alpha \in \mathbb{R},$$

будем обозначать символ Похгаммера. Рассмотрим обобщенную гипергеометрическую функцию [11]

$${}_mF_n \left[\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix}; z \right] = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \cdots (a_m)_k}{(b_1)_k (b_2)_k \cdots (b_n)_k} \frac{z^k}{k!}.$$

Обозначим

$$\Delta(k, a) = \frac{a}{k}, \frac{a+1}{k}, \dots, \frac{a+k-1}{k};$$

$$S_\rho(A, B, C; x) = \sum_{n=0}^{A-1} \frac{x^n}{n!} \Gamma\left(C - \frac{B}{A}n\right) \times \\ \times {}_1F_{A+B} \left[\begin{matrix} 1; \Delta(A, 1+n), \\ \Delta\left(B, 1-C + \frac{B}{A}n\right) \end{matrix}; \left(\frac{(-1)^B}{A^A B^B} x^A\right) \right],$$

$$x \in \mathbb{R}, \quad (2)$$

где A и B — взаимно простые натуральные числа, а $C > 0$.

Теорема 5. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$, причем $uv < 0$ и для некоторых $\gamma > 0$ и взаимно простых натуральных чисел ad и bc выполнены равенства $|v| = \gamma a/b$ и $|u| = \gamma c/d$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность:

$$f_\rho(x) = \frac{x^{-1}}{\Gamma(p)\Gamma(q)} \times \\ \times \left[|v| \left(\frac{\alpha x}{\theta}\right)^{vq} S_\rho\left(bc, ad, p - \frac{ad}{bc}q; -\left(\frac{\alpha x}{\theta}\right)^v\right) + \right. \\ \left. + |u| \left(\frac{\alpha x}{\theta}\right)^{-up} S_\rho\left(ad, bc, q - \frac{bc}{ad}p; -\left(\frac{\alpha x}{\theta}\right)^{-u}\right) \right],$$

$$pbc \neq qad.$$

Доказательство. Без ограничения общности будем считать числа a, b, c и d положительными.

Заметим [11], что для взаимно простых натуральных чисел m и n и положительных чисел δ, t и s

$$\int_0^\infty x^{\delta-1} \exp\{-tx^{-m/n} - sx\} dx = \\ = s^{-\delta} \sum_{l=0}^{n-1} \frac{(-1)^l}{l!} (ts^{m/n})^l \Gamma\left(\delta - \frac{m}{n}l\right) \times \\ \times {}_1F_{m+n} \left[\begin{matrix} 1; \Delta(n, 1+l), \\ \Delta\left(m, 1-\delta + \frac{ml}{n}\right) \end{matrix}; z \right] + \\ + t^{\delta n/m} \frac{n}{m} \sum_{k=0}^{m-1} \frac{(-1)^k}{k!} (t^{n/m}s)^k \Gamma\left(-\frac{\delta+k}{m/n}\right) \times \\ \times {}_1F_{m+n} \left[\begin{matrix} 1; \Delta(m, 1+k), \\ \Delta\left(n, 1+\frac{\delta+k}{n/m}\right) \end{matrix}; z \right], \quad (3)$$

где

$$z = (-1)^{m+n} \left(\frac{s}{m}\right)^m \left(\frac{t}{n}\right)^n.$$

Для плотности ρ имеем:

$$f_\rho(x) = \frac{|uv|x^{vq-1}}{\theta^{vq}\alpha^{up}\Gamma(p)\Gamma(q)} \times \\ \times \int_0^\infty y^{-up-vq-1} \exp\left\{-\left(\frac{x}{\theta}\right)^v y^{-v} - (\alpha y)^{-u}\right\} dy.$$

В случае $pbc > qad$ получаем:

$$f_\rho(x) = \frac{|v|x^{vq-1}}{\theta^{vq}\alpha^{up}\Gamma(p)\Gamma(q)} \times \\ \times \int_0^\infty z^{p+qv/u-1} \exp\left\{-\left(\frac{x}{\theta}\right)^v z^{-ad/(bc)} - \alpha^{-u}z\right\} dz.$$

Используя формулу (3) и обозначение (2), получим утверждение теоремы.

Аналогично рассматривается случай $qad > pbc$. Теорема доказана.

Введем обозначение

$$T_\rho(A, B, C; x) = \sum_{k=0}^{A-1} \frac{x^k}{k!} \Gamma\left(C + k\frac{B}{A}\right) \times \\ \times {}_{B+1}F_A \left[\begin{matrix} 1, \Delta\left(B, C + \frac{kB}{A}\right) \\ \Delta(A, 1+k) \end{matrix}; \left(\frac{B^B}{A^A} x^A\right) \right],$$

$$x \in \mathbb{R}, \quad (4)$$

где A и B — взаимно простые натуральные числа, а $C > 0$.

Теорема 6. Пусть независимые случайные величины λ и μ имеют соответственно распределения $GG(v, q, \theta)$ и $GG(u, p, \alpha)$, причем $uv > 0$ и для некоторых $\gamma > 0$

и взаимно простых натуральных чисел ad и bc выполнены равенства $|v| = \gamma a/b$ и $|u| = \gamma c/d$. Тогда их отношение $\rho = \lambda/\mu$ при $x > 0$ имеет плотность:

$$f_\rho(x) = \begin{cases} \frac{|v|\alpha^{vq}x^{vq-1}}{\theta^{vq}\Gamma(p)\Gamma(q)} T_\rho\left(bc, ad, \frac{vq}{u} + p; -\left(\frac{\alpha x}{\theta}\right)^v\right), & |v| < |u|; \\ \frac{|u|\alpha^{up}x^{-up-1}}{\theta^{up}\Gamma(p)\Gamma(q)} T_\rho\left(ad, bc, \frac{up}{v} + q; -\left(\frac{\alpha x}{\theta}\right)^{-u}\right), & |v| > |u|. \end{cases}$$

Доказательство. Заметим, что для

$$U = \int_0^\infty x^{\delta-1} e^{-tx^r - sx} dx,$$

где $t, s > 0, r = m/n$, а m и n — взаимно простые натуральные числа, справедливы соотношения [11]:

$$U = s^{-\delta} \sum_{k=0}^{n-1} \frac{(-ts^{-r})^k}{k!} \Gamma(\delta + rk) \times \\ \times {}_{m+1}F_n \left[\begin{matrix} 1, \Delta(m, \delta + rk); \\ \Delta(n, 1 + k) \end{matrix} \right] (-1)^n z, \\ 0 < r < 1; \quad (5)$$

$$U = \frac{t^{-\delta/r}}{r} \sum_{k=0}^{m-1} \frac{(-t^{-1/r} s)^k}{k!} \Gamma\left(\frac{\delta + k}{r}\right) \times \\ \times {}_{n+1}F_m \left[\begin{matrix} 1, \Delta\left(n, \frac{\delta + k}{r}\right); \\ \Delta(m, 1 + k) \end{matrix} \right] (-1)^m z^{-1}, \\ r > 1, \quad (6)$$

где

$$z = \left(\frac{m}{s}\right)^m \left(\frac{t}{n}\right)^n.$$

Аналогично доказательству теоремы 5, используя для плотности $f_\rho(x)$ преобразование (5) в случае $|u| < |v|$, формулу (6) в случае $|v| < |u|$ и обозначение (4), получаем утверждение теоремы.

Теоремы 1 и 6 дают возможность выявить еще одно свойство гамма-экспоненциальной функции.

Следствие 3. Для взаимно простых натуральных $0 < c < m < n$, действительных $c > 0$ и $x > 0$ выполнено соотношение:

$$Ge_{m/n, c}(-x) = T_\rho(n, m, c; -x).$$

3 Заключение

Данная работа завершает этап описания основных вероятностных характеристик индекса баланса факторов, имеющих обобщенные гамма-распределения. Результаты формулируются в терминах специальных гамма-функции, гамма-экспоненциальной функции, функций Фокса и Макдональда, а также обобщенной гипергеометрической функции. Несмотря на кажущуюся громоздкость результатов, полученные соотношения на практике дают существенное увеличение скорости вычислений характеристик при конкретных значениях параметров по сравнению с численными методами.

Полученные результаты могут быть легко переформулированы для масштабных смесей обобщенных гамма-распределений и могут найти широкое применение в естественно-научных моделях, использующих для описания процессов и явлений распределения с положительным неограниченным носителем.

Литература

1. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
2. Кудрявцев А. А. Априорное обобщенное гамма-распределение в байесовских моделях баланса // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 27–33.
3. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
4. Крицкий С. Н., Менкель М. Ф. О приемах исследования случайных колебаний речного стока // Труды НИУ ГУГМС. Сер. IV, 1946. Вып. 29. С. 3–32.
5. Крицкий С. Н., Менкель М. Ф. Выбор кривых распределения вероятностей для расчетов речного стока // Известия АН СССР. Отд. техн. наук, 1948. № 6. С. 15–21.
6. Золотарев В. М. Преобразования Меллина–Стилтьеса в теории вероятностей // Теория вероятностей и её применения, 1957. Т. 2. Вып. 4. С. 444–469.
7. Galambos J., Simonelli I. Products of random variables: Applications to problems of physics and to arithmetical functions. — New York, NY, USA: Marcel Dekker, Inc., 2004. 315 p.
8. Прудников А. П., Брычков Ю. А., Маричев О. И. Интегралы и ряды: в 3 т. Т. 3: Специальные функции. Дополнительные главы. — 2-е изд., испр. — М.: Физматлит, 2003. 688 с.
9. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. — М.: Наука, 1971. 1108 с.

10. Ayianda K. C. Gamma and related distributions. — Nairobi, Kenya: School of Mathematics, University of Nairobi, 2013. PhD Thesis. 148 p.
11. Прудников А. П., Брычков Ю. А., Маричев О. И. Интегралы и ряды: в 3 т. Т. 1: Элементарные функции. — 2-е изд., испр. — М.: Физматлит, 2002. 632 с.

Поступила в редакцию 18.06.2020

PROBABILISTIC CHARACTERISTICS OF BALANCE INDEX OF FACTORS WITH GENERALIZED GAMMA DISTRIBUTION

E. N. Arutyunov¹, A. A. Kudryavtsev², and Iu. N. Nedolivko²

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

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: The main probabilistic characteristics of balance index in Bayesian formulation, assuming that negative and positive factors have *a priori* generalized gamma distribution, are given. The formulation of this problem is equivalent to a study of generalized gamma laws scale mixture characteristics. Special attention is paid to the case in which the factors distributions have shape parameters of different signs. Moment characteristics and different presentation of density in terms of gamma-exponential function, H-function, Macdonald function, and generalized hypergeometric function are given. The analysis method is based on Mellin transform and its inverse transform. New properties of gamma-exponential function are given. The obtained results can be widely applied within the natural science models that use distributions with positive unlimited support to describe processes and phenomena.

Keywords: Bayesian approach; generalized gamma distribution; gamma-exponential function; balance models; Mellin transform; H-function; hypergeometric function

DOI: 10.14357/19922264210109

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 20-07-00655).

References

1. Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
2. Kudryavtsev, A. A. 2019. Apriornoe obobshchennoe gamma-raspredelenie v bayesovskikh modelyakh balansa [A priori generalized gamma distribution in Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):27–33.
3. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
4. Kritsky, S. N., and M. F. Menkel. 1946. O priemakh issledovaniya sluchaynykh kolebaniy rechnogo stoka [Methods of investigation of random fluctuations of river flow]. *Trudy NIU GUGMS. Ser. IV* [Proceedings of GUGMS Research Institutions. Ser. IV] 29:3–32.
5. Kritsky, S. N., and M. F. Menkel. 1948. Vybor krivykh raspredeleniya veroyatnostey dlya raschetov rechnogo stoka [Selection of probability distribution curves for river flow calculations]. *Izvestiya AN SSSR. Otd. tekhn. nauk* [Herald of the Russian Academy of Sciences. Technical Sciences] 6:15–21.
6. Zolotarev, V. M. 1957. Mellin–Stieltjes transformations in probability theory. *Theor. Probab. Appl.* 2(4):433–460.
7. Galambos, J., and I. Simonelli. 2004. *Products of random variables: Applications to problems of physics and to arithmetical functions*. New York, NY: Marcel Dekker, Inc. 315 p.
8. Prudnikov, A. P., Yu. A. Brychkov, and O. I. Marichev. 2003. *Integraly i ryady: v 3 t. T. 3. Spetsial'nye funktsii. Dopolnitel'nye glavy* [Integrals and series. In 3 vols. Vol. 3. Special functions. Additional chapters]. 2nd ed. Moscow: Fizmatlit. 688 p.
9. Gradshteyn, I. S., and I. M. Ryzhik. 1971. *Tablitsy integralov, summ, ryadov i proizvedeniy* [Tables of integrals, sums, series, and products]. Moscow: Nauka. 1108 p.
10. Ayianda, K. C. 2013. Gamma and related distributions. Nairobi, Kenya: School of Mathematics, University of Nairobi. PhD Thesis. 148 p.

11. Prudnikov, A. P., Yu. A. Brychkov, and O. I. Marichev. 2002. *Integraly i ryady*: v 3 t. T. 1. *Elementarnye funktsii* [Integrals and series: In 3 vols. Vol. 3. Elementary functions]. 2nd ed. Moscow: Fizmatlit. 632 p.

Received June 18, 2020

Contributors

Arutyunov Evgeny N. (b. 1952) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; enapoleon@mail.ru

Kudryavtsev Alexey A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

Nedolivko Iuliia N. (b. 1998) — student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; mouse98@mail.ru

НЕАСИМПТОТИЧЕСКИЙ АНАЛИЗ СТАТИСТИКИ БАРТЛЕТТА–НАНДА–ПИЛАЯ ДЛЯ ДАННЫХ БОЛЬШОЙ РАЗМЕРНОСТИ

А. А. Липатев¹

Аннотация: Представлены вычислимые оценки скорости сходимости нормированной статистики Бартлетта–Нанда–Пилая к стандартному нормальному распределению при условии, что размерность данных возрастает пропорционально объему выборки. Приведенный результат позволяет корректно вычислять p -значения в прикладных задачах многомерного анализа данных. Задачи в постановке, когда число анализируемых признаков сравнимо с объемом выборки, все чаще возникают в области обработки сигналов. Доказательство базируется существенным образом на нормальности распределения элементов рассматриваемых матриц с распределением Уишарта. Для случайных величин, представляющих собой матричные следы произведения и квадратов матриц с нормированным распределением Уишарта, находятся удобные оценки сверху для $1 - F$, где F — функция распределения соответствующего следа матрицы. Применяя свойства обратных матриц и неотрицательно определенных матриц, статистика Бартлетта–Нанда–Пилая ограничивается сверху комбинацией из упомянутых выше следов матриц.

Ключевые слова: точность приближений; многомерный дисперсионный анализ; вычислимые оценки; статистика Бартлетта–Нанда–Пилая; данные большой размерности

DOI: 10.14357/19922264210110

1 Введение

В большом числе прикладных задач исследователи анализируют многомерные данные, в которых количество p признаков сравнимо с числом n наблюдений. Для анализа данных фиксированной размерности существует множество статистических процедур, уже ставших классическими. Однако часто нет возможности использовать традиционную статистическую процедуру, лишь устремив в ней число признаков к бесконечности, так как при этом изменяется предельное распределение статистики критерия (см. [1, разд. 6.3.4]).

Цель данной работы — нахождение вычислимых оценок точности аппроксимации статистики Бартлетта–Нанда–Пилая (Bartlett–Nanda–Pillai test) нормальным распределением в модели многомерного дисперсионного анализа (MANOVA — multivariate analysis of variance) для данных большой размерности, когда отношение числа признаков к числу наблюдений p/n стремится к некоторой константе из интервала $(0, 1)$.

Результаты, касающиеся распределений статистик, возникающих в модели MANOVA при условии, что нулевая гипотеза верна, оказываются полезны в области обработки сигналов. Например, в [2] показано, каким образом результаты из MANOVA и обработки сигналов могут сво-

диться к спектру определенной матрицы $E^{-1}H$. В статье [3] приводится пример применения статистики Лоули–Хотеллинга, родственной статистике Бартлетта–Нанда–Пилая, в контексте обработки данных радаров с синтезированной апертурой.

В разд. 2 сформулирован основной результат работы — теорема 1. Теорема 2 является вспомогательной, но при этом представляет самостоятельный интерес. В разд. 4 даны доказательства основных теорем, которые опираются на леммы из разд. 3.

2 Постановка задачи и основной результат

В рамках многомерного дисперсионного анализа исследуется следующая многомерная линейная модель: $X = Q\mathbb{B} + \mathcal{E}$, где X — случайная матрица наблюдений размера $N \times p$; Q — неслучайная матрица плана эксперимента размера $N \times k$; \mathbb{B} — неслучайная матрица $k \times p$ регрессионных коэффициентов; \mathcal{E} — матрица ошибок $N \times p$ с распределением $N_{N \times p}(O, I_N \otimes \Sigma)$.

Рассмотрим следующую линейную гипотезу: $H_0 : C\mathbb{B} = O$, где C — известная матрица размера $q \times k$ ранга q . Статистики критериев, инвариантные относительно некоторой группы аффинных

¹Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, кафедра математической статистики, allipatev@cs.msu.ru

преобразований, оказываются функциями от ненулевых собственных значений матрицы $S_h S_e^{-1}$, где

$$\left. \begin{aligned} S_h &= \hat{\mathbb{B}}^T C^T \left(C (Q^T Q)^{-1} C^T \right)^{-1} C \hat{\mathbb{B}}; \\ S_e &= (X - Q \hat{\mathbb{B}})^T (X - Q \hat{\mathbb{B}}) \end{aligned} \right\} \quad (1)$$

при $\hat{\mathbb{B}} = (Q^T Q)^{-1} Q^T X$ (см. [4, гл. 8]). Одной из наиболее известных инвариантных статистик является статистика Бартлетта–Нанда–Пилая: $V_{\text{BNP}} = (n + q) \text{tr} S_h (S_h + S_e)^{-1}$. В дальнейшем предполагаем, что гипотеза H_0 верна.

В [5] рассмотрен случай большого объема выборки, т. е. выполнено условие **A1**:

$$\mathbf{A1} : p \text{ и } q \text{ фиксированы, } n \rightarrow \infty,$$

и получены неасимптотические оценки точности аппроксимации функции распределения статистики Бартлетта–Нанда–Пилая:

$$\begin{aligned} \mathbf{P}\{V_{\text{BNP}} < x\} &= G_a(x) + \frac{3a}{4n} \{G_a(x) - \\ &- 2G_{a+2}(x) + G_{a+4}(x)\} + O(n^{-2}), \end{aligned}$$

где $a = pq$; G_a — функция χ^2 -распределения с a степенями свободы. В [6] для остаточного члена найдена оценка сверху.

В [7] рассмотрен случай большой размерности данных, т. е. выполнено условие **A2**:

$$\begin{aligned} \mathbf{A2} : q \text{ фиксировано, } p \rightarrow \infty, n \rightarrow \infty, \\ \frac{p}{n} \rightarrow c \in (0; 1), \end{aligned}$$

и получено следующее приближение:

$$\begin{aligned} \mathbf{P}\left(\frac{1}{\sigma} T_{\text{BNP}} < z\right) &= \\ &= \Phi(z) - \phi(z) \left[\frac{1}{\sqrt{p}} \left\{ \frac{1}{\sigma} b_1 + \frac{1}{\sigma^3} b_3 H_2(z) \right\} + \right. \\ &+ \left. \frac{1}{p} \left\{ \frac{1}{\sigma^2} b_2 H_1(z) + \frac{1}{\sigma^4} b_4 H_3(z) + \frac{1}{\sigma^6} b_6 H_5(z) \right\} \right] + \\ &+ O\left(\frac{1}{p\sqrt{p}}\right), \end{aligned}$$

где $T_{\text{BNP}} = \sqrt{p}(1 + m^{-1}p) \{p^{-1}V_{\text{BNP}} - q\}$; $\Phi(z)$ и $\phi(z)$ — соответственно функция распределения и плотность распределения стандартного нормального закона; $m = n - p + q$; $r = p/m$; $\sigma = \sqrt{2q(1+r)}$; $b_i = b_i(r, q)$ суть некоторые функции от r и q ; $H_i(z)$ — полиномы Эрмита. При этом результат имел именно асимптотический вид, верхние оценки остаточного члена не находились.

Основной результат данной работы — две теоремы, дающие оценку точности аппроксимации распределения статистики Бартлетта–Нанда–Пилая нормальным распределением для данных большой размерности, т. е. при выполнении условия **A2**:

Теорема 1. При всех $m > M = M(r, q)$ справедливо неравенство

$$\sup_z \left| \mathbf{P}\left(\frac{T_{\text{BNP}}}{\sqrt{2q(1+r)}} < z\right) - \Phi(z) \right| \leq \frac{K_2(r, q) \ln m}{\sqrt{m}},$$

где $K_2(r, q)$ — вычисляемая функция от r и q .

Отметим, что результат теоремы 1 на логарифмический множитель уступает результату из [7], но превосходит последний в том, что для ошибки погрешности дается вычисляемая оценка сверху. При этом само доказательство является новым.

Теорема 2. Пусть матрицы U и V суть нормированные варианты матриц B и W :

$$U = \frac{B - pI_q}{\sqrt{p}}; \quad V = \frac{W - mI_q}{\sqrt{m}}, \quad (2)$$

где B и W независимы и имеют распределения Уишарта $W_q(p, I_q)$ и $W_q(m, I_q)$ с $m = n - p + q$ соответственно. Если $\text{tr}(\sqrt{r}U + V)^2 < (r + 1)m$, то выполнено следующее неравенство:

$$\begin{aligned} & \left| \sqrt{m}(r + 1) \left((r + 1) \text{tr} B (B + W)^{-1} - rq \right) - \right. \\ & \quad \left. - (\sqrt{r} \text{tr} U - r \text{tr} V) \right| \leq \\ & \leq \frac{(r + 1) \sqrt{r} (|\text{tr} UV| + \sqrt{r} \text{tr} U^2)}{(r + 1) \sqrt{m} - \text{tr}(\sqrt{r}U + V)^2 / \sqrt{m}} + \\ & + \left(rq(r + 1) + \frac{(\sqrt{r} \text{tr} U - r \text{tr} V)}{\sqrt{m}} \right) \times \\ & \quad \times \frac{\text{tr}(\sqrt{r}U + V)^2}{(r + 1) \sqrt{m} - \text{tr}(\sqrt{r}U + V)^2 / \sqrt{m}}. \quad (3) \end{aligned}$$

Заметим, что вероятность события, противоположного событию $\text{tr}(\sqrt{r}U + V)^2 < (r + 1)m$, фигурирующему в теореме 2, имеет порядок $O(1/\sqrt{m})$, как это станет ясно из результатов разд. 3.

3 Вспомогательные утверждения

В этой части приведены вспомогательные утверждения, используемые в доказательствах теорем 1 и 2.

Введем дополнительные случайные величины:

$$\left. \begin{aligned} Z_1 &= \text{tr} UV; & Z_3 &= \text{tr} U - \sqrt{r} \text{tr} V; \\ Z_2 &= \text{tr} V^2; & Z_4 &= \text{tr} U^2, \end{aligned} \right\} \quad (4)$$

где случайные матрицы U и V определены в (2).

Положим

$$B = B(q, r, m) = 4 (q^2 + \sqrt{r}) (\sqrt{\ln m} + \sqrt{\ln p})^2.$$

Определим также для $i = 1, 2, 3, 4$ и натуральных m случайные события $A_{i,m}$ как

$$A_{i,m} = \{\omega : |Z_i(\omega)| \leq B\}.$$

Положим

$$Z = \frac{(r+1)\sqrt{r}(|\text{tr} UV| + \sqrt{r} \text{tr} U^2)}{(r+1)\sqrt{m} - S_Z/\sqrt{m}} + \frac{rq(r+1) + \sqrt{r/m}|\text{tr} U - \sqrt{r} \text{tr} V|}{(r+1)\sqrt{m} - S_Z/\sqrt{m}} S_Z,$$

где $S_Z = (r \text{tr} U^2 + 2\sqrt{r}|\text{tr} UV| + \text{tr} V^2)$.

Ясно, что существует натуральное $M_1 = M_1(r, q, c)$ такое, что при всех $m \geq M_1$ и $\omega \in \cap_{i=1}^4 A_{i,m}$ выполняется

$$\begin{aligned} Z(\omega) &\leq \left((r+1)\sqrt{r}(1+\sqrt{r})B + \left(rq(r+1) + \sqrt{r} \frac{B}{\sqrt{m}} \right) B(1+\sqrt{r})^2 \right) / \left((r+1)\sqrt{m} - \frac{B(1+\sqrt{r})^2}{\sqrt{m}} \right) \leq \\ &\leq 16 \frac{(1+\sqrt{r})^2 (2\sqrt{r} + r(rq+q+1)) (q^2 + \sqrt{r})}{r+1} \times \frac{(\ln m + \ln \sqrt{r})}{\sqrt{m}}. \end{aligned} \quad (5)$$

Оценим вероятности $\mathbf{P}(A_{i,m}^c)$ для $i = 1, 2, 3, 4$. Согласно леммам 1 и 2 из [8] справедливо следующее неравенство:

$$\begin{aligned} \mathbf{P}(|Z_1| > B) + \mathbf{P}(Z_2 > B) + \mathbf{P}(|Z_3| > B) + \\ + \mathbf{P}(Z_4 > B) &\leq 25,8 q^2 \frac{1+1/\sqrt{r}}{\sqrt{m}}. \end{aligned} \quad (6)$$

Лемма 1. Пусть случайные величины T, Y и Z определены на одном вероятностном пространстве $(\Omega, \mathbf{A}, \mathbf{P})$, при этом распределение Y является абсолютно непрерывным с ограниченной плотностью $f_Y(z)$. Предположим, что для некоторого события $A \in \mathbf{A}$ при всех $\omega \in A$ выполнено следующее соотношение:

$$|T(\omega) - Y(\omega)| \leq Z(\omega) \leq a$$

с некоторой положительной постоянной a . Тогда справедливо неравенство:

$$\begin{aligned} \sup_x |\mathbf{P}(T < x) - \mathbf{P}(Y < x)| &\leq \\ &\leq \mathbf{P}(A^c) + a \sup_x f_Y(x). \end{aligned} \quad (7)$$

Доказательство леммы 1. См. лемму 3 в [8]. \square

В следующих двух леммах приводятся два известных результата о скорости сходимости в центральной предельной теореме для независимых одинаково распределенных случайных величин. Первый из результатов относится к случайным величинам без ограничений на тип распределения. Второй результат относится к случайной величине с распределением χ^2 , рассматриваемой как сумма независимых одинаково распределенных случайных величин с известным распределением. Согласно [9], справедлива следующая лемма.

Лемма 2. Пусть случайные величины ξ_1, ξ_2, \dots независимы и одинаково распределены, выполнено $\mathbf{D}\xi_1 = \sigma^2 > 0$ и существует $\mathbf{E}|\xi_1|^3 < \infty$. Тогда для нормированной суммы $T_n = (S_n - \mathbf{E}S_n)/\sqrt{\mathbf{D}S_n}$ выполнено неравенство:

$$\sup_x |F_{T_n}(x) - \Phi(x)| \leq 0,4748 \frac{\mathbf{E}|\xi_1 - \mathbf{E}\xi_1|^3}{\sigma^3 \sqrt{n}}.$$

Случайная величина с функцией распределения $G_p(x)$, имеющая χ^2 -распределение с p степенями свободы, может быть представлена в виде суммы p независимых одинаково распределенных случайных величин с χ^2 -распределением с одной степенью свободы. Этот факт позволяет дать более точные оценки точности аппроксимации нормальным распределением, чем те, которые можно получить в общем случае с помощью неравенства Берри–Ессеена, а именно: имеет место следующий результат (см. лемму 2 в [10] при $\lambda = 0,5$).

Лемма 3. Для всех $\lambda \in (0; \sqrt{3} - 1)$ и целых $p > 1$ выполнено

$$\sup_x \left| G_p(p + x\sqrt{2p}) - \Phi(x) \right| \leq \frac{6,22}{\sqrt{p}}.$$

Лемма 4. Для любых случайных величин X и Y и любого действительного числа $a > 0$ справедливы неравенства:

$$\mathbf{P}(|X + Y| \geq 2a) \leq \mathbf{P}(|X| \geq a) + \mathbf{P}(|Y| \geq a);$$

$$\mathbf{P}(|X \cdot Y| \geq a^2) \leq \mathbf{P}(|X| \geq a) + \mathbf{P}(|Y| \geq a).$$

Доказательство леммы 4 очевидным образом вытекает из рассуждений от противного. \square

Лемма 5. Если случайные величины X_1, \dots, X_k независимы и таковы, что $|\mathbf{P}(X_j \leq x) - \Phi(x)| \leq D_j$ при всех x и $j = 1, \dots, k$ с некоторыми постоянными D_1, \dots, D_k , то

$$\left| \mathbf{P} \left(\sum_{j=1}^k c_j X_j \leq x \right) - \Phi(x) \right| \leq \sum_{j=1}^k D_j,$$

где c_1, \dots, c_k суть произвольные постоянные, для которых $c_1^2 + \dots + c_k^2 = 1$.

Доказательство леммы 5 см., например, в теореме 3.1 в [11]. \square

4 Доказательства теорем 1 и 2

Начнем с доказательства теоремы 2, поскольку неравенство (3) является ключевым в доказательстве теоремы 1.

Доказательство теоремы 2. Воспользовавшись матричным равенством

$$(I + A)^{-1} - (I - A) = A^2 (I + A)^{-1},$$

из определения (2) получаем

$$\begin{aligned} (B + W)^{-1} &= (\sqrt{p}U + pI_q + \sqrt{m}V + mI_q)^{-1} = \\ &= \frac{1}{p+m} \left(I_q - \frac{1}{p+m} (\sqrt{p}U + \sqrt{m}V) + \right. \\ &\quad \left. + \frac{(\sqrt{p}U + \sqrt{m}V)^2}{p+m} (B + W)^{-1} \right), \end{aligned}$$

далее

$$\begin{aligned} &\sqrt{m}(r+1) \left((r+1)B(B+W)^{-1} - rI_q \right) - \\ & - (\sqrt{r}U - rV) = \frac{\sqrt{r}}{\sqrt{m}}U(\sqrt{r}U + V) + \\ & + \frac{1}{\sqrt{m}}B(\sqrt{r}U + V)^2(B+W)^{-1}. \end{aligned}$$

Отсюда для следов этих матриц имеем следующее неравенство:

$$\begin{aligned} &\left| \sqrt{m}(r+1) \left((r+1) \operatorname{tr} B(B+W)^{-1} - rq \right) - \right. \\ &\quad \left. - (\sqrt{r} \operatorname{tr} U - r \operatorname{tr} V) \right| \leq \\ &\leq \frac{1}{\sqrt{m}} \sqrt{r} (|\operatorname{tr} UV| + \sqrt{r} \operatorname{tr} U^2) + \\ &+ \frac{1}{\sqrt{m}} \left| \operatorname{tr} \left[B(\sqrt{r}U + V)^2 (B+W)^{-1} \right] \right| \leq \\ &\leq \frac{1}{\sqrt{m}} \sqrt{r} (|\operatorname{tr} UV| + \sqrt{r} \operatorname{tr} U^2) + \\ &+ \frac{1}{\sqrt{m}} \operatorname{tr} (\sqrt{r}U + V)^2 \operatorname{tr} B(B+W)^{-1}. \quad (8) \end{aligned}$$

Для получения предпоследнего неравенства использованы симметричность и неотрицательная определенность обеих случайных матриц $(\sqrt{r}U + V)^2$ и $B(B+W)^{-1}$, поскольку для симметричных неотрицательно определенных матриц X

и Y выполнено соотношение (см. [12]) $\operatorname{tr} XY \leq \operatorname{tr} X \operatorname{tr} Y$.

Видно, что случайная величина $\operatorname{tr} B(B+W)^{-1}$ фигурирует в крайней левой и крайней правой частях неравенства (8). Преобразуя полученное неравенство, получаем, что при $\operatorname{tr} (\sqrt{r}U + V)^2 < (r+1)m$ выполнено (3). Тем самым доказательство теоремы 2 завершено. \square

Переходим к доказательству теоремы 1.

Доказательство теоремы 1. Используя лемму 1 из [7], перейдем к представлению статистики Барлетта–Нанда–Пилая

$$\begin{aligned} T_{\text{BNP}} &= \\ &= \sqrt{p} \left(1 + \frac{p}{m} \right) \left\{ \left(1 + \frac{m}{p} \right) \operatorname{tr} \left[S_h (S_h + S_e)^{-1} \right] - q \right\} \end{aligned}$$

в терминах матриц B и W размера $q \times q$ вместо матриц S_h и S_e размера $p \times p$, где S_h и S_e определены в (1), а матрицы B и W независимы и имеют распределения Уишарта $W_q(p, I_q)$ и $W_q(m, I_q)$ с $m = n - p + q$ соответственно. При этом будем пользоваться следующим соотношением (см. [7]):

$$\operatorname{tr} S_h (S_h + S_e)^{-1} = \operatorname{tr} B (B + W)^{-1}.$$

Согласно (3) для Z_1, Z_2, Z_3 и Z_4 (см. определение в (4)) при $rZ_4 + 2\sqrt{r}|Z_1| + Z_2 < m$ имеем:

$$\begin{aligned} &|\sqrt{r}T_{\text{BNP}} - (\sqrt{r} \operatorname{tr} U - r \operatorname{tr} V)| = \\ &= \left| \sqrt{m}(r+1) \left((r+1) \operatorname{tr} B(B+W)^{-1} - rq \right) - \right. \\ &\quad \left. - (\sqrt{r} \operatorname{tr} U - r \operatorname{tr} V) \right| \leq \\ &\leq \left((r+1) \sqrt{r} (|Z_1| + \sqrt{r} Z_4) + \right. \\ &\quad \left. + \left(rq(r+1) + \frac{\sqrt{r}|Z_3|}{\sqrt{m}} \right) (rZ_4 + 2\sqrt{r}|Z_1| + \right. \\ &\quad \left. + Z_2) \right) / \left((r+1) \sqrt{m} - \frac{rZ_4 + 2\sqrt{r}|Z_1| + Z_2}{\sqrt{m}} \right). \end{aligned}$$

Следовательно, в силу (5) и (7) при всех $m \geq M_1$ получаем

$$\begin{aligned} &\sup_z \left| \mathbf{P} \left(\frac{T_{\text{BNP}}}{\sqrt{2q(1+r)}} < z \right) - \right. \\ &\quad \left. - \mathbf{P} \left(\frac{\operatorname{tr} U - \sqrt{r} \operatorname{tr} V}{\sqrt{2q(1+r)}} < z \right) \right| \leq \\ &\leq \sum_{i=1}^4 \mathbf{P} (|Z_i| > B) + K_4(r, q) \frac{\ln m}{\sqrt{m}} \sup_x f(x), \quad (9) \end{aligned}$$

где $f(x)$ — плотность случайной величины $(\text{tr } U - \sqrt{r} \text{tr } V) / \sqrt{2q(1+r)}$; $K_4(r, q)$ — некоторая вычисляемая функция от r и q .

Отметим, что, поскольку матрицы B и W независимы, матрицы U и V также независимы между собой. Известно (см., например, гл. 2 в [1]), что $\text{tr } B$ и $\text{tr } W$ имеют χ^2 -распределения с pq и mq степенями свободы соответственно. Известно также, что плотность χ^2 -распределения с $k \geq 3$ степенями свободы ограничена сверху величиной $1/(2\sqrt{\pi(k-2)})$. Поэтому для плотности $f(x)$ справедлива равномерная оценка

$$f(x) \leq \min \left(\frac{\sqrt{p}}{\sqrt{(pq-2)}}, \frac{\sqrt{m}}{\sqrt{r(mq-2)}} \right) \frac{\sqrt{q(1+r)}}{\sqrt{2\pi}}. \quad (10)$$

Объединяя утверждения лемм 3 и 5 и соотношения (2), (6), (9) и (10), получаем утверждение теоремы 1. \square

Литература

1. *Fujikoshi Y., Ulyanov V. V., Shimizu R.* Multivariate statistics: High-dimensional and large-sample approximations. — Hoboken, NJ, USA: John Wiley & Sons, 2010. 512 p.
2. *Johnstone I. M., Nadler B.* Roy's largest root test under rank-one alternatives // *Biometrika*, 2017. Vol. 104. No. 1. P. 181–193.
3. *Akbari V., Anfinson S. N., Doulgeris A. P., Eltoft T., Moser G., Serpico S. B.* Polarimetric SAR change detection with the complex Hotelling–Lawley trace statistic // *IEEE T. Geosci. Remote*, 2016. Vol. 54. Iss. 7. P. 3953–3966.
4. *Anderson T. W.* An introduction to multivariate analysis. — 3rd ed. — Hoboken, NJ, USA: John Wiley & Sons, 2003. 742 p.
5. *Muirhead R. J.* Asymptotic distributions of some multivariate tests // *Ann. Math. Stat.*, 1970. Vol. 41. No. 3. P. 1002–1010.
6. *Лунатьев А. А., Ульянов В. В.* Вычисляемые оценки точности приближений для статистики Бартлетта–Нанда–Пиллай // *Математические труды*, 2016. Т. 19. № 2. С. 109–118.
7. *Wakaki H., Fujikoshi Y., Ulyanov V. V.* Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large // *Hiroshima Math. J.*, 2014. Vol. 44. No. 3. P. 247–259.
8. *Лунатьев А. А., Ульянов В. В.* Неасимптотический анализ статистики Лоули–Хотеллинга для данных большой размерности // *Записки научных семинаров ПОМИ*, 2019. Т. 486. С. 178–189.
9. *Shevtsova I. G.* On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands // *arXiv.org*, 2011. arXiv:1111.6554 [math.PR]. 7 p.
10. *Кавагучи Ю., Ульянов В. В., Фуджикоши Я.* Приближения для статистик, описывающих геометрические свойства данных большой размерности, с оценками ошибок // *Информатика и её применения*, 2010. Т. 4. Вып. 1. С. 22–27.
11. *Ulyanov V. V., Wakaki H., Fujikoshi Y.* Berry–Esseen bound for high dimensional asymptotic approximation of Wilks' Lambda distribution // *Stat. Probabil. Lett.*, 2006. Vol. 76. No. 12. P. 1191–1200.
12. *Coope I. D.* On matrix trace inequalities and related topics for products of Hermitian matrices // *J. Math. Anal. Appl.*, 1949. Vol. 188. No. 3. P. 999–1001.

Поступила в редакцию 07.01.2020

NONASYMPTOTIC ANALYSIS OF BARTLETT–NANDA–PILLAI STATISTIC FOR HIGH-DIMENSIONAL DATA

A. A. Lipatiev

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

Abstract: The author gets the computable error bounds for normal approximation of Bartlett–Nanda–Pillai statistic when dimensionality grows proportionally to the sample size. This result enables one to get more precise calculations of the p-values in applications of multivariate analysis. In practice, more and more often, analysts encounter situations when the number of factors is large and comparable with the sample size. The examples include signal processing. The proof is essentially based on the normality of the distribution of the elements of the matrices under consideration with the Wishart distribution. For random variables that are the matrix traces of the product and squares of matrices with the normalized Wishart distribution, convenient upper bounds for $1 - F$ are found where F is the distribution function of the corresponding matrix trace. Applying the properties of inverse matrices and positive semidefinite matrices, the Bartlett–Nanda–Pillai statistic is bounded from above by a combination of the above-mentioned matrix traces.

Keywords: computable estimates; accuracy of approximation; MANOVA; computable error bounds; Bartlett–Nanda–Pillai statistic; high-dimensional data

DOI: 10.14357/19922264210110

References

1. Fujikoshi, Y., V. V. Ulyanov, and R. Shimizu. 2010. *Multivariate statistics: High-dimensional and large-sample approximations*. Hoboken, NJ: John Wiley & Sons. 512 p.
2. Johnstone, I. M., and B. Nadler. 2017. Roy's largest root test under rank-one alternatives. *Biometrika* 104(1):181–193.
3. Akbari, V., S. N. Anfinsen, A. P. Doulgeris, T. Eltoft, G. Moser, and S. B. Serpico. 2016. Polarimetric SAR change detection with the complex Hotelling–Lawley trace statistic. *IEEE T. Geosci. Remote* 54(7):3953–3966.
4. Anderson, T. W. 2003. *An introduction to multivariate analysis*. 3rd ed. Hoboken, NJ: John Wiley & Sons. 742 p.
5. Muirhead, R. J. 1970. Asymptotic distributions of some multivariate tests. *Ann. Math. Stat.* 41(3):1002–1010.
6. Lipatiev, A. A., and V. V. Ulyanov. 2017. On computable estimates for accuracy of approximation for the Bartlett–Nanda–Pillai statistic. *Siberian Adv. Math.* 27(3):153–159.
7. Wakaki, H., Y. Fujikoshi, and V. V. Ulyanov. 2014. Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large. *Hiroshima Math. J.* 44(3):247–259.
8. Lipatiev, A. A., and V. V. Ulyanov. 2019. Neasimptoticheskiy analiz statistiki Louli–Khotellinga dlya dannykh bol'shoy razmernosti [Nonasymptotic analysis of Lawley–Hotelling statistic for high dimensional data]. *Zapiski nauchnykh seminarov POMI* [POMI Notes of Scientific Seminars] 486:178–189.
9. Shevtsova, I. G. 2011. On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands. arXiv:1111.6554 [math.PR]. Available at: <https://arxiv.org/pdf/1111.6554> (accessed December 16, 2020).
10. Kawaguchi, Yu., V. V. Ulyanov, and Ya. Fujikoshi. 2010. Priblizheniya dlya statistik, opisuyayushchikh geometricheskie svoystva dannykh bol'shoy razmernosti, s otsenkami oshibok [Asymptotic distributions of basic statistics in geometric representation for high-dimensional data and their error bounds]. *Informatika i ee Primeneniya — Inform. Appl.* 4(1):22–27.
11. Ulyanov, V. V., H. Wakaki, and Y. Fujikoshi. 2006. Berry–Esseen bound for high dimensional asymptotic approximation of Wilks' Lambda distribution. *Stat. Probabil. Lett.* 76(12):1191–1200.
12. Coope, I. D. 1949. On matrix trace inequalities and related topics for products of Hermitian matrices. *J. Math. Anal. Appl.* 188(3):999–1001.

Received January 7, 2020

Contributor

Lipatiev Alexander A. (b. 1988) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; allipatev@cs.msu.ru

АРХИТЕКТУРА РАСПРЕДЕЛЕННОГО РЕШЕНИЯ ЗАДАЧ АНАЛИЗА ДАННЫХ В ОБЛАСТИ НЕЙРОФИЗИОЛОГИИ*

Д. О. Брюхов¹, С. А. Ступников², Д. Ю. Ковалёв³, И. А. Шанин⁴

Аннотация: С ростом объема и разнообразия нейрофизиологических данных происходит и рост интереса к применению методов информатики, таких как статистический анализ, машинное обучение, нейронные сети, для анализа этих данных. Появляется потребность в создании инфраструктур, обеспечивающих как хранение большого объема данных в области нейрофизиологии, так и их распределенную обработку и анализ. В данной статье предлагается архитектура средств решения задач на основе технологий распределенного хранения и анализа больших данных Hadoop и высокопроизводительных вычислений с применением графических ускорителей.

Ключевые слова: нейрофизиология; нейроинформатика; интенсивное использование данных; инфраструктуры решения задач; анализ данных

DOI: 10.14357/19922264210111

1 Введение

Понимание работы человеческого мозга становится одной из основных научных задач в настоящее время. Растет количество и качество оборудования в области нейрофизиологии. В связи с этим растет и объем получаемых данных.

Традиционные методы анализа данных не справляются с обработкой большого объема нейрофизиологических данных. В связи с этим происходит рост интереса к использованию методов информатики, таких как статистический анализ, машинное обучение (и, в частности, нейронные сети глубокого обучения). Появляется потребность в создании инфраструктур, позволяющих как хранить большой объем данных, так и обрабатывать и анализировать эти данные.

Первоначально такие инфраструктуры представляли собой веб-каталоги данных и программных средств, где ученые могли делиться своими данными с другими учеными, находить и использовать данные и программы в своих исследованиях. В последнее время стали развиваться инфраструктуры, предоставляющие исследователям средства высокопроизводительных вычислений для обработки и анализа данных в области нейрофизиологии.

В данной статье предлагается архитектура распределенного решения задач в области нейрофизиологии, объединяющая технологии распределенного хранения и анализа больших данных Hadoop и технологии высокопроизводительных вычислений с применением графических ускорителей. Hadoop позволяет хранить данные большого объема (например, нейроизображения) и выполнять распределенную параллельную обработку этих данных на вычислительном кластере.

Ряд методов, включающих сложные вычисления, требует использования графических ускорителей для эффективного выполнения (например, построение моделей в методах машинного обучения). Такие вычисления выполняются на отдельных серверах с мощными графическими ускорителями.

В рамках данной статьи был проведен анализ существующих инфраструктур в области нейрофизиологии, предложена архитектура средств распределенной обработки нейрофизиологических данных, приведен пример использования этой архитектуры при решении задачи поиска значимых различий нелинейной функциональной связности головного мозга для мужчин и женщин.

* Работа выполнена при финансовой поддержке РФФИ (проект 18-29-22096).

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

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

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, dm.kovalev@gmail.com

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ivan.shanin@gmail.com

2 Инфраструктуры решения задач в области нейрофизиологии

В последнее время стали активно разрабатываться инфраструктуры для запуска различных программ обработки и анализа нейрофизиологических данных в одной среде, позволяющие встраивать существующие программы и библиотеки в создаваемые пользователями программы, решающие конкретные задачи. Ниже рассмотрены основные современные инфраструктуры для решения задач в области нейрофизиологии.

XNAT [1] — это открытая информационная платформа для работы с нейроизображениями, разработанная исследовательской группой по нейроринформатике в Вашингтонском университете. Она облегчает общие задачи управления, обеспечения производительности и качества обработки нейроизображений и связанных данных. XNAT позволяет программировать сложные потоки работ с несколькими уровнями автоматизации.

NeuroImaging Tools & Resources Collaboratory (NITRC) [2] — это бесплатный веб-ресурс, который предлагает информацию о постоянно расширяющемся наборе программного обеспечения и данных для нейроринформатики. Он состоит из трех компонентов: реестра ресурсов, репозитория

изображений и вычислительной среды — виртуальной машины, содержащей предустановленный набор программных средств для работы с нейроизображениями.

Программная платформа MIRF (Medical Images Research Framework) [3] — это платформа с открытым исходным кодом для быстрой разработки приложений для обработки медицинских изображений. Основным сценарием использования MIRF является создание пользователем конвейеров — последовательности обработчиков исходных данных.

В рамках проекта Европейского Союза Human Brain Project (HBP) [4] разрабатывается архитектура технологии хранения информации и коммуникации с исследователями ICT (Information and Communication Technology) [5]. Основные компоненты ICT — платформа нейроринформатики NIP и высокопроизводительная аналитическая и вычислительная платформа HPAC. Компонент COLLAB обеспечивает взаимодействие как между компонентами ICT, так и взаимодействие исследователей с ICT.

В таблице приведено сравнение существующих инфраструктур решения задач в области нейрофизиологии.

Можно отметить, что развивающиеся в настоящее время инфраструктуры ориентированы на предоставление возможности повторного исполь-

Сравнение инфраструктур решения задач в области нейрофизиологии

Инфраструктура	XNAT	NTRC	MIRF	HBP ICT	Предлагаемая архитектура
Вид	Веб-приложение	Набор программ	Вычислительная среда	Вычислительная среда	Вычислительная среда
Распределенные технологии		Облако		Облако, HPC	Nadoop, HPC
Способ хранения данных	Файлы	Файлы	Файлы	Файлы	Файлы, базы данных
Распределенное хранение					Да (HDFS)
Распределенные вычисления				Да	Да
Графические ускорители	Да	Да	Да	Да	Да
Набор встроенных библиотек	Да	Да	Да	Да	Да
Атласы				Да	Да
Обработка изображений	Да	Да	Да	Да	Да
Визуализация изображений	Да	Да	Да	Да	
Анализ изображений	Да	Да	Да	Да	Да
Методы машинного обучения	Да	Да	Да	Да	Да
Программная архитектура	Отдельные программы	Отдельные программы	Конвейеры	Программа организуется как сервис (SaaS)	Модули
Взаимодействие с другими программами	Расширения (плагины)	Общие входные/выходные данные	Вызов через API программы	Вызов как сервис (SaaS)	Вызов через API программы
Поддержка баз данных					Да

зования написанных программ и библиотек для решения новых задач в области нейрофизиологии. К основным недостаткам рассмотренных инфраструктур можно отнести отсутствие поддержки распределенного хранения большого объема данных и распределенной параллельной обработки этих данных. Предлагаемая в данной работе (см. разд. 3) архитектура решения задач нацелена на преодоление этих недостатков. Она также нацелена на повторное использование уже реализованных методов анализа нейрофизиологических данных. В архитектуре предусмотрено использование различных баз данных для эффективного хранения часто используемых типов данных, например временных рядов с данными по регионам мозга.

3 Программная архитектура распределенного решения задач анализа данных в области нейрофизиологии

Задача исследования работы головного мозга становится все более актуальной. Растет объем исследований головного мозга, развиваются методы анализа получаемых данных, включая нейроизображения. В работе [6] было продемонстрировано многообразие форматов представления нейроизображений, программных средств обработки и анализа нейроизображений. Активно развиваются методы глубокого обучения, появляется все больше исследований, использующих нейронные сети для анализа нейроизображений. В связи с этим встает вопрос о создании архитектуры, поддерживающей хранение больших объемов данных и методы распределенной обработки этих данных.

Предлагаемая в рамках данной работы архитектура основана на технологиях распределенного хранения и анализа больших данных Hadoop и высокопроизводительных вычислений с применением графических ускорителей. В качестве платформы распределенных вычислений в Hadoop был выбран фреймворк Spark. В зависимости от конкретной задачи анализ нейрофизиологических данных может выполняться как на одной из этих платформ, так и совместно на обеих платформах.

Spark обычно применяется в случаях, когда требуется параллельная обработка большого числа файлов с данными в области нейрофизиологии. При распределенном анализе рекомендуется использовать методы статистического анализа, в частности регрессионный анализ, проверку статистических гипотез, детектирование аномалий. На

Spark также можно реализовывать методы машинного обучения для небольших моделей с малым числом параметров (признаков). Если для анализа данных применяются сложные методы, требующие большого объема вычислений, то эти методы рекомендуется выполнять на графических ускорителях. К таким методам относятся методы глубокого обучения, включая сверточные и рекуррентные нейронные сети, автокодировщики и другие эффективные методы машинного обучения, такие как градиентный бустинг и метод опорных векторов. Для решения ряда задач может применяться комбинация этих подходов, например модели в методах машинного обучения или в нейронных сетях могут строиться на данных из обучающей выборки на отдельном сервере с графическими картами, а дальнейшее применение этих моделей и оценка качества построенных моделей осуществляется в Spark.

Программы анализа данных могут быть реализованы на любых языках программирования, поддерживающих Spark, например Python, Java, Scala, Julia. В реализации предлагаемой в этой статье архитектуры распределенного решения задач будет использован язык Python, также будут использованы библиотеки методов обработки и анализа данных, написанные на Python.

На рис. 1 представлена схема архитектуры решения задач в области нейрофизиологии. В архитектуре выделяются три уровня: уровень данных, уровень поддержки приложений и уровень приложений.

На *уровне данных* обеспечивается надежное хранение большого объема нейрофизиологических данных, включая исходные данные (например, нейроизображения), промежуточные данные (обработанные данные, сгенерированные модели) и результаты работы модулей анализа данных (предсказания, результаты проверки гипотез, результаты классификации данных). Данные могут храниться как в виде файлов в файловой системе, так и в базах данных. Исходные данные загружаются и хранятся в распределенной файловой системе HDFS (Hadoop distributed file system). HDFS автоматически создает несколько копий данных, что обеспечивает надежность их хранения на случай отказа оборудования. Для хранения данных, представленных в виде многомерных временных рядов (например, данные, извлеченные по регионам мозга) используется база данных OpenTSDB.

Часть данных, необходимых для выполнения программ с использованием графических ускорителей, хранится в локальной файловой системе этого сервера. На *уровне поддержки приложений* устанавливаются программы и библиотеки, используемые

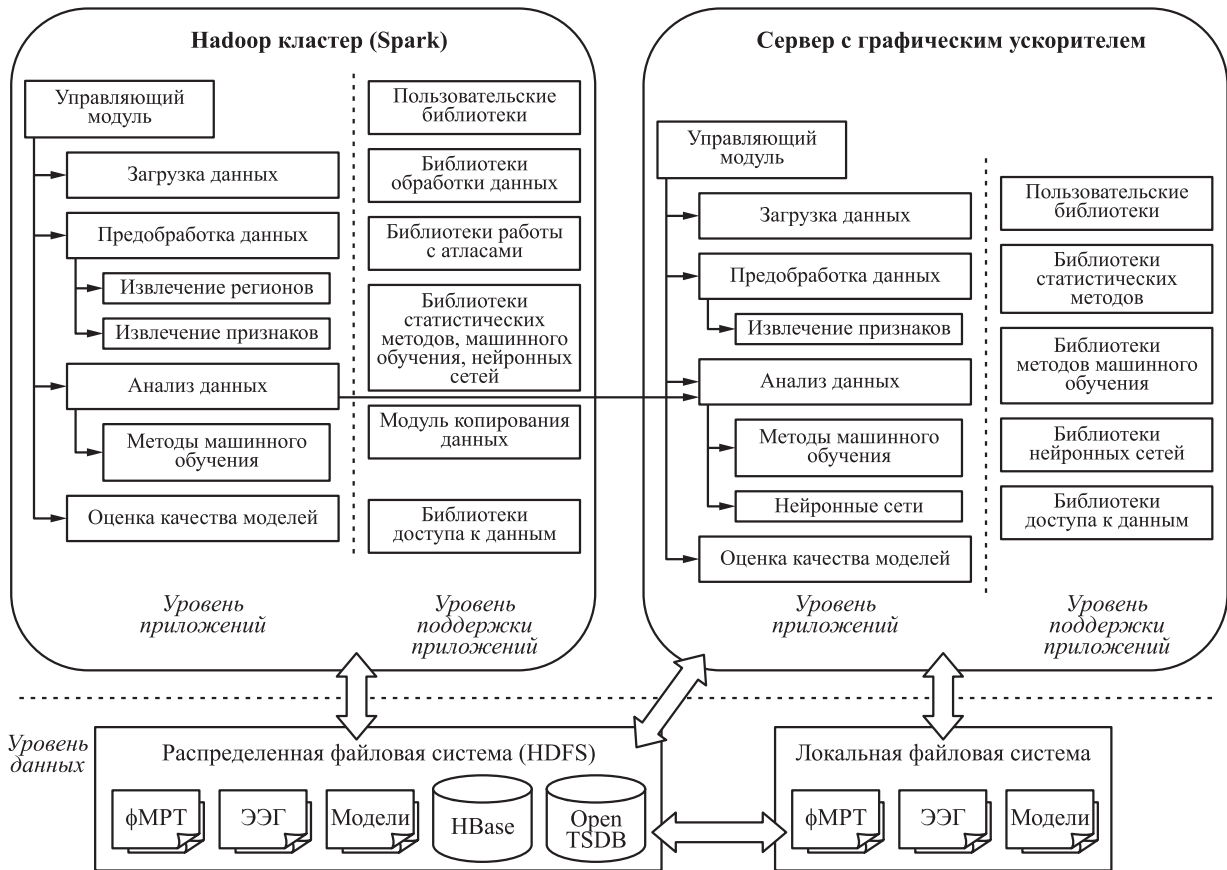


Рис. 1 Общая архитектура распределенного решения задач анализа данных в области нейрофизиологии

для обработки и анализа нейрофизиологических данных. Разработанные программы или их части могут быть оформлены в виде отдельных библиотек и затем повторно использованы при разработке новых программных средств. Такие программы формируют *пользовательские библиотеки*. Примерами таких модулей могут служить: извлечение регионов мозга по заданному атласу, выявление и классификация артефактов в электроэнцефалограмме (ЭЭГ).

Библиотеки обработки данных включают библиотеки по работе (чтению и записи) с различными форматами нейрофизиологических изображений и простыми операциями над ними. Примерами таких библиотек служат NiPy и Nibabel для языка Python. *Библиотеки работы с атласами* используются для извлечения регионов мозга на основе существующих атласов. Для работы с атласами можно использовать как специализированные библиотеки (например, AtlasReader), так и общие библиотеки по анализу нейрофизиологических данных, поддерживающих работу с атласами (например, Nilearn). *Библиотеки доступа к данным* используются для работы с базами данных, в которых хранятся промежуточные данные и резуль-

таты работы программ. Например, библиотека orpentsdb.py используется для организации доступа из программы, написанной на языке Python, к базе данных OpenTSDB. Библиотека happybase используется для связи с базой данных HBase, для хранения вычисленных нелинейных функций зависимости одних регионов мозга от других. *Модуль копирования данных* предназначен для поддержки копирования данных между распределенной файловой системой HDFS и локальной файловой системой сервера с графическим ускорителем.

Библиотеки статистических методов, методов машинного обучения и нейронных сетей используются для выполнения анализа нейрофизиологических данных. Примерами таких библиотек могут служить: Nitime (для анализа временных рядов), Nilearn (для статистического исследования данных нейровизуализации), MNE-Python (для статистического анализа и методов машинного обучения над данными магнитно-резонансной томографии (МРТ) и ЭЭГ), tsfresh (для анализа временных рядов), scikit-learn (для методов машинного обучения и нейронных сетей), Keras (для нейронных сетей), gplearn (для методов генетического программирования).

ния), XGBoost (для метода градиентного бустинга на деревьях), CatBoost (для метода градиентного бустинга на деревьях). Все библиотеки, используемые в Spark, устанавливаются на каждый узел кластера.

Уровень приложений содержит пользовательские программы для обработки и анализа нейрофизиологических данных. Программы могут быть реализованы либо только на Spark, либо только на графических ускорителях, либо в общей архитектуре. Программы имеют модульную структуру. Центральным является *управляющий модуль*, который отвечает за организацию процесса обработки и анализа данных, взаимодействуя с остальными модулями. Данные могут быть загружены в систему до начала работы программы или может быть использован *модуль загрузки данных* для динамической загрузки данных из внешних ресурсов.

Модуль предобработки данных производит предварительную обработку данных, включая очистку данных, преобразование к формату, используемому при дальнейшем анализе этих данных, разделение данных по заданному критерию (например, для разных групп субъектов), отбор значимых признаков. Особенно следует выделить *подмодуль выявления регионов*, осуществляющий преобразование исходных 4-мерных фМРТ-изображений (фМРТ — функциональная МРТ) в набор многомерных временных рядов для каждого региона мозга. Регионы мозга определяются на основании применения атласов регионов (например, Harvard–Oxford, Automated Anatomical Labeling). Полученные временные ряды сохраняются в базе данных OpenTSDB. Подмодуль *извлечения признаков* выявляет значимые признаки в исходных данных, включая статистики для описания временных рядов. Отобранные признаки подаются на вход методов анализа данных, включая методы машинного обучения и нейронные сети.

Модуль анализа данных использует методы информатики, такие как статистический анализ, машинное обучение, нейронные сети, для анализа нейрофизиологических данных. Модуль может выполняться как на платформе Spark для методов параллельной обработки входных данных (например, параллельная обработка данных по разным объектам исследования), так и на сервере с графическим ускорителем для вычислительно сложных методов, хорошо распараллеливаемых на графических ускорителях (например, нейронных сетей). Связь между модулями, работающими на Spark и на сервере с графическим ускорителем, осуществляется посредством удаленного вызова скрипта на сервере с графическим ускорителем, запускающего необходимый программный модуль анализа данных. Программа может содержать несколько мо-

дулей предобработки и анализа данных, например, для сравнения результатов применения различных методов или использования ансамбля методов для получения более точных результатов.

В *модуле оценки качества* происходит вычисление известных метрик, таких как доля правильно классифицированных объектов, точность и полнота, для построенных моделей.

4 Пример задачи анализа данных в области нейрофизиологии

В качестве примера задачи, реализуемой в рамках представленной архитектуры, рассмотрим задачу поиска значимых различий нелинейной функциональной связности головного мозга для мужчин и женщин в состоянии покоя. Изучение гендерных различий в работе мозга — одна из важных областей когнитивного анализа мозга. Понимание роли гендерных эффектов может помочь в развитии специализированных методов лечения, различающихся для мужчин и женщин. В работе [7] было рассмотрено применение моделирования нелинейной функциональной связности головного мозга для поиска различий между мужчинами и женщинами.

На рис. 2 представлена предлагаемая в работе программная архитектура средств реализации этой задачи, соответствующая общей архитектуре (см. рис. 1). Архитектура включает следующие основные компоненты, соответствующие этапам решения задачи.

Извлечение регионов головного мозга. На данном этапе происходит преобразование 4-мерного изображения головного мозга в двумерный массив, где одним из измерений выступают регионы головного мозга, а другим — время. Извлечение регионов происходит с использованием вероятностного атласа Harvard–Oxford [8]. Полученные данные сохраняются в виде временных рядов в базе данных OpenTSDB.

Построение аналитических функций с применением метода генетического программирования [9]. Генетическое программирование — это метод, при помощи которого можно восстановить нелинейную функциональную связь. Метод основан на идее биологической эволюции. Существенный недостаток этого метода заключается в том, что его вычислительная стоимость растет экспоненциально с увеличением размерности, поэтому данный этап выполняется на компьютере с графическим ускорителем. Входные данные берутся из базы данных OpenTSDB. Вычисленные нелинейные функции зависимости одних регионов мозга от других

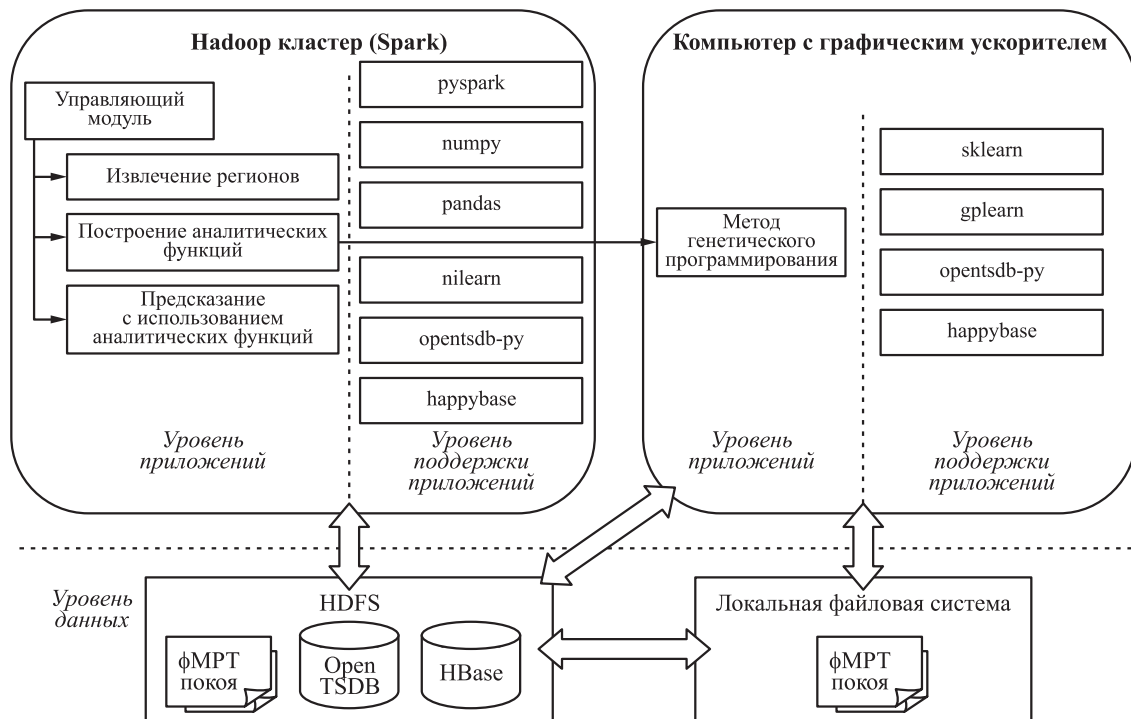


Рис. 2 Архитектура средств поиска значимых различий нелинейной функциональной связности головного мозга для мужчин и женщин

как для мужчин, так и для женщин сохраняются в базу данных HBase.

Проведение статистического анализа для сравнения полученных функций и вывод о схожести или различии того или иного региона головного мозга для мужчин и женщин. Включает в себя построение предсказаний с использованием функций, построенных отдельно для мужчин и отдельно для женщин, и проверку гипотезы о том, что ошибка между предсказаниями функций равна нулю. Входные данные для выполнения этого этапа берутся из базы данных HBase.

На основе этой архитектуры был разработан прототип поиска значимых различий нелинейной функциональной связности головного мозга для мужчин и женщин. Основная часть прототипа была реализована на языке Python в среде Spark с применением библиотеки `pyspark`, а метод генетического программирования был реализован на компьютере с графическим ускорителем. Для извлечения регионов мозга на основе атласа Harvard-Oxford использовалась библиотека `nilearn`. Для обработки полученных данных по регионам мозга использовались библиотеки `numpy` и `pandas`. Полученные данные в виде временных рядов сохранялись в базу данных OpenTSDB при помощи библиотеки `opentsdb-py`. При анализе этих временных рядов с помощью методов генетического программирования

использовались библиотеки `gplearn` и `sklearn`. Данные по полученным в результате анализа аналитическим функциям записывались в базу данных HBase при помощи библиотеки `happybase`. Разработанный прототип был опробован на данных из проекта Human Connectome Project (HCP) [10].

5 Заключение

Рост объема данных в области нейрофизиологии приводит к тому, что использование одного компьютера уже недостаточно для их обработки и анализа. Анализ существующих систем выявляет потребность создания архитектуры, позволяющей как хранить большой объем данных, так и обрабатывать эти данные с применением распределенных технологий. В статье предложена программная архитектура распределенного решения задач в области нейрофизиологии над данными больших объемов, объединяющая технологии распределенного хранения и анализа больших данных Hadoop и технологии высокопроизводительных вычислений с применением графических ускорителей. Хранение данных осуществляется в распределенной файловой системе HDFS. Параллельная распределенная обработка и анализ нейрофизиологических данных выполняются на Hadoop/Spark-кластере. Выпол-

нение вычислительно сложных алгоритмов, требующих графических ускорителей, осуществляется на отдельных серверах, содержащих графические карты.

Предложенная архитектура была использована при решении задач по обработке нейрофизиологических изображений, включая анализ функциональной связности отделов мозга на основе данных фМРТ состояния покоя, выявления шаблонов гендерных различий в работе мозга на основе данных фМРТ состояния покоя, определения активности человека с использованием фМРТ действия, обнаружения артефактов в ЭЭГ.

Литература

1. Marcus D., Olsen T. R., Ramaratnam M., Buckner R. L. The extensible neuroimaging archive toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data // *Neuroinformatics*, 2007. Vol. 5. P. 11–34.
2. NITRC. <https://www.nitrc.org>.
3. Musatian S., Lomakin A., Chizhova A. Medical images research framework // *CEUR Workshop Procee.*, 2019. Vol. 2372. P. 60–66.
4. Human brain project. <https://www.humanbrainproject.eu>.
5. Amunts K., Ebell C., Muller J., Telefont M., Knoll A., Lippert L. The human brain project: Creating a European research infrastructure to decode the human brain // *Neuron*, 2016. Vol. 92. P. 574–581.
6. Брюхов Д. О., Ступников С. А., Ковалёв Д. Ю., Шанин И. А. Нейрофизиология как предметная область для решения задач с интенсивным использованием данных // *Информатика и её применения*, 2020. Т. 14. Вып. 1. С. 40–47.
7. Kovalev D., Tirikov E., Sergeev D., Ponomareva N. V. Methods and tools for analyzing human brain signals based on functional magnetic resonance imaging data // *CEUR Workshop Procee.*, 2020. Vol. 2790. P. 214–229.
8. Desikan R. S., Segonne F., Fischl B., et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest // *NeuroImage*, 2006. Vol. 31. No. 3. P. 968–980.
9. Allgaier N., Banaschewski T., Barker G., et al. Nonlinear functional mapping of the human brain // *arXiv.org*, 2015. arXiv:1510.03765. 21 p.
10. Elam J. S., Van Essen D. Human connectome project // *Encyclopedia of computational neuroscience* / Eds. D. Jaeger, R. Jung. — New York, NY, USA: Springer, 2013. doi: 10.1007/978-1-4614-7320-6_592-1.

Поступила в редакцию 27.12.2020

AN ARCHITECTURE FOR DISTRIBUTED DATA ANALYSIS PROBLEM SOLVING IN NEUROPHYSIOLOGY

D. O. Briukhov, S. A. Stupnikov, D. Yu. Kovalev, and I. A. Shanin

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

Abstract: The growth of volume and variety of data in the field of neurophysiology increases the need of the application of computer science methods such as statistical analysis, machine learning, and neural networks for the data analysis. Infrastructures providing storage of a large volume of data in neurophysiology as well as data distributed processing and analysis are required. This article proposes a software architecture for the problem solving based on the Hadoop distributed storage and analysis framework and GPU-assisted high-performance computing technologies.

Keywords: neurophysiology; neurophysiological resources; neuroinformatics; data intensive research; problem solving infrastructure; analysis of neurophysiological data

DOI: 10.14357/19922264210111

Acknowledgments

This research was financially supported by the Russian Foundation for Basic Research (project 18-29-22096).

References

1. Marcus, D., T.R. Olsen, M. Ramaratnam, and R. L. Buckner. 2007. The extensible neuroimaging archive toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5:11–34.
2. NITRC. Available at: <https://www.nitrc.org/> (accessed January 14, 2021)

3. Musatian, S., A. Lomakin, and A. Chizhova. 2019. Medical images research framework. *CEUR Workshop Procee.* 2372:60–66.
4. Human brain project. Available at: <https://www.humanbrainproject.eu> (accessed January 14, 2021).
5. Amunts, K., C. Ebell, J. Muller, M. Telefont, A. Knoll, and L. Lippert. 2016. The human brain project: Creating a European research infrastructure to decode the human brain. *Neuron* 92:574–581.
6. Bryukhov, D. O., S. A. Stupnikov, D. Yu. Kovalev, and I. A. Shanin. 2020. Neyrofiziologiya kak predmetnaya oblast' dlya resheniya zadach s intensivnym ispol'zovaniem dannykh [Neurophysiology as a subject domain for data intensive problem solving]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):40–47.
7. Kovalev, D., D. Sergeev, E. Tirikov, and N. Ponomareva. 2020. Methods and tools for analyzing human brain signals based on functional magnetic resonance imaging data. *CEUR Workshop Procee.* 2790:214–229.
8. Desikan, R. S., F. Segonne, B. Fischl, *et al.* 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31(3):968–980.
9. Allgaier, N., T. Banaschewski, G. Barker, *et al.* 2015. Nonlinear functional mapping of the human brain. 21 p. Available at: <https://arxiv.org/abs/1510.03765> (accessed January 14, 2021).
10. Elam, J. S., and D. Van Essen. 2013. Human connectome project. *Encyclopedia of computational neuroscience*. Eds. D. Jaeger and R. Jung. New York, NY: Springer. doi: 10.1007/978-1-4614-7320-6_592-1.

Received December 27, 2020

Contributors

Briukhov Dmitry O. (b. 1971) — Candidate of Science (PhD) in technology, 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; dbriukhov@ipiran.ru

Stupnikov Sergey A. (b. 1978) — Candidate of Science (PhD) in technology, 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; sstupnikov@ipiran.ru

Kovalev Dmitry Y. (b. 1988) — 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; dkovalev@ipiran.ru

Shanin Ivan A. (b. 1991) — junior 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; v08shanin@gmail.com

ПРЕДСТАВЛЕНИЕ НОВЫХ ЛЕКСИКОГРАФИЧЕСКИХ ЗНАНИЙ В ДИНАМИЧЕСКИХ КЛАССИФИКАЦИОННЫХ СИСТЕМАХ*

А. А. Гончаров¹, И. М. Зацман², М. Г. Кружков³

Аннотация: Характерная особенность динамических классификационных систем (ДКС) состоит в том, что в процессе применения этих систем в них в любой момент времени могут добавляться новые рубрики и/или изменяться дефиниции существующих рубрик, включая перераспределение смыслового содержания между ними. С одной стороны, эта особенность ДКС дает возможность оперативно отражать в них новое знание и сразу начинать его использовать, например в процессе лингвистического аннотирования. С другой стороны, если некоторая рубрика использовалась при аннотировании, а затем была изменена, то аннотации с этой рубрикой, сформированные до внесения изменений, в ряде случаев должны быть реклассифицированы. Статья преследует двоякую цель, которая состоит, во-первых, в сопоставлении подходов к классификации сущностей на основе (1) ДКС и (2) онтологий, изменяемых во времени, а во-вторых, в описании специфики представления новых лексикографических знаний в ДКС.

Ключевые слова: динамическая классификационная система; версионные онтологии; лингвистическое аннотирование; реклассификация аннотаций

DOI: 10.14357/19922264210112

1 Введение

Характерная особенность ДКС состоит в том, что в процессе применения в них могут добавляться новые рубрики и/или изменяться дефиниции существующих рубрик, включая перераспределение смыслового содержания между ними. В отличие от версионных классификационных систем, для которых установлен период их обновления (например, рубрики Международной патентной классификации могут меняться не чаще, чем раз в квартал [1]), добавления и изменения в ДКС в случае необходимости могут быть сделаны в любой момент времени. С одной стороны, эта особенность ДКС позволяет оперативно отражать в них новое знание и сразу начинать его использовать, например в процессе лингвистического аннотирования [2]. С другой стороны, если некоторая рубрика использовалась при аннотировании, а затем была изменена, то аннотации с этой рубрикой, сформированные до внесения изменений, в ряде случаев должны быть реклассифицированы [3].

Цель статьи состоит в сопоставлении подходов к классификации сущностей на основе (1) ДКС и (2) онтологий, изменяемых во времени, а также в описании специфики представления новых

лексикографических знаний в ДКС. В качестве примера ДКС в статье рассматривается фасетная классификация (ФК) надкорпусной базы данных (НБД) [4–6]. В проекте по гранту № 20-012-00166 НБД используется для аннотирования таких сущностей, как употребления немецких модальных глаголов (НМГ) в параллельных текстах [7–9], в процессе которого могут (1) обнаруживаться новые значения НМГ, (2) добавляться новые рубрики для этих значений в момент их обнаружения [10, 11] и (3) изменяться дефиниции рубрик ФК [3].

2 Онтологии, изменяемые во времени

Методы аннотирования сущностей с использованием, с одной стороны, ДКС и, с другой стороны, онтологий, изменяемых во времени, во многом схожи. При аннотировании сущностей с помощью ДКС им присваиваются рубрики, в том числе добавленные или измененные непосредственно в процессе формирования какой-либо аннотации. При классификации сущностей с помощью онтологий для них устанавливаются атрибуты, представляющие собой ссылки на концепты онтологий. С те-

* Работа выполнена в Институте проблем информатики ФИЦ ИУ РАН при поддержке РФФИ (проект 20-012-00166).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

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

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, magnit75@yandex.ru

чением времени могут меняться структура и наполнение как ДКС, так и онтологии. Это может быть связано с изменениями (1) самой предметной области, (2) экспертных знаний об этой области и/или (3) стандартов ее описания [12].

Кроме этого, при проведении исследований научные коллективы могут пользоваться теми классификационными системами и онтологиями для своей предметной области, которые находятся в открытом доступе, а затем модифицировать их в зависимости от целей и задач выполняемых проектов. В силу перечисленных сходств работы, связанные с эволюцией и версионностью онтологий, представляют значительный интерес для разработки средств актуализации ДКС.

В [12] описывается подход к управлению версионностью онтологий и предлагается система нумерации версий онтологий, позволяющая определять, обладают ли версии свойством обратной совместимости, или эквивалентности (с точностью до синтаксических различий). Обратная совместимость важна потому, что в случае ее сохранения снимается необходимость в реклассификации ранее сформированных аннотаций (об этой проблеме применительно к ДКС на примере ФК НБД см. [3]). Сохранять историю изменений понятий онтологии предлагается либо в отдельной онтологии, либо в выделенных для этого классах исходной онтологии. В [13] представлены инструменты, позволяющие выделять и визуализировать структурные различия между версиями одной и той же онтологии.

В [14] показано, как с помощью онтологий могут фиксироваться временные данные, знания, правила и отношения, и рассмотрены подходы к описанию изменений в онтологиях. Темпоральная дескриптивная логика (Temporal Description Logics) служит основой дополнения онтологии логическими функциями для работы со временем, позволяющими выражать и обрабатывать такие концепты, как «постоянно в прошлом» или «в некоторый момент в будущем». Для регистрации фактов, относящихся к определенному периоду времени, используются подходы, позволяющие обходить ограничения большинства онтологий, опирающихся на язык Web Ontology Language (OWL): реификация (Reification), четырехмерные (4D-fluent) онтологии и переход к n -арным отношениям (N -ary relations). Поскольку по умолчанию отношения в онтологиях OWL являются бинарными, тогда как для фиксации временного интервала события необходим по крайней мере один дополнительный аргумент, все перечисленные подходы характеризуются тем, что для описания временных фактов в онтологиях каждый раз создается новая сущность.

На основе вышеназванных подходов разработана структура SOWL (spatio-temporal OWL), предназначенная для включения пространственно-временной информации в онтологии OWL, а также инструмент CHRONOS, упрощающий создание и редактирование данных об изменениях в онтологиях, использующих эту структуру. Кроме того, для извлечения информации из SOWL-онтологий создан язык запросов, описанный в [15].

В [16] дается обзор процессов и алгоритмов, связанных с развитием онтологий (ontology evolution). Комплекс таких процессов описан как единый и непрерывный цикл, который можно разделить на несколько этапов:

- (1) выявление потребности в изменениях;
- (2) формулировка предлагаемых изменений;
- (3) оценка адекватности предлагаемых изменений;
- (4) оценка последствий реализации изменений;
- (5) внедрение изменений.

Каждый этап рассматривается отдельно с учетом опыта, полученного в ходе других исследований.

Следует отметить ряд особенностей подхода, предлагаемого в настоящей работе, в сравнении с [12–16]. Во-первых, он нацелен не на описание любых изменений онтологий во времени, а на фиксирование дополнений, изменений порядка и смыслового содержания рубрик самой ДКС. Таким образом, он ближе всего к работам, где описываются подходы к управлению различными версиями онтологий и развитию онтологий; работы же, посвященные описанию темпоральных сущностей в онтологиях, как правило, имеют иную целевую направленность.

Во-вторых, хотя структура онтологии обычно сложнее, чем структура ДКС, это отличие не имеет принципиального значения для задач классификации — описываемую здесь ФК в качестве примера ДКС можно рассматривать как потенциальную составляющую онтологии знаний по лексикографии немецкого языка. Узкая направленность ФК позволяет уделить больше внимания непосредственно теме исследования в рамках упомянутого проекта, что немаловажно, поскольку выделение значений НМГ является нетривиальной задачей, и помимо таких операций, как слияние или разделение значений, порой требуется перенести часть компонентов смыслового содержания из одного значения в другое, изменить нумерацию значений в соответствии с последовательностью их описания в словарных статьях, предназначенных для включения в словарь [17], и т. д.

В-третьих, описываемая ФК вложена в НБД — объекты аннотирования и сама ФК физически располагаются в одной и той же базе данных, благодаря чему разработчикам не нужно беспокоиться о том, что какие-либо внешние данные окажутся несовместимыми с той или иной версией ФК. Однако при внесении изменений в ФК важно в случае необходимости сразу же вносить соответствующие изменения в аннотации, сформированные ранее. Если возможно, это делается автоматически, в противном случае система должна пометать аннотации, затронутые изменениями, чтобы их реклассифицировали эксперты [3].

3 Динамическая классификационная система

В отличие от упомянутых выше решений по поддержке версионности онтологий в предлагаемом подходе и его реализации в проекте не предусмотрена нумерация версий или создание новой версии ФК после вносимых изменений. Вместо этого для каждой рубрики ФК в НБД фиксируются все ее изменения и для них проставляются временные штампы, а все старые версии сохраняются. Поэтому можно проследить историю эволюции каждой рубрики, включая ее взаимодействие с другими рубриками, а также восстановить состояние ФК на любой момент времени.

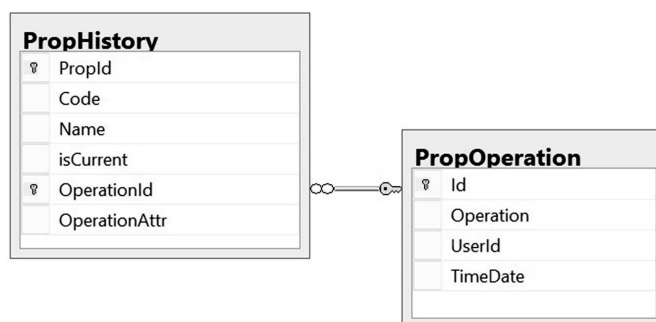
В структуре НБД за сохранение изменений отвечают две связанные между собой таблицы: в первой

(PropHistory) хранятся все — текущие и устаревшие — состояния рубрик ФК (которые в проекте соответствуют значениям НМГ), а во второй (PropOperation) — все операции изменения, которые применялись к этим рубрикам. Структура таблиц и связь между ними показаны на рисунке.

Различаются следующие виды операций.

1. CREATE — создание новой рубрики.
2. REORDER — изменение кода рубрики (соответствует номеру значения НМГ по словарю [17]).
3. REVISE — изменение дефиниции рубрики (не затрагивающее никакие другие рубрики).
4. MERGE — слияние дефиниций двух рубрик, в результате которого одна из рубрик удаляется.
5. DELETE — удаление рубрики.
6. SPLIT — разделение дефиниции одной рубрики на две, в результате чего создается новая рубрика, а дефиниция исходной рубрики перераспределяется между старой и новой рубриками.
7. REDISTR (от англ. *redistribute*) — изменение дефиниций двух рубрик, предусматривающее перенос части компонентов смыслового содержания из одной рубрики в другую.

В то время как для понимания операций 1–6 достаточно приведенных определений, операция REDISTR заслуживает более детального рассмотрения. Для этого понадобятся условные обозначения, введенные в [3], а именно:



Таблицы для хранения истории изменений рубрик ФК в НБД. Поля таблицы **PropHistory**: **PropId** — уникальный идентификатор рубрики (соответствующий значению НМГ); **Code** — краткое обозначение рубрики (для значений НМГ обычно имеет следующий вид: модальный глагол, дефис, номер значения, например, «sollen-01»); **Name** — дефиниция рубрики; **isCurrent** — признак актуальности данного состояния рубрики (1 — актуально, 0 — не актуально); **OperationId** — идентификатор операции, в результате которой рубрика приняла вид, соответствующий данной строке таблицы; **OperationAttr** — атрибут, который присваивается рубрике в рамках операции с Id, указанным в поле OperationId (см. об атрибутах и их значениях в операциях в табл. 3). Поля таблицы **PropOperation**: **Id** — идентификатор операции (обеспечивает связь между операциями и рубриками, которые они затрагивают); **Operation** — задает тип операции (допустимые значения: CREATE, DELETE, MERGE, SPLIT, REORDER, REDISTR, REVISE); **UserId** — идентификатор пользователя, осуществившего операцию; **TimeDate** — дата и время выполнения операции

Таблица 1 Пример выделения компонентов смыслового содержания дефиниции рубрики

Рубрика	Id	Код	Дефиниция рубрики без структурного выделения компонентов смыслового содержания	Дефиниция рубрики, в которой выделены компоненты ее смыслового содержания
X	482	sollen-01	Обязанность что-л. делать по чье-му-л. указанию, по закону, по правилам и т. п.: должен. Моральный запрет (под отрицанием): нельзя	a. Обязанность что-л. делать по чьему-л. указанию, по закону, по правилам и т. п.: должен b. Моральный запрет (под отрицанием): нельзя

- X, Y, ... — рубрики ФК, обозначающие смысловые значения НМГ;
- def_X, def_Y, \dots — дефиниции рубрик;
- $S_{def_X}, S_{def_Y}, \dots$ — смысловое содержание дефиниций рубрик.

Кроме того, для обозначения сущностей, которые были каким-то образом изменены в результате выполнения операции, используется индекс «ch» (от англ. *changed*).

Операция REDISTR выполняется только в том случае, если def_X такова, что в S_{def_X} — смысловом содержании дефиниции рубрики (соответствующей значению НМГ) — выделены компоненты (соответствующие подзначениям внутри значения НМГ¹). В табл. 1 приводится пример рубрики X, в дефиниции которой структурно выделены компоненты ее смыслового содержания, соответствующие подзначениям, — a и b.

Если представить S_{def_X} , компонентами которого являются a и b, в виде множества $S_{def_X}\{a, b\}$, то изменение def_X может быть таким, что набор элементов этого множества: (1) сократится: $S_{def_X}^1 = \{a\}$; (2) увеличится: $S_{def_X}^2 = \{a, b, c, \dots\}$.

Более того, возможна ситуация, когда изменение набора элементов затронет не только def_X , а одновременно def_X и def_Y . Рубрики до внесения изменений обозначим как X и Y, а после их внесения — как X_{ch} и Y_{ch} . Если изменение рубрик X и Y таково, что $S_{def_X} \cap S_{def_Y}^{ch} \neq \emptyset$ (смысловое содержание def_X и смысловое содержание измененной def_Y имеют один или более общих компонентов) и/или $S_{def_Y} \cap S_{def_X}^{ch} \neq \emptyset$ (смысловое содержание def_Y и смысловое содержание измененной def_X имеют один или более общих компонентов), то оно может быть описано с помощью операции REDISTR (X, Y). Экспертная реклассификация требуется для тех аннотаций, которые до выполнения операции REDISTR содержали: (1) рубрику X, смысловое содержание которой в результате выполнения операции REDISTR сужается; (2) рубрику X

или Y, если перенос компонентов смыслового содержания дефиниций осуществляется как из X в Y, так и из Y в X.

Ниже приводится пример выполнения операции REDISTR для рубрик с постоянными номерами (id) 482 и 484, которые позволяют отслеживать историю изменений рубрики (табл. 2). В данном примере: X — рубрика 482 до выполнения операции, причем $S_{def_X} = \{a, b\}$; Y — рубрика 484 до выполнения операции, причем $S_{def_Y} = \{m, n, o\}$; X_{ch} — рубрика 482 после выполнения операции, причем $S_{def_X}^{ch} = \{a, b, n, o\}$; Y_{ch} — рубрика 484 после выполнения операции, причем $S_{def_Y}^{ch} = \{m\}$.

Поскольку $S_{def_Y} \cap S_{def_X}^{ch} = \{n, o\}$ и $S_{def_X} \cap S_{def_Y}^{ch} = \emptyset$, смысловое содержание рубрики 482 после внесения изменения расширяется, а смысловое содержание рубрики 484 — сужается. Следовательно, аннотации, которые до выполнения операции REDISTR содержали рубрику 482, после ее выполнения не требуют экспертной реклассификации и автоматически обозначаются кодом «sollen-01», тогда как аннотации, которые до выполнения операции REDISTR содержали рубрику 484, после ее выполнения требуют экспертной реклассификации и поэтому автоматически помечаются тегом «TBR-R» (от англ. *To Be Reclassified because of Redistribution*).

Может показаться, что введение операции REDISTR не оправданно, так как рассмотренное перераспределение компонентов значений между 482 и 484 можно описать последовательностью операций SPLIT (MERGE (X, Y)). Однако при таком подходе объем реклассификации может существенно возрасти. Объем реклассификации будет тот же, лишь если верно одновременно и $S_{def_X} \cap S_{def_Y}^{ch} \neq \emptyset$, и $S_{def_Y} \cap S_{def_X}^{ch} \neq \emptyset$, т. е. осуществляется перенос компонентов смыслового содержания одновременно и из S_{def_X} в S_{def_Y} , и из S_{def_Y} в S_{def_X} . Если же верно или $S_{def_X} \cap S_{def_Y}^{ch} \neq \emptyset$, а $S_{def_Y} \cap S_{def_X}^{ch} = \emptyset$, или $S_{def_Y} \cap S_{def_X}^{ch} \neq \emptyset$, а $S_{def_X} \cap S_{def_Y}^{ch} = \emptyset$ (т. е.

¹С точки зрения смыслового содержания в дефиниции рубрики могут быть выделены и более мелкие части, соответствующие частям подзначений. Данная ситуация не рассматривается в рамках настоящей статьи.

Таблица 2 Исходные данные и результат выполнения операции REDISTR

Рубрика	Id	Код	Дефиниции двух рубрик, в которых структурно выделены компоненты их смыслового содержания, до и после выполнения операции REDISTR
X	482	sollen-01	<i>a.</i> Обязанность что-л. делать по чьему-л. указанию, по закону, по правилам и т. п.: должен <i>b.</i> Моральный запрет (под отрицанием): нельзя
Y	484	sollen-03	<i>m.</i> Желательность по мнению говорящего (в формах <i>praet conj</i> и <i>ppr conj</i>): следовало (бы), нужно было (бы), должно было (бы) <i>n.</i> Совет, рекомендация (только в формах <i>praet conj</i>) <i>o.</i> Нежелательность (под отрицанием): не следовало (бы), нельзя
X _{ch}	482	sollen-01	<i>a.</i> Обязанность что-л. делать по чьему-л. указанию, по закону, по правилам и т. п.: должен <i>b.</i> Моральный запрет (под отрицанием): нельзя <i>n.</i> Совет, рекомендация (только в формах <i>praet conj</i>) <i>o.</i> Нежелательность (под отрицанием): не следовало (бы), нельзя
Y _{ch}	484	sollen-03	<i>m.</i> Желательность по мнению говорящего (в формах <i>praet conj</i> и <i>ppr conj</i>): следовало (бы), нужно было (бы), должно было (бы)

Примечания. Расшифровка используемых в таблице сокращений:

- *praet conj* — форма прошедшего времени (лат. *praeteritum*) сослагательного наклонения (лат. *conjunctivus*);
- *ppr conj* — форма предпрошедшего времени (лат. *plusquamperfectum*) сослагательного наклонения (лат. *conjunctivus*).

Примеры употребления глагола *sollen* по словарю [17], иллюстрирующие каждый из компонентов смыслового содержания дефиниций (формы глагола *sollen* выделены полужирным шрифтом):

- a.* ich **soll** heute noch in die Stadt fahren — я должен сегодня еще поехать в город;
- b.* du **sollst** nicht töten! — не убий! (*библейская заповедь*);
- m.* das **sollte** sie doch wissen — это она (вообще-то) должна была (бы) знать;
- n.* Sie **sollten** mit dem Rauchen aufhören — вам следует бросить курить;
- o.* das **sollte** man nie tun — этого не следует делать.

Таблица 3 Соответствия между операциями, атрибутами и рубриками

Операция	Атрибут	Рубрика
CREATE (создается новая рубрика X)	A	Создаваемая рубрика X
	C	Рубрики, которые надо перенумеровать после выполнения операции, чтобы освободить в нумерации нужную позицию для X
REORDER (код рубрики X изменяется)	A	Рубрика X, код которой изменяется
	C	Рубрики, которые надо перенумеровать после выполнения операции, чтобы освободить в нумерации нужную позицию для X
REVISE (изменяется def _X)	A	Рубрика X, дефиниция которой изменяется
MERGE (объединяются def _X и def _Y)	A	Рубрика X, причем def _X поглощает def _Y , а X остается в базе данных
	B	Рубрика Y, причем def _Y включается в def _X , а Y удаляется из базы данных
	C	Рубрики, которые напрямую не затрагиваются операцией MERGE, но которые надо перенумеровать после выполнения операции, чтобы заполнить пробелы в нумерации, образовавшиеся из-за удаления Y
DELETE (рубрика X удаляется)	A	Удаляемая рубрика X
	C	Рубрики, которые надо перенумеровать после выполнения операции, чтобы заполнить пробелы в нумерации, образовавшиеся из-за удаления X
SPLIT (def _X делится на две части — def _{X1} и def _{X2})	A	Рубрика X, дефиниция которой делится на две части, причем def _{X1} становится новой дефиницией X
	B	Рубрика Y, которая создается в базе данных, причем def _{X2} становится дефиницией Y
REDISTR (def _X и def _Y изменяются так, что происходит перераспределение компонентов смыслового содержания между S _{def_X} и S _{def_Y})	A	Рубрика X, если S _{def_X} расширяется за счет переноса компонентов смыслового содержания из S _{def_Y} , причем ни один компонент S _{def_X} не переносится в S _{def_Y}
	B	Рубрика Y, если S _{def_Y} сужается за счет переноса компонентов смыслового содержания в S _{def_X} , причем ни один компонент S _{def_X} не переносится в S _{def_Y}
	AB	Рубрики X и Y, если одновременно осуществляется перенос компонентов смыслового содержания из S _{def_X} в S _{def_Y} и из S _{def_Y} в S _{def_X}

смысловое содержание дефиниции одной рубрики расширяется, а другой — сужается), то в таком случае реклассификация нужна только для аннотаций, содержащих рубрику с дефиницией, смысловое содержание которой в результате изменения сужается (в примере из табл. 2 это аннотации, содержащие рубрику Y). Таким образом, введение операции REDISTR оправданно, так как позволяет сократить объем реклассификации.

Рассмотренный пример (см. табл. 2) показывает, что операция REDISTR по-разному влияет на рубрики 482 и 484: смысловое содержание дефиниции рубрики 482 расширяется, а рубрики 484 — сужается. Для того чтобы на основе таблиц НБД с историей изменения рубрик иметь возможность определять, как именно операция повлияла на некоторую рубрику, каждой из рубрик, затронутых операцией, присваивается атрибут. В этом примере рубрике 482 будет присвоен атрибут А, а рубрике 484 — В. Все возможные атрибуты и их смысловое содержание для каждой из семи операций приведены и расшифрованы в табл. 3.

4 Заключение

Предлагаемый подход к ведению ДКС дает возможность сохранять всю информацию об изменениях рубрик ФК, фиксировать виды изменений, время их внесения и данные о пользователе, который их внес. Он позволяет отслеживать совершенные изменения в хронологическом порядке, а также, при необходимости, восстанавливать состояние ДКС на любой момент времени в прошлом.

Литература

1. Зацман И. М., Косарик В. В., Курчавова О. А. Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
2. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1468 p.
3. Гончаров А. А., Зацман И. М., Кружков М. Г. Эволюция классификаций в надкорпусных базах данных // Информатика и её применения, 2020. Т. 14. Вып. 4. С. 108–116.
4. Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А. Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
5. Зализняк А., Зацман И. М., Инькова О. Ю. Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–108.
6. Зацман И. М., Кружков М. Г. Надкорпусная база данных коннекторов: развитие системы терминов проектирования // Системы и средства информатики, 2018. Т. 28. № 4. С. 156–167.
7. Добровольский Д. О., Зализняк Анна А. Немецкие конструкции с модальными глаголами и их русские соответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. С. 172–184.
8. Добровольский Д. О. Немецкие модальные глаголы в параллельном корпусе и задачи двуязычной лексикографии // Германские языки: текст, корпус, перевод. — М.: Институт языкознания РАН, 2020. С. 103–116.
9. Добровольский Д. О., Зализняк Анна А. Русские конструкции с потенциально модальным значением по данным параллельных корпусов // Труды Института русского языка им. В. В. Виноградова, 2020. № 3. С. 35–49.
10. Zatsman I. Finding and filling lacunas in linguistic typologies // 15th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera: Institute of Knowledge Asset Management, 2020. P. 780–793.
11. Zatsman I. Three-dimensional encoding of emerging meanings in AI-systems // 21st European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2020. P. 878–887.
12. Klein M., Fensel D., De A. Ontology versioning on the Semantic Web // 1st Conference (International) on Semantic Web Working Proceedings. — Stanford, CA, USA: Stanford University, 2001. P. 75–91.
13. Noy N., Kunnatur S., Klein M., Musen M. Tracking changes during ontology evolution // International Semantic Web Conference / Eds. S. A. McIlraith, D. Plexousakis, F. van Harmelen. — Lecture notes in computer science ser. — Springer, 2004. Vol. 3298. P. 259–273.
14. Preventis A., Petrakis E. G. M., Batsakis S. CHRONOS Ed: A tool for handling temporal ontologies in protégé // Int. J. Artif. Intell. T., 2014. Vol. 23. No. 4. P. 1460018-1–1460018-26. doi: 10.1142/S0218213014600185.
15. Stravoskoufos K., Petrakis E., Mainas N., Batsakis S., Samoladas V. SOWL QL: Querying spatio-temporal ontologies in OWL // J. Data Semantics, 2016. Vol. 5. No. 4. P. 249–269.
16. Zablith F., Antoniou G., D'Aquin M., Flouris G., Kondylakis H., Motta E., Plexousakis D., Sabou M. Ontology evolution: A process-centric survey // Knowl. Eng. Rev., 2015. Vol. 30. No. 1. P. 45–75.
17. Немецко-русский словарь актуальной лексики / Под ред. Д. О. Добровольского. — М.: Лексрус, 2021 (в печати).

Поступила в редакцию 12.01.2021

REPRESENTATION OF NEW LEXICOGRAPHICAL KNOWLEDGE IN DYNAMIC CLASSIFICATION SYSTEMS

A. A. Goncharov, I. M. Zatsman, and M. G. Kruzhkov

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

Abstract: The distinctive feature of dynamic classification systems is that new categories may be introduced in the course of their use or definitions of existing categories may be modified, including cases of rearranging semantic content between categories. On one hand, this feature of dynamic classification systems provides a possibility to integrate new knowledge on-the-fly and to start using it immediately for linguistic annotation. On the other hand, if a category is changed, then, in some cases, the annotations it has been previously applied to will have to be reclassified. This paper has a twofold purpose, which is, first, to compare approaches to classification of entities based on (i) dynamic classification systems and (ii) ontologies that change over time; and then, second, to describe how new lexicographical knowledge is represented in dynamic classification systems.

Keywords: dynamic classification system; ontology versioning; linguistic annotation; reclassification of annotations

DOI: 10.14357/19922264210112

Acknowledgments

The study has been conducted at the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences (FRC CSC RAS) with financial support of the Russian Foundation for Basic Research (grant No. 20-012-00166).

References

1. Zatsman, I. M., V. V. Kosarik, and O. A. Kurchavova. 2008. Zadachi predstavleniya lichnostnykh i kollektivnykh konseptov v tsifrovoy srede [Representation of individual and collective concepts in digital medium]. *Informatika i ee Primeneniya — Inform. Appl.* 2(3):54–69.
2. Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1468 p.
3. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2020. Evolyutsiya klassifikatsiy v nadkorpornykh bazakh dannykh [Evolution of classifications in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 14(4):108–116.
4. Zatsman, I. M., O. Yu. Inkova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie kross-yazykovykh znaniy o konnektorakh v nadkorpornykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
5. Zaliznyak, A. A., I. M. Zatsman, and O. Yu. In'kova. 2017. Nadkorpornaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database of connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–108.
6. Zatsman, I. M., and M. G. Kruzhkov. 2018. Nadkorpornaya baza dannykh konnektorov: razvitie sistemy terminov proektirovaniya [Supracorpora database of connectives: Design-oriented evolution of the term system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):156–167.
7. Dobrovolskiy, D. O., and Anna A. Zaliznyak. 2018. Nemetskie konstruksii s modal'nymi glagolami i ikh russkie sootvetstviya: proekt nadkorpornoy bazy dannykh [German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Komp'yuternaya lingvistika i intellektual'nyye tekhnologii: po mat-lam Mezh-dunar. konf. "Dialog"* [Computational Linguistics and Intellectual Technologies. Papers from the Annual Conference (International) “Dialogue”]. Moscow: RSHI. 17(24):172–184.
8. Dobrovolskiy, D. O. 2020. Nemetskie modal'nye glagoly v paralle'nom korpuse i zadachi dvuyazychnoy leksikografii [German modal verbs in a parallel corpus and bilingual lexicography tasks]. *Germanskie yazyki: tekst, korpus, perevod* [German languages: Text, corpus, translation]. Moscow: Institute of Linguistics RAS. 103–116.
9. Dobrovolskiy, D. O., and Anna A. Zaliznyak. 2020. Russkie konstruksii s potentsial'no modal'nym znacheniem po dannym paralle'nykh korpusov [Russian constructions with potentially modal meanings: An analysis based on parallel corpus data]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [V. V. Vinogradov Russian Language Institute Proceedings]. 35–49.
10. Zatsman, I. 2020. Finding and filling lacunas in linguistic typologies. *15th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera: Institute of Knowledge Asset Management. 780–793.
11. Zatsman, I. 2020. Three-dimensional encoding of emerging meanings in AI-systems. *21st European Conference on*

- Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 878–887.
12. Klein, M., D. Fensel, and A. De. 2001. Ontology versioning on the semantic web. *1st Conference (International) on Semantic Web Working Proceedings*. Stanford, CA: Stanford University. 75–91.
 13. Noy, N., S. Kunnatur, M. Klein, and M. Musen. 2004. Tracking changes during ontology evolution. *International Semantic Web Conference*. Eds. S. A. McIlraith, D. Plexousakis, and F. van Harmelen. Lecture notes in computer science ser. Springer. 3298:259–273.
 14. Preventis, A., E. G. M. Petrakis, and S. Batsakis. 2014. CHRONOS Ed: A tool for handling temporal ontologies in protégé. *Int. J. Artif. Intell. T.* 23(4):1460018. 26 p. doi: 10.1142/S0218213014600185.
 15. Stravoskoufos, K., E. Petrakis, N. Mainas, S. Batsakis, and V. Samoladas. 2016. SOWL QL: Querying spatio-temporal ontologies in OWL. *J. Data Semantics* 5(4):249–269.
 16. Zablith F., G. Antoniou, M. D’Aquin, G. Flouris, H. Kondylakis, E. Motta, D. Plexousakis, and M. Sabou. 2015. Ontology evolution: A process-centric survey. *Knowl. Eng. Rev.* 30(1):45–75.
 17. Dobrovol’skiy, D.O., ed. 2021 (in press). *Nemetsko-russkiy slovar’ aktual’noy leksiki* [German–Russian dictionary of actual vocabulary]. Moscow: Leksrus.

Received January 12, 2021

Contributors

Goncharov Alexander A. (b. 1994) — junior 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; a.gonch48@gmail.com

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

Kruzhkov Mikhail G. (b. 1975) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ АКТУАЛИЗАЦИЯ СЛОВАРНЫХ СТАТЕЙ ДВУЯЗЫЧНЫХ СЛОВАРЕЙ И МЕДИЦИНСКОЙ ТЕРМИНОЛОГИИ: СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ*

И. М. Зацман¹

Аннотация: Сопоставляются два подхода к целенаправленному извлечению нового знания из текстовых данных. Первый подход относится к предметной области лексикографии и ориентирован на извлечение новых значений слов из текстов в интересах пополнения словарных статей двуязычных словарей. Второй подход относится к медицинской науке и ориентирован на извлечение новых значений терминов в интересах актуализации терминологического портрета болезни, включающего дефиниции терминов с отражением их динамики во времени, отношения между терминами, контексты их использования и ссылки на источники контекстов. Эти подходы сравниваются по следующим позициям: *проблема*, для решения которой извлекается новое знание, *цель* его извлечения, *источник* концептов нового знания, *эталон*, сравнение с которым используется как критерий новизны, *связь* концепта с его источником и *динамика* концептов. Цель статьи состоит в описании результатов сопоставительного анализа обоих подходов. Результаты анализа предлагается позиционировать как исходные данные для создания концепции системы искусственно-естественного интеллекта (ИЕИ-системы) для целенаправленного извлечения нового знания из данных большого объема, применимой в разных предметных областях.

Ключевые слова: генерация нового знания; извлечение знания из текстов; искусственный интеллект; система искусственно-естественного интеллекта

DOI: 10.14357/19922264210113

1 Введение

Сопоставляются два подхода к извлечению нового знания из текстовых данных. Первый подход относится к предметной области лексикографии и ориентирован на извлечение новых значений языковых единиц (ЯЕ) из параллельных текстов художественных произведений в интересах пополнения словарных статей двуязычных словарей. Второй подход относится к медицинской науке и ориентирован на извлечение новых значений терминов в интересах актуализации терминологического портрета болезни².

Разработка первого подхода для целенаправленного извлечения новых значений ЯЕ и их переводов из параллельных текстов в интересах создания новых версий словарных статей о немецких модальных глаголах (НМГ) является целью проекта по гранту РФФИ № 20-012-00166, который в на-

стоящее время выполняется в Институте проблем информатики ФИЦ ИУ РАН [1, 2]. Предлагаемый подход, описанию которого посвящен второй раздел статьи, предполагает применение двух методов.

Первый метод служит для нахождения лакун в типологии значений НМГ и их переводов на русский язык (см. структуру типологии на рис. 3 в [3]). Для нахождения лакун применяется процедура лингвистического аннотирования предложений с НМГ [4, 5].

Второй метод предназначен для заполнения лакун, обнаруженных с помощью первого метода, что обеспечивает целенаправленное пополнение существующей типологии [5, 6].

Разработка второго подхода и его применение для извлечения из текстов новых значений медицинских терминов в интересах актуализации терминологического портрета исследуемой болезни

* Исследование выполнено при поддержке РФФИ (проект 20-012-00166) и по гранту РФФИ и Государственного фонда естественных наук Китая № 21-57-53018.

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

² В рамках второго подхода портрет болезни — это множество аннотаций терминов, используемых при описании болезни. Аннотация содержит: (1) структурированную дефиницию значения термина, в которой могут быть выделены его подзначения; (2) ретроспективу ее изменений во времени; (3) отношения с другими терминами; (4) контексты использования термина; (5) его синонимы; (6) ссылки на источники контекстов.

является одной из задач проекта по гранту РФФИ и Государственного фонда естественных наук Китая № 21-57-53018. Цель проекта состоит в разработке теоретических основ создания интеллектуальной системы извлечения из текстовых данных знания о проблеме раннего выявления и прогнозирования заболеваний с использованием регулярно актуализируемых терминологических портретов болезней, содержащих аннотации терминов.

Второй подход, описанию которого посвящен третий раздел статьи, основан на обработке научных текстов по медицинской науке с целью формирования и итерационного обновления терминологического портрета болезни. До начала первой итерации экспертами формируется исходная версия терминологического портрета на основе существующих словарей медицинских терминов и их дефиниций [7–9]. На первой итерации исходная версия портрета используется для программного поиска и извлечения новых терминов из коллекции текстов большого объема в целях его обновления, в результате которого формируется первая версии портрета. На второй и последующих итерациях используются уже обновленные портреты для извлечения новых терминов. Итерационный процесс завершается экспертами, если прекращается значимое с их точки зрения обновление портрета. Источником новых терминов служат коллекции, содержащие аннотации научных статей по медицинской науке и их полные тексты. В качестве одного из наиболее известных примеров такой коллекции можно привести базу данных PubMed, содержащую более 30 млн записей и обеспечивающую развитые возможности поиска и формирования статистических данных, включая получение распределения статей с заданным термином по годам их публикации [10].

Цель статьи состоит в описании результатов сопоставительного анализа двух подходов, используемых в лингвистике и медицинской науке. Результаты сопоставления предлагается позиционировать как исходные данные для создания концепции ИЕИ-системы для целенаправленного извлечения нового знания из данных большого объема.

2 Первый подход

Основная идея целенаправленного извлечения новых значений ЯЕ и их переводов из параллельных текстов заключается в следующем. Сначала формулируется проблема, например обновление двуязычных словарей за счет извлечения новых значений ЯЕ некоторого класса. В проекте по гранту РФФИ № 20-012-00166 исследуемыми ЯЕ служат 6 НМГ (*dürfen, können, mögen, müssen, sollen* и *wollen*), новыми извлеченными значениями которых пополняется немецко-русский словарь [11]. Вариант этого же словаря по состоянию на начало реализации проекта был выбран как эталон, отражающий современный уровень знания о НМГ. Информация его словарных статей о значениях НМГ и их переводах была представлена в виде исходной версии лингвистической типологии, структура которой приведена на рис. 3 в [3].

После формулировки проблемы (обновление двуязычного словаря), определения эталона и формирования исходной версии лингвистической типологии экспертами выбирается массив параллельных текстов как потенциальных источников нового знания о ЯЕ исследуемого класса. Для поиска новых значений НМГ использовались тексты, полученные из параллельного немецкого подкорпуса Национального корпуса русского языка [12]. Эти тексты включают 2,6 млн словоупотреблений: 1,4 млн словоупотреблений в оригинальных текстах на немецком языке, включая 16 268 употреблений НМГ, и 1,2 млн словоупотреблений в их переводах на русский язык (см. табл. 1 с примерами предложений и их переводами, которые содержат 3 из 6 исследуемых НМГ, выделенных полужирным шрифтом).

Пополнение исходной версии типологии происходит в процессе обработки параллельных текстов с использованием упомянутых в первом разделе методов и разработанной на их основе информационной технологии поиска новых значений НМГ [13]. Сначала в параллельных текстах программно осуществляется поиск предложений, которые содержат исследуемые ЯЕ.

Таблица 1 Примеры предложений немецко-русских параллельных текстов

Sollte jetzt etwa eine Predigt stattfinden? [Franz Kafka. Der Prozess (1914)]	Неужели сейчас кто-то будет читать проповедь? [Франц Кафка. Процесс (Р. Райт-Ковалева, 1965)]
Warum mußte er diese Demoiselle Stüwing heiraten und den. . . Laden. . . [Thomas Mann. Buddenbrooks (1896–1900)]	Зачем ему понадобилось жениться на этой мадемуазель Штювинг с ее. . . лавкой? [Томас Манн. Будденброки (Н. Ман, 1953)]
Er durfte nun eine Weile lang guten Gewissens ruhen. [Patrick Süskind. Das Parfum: Die Geschichte eines Mörders (1985)]	Теперь он имел право некоторое время отдыхать. [Патрик Зюскинд. Парфюмер: История одного убийцы (Э. Венгерова, 1992)]

Для предложений и их переводов, найденных по леммам НМГ, заполняются поля структурированных аннотаций (см. примеры двух аннотаций в табл. 2 и 3 в [14]), включая поле значения исследуемой ЯЕ в найденном предложении. В целях его заполнения лингвист сопоставляет смысловое содержание ЯЕ в найденном предложении с дефинициями значений ЯЕ, уже имеющимися в текущей версии типологии. Если такая дефиниция в ней найдена, то лингвист проставляет в аннотации код этой дефиниции (см. код «sollen-3» в табл. 2 в [14]). Если она отсутствует, то в этом поле аннотации проставляется специальный код с меткой «х» (см. код «sollen-х» в табл. 3 в [14]), говорящий о том, что обнаружена лакуна в текущей версии типологии и поэтому процесс формирования аннотации не может быть завершен без пополнения типологии. Коды с литерой «х» помечают неполные (= незавершенные) аннотации относительно *текущего состояния типологии*.

На основе сформулированных положений по извлечению новых значений ЯЕ были разработаны два метода [15]:

- (1) для обнаружения понятийных лакун в текущей версии типологии значений НМГ;
- (2) для целенаправленного заполнения обнаруженных лакун.

Каждая итерация первого метода включает следующие шаги (итерация выполняется для одного употребления исследуемой ЯЕ):

- в параллельных текстах осуществляется поиск предложения, которое содержит исследуемую ЯЕ;
- в найденном предложении с ЯЕ (см. их примеры, найденные по леммам НМГ *sollen*, *müssen* и *dürfen*, в первом столбце табл. 1) в результате семантического анализа лингвистами определяется значение ЯЕ в контексте этого предложения;
- для найденного предложения и его перевода формируется структурированная аннотация контекста НМГ, которая содержит код дефиниции значения ЯЕ (по текущей версии типологии) или код с меткой «х», который обозначает аннотацию, процесс формирования которой не завершен из-за отсутствия в типологии обнаруженного значения ЯЕ, т. е. в ней обнаружена понятийная лакуна (см. табл. 2 и 3 в [14] соответственно), а также гиперссылка на источник контекста с найденной ЯЕ;
- если сформирована незавершенная аннотация с кодом с меткой «х», то она передается на дальнейшую обработку с помощью второго метода;

– аннотированное предложение помечается как обработанное первым методом и осуществляется переход к первому шагу.

Незавершенные аннотации обрабатываются экспертами высокой квалификации, имеющими, как правило, степень доктора филологических наук, с помощью второго метода заполнения обнаруженных понятийных лакун в типологии. Обработка незавершенных аннотаций, которых, как правило, на три порядка меньше, чем число употреблений ЯЕ, включает следующие шаги:

- для нового значения ЯЕ (т. е. отсутствующего в текущей версии типологии) лингвистами формулируется новая дефиниция или увеличивается объем значения уже существующей дефиниции [16];
- если формулируется новая дефиниция, то она обозначается новым кодом и добавляется в текущую версию типологии;
- в незавершенной аннотации новый код заменяет код с меткой «х», и она становится завершенной;
- если увеличивается объем значения уже существующей дефиниции, то ее код проставляется в незавершенной аннотации;
- вариант перевода нового значения добавляется в типологию.

В заключение описания первого подхода и двух его методов отметим, что семантический анализ незавершенных аннотаций придает целенаправленность процессу заполнения лакун в лингвистической типологии. Это обеспечивается тем, что сначала фиксируются понятийные лакуны в типологии с помощью простановки в незавершенных аннотациях кодов с меткой «х», а затем формируются дефиниции новых значений исследуемых ЯЕ, что и заполняет лакуны в типологии *согласно цели проекта*.

3 Второй подход

Второй подход к обнаружению концептов нового знания относится к медицинской науке и ориентирован на извлечение концептов нового знания из научных текстов в этой предметной области. Сначала формулируется проблема, например раннее выявление и прогнозирование заболеваний с использованием регулярно актуализируемых терминологических портретов болезней. Затем выбирается эталонный словарь, отражающий современный уровень знания о терминах медицинской науки, например [8]. После формулировки проблемы и выбора эталонного словаря определяется коллекция

научных текстов как потенциальных источников нового знания для актуализации портрета, например PubMed [10]. Второй подход основан на итерационном выполнении четырех стадий обработки научных текстов коллекции большого объема. Первые две стадии предназначены для автоматической их обработки, которая существенно сужает круг потенциальных источников нового знания для актуализации портрета экспертами. Существенное сужение необходимо обеспечить на первых двух стадиях из-за последующего подключения экспертов к извлечению нового знания на каждой итерации. Эксперты выполняют две последние стадии семантического анализа потенциальных источников и актуализацию терминологического портрета болезни.

До начала первой итерации автоматической обработки научных текстов коллекции экспертами в области медицинской науки формируется исходная версия терминологического портрета болезни, который включает аннотации, состоящие из 6 компонентов, перечисленных в сноске первого раздела статьи. Эксперты используют, с одной стороны, термины словарей и классификаций [7–9, 17], с другой стороны, термины и их контексты из экспертной выборки статей об исследуемой болезни.

Как было отмечено в разд. 1, на первой итерации исходная версия портрета используется для извлечения из коллекции потенциально новых терминов (далее — термины-кандидаты) и их контекстов в целях пополнения портрета. Потом эксперты выполняют семантический анализ контекстов терминов-кандидатов и отбирают новые термины. Они добавляются в исходную версию портрета, и в конце первой итерации эксперты формируют его первую версию. На второй и последующих итерациях используются уже пополненные портреты для извлечения новых терминов-кандидатов и их контекстов.

Рассмотрим подробнее четыре стадии каждой итерации.

Цель первой стадии автоматической обработки коллекции состоит в создании предметно-ориентированного массива (ПОМ) научных текстов, которые описывают исследуемую болезнь на всех стадиях заболевания, включая ранние. Эти тексты позиционируются как потенциальные источники нового знания об этой болезни, т. е. неотраженного в медицинских классификациях, словарях и справочниках, используемых экспертами проекта. Предметно-ориентированный массив в простейшем случае может быть сформирован в результате поиска по названию болезни и его синонимам. Например, если в коллекции PubMed, содержащей более 30 млн записей [10], задать поиск по запросу «*covid-19*», то будут найдены 87 737 текстов (по

состоянию на 16.01.2021), а если поисковый запрос расширить за счет синонима: («*covid-19*») OR («*coronavirus disease 2019*»), то поисковая выдача увеличится до 88 365 текстов.

На второй стадии каждой итерации из ПОМ научных текстов извлекаются термины-кандидаты. Для их поиска в ПОМ используются термины текущей версии портрета. На основе *поисковых запросов с этими терминами* сначала из ПОМ извлекаются тексты. Затем с помощью лингвистического процессора [18–20] из этих текстов извлекаются те термины, которых нет в текущей версии портрета (т. е. термины-кандидаты), каждый с одним или несколькими контекстами и гиперссылками на тексты-источники этих контекстов. Из этих же текстов дополнительно извлекаются новые контексты терминов портрета, которых нет в его текущей версии, и гиперссылки на тексты-источники этих контекстов. В конце второй стадии каждой итерации термины-кандидаты объединяются с терминами текущей версии портрета в рамках единого массива терминов (ЕМТ) вместе с их контекстами и гиперссылками. Такое объединение позволяет в последующей обработке актуализировать значения терминов текущей версии портрета, если были извлечены их новые контексты из ПОМ.

На третьей стадии выполняется экспертный анализ контекстов каждого термина ЕМТ, чтобы сформировать четыре группы терминов и разделить ЕМТ на четыре соответствующих им подмассива ЕМТ1–ЕМТ4 на основе четырех критериев (первый критерий для первой группы и т. д.):

- (1) термин не имеет отношения к предметной области исследуемой болезни или является общим для терминологических портретов разных болезней;
- (2) контексты термина говорят о новом синониме одного из терминов текущей версии терминологического портрета болезни и/или выбранного эталона;
- (3) новые контексты термина текущей версии портрета, включенные в ЕМТ, говорят о необходимости актуализировать дефиницию его значения;
- (4) термин в контекстах выражает концепт нового знания, т. е. это его значение не представлено ни в текущей версии портрета болезни, ни в выбранном эталоне.

На этой стадии объектами экспертного анализа служат контексты терминов ЕМТ, извлеченные на второй стадии из ПОМ, сформированного на первой стадии каждой итерации. Если в процессе анализа экспертам требуется расширить некоторый

контекст, то используется гиперссылка на текст-источник этого контекста.

Цель четвертой стадии каждой итерации состоит в создании экспертами новой версии терминологического портрета болезни добавлением новых или обновлением уже существующих аннотаций его текущей версии. При создании новой версии портрета:

- термин второй группы добавляется как синоним в соответствующую аннотацию, *если она уже есть в портрете*; в противном случае из эталонного словаря берется термин с дефиницией и на ее основе создается новая аннотация термина с добавлением обнаруженного синонима и его контекстов;
- на основе контекстов термина третьей группы экспертами создается новый вариант его дефиниции, что обновляет существующую аннотацию этого термина;
- для каждого термина четвертой группы создается его аннотация.

В заключение этого раздела отметим, что третья стадия может быть автоматизирована (но не на начальных итерациях), если спроектировать искусственную нейронную сеть для распределения терминов ЕМТ по четырем группам. Предварительно необходимо создать представительный набор подмассивов ЕМТ1–ЕМТ4, который может быть использован для обучения нейронной сети, чтобы потом применить ее на последующих итерациях для автоматического формирования групп терминов и подмассивов.

4 Сопоставление подходов

Рассмотренные подходы сопоставим по следующим позициям: *проблема*, для решения которой извлекается новое знание; *цель* его извлечения; *источник* концептов нового знания; *эталон*, сравнение с которым используется как критерий новизны; *связь* концепта с его источником; *динамика* концептов (изменение их смыслового содержания во времени) (табл. 2).

В последнюю колонку табл. 2 включены исходные положения, на основе которых планируется создать концепцию системы, которую предлагается назвать *системой искусственно-естественного интеллекта*, для целенаправленного извлечения нового знания из данных большого объема с подключением экспертов к его извлечению. Основная идея концепции состоит в том, чтобы эту систему можно было адаптировать для применения в целом ряде предметных областей с учетом их специфики.

Подключение экспертов к извлечению нового знания [21, 22] предполагает, что функционирование ИЕИ-системы будет охватывать ментальную, информационную и цифровую среды, включая границы между ними с интерфейсами второго и третьего порядка [23]. Последние описывают взаимодействие сущностей всех трех сред с помощью полисредовых объектов ИЕИ-систем, объединяющих: (1) концепт ментальной среды знаний экспертов; (2) выражающее его слово как последовательность литер, принадлежащей информационной среде; (3) компьютерный код пары (концепт, литеры) цифровой среды. Понятие полисредовых

Таблица 2 Сравнительный анализ двух подходов

Позиции сравнения	Проект по лексикографии	Проект по медицинской науке	Исходные положения как обобщение
Проблема	Обновление двуязычных словарей	Раннее выявление и прогнозирование заболеваний	Формулировка проблемы
Цель	Актуализация описаний значений НМГ	Актуализация системы терминов и их дефиниций для описания болезни	Формулировка цели извлечения концептов нового знания
Источник	Параллельные тексты художественных произведений	Коллекция научных текстов по медицинской науке	Определение потенциального источника нового знания
Эталон	Немецко-русский словарь с дефинициями значений НМГ	Медицинский словарь с дефинициями терминов	Определение эталона, который используется как критерий новизны
Связь	Гиперссылки на тексты художественных произведений	Гиперссылки на тексты по медицинской науке	Определение связей концептов с их источниками
Динамика	Создается новый вариант существующей дефиниции значения НМГ с сохранением ее предыдущих вариантов	Создается новый вариант существующей дефиниции термина с сохранением ее предыдущих вариантов	Отслеживание динамики концептов и создание ретроспективы вариантов их дефиниций

объектов может служить теоретической основой разработки кодовых таблиц для пар (концепт, литеры). Они обеспечат в ИЕИ-системах присвоение разных компьютерных кодов парам, содержащим один и тот же термин с идентификаторами разных его значений (= концептов), выраженных их дефинициями, с учетом их изменений. Это отличает ИЕИ-системы от систем искусственного интеллекта, в которых, как правило, используются традиционные кодовые таблицы для цифрового кодирования только самих слов как последовательностей литер без одновременного учета кодами значения слова и его изменения во времени.

5 Заключение

Существующие качественные подходы к описанию процессов генерации нового знания [24] носят объяснительный характер и не являются алгоритмо-ориентированными. Рассмотренные два подхода таковыми являются, что проверено при проектировании на основе первого подхода алгоритмов, программ и технологии целенаправленного извлечения экспертами концептов нового знания о значениях НМГ [3, 5, 6]. Сохранение свойства алгоритмо-ориентированности — ключевое требование к созданию концепции ИЕИ-систем на основе исходных положений, перечисленных в табл. 2.

Литература

1. Добровольский Д. О. Немецкие модальные глаголы в параллельном корпусе и задачи двуязычной лексикографии // Германские языки: текст, корпус, перевод. — М.: Институт языкознания РАН, 2020. С. 103–116.
2. Добровольский Д. О., Зализняк Анна А. Русские конструкции с потенциально модальным значением по данным параллельных корпусов // Труды Института русского языка им. В. В. Виноградова, 2020. № 3. С. 35–49.
3. Zatsman I. Finding and filling lacunas in knowledge systems // 20th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2019. Vol. 2. P. 1143–1151.
4. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1468 p.
5. Зацман И. М. Целенаправленное развитие систем лингвистических знаний: выявление и заполнение лакун // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 91–98.
6. Зацман И. М. Проблемно-ориентированная верификация полноты темпоральных онтологий и заполнение понятийных лакун // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 119–128.
7. Medical dictionary of health terms. — Harvard Health Publishing, 2011. <https://www.health.harvard.edu/a-through-c>.
8. Dorland's illustrated medical dictionary. — 33rd ed. — Philadelphia, PA, USA: Elsevier Saunders, 2019. 2144 p.
9. MedTerms medical dictionary. <https://www.medicinenet.com/medterms-medical-dictionary/article.htm>.
10. PubMed. National Library of Medicine. <https://pubmed.ncbi.nlm.nih.gov>.
11. Немецко-русский словарь актуальной лексики / Под ред. Д. О. Добровольского. — М.: Лексрус, 2021 (в печати).
12. Параллельный немецкий подкорпус Национального корпуса русского языка. <http://www.ruscorpora.ru/search-para-de.html>.
13. Zatsman I. Finding and filling lacunas in linguistic typologies // 15th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera: Institute of Knowledge Asset Management, 2020. P. 780–793.
14. Зацман И. М. Стадии целенаправленного извлечения знаний, имплицитированных в параллельных текстах // Системы и средства информатики, 2018. Т. 28. № 3. С. 175–188.
15. Zatsman I. Three-dimensional encoding of emerging meanings in AI-systems // 21st European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2020. P. 878–887.
16. Гончаров А. А., Зацман И. М., Кружков М. Г. Эволюция классификаций в надкорпусных базах данных // Информатика и её применения, 2020. Т. 14. Вып. 4. С. 108–116.
17. Международная статистическая классификация болезней и проблем, связанных со здоровьем. Десятый пересмотр. Т. 1. Ч. 1. <https://apps.who.int/iris/bitstream/handle/10665/85974/5225032680.pdf?sequence=1&isAllowed=y>.
18. Vázquez M., Oliver A. Improving term candidates' selection using terminological tokens // Terminology, 2018. Vol. 24. No. 1. P. 122–147.
19. Khakimova A., Yang X., Zolotarev O., Berberova M., Char-nine M. Tracking knowledge evolution based on the terminology dynamics in 4P-medicine // Int. J. Env. Res. Pub. He., 2020. Vol. 17. Iss. 20. Art. ID: 7444.
20. Oliver A., Vázquez M. TermEval 2020: Using TSR filtering method to improve automatic term extraction // 6th Workshop (International) on Computational Terminology Proceedings. — Paris: European Language Resources Association, 2020. P. 106–113.

21. Jarrahi M. H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making // *Bus. Horizons*, 2018. Vol. 61. No. 4. P. 577–586.
22. Trunk A., Birkel H., Hartmann E. On the current state of combining human and artificial intelligence for strategic organizational decision making // *Bus. Res.*, 2020. Vol. 13. No. 3. P. 875–919.
23. Зацман И. М. Интерфейсы третьего порядка в информатике // *Информатика и её применения*, 2019. Т. 13. Вып. 3. С. 82–89.
24. Wierzbicki A. P., Nakamori Y. Basic dimensions of creative space // *Creative space: Models of creative processes for knowledge civilization age* / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin: Springer Verlag, 2006. P. 59–90.

Поступила в редакцию 16.01.2021

PROBLEM-ORIENTED UPDATING OF DICTIONARY ENTRIES OF BILINGUAL DICTIONARIES AND MEDICAL TERMINOLOGY: COMPARATIVE ANALYSIS

I. M. Zatsman

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

Abstract: Two approaches to the goal-oriented discovery of new knowledge from text data are compared. The first approach relates to the subject domain of lexicography. It focuses on extracting new meanings of linguistic units from texts to replenish the dictionary entries of bilingual dictionaries. The second approach relates to medical science and focuses on discovering new meanings of terms to update a disease’s description in the form of its terminological portrait. The portrait includes definitions of terms with reflecting their dynamics over time, relationships between terms, contexts of their use, and links to sources of contexts. These approaches are compared in the following positions: the problem for the solution of which new knowledge is discovered, the purpose of its discovery, sources of concepts of new knowledge, the standard, comparison with which uses as the criterion of concepts’ novelty, concept-source linkages, and concept dynamics. The purpose of the paper is to describe the outcomes of the comparative analysis of the approaches. It is proposed to position analysis outcomes as initial data for creating the conception of a human-artificial intelligence system for goal-oriented discovery of new knowledge from big data which is applicable in different subject domains.

Keywords: new knowledge generation; discovering knowledge from texts; artificial intelligence; human-artificial intelligence system

DOI: 10.14357/19922264210113

Acknowledgments

The reported study was funded by RFBR, project number 20-012-00166, and by RFBR and NSFC, project number 21-57-53018.

References

1. Dobrovolskiy, D. O. 2020. Nemetskie modal’nye glagoly v parallel’nom korpuse i zadachi dvuyazychnoy leksikografii [German modal verbs in a parallel corpus and bilingual lexicography tasks]. *Germanskije yazyki: tekst, korpus, perevod* [German languages: Text, corpus, translation]. 103–116.
2. Dobrovolskiy, D. O., and Anna A. Zaliznyak. 2020. Russkie konstruksii s potentsial’no modal’nym znacheniem po dannym parallel’nykh korpusov [Russian constructions with potentially modal meanings: An analysis based on parallel corpus data]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [V. V. Vinogradov Russian Language Institute Proceedings]. 3:35–49.
3. Zatsman, I. 2019. Finding and filling lacunas in knowledge systems. *20th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 2:1143–1151.
4. Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1468 p.
5. Zatsman, I. M. 2019. Tselenapravlennoe razvitie sistem lingvisticheskikh znaniy: vyyavlenie i zapolnenie lakun [Goal-oriented development of linguistic knowledge systems: Identifying and filling of lacunae]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):91–98.
6. Zatsman, I. 2020. Problemno-orientirovannaya verifikatsiya polnoty temporal’nykh ontologii i zapolnenie po-

- nyatiynykh lakun [Problem-oriented verifying the completeness of temporal ontologies and filling conceptual lacunas]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):119–128.
7. Medical dictionary of health terms. Available at: <https://www.health.harvard.edu/a-through-c> (accessed January 14, 2021).
 8. *Dorland's illustrated medical dictionary*. 2019. 33rd ed. Philadelphia, PA: Elsevier Saunders. 2144 p.
 9. MedTerms medical dictionary. Available at: <https://www.medicinenet.com/medterms-medical-dictionary/article.htm> (accessed January 14, 2021).
 10. National library of medicine. Available at: <https://pubmed.ncbi.nlm.nih.gov> (accessed January 14, 2021).
 11. Dobrovolskiy, D. O., ed. 2021 (in press). *Nemetsko-russkiy slovar' aktual'noy leksiki* [German–Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
 12. Parallelnyy nemetskiy podkorpus Natsional'nogo korpusa russkogo yazyka [Parallel German subcorpus of the National Russian corpus]. Available at: <http://www.ruscorpora.ru/searchpara-de.html> (accessed January 14, 2021).
 13. Zatsman, I. 2020. Finding and filling lacunas in linguistic typologies. *15th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera: Institute of Knowledge Asset Management. 780–793.
 14. Zatsman, I. 2018. Stadii tselenapravlennoogo izvlecheniya znaniy, implitsirovannykh v parallelnykh tekstakh [Stages of goal-oriented discovery of knowledge implied in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):175–188.
 15. Zatsman, I. 2020. Three-dimensional encoding of emerging meanings in AI-systems. *21st European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 878–887.
 16. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2020. Evolyutsiya klassifikatsiy v nadkorpusnykh bazakh dannykh [Evolution of classifications in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 14(4):108–116.
 17. *Mezhdunarodnaya statisticheskaya klassifikatsiya bolezney i problem, svyazannykh so zdorov'em. Desyatyy peresmotr* [International statistical classification of diseases and related health problems. Tenth Revision]. 1(1). Available at: <https://apps.who.int/iris/bitstream/handle/10665/85974/5225032680.pdf?sequence=1&isAllowed=y> (accessed January 14, 2021).
 18. Vázquez, M., and A. Oliver. 2018. Improving term candidates' selection using terminological tokens. *Terminology* 24(1):122–147.
 19. Khakimova, A., X. Yang, O. Zolotarev, M. Berberova, and M. Charnine. 2020. Tracking knowledge evolution based on the terminology dynamics in 4P-Medicine. *Int. J. Env. Res. Pub. He.* 17(20):7444. 19 p.
 20. Oliver, A., and M. Vázquez. 2020. TermEval 2020: Using TSR filtering method to improve automatic term extraction. *6th Workshop (International) on Computational Terminology Proceedings*. Paris: European Language Resources Association. 106–113.
 21. Jarrahi, M. H. 2018. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Bus. Horizons* 61(4):577–586.
 22. Trunk, A., H. Birkel, and E. Hartmann. 2020. On the current state of combining human and artificial intelligence for strategic organizational decision making. *Bus. Res.* 13(3):875–919.
 23. 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.
 24. Wierzbicki, A. P., and Y. Nakamori. 2006. Basic dimensions of creative space. *Creative space: Models of creative processes for knowledge civilization age*. Eds. A. P. Wierzbicki and Y. Nakamori. Berlin: Springer Verlag. 59–90.

Received January 16, 2021

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

МОДЕЛИРОВАНИЕ СТОХАСТИЧЕСКОЙ ДИНАМИКИ ИЗМЕНЕНИЯ СОСТОЯНИЙ УЗЛОВ И ПЕРКОЛЯЦИОННЫХ ПЕРЕХОДОВ В СОЦИАЛЬНЫХ СЕТЯХ С УЧЕТОМ САМООРГАНИЗАЦИИ И НАЛИЧИЯ ПАМЯТИ

Д. О. Жуков¹, Т. Ю. Хватова², А. Д. Зальцман³

Аннотация: Обсуждаются вопросы использования подходов теоретической информатики и применение ее приложений для анализа и моделирования процессов в социотехнических системах (социальных сетях). Разработана стохастическая модель динамики изменения состояний (настроений или мнений) пользователей (узлов) и достижения порога перколяции в социальной сети, имеющей случайные связи между узлами. Модель показывает возможность скачкообразных переходов между состояниями (мнений, настроений и т. д.) узлов в социальной сетевой структуре в течение короткого времени без внешнего воздействия, что может быть связано с памятью о предыдущих состояниях и самоорганизацией. При создании модели были рассмотрены схемы вероятностей переходов между возможными состояниями узлов с учетом предыдущих шагов (немарковские процессы с наличием памяти) и выведено нелинейное дифференциальное уравнение второго порядка, которое содержит член, отвечающий за возможность самоорганизации, а также сформулирована и решена граничная задача для определения функции плотности вероятности нахождения системы в определенном состоянии с течением времени. Разработанная модель может быть связана с полученными ее авторами ранее результатами описания процессов в социальных сетевых структурах с помощью теории перколяции (определение времени достижения пороговых значений доли узлов сети, при котором мнения или предпочтения могут беспрепятственно распространяться по сети в целом).

Ключевые слова: стохастическая динамика; состояния узлов социальной сети; самоорганизация систем; процессы с памятью; перколяция в социальных сетях

DOI: 10.14357/19922264210114

1 Введение

Отличительной особенностью динамики явлений в социотехнических и социальных системах является активное воздействие на них человеческого фактора, который, с одной стороны, вносит неопределенность и создает стохастичность, а с другой стороны, создает возможности для самоорганизации, позволяет говорить о наличии памяти и придает динамике процессов существенно нелинейный характер.

Для моделирования нелинейной динамики самоорганизующихся социальных систем с памятью можно и нужно применять методы и средства теоретической информатики и кибернетики, которая, по определению Роберта Винера, является наукой об управлении не только техническими, но и биологическими системами.

Использование методов теоретической информатики, разработанных в ней принципов моделирования и ее приложений может позволить полу-

чить качественно новые результаты для описания сложных социальных, экономических и социотехнических систем, а также создать новые методики прогнозирования поведения людей в социальных и социотехнических системах.

2 Обзор некоторых моделей описания динамики процессов в социальных сетевых структурах

Многие из существующих теоретических подходов к описанию социальных сетевых систем имеют много общего с кинетическим описанием физических систем и распространением вирусов в компьютерных сетях. Однако эти модели в основном рассматривают цепные явления, где макроскопическая доля узлов с определенным состоянием в сети

¹МИРЭА — Российский технологический университет, zhukov_do@mirea.ru

²Санкт-Петербургский политехнический университет Петра Великого, khvatova.ty@spbstu.ru

³МИРЭА — Российский технологический университет, ad.zaltcman@gmail.com

быстро возникает из некоторого микроскопического состояния, захватывающего все новые и новые узлы.

В более сложных моделях взаимодействие пользователей социальных сетей описывается теорией многоагентных систем [1–3], а также аппаратом теории клеточных автоматов [4, 5]. В этих моделях на основании некоторых правил переходов агенты принимают определенные состояния, образуют связанную по своим свойствам группу, могут сотрудничать, чтобы решить некую задачу или достигнуть определенной цели [1], а временная логика поведения агентов может зависеть от динамически меняющихся условий [2].

В работе [5] было изучено влияние структуры сетей (случайные структуры, маленькие миры, цикл, колесо, звезда, иерархическая) и правил поведения клеточных автоматов на динамику процессов в социальных сетях. При одинаковых правилах взаимодействия клеток динамика процессов зависит от топологии сети (наибольшая скорость наблюдается в регулярных структурах, а наименьшая — в неупорядоченных).

Для описания процессов в социальных сетях также применяются стохастические подходы, учитывающие зависимости изменения состояния узлов от времени. В работе [6] описана модель смешанного членства в стохастически формирующихся группах, основанная на попарном рассмотрении присутствия или отсутствия связей между объектами. Анализ вероятностных изменений связей требует специальных предположений, например независимости или предположения непостоянства данной связи (смешанного членства в стохастически формирующихся группах). Данная модель позволяет описать динамику кластеризации членов по группам и изменение их численности.

Другим направлением анализа и прогнозирования динамики процессов в сложных социальных системах является использование нестационарных временных рядов. Традиционный подход к их анализу основан на том, чтобы с помощью применения линейных методов свести их к стационарным (например, авторегрессионные интегрированные модели скользящего среднего — ARIMA, autoregressive integrated moving average [7]). Эти модели оперируют не функциями распределения, а непосредственно элементами временного ряда. Ряды, не укладывающиеся в рамки регрессионного анализа, чаще всего изучаются адаптивными эвристическими методами, в которых ряды на некоторой длине описываются стационарной моделью типа регрессии или авторегрессии, а параметры модели пересчитываются с учетом новой информации или с учетом сравнения предсказанного значения с фактом. Не-

достаток этих подходов заключается в том, что длина участка возможной стационарности не является известной величиной. При исследовании стационарных случайных процессов, согласно теореме Гливленко [8] (о сходимости эмпирической вероятности к теоретическому распределению), чем больше учтено наблюдаемых значений, тем точнее будут получены теоретические характеристики распределения. Для нестационарных временных рядов данное условие, в силу их специфики, не может быть выполнено, что затрудняет возможности прогнозирования.

Следует отметить, что ни одна из существующих моделей не рассматривает самоорганизацию и возможность наличия памяти. Поэтому можно сделать вывод о том, что для прогнозирования динамики процессов в социотехнических системах, имеющих сетевую структуру, требуется продолжение разработки их моделей с учетом структурных свойств, самоорганизации и наличия памяти.

3 Постановка задач исследования

С позиций структурного подхода социотехнические системы представляют собой случайную сеть взаимосвязей и взаимодействий пользователей, которая приводит к нелинейной динамике изменения состояний узлов. При моделировании нелинейных динамических процессов в социальных сетях необходимо ответить минимум на два важных вопроса. Во-первых, как учесть стохастичность, неопределенность, самоорганизацию процессов и наличие памяти и как они влияют на наблюдаемые явления. Во-вторых, как структура сетей влияет на динамику процессов и как она может быть связана со стохастичностью, неопределенностью, самоорганизацией процессов и наличием памяти. Ответы на эти вопросы могут позволить создать эффективные алгоритмы мониторинга состояния социальных сетей.

В теории перколяции (теория вероятностей на графах) изучают решение задач узлов и связей для сетей с различной структурой. При решении задачи связей определяют долю связей, которую нужно разорвать, чтобы сеть распалась минимум на две несвязанные части (или, наоборот, долю проводящих связей в сети, когда в целом между любыми произвольными узлами появляется проводимость). В задаче узлов определяют среднюю долю заблокированных узлов, при которой сеть распадется на не связанные между собой кластеры, внутри которых сохраняются связи (или, наоборот, долю проводящих узлов, когда проводимость возникает). Доля заблокированных узлов (в задаче узлов) или разорван-

ных связей (в задаче связей), при которой исчезает проводимость (или, наоборот, появляется) между двумя произвольно выбранными узлами сети, называется порогом перколяции (протекания) [9].

Использование понятия долей заблокированных узлов или связей эквивалентно понятию вероятности нахождения случайно выбранного узла (или связи) в заблокированном (разорванном) состоянии. Поэтому величина порога перколяции определяет вероятность передачи информации через всю сеть в целом, если задана средняя вероятность блокирования узла или связи.

Величину порога перколяции для случайной сетевой структуры можно определить теоретически методами численного моделирования или экспериментально при изучении реальных сетей найти с помощью инструментов социального сетевого анализа (SNA — social network analysis).

Если для случайной сети социальных связей порог ее перколяции известен, то, описав механизмы перехода ее узлов в заблокированное или проводящее состояние, можно определить время его достижения, а следовательно, спрогнозировать динамику распространения определенных мнений или взглядов.

4 Перколяционные свойства случайных сетевых структур

Исследования [1, 10–14] показывают, что пороги перколяции случайных сетей зависят от среднего числа связей в расчете на один узел (плотности) сети. Для задачи связей имеется линейная зависимость: $y = -6,581z - 0,203$; а для задачи узлов: $y = 4,39z - 2,41$. Здесь $z = 1/x$, где x — плотность связей; y — натуральный логарифм доли разорванных связей (или узлов в задаче узлов), при которой исчезает проводимость всей сети в целом [10–13, 15–17].

Полученные ранее результаты [10–13, 15–17] о перколяционных свойствах случайных сетей позволяют сделать ряд очень важных выводов. Например, о наличии насыщения порога перколяции, о роли увеличения плотности связей в информационном влиянии сети и ряд других. Следует отметить, что динамика изменения состояния узлов сетей в совокупности с их перколяционными свойствами была рассмотрена в работе [14], где исследовалось распространение компьютерных вирусов. Однако влияние процессов самоорганизации и наличия памяти на динамику изменения состояния узлов и достижение порогов перколяции с течением времени исследовано не было.

5 Стохастическая динамика переходов между состояниями в сетях социальных связей и достижение порога перколяции с учетом памяти и самоорганизации

5.1 Построение разностных вероятностных схем переходов между состояниями

Будем описывать социальную сеть как систему, состояния которой в любой момент времени могут быть заданы параметром, принимающим непрерывные случайные значения с недетерминированным законом распределения. Например, это может быть доля пользователей (узлов сети), разделяющих и пропагандирующих определенные взгляды или настроения.

Все множество состояний будем обозначать как X . Состояние, наблюдаемое в момент времени t , можно обозначить как x_i ($x_i \in X$).

Введем интервал времени τ_0 , за который возможно изменение состояния x_i . В данном случае любое значение текущего времени $t = h\tau_0$, где h — номер шага перехода между состояниями (процесс перехода между состояниями становится квазинепрерывным с бесконечно малым временным интервалом τ_0); $h = 0, 1, 2, \dots, N$. Текущее состояние x_i на шаге h после перехода на шаг $h + 1$ может за счет случайно возникающих факторов увеличиваться на некоторую величину ε или уменьшаться на величину ξ , т. е. оказаться равным $x_i + \varepsilon$ или $x_i - \xi$ соответственно. Введем понятие вероятности нахождения системы в том или ином состоянии: после некоторого числа шагов h про описываемую систему можно сказать, что $P(x - \varepsilon, h)$ — вероятность того, что она находится в состоянии $(x - \varepsilon)$; $P(x, h)$ — вероятность того, что она находится в состоянии x ; $P(x + \xi, h)$ — вероятность того, что она находится в состоянии $(x + \xi)$.

После каждого шага состояние x_i (далее индекс i для краткости опустим) может изменяться на величину ε или ξ . Вероятность $P(x, h + 1)$ того, что на следующем, $(h + 1)$ -м, шаге система (или процесс) окажется в состоянии x описывается уравнением (рис. 1):

$$P(x, h + 1) = P(x - \varepsilon, h) + P(x + \xi, h) - P(x, h). \quad (1)$$

Поясним уравнение (1) и представленную на рис. 1 схему. Вероятность перехода в состояние x на

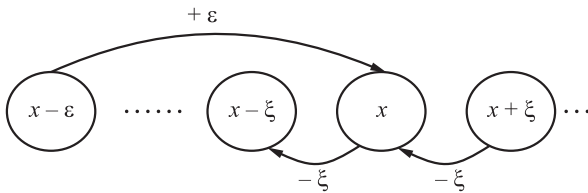


Рис. 1 Схема возможных переходов между состояниями системы (или процесса) на $(h + 1)$ -м шаге

$(h + 1)$ -м шаге $P(x, h + 1)$ определяется суммой вероятностей перехода в это состояние из состояний $(x - \varepsilon)$: $P(x - \varepsilon, h)$ и $(x + \xi)$: $P(x + \xi, h)$, в которых находилась система на шаге h за вычетом вероятности $P(x, h)$ перехода системы из состояния x (в котором она находилась на шаге h) в любое другое состояние на $(h + 1)$ -м шаге. В реальности в социальной сети всегда остается память о предыдущих состояниях. Для учета этого определим вероятности $P(x - \varepsilon, h)$, $P(x + \xi, h)$ и $P(x, h)$ через состояния на $(h - 1)$ -м шаге. Схемы соответствующих переходов можно изобразить аналогично схеме, представленной на рис. 1, и получить для вероятности перехода следующее алгебраическое уравнение:

$$\begin{aligned}
 P(x, h + 2) &= \\
 &= \{P(x - 2\varepsilon, h) + P(x - \varepsilon + \xi, h) - P(x - \varepsilon, h)\} + \\
 &+ \{P(x + \xi - \varepsilon, h) + P(x + \xi, h) - P(x + \xi, h)\} - \\
 &- P(x - \varepsilon, h) - P(x + \xi, h - 1) + P(x, h).
 \end{aligned}$$

Далее, учитывая, что $t = h\tau_0$, перейдем от h к t , а затем проведем соответствующие разложения в ряд Тейлора и получим:

$$\frac{dP(x, t)}{dt} = a \frac{d^2P(x, t)}{dx^2} - b \frac{dP(x, t)}{dx} - c \frac{d^2P(x, t)}{dt^2}, \quad (2)$$

где

$$a = \frac{\varepsilon^2 - \varepsilon\xi + \xi^2}{\tau_0}; \quad b = \frac{\varepsilon - \xi}{\tau_0}; \quad c = \tau_0.$$

Член уравнения $dP(x, t)/dx$ описывает упорядоченный переход либо в состояние, когда оно увеличивается ($\varepsilon > \xi$), либо когда оно уменьшается ($\varepsilon < \xi$); член $d^2P(x, t)/dx^2$ описывает случайное изменение состояния (неопределенность изменения). Член $dP(x, t)/dt$ определяет скорость общего изменения состояния системы с течением времени; член $d^2P(x, t)/dt^2$ описывает процесс, при котором состояния сами становятся источниками возникновения других состояний (*самоорганизация* и ускорение упорядоченных и случайных переходов).

Сравним полученный результат с существующими моделями анализа и описания поведения

нестационарных временных рядов. В настоящее время в качестве аппроксимаций выборочных распределений чаще всего используются диффузионные уравнения, включая нелинейную диффузию [18]:

$$\frac{\partial \rho(x, t)}{\partial t} = \frac{D(t) \partial^2 \rho^{(n-1)/(n+1)}(x, t)}{\partial x^2},$$

где n — числовой параметр модели; $(n - 1)/(n + 1)$ — показатель степени функции плотности распределения $\rho(x, t)$ (это уравнение учитывает только случайные переходы); уравнение Лиувилля [18]:

$$\frac{\partial \rho(x, t)}{\partial t} = - \frac{\partial \{U(x, t) \rho(x, t)\}}{\partial x},$$

которое определяет упорядоченный перенос; уравнение Фоккера–Планка [19]:

$$\frac{\partial \rho(x, t)}{\partial t} = \frac{D(t)}{2} \frac{\partial^2 \rho(x, t)}{\partial x^2} - \frac{\partial \{U(x, t) \rho(x, t)\}}{\partial x},$$

где $U(x, t)$ — скорость «сноса»; $D(t)$ — коэффициент диффузии (это уравнение учитывает не только случайное изменение (член $\partial^2 \rho(x, t)/\partial x^2$), но и упорядоченные переходы (член $\partial \{U(x, t) \rho(x, t)\}/\partial x$), или «снос») и ряд других. Однако ни одна из этих моделей не рассматривает самоорганизацию и память. Разработанная авторами модель обобщает другие модели, и при равенстве нулю некоторых коэффициентов в уравнении (2) оно переходит в уравнения Лиувилля или Фоккера–Планка, которые выступают ее частными случаями.

5.2 Формулировка и решение краевой задачи

Считая функцию $P(x, t)$ непрерывной, можно перейти от вероятности $P(x, t)$ (уравнение (2)) к плотности вероятности $\rho(x, t) == dP(x, t)/dx$ и сформулировать граничную задачу, решение которой и будет описывать процесс перехода между состояниями. Предположим, что необходимо, чтобы доля пользователей (узлов) социальной сети, имеющих негативное мнение, не превышала определенного значения (т. е. величина доли негативных настроений должна находиться на отрезке от 0 до величины порога перколяции l для данной сети).

Первое граничное условие. Состояние $x = 0$ определяет полное отсутствие негативных мнений (доля равна 0). Сама вероятность обнаружить такое состояние может быть отлична от 0, однако плотность вероятности, определяющую поток в состоянии $x = 0$, необходимо положить равной 0 (состояния системы не могут быть отрицательными): $\rho(x, t)_{x=0} = 0$.

Второе граничное условие. Вероятность обнаружить состояние с максимально возможной долей негативно настроенных пользователей $x = L = 1$ отлична от 0. Однако плотность вероятности, определяющая поток в этом состоянии, необходимо положить равной 0 (величина состояния не может быть больше, чем максимально возможная доля): $\rho(x, t)_{x=L} = 0$.

Поскольку в момент времени $t = 0$ состояние системы уже может быть равно некоторому значению x_0 , то начальное условие зададим в виде:

$$\rho(x, t = 0) = \delta(x - x_0) = \begin{cases} \int \delta(x - x_0) dx = 1, & x = x_0; \\ 0, & x \neq x_0. \end{cases}$$

Второе начальное условие можно задать в виде:

$$\left. \frac{\partial G(x, t)}{\partial t} \right|_{t=0} = 0,$$

так как начальное условие содержит дельта-функцию; кроме того, ее наличие приводит к тому, что решение для $\rho(x, t)$ разбивается на две области при $x > x_0$ и при $x \leq x_0$. Используя методы операционного исчисления для плотности вероятности $\rho_1(x, t)$ и $\rho_2(x, t)$ обнаружения состояния системы в одном из значений на отрезке от 0 до L , можно получить следующую систему уравнений:

$$\begin{aligned} \text{при } x \geq x_0 : \rho_1(x, t) = & -\frac{2}{L} e^{-t/(2\tau_0)} e^{k(x-x_0)} \times \\ & \times \sum_{n=1}^{\infty} \frac{\sin(\pi n x_0/L) \sin(\pi n(L-x)/L)}{\cos(\pi n)} \times \\ & \times \operatorname{ch} \left(\frac{t}{\tau_0} \sqrt{\frac{k\varepsilon\xi}{2(\varepsilon-\xi)} - \frac{\pi^2 n^2(\varepsilon-\xi)}{2kL^2}} \right); \end{aligned}$$

$$\begin{aligned} \text{при } x < x_0 : \rho_2(x, t) = & -\frac{2}{L} e^{-t/(2\tau_0)} e^{k(x-x_0)} \times \\ & \times \sum_{n=1}^{\infty} \frac{\sin(\pi n(L-x_0)/L) \sin(\pi n x/L)}{\cos(\pi n)} \times \\ & \times \operatorname{ch} \left(\frac{t}{\tau_0} \sqrt{\frac{k\varepsilon\xi}{2(\varepsilon-\xi)} - \frac{\pi^2 n^2(\varepsilon-\xi)}{2kL^2}} \right), \end{aligned}$$

где

$$k = \frac{\varepsilon - \xi}{2(\varepsilon^2 - \varepsilon\xi + \xi^2)}.$$

Если вычислить интеграл

$$P(l, t) = \int_0^{x_0} \rho_2(x, t) dx + \int_{x_0}^l \rho_1(x, t) dx, \quad (3)$$

то функция $P(l, t)$ будет задавать вероятность того, что состояние системы к моменту времени t будет находиться на отрезке от 0 до l , т.е. порог перколяции l не будет достигнут. Соответственно, вероятность $Q_i(t)$ того, что порог перколяции l окажется к моменту времени t достигнутым или превзойденным, будет равна:

$$Q(l, t) = 1 - P(l, t). \quad (4)$$

5.3 Моделирование динамики и самоорганизации состояний узлов социальной сети

При анализе модели необходимо задать приемлемые величины значений порогов перколяции случайной сети. Плотность связей можно определить экспериментально, а затем, используя зависимости величины порогов перколяции от среднего числа связей, приходящегося на один узел [10–13, 15–17], рассчитать их допустимые величины (см. уравнения в разд. 4).

Для моделирования примем, что начальная доля x_0 негативных мнений равна 5% ($x_0 = 0,05$), величину τ_0 примем равной одной условной единице времени ($\tau_0 = 1$), $\varepsilon = 0,02$ (2%) и $\xi = 0,01$ (1%). Результаты моделирования времени достижения порога перколяции (4) с использованием (3) при заданном выше в качестве примера наборе параметров модели представлены в графическом виде на рис. 2. Кривые 1 и 2 на рис. 2 показывают, что чем ближе значение величины начального состояния системы x_0 в момент времени $t = 0$ к пороговому значению, тем быстрее возрастает вероятность перехода и тем сильнее вероятность его достижения приближается к единице. Кривая 4 на рис. 2 показывает, что при большой разности между величиной порогового значения и x_0 вероятность его достижения имеет осциллирующий характер, при этом она сначала снижается с течением времени, а затем показывает рост, причем чем дальше значение величины x_0 от порогового значения, тем сильнее проявляются осцилляции. Кроме того, существует и отличная от нуля вероятность достижения порогового значения при $t = 0$ (мгновенная реализация).

Рост вероятности перехода через пороговое значение имеет ступенчатый характер, а протяженность ступени во времени зависит от того, насколько начальная величина состояния системы x_0 близка к пороговому значению. Процесс достижения порогового значения имеет протяженное во времени плато, величина которого (в единицах вероятности) зависит от x_0 .

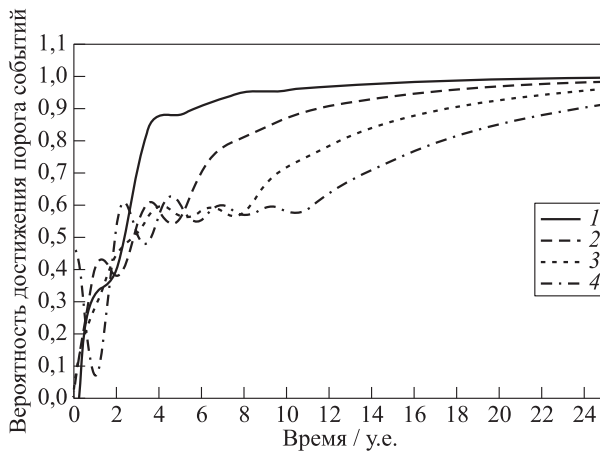


Рис. 2 Графическое представление результатов моделирования преодоления порогов перколяции для распространения негативных мнений в социальной сети: $1 - l = 0,1$; $2 - 0,15$; $3 - 0,20$; $4 - l = 0,25$

Увеличение значения величин ε и ξ (при выполнении условия $\varepsilon > \xi$) изменяет величину плато (горизонтальный участок зависимости вероятности перехода через пороговое значение до второго участка резкого роста) на рис. 2, однако общая зависимость вероятности перехода от времени качественно не изменяется.

Ход кривых на рис. 2 показывает возможность роста вероятности достижения порогового значения состояния системы практически сразу после начала процесса. Вероятность перехода через пороговое значение отлична от нуля уже после первого шага и нелинейно возрастает с течением времени. Это следствие того, что не только величины ε и ξ определяют изменение состояния x , но и сами состояния x служат источником изменения вследствие наличия памяти о предыдущих состояниях и самоорганизации, за которую отвечает в дифференциальном уравнении член $d^2P(x, t)/dt^2$. Арифметический расчет показывает, что число шагов (обозначим его как q_0), за которое можно достичь порогового значения l , должно быть не меньше чем $q = (l - x_0)/(\varepsilon - \xi)$. Например, для пороговых значений состояния системы $l = 0,1$ и $0,2$ при ее начальном состоянии $x_0 = 0,05$, $\varepsilon = 0,02$ и $\xi = 0,01$ для q получим соответственно 5 и 15. Результаты (см. рис. 2) показывают, что это не так, т. е. происходит самоорганизация.

При равенстве ε и ξ (например, $\varepsilon = \xi = 0,02$) характер хода кривых, описывающих вероятность достижения пороговых значений, изменяется (рис. 3). В частности, не наблюдается протяженного во времени плато с последующим плавным ростом вероятности достижения пороговых значе-

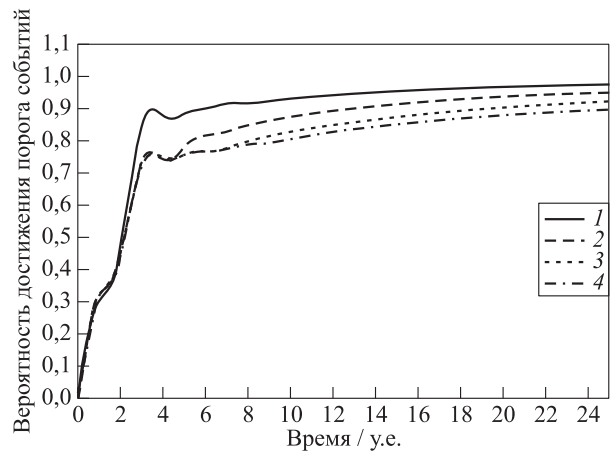


Рис. 3 Графическое представление результатов моделирования преодоления порогов перколяции для распространения негативных мнений в социальной сети при $\varepsilon = \xi = 0,02$: $1 - l = 0,1$; $2 - 0,15$; $3 - 0,20$; $4 - l = 0,25$

ний до единицы, а рост вероятности имеет характер резкого скачка. Это связано с тем, что коэффициент b в уравнении (2) окажется равен нулю и упорядоченные переходы будут невозможны, а член $d^2\rho(x, t)/dt^2$ будет ускорять только случайные переходы $d^2\rho(x, t)/dx^2$.

6 Алгоритм мониторинга состояний социальной сетевой структуры

Разработанная модель позволяет создать практически реализуемый алгоритм мониторинга состояния социальной сети.

1. Определяем с помощью социологического мониторинга плотность сети и долю узлов x_0 с определенным мнением или настроением (состояние узла) в данный момент времени $t = 0$.
2. Спустя одну выбранную условную единицу времени $\tau = 1$ (например, одна неделя) снова находим долю узлов с определенным мнением или настроением в данный момент времени x_1 ($t = 0 + \tau$). Находим величину $\varepsilon = x_1 - x_0$, а величину ξ считаем равной 0. Если $\varepsilon < 0$, то считаем $\xi = x_1 - x_0$, а $\varepsilon = 0$.
3. На основании данных о среднем числе связей рассчитываем порог перколяции данной сети. Используя уравнения (3) и (4), по определенным в пп. 1 и 2 значениям параметров x_0 , ε , ξ и порогу перколяции l моделируем поведение от условного времени вероятности перехода поро-

га перколяции и определяем допустимый лимит времени для изменения ситуации.

7 Заключение

Создана новая модель описания динамики изменения состояния узлов и перколяционных переходов в социальных сетях с учетом самоорганизации и наличия памяти, которая вносит значительный вклад в развитие теории управления сложными системами.

Результаты анализа разработанной модели могут быть связаны с полученными ранее результатами описания процессов в социальных сетевых структурах с помощью теории перколяции (это необходимо для определения времени достижения пороговых значений доли узлов социальной сети, когда определенные мнения или предпочтения могут беспрепятственно распространяться по сети в целом).

Полученные результаты существенно отличаются от применяемых в настоящее время моделей для описания нестационарных процессов на основе теории хаоса, диффузионных подходов, уравнений Лиувилля и Фоккера–Планка. Все это в целом представляется абсолютно новым и оригинальным, а также вносит вклад в развитие теории управления сложными системами.

Разработанная модель во взаимосвязи с методами теории перколяции существенно расширяет возможности применения уравнений математической физики и теоретической информатики для моделирования социальных систем.

Литература

1. Gasser L. Social conceptions of knowledge and action: DAI foundations and open system semantics // *Artif. Intell.*, 1991. Vol. 47. No. 1-3. P. 107–138.
2. Jennings N. R., Faratin P., Lomuscio A. R., Parsons S., Sierra C., Wooldridge M. Automated negotiation: Prospects, methods and challenges // *Group Decis. Negot.*, 2001. Vol. 10. No. 2. P. 199–215.
3. Plikynas D., Raudys A., Raudys S. Agent-based modelling of excitation propagation in social media groups // *J. Exp. Theor. Artif. In.*, 2015. Vol. 27. No. 4. P. 373–388.
4. Hay J., Flynn D. The effect of network structure on individual behavior // *Complex Systems*, 2014. Vol. 23. No. 4. P. 295–311.
5. Hay J., Flynn D. How external environment and internal structure change the behavior of discrete systems // *Complex Systems*, 2016. Vol. 25. No. 1. P. 39–49.
6. Airoidi E. M., Blei D. M., Fienberg S. E., Xing E. P. Mixed membership stochastic blockmodels // *J. Mach. Learn. Res.*, 2008. Vol. 9. P. 1981–2014.
7. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление / Пер. с англ. — М.: Мир, 1974. 553 с. (Box G. E. P., Jenkins G. M. Time series analysis: Forecasting and control. — Holden-day, 1970. 553 p.)
8. Гнеденко Б. В. Курс теории вероятностей. — М.: Физматлит, 1961. 406 с.
9. Grimmett G. R. Percolation. — New York, NY, USA: Springer-Verlag, 1989. 296 p.
10. Zhukov D., Lesko S. Percolation models of information dissemination in social networks // *IEEE Conference (International) on Smart City/SocialCom/SustainCom Together with DataCom Proceedings.* — IEEE, 2015. P. 213–216.
11. Khvatova T., Block M., Zhukov D., Lesko S. Studying the structural topology of the knowledge sharing network // *11th European Conference on Management Leadership and Governance Proceedings.* — Lisbon, Portugal: Academic Conferences and Publishing International Ltd., 2015. P. 20–27.
12. Khvatova T. Yu., Zaltsman A. D., Zhukov D. O. Information processes in social networks: Percolation and stochastic dynamics // *CEUR Workshop Procee.*, 2017. Vol. 2064. P. 277–288.
13. Zhukov D., Khvatova T., Lesko S., Zaltsman A. Managing social networks: Applying percolation theory methodology to understand individuals' attitudes and moods // *Technol. Forecast. Soc.*, 2018. Vol. 129. P. 297–307.
14. Лесько С. А., Алёшкин С., Филатов В. В. Стохастические и перколяционные модели динамики блокировки вычислительных сетей при распространении эпидемий эволюционирующих компьютерных вирусов // *Российский технологический ж.*, 2019. Т. 7. № 3(29). С. 7–27.
15. Khvatova T., Block M., Zhukov D., Lesko S. How to measure trust: The percolation model applied to intraorganisational knowledge sharing networks // *J. Knowl. Manag.*, 2016. Vol. 20. No. 5. P. 918–935.
16. Жуков Д. О., Хватова Т. Ю., Лесько С. А., Зальцман А. Д. Влияние плотности связей на кластеризацию и порог перколяции при распространении информации в социальных сетях // *Информатика и её применения*, 2018. Т. 12. Вып. 2. С. 90–97.
17. Zhukov D. O., Khvatova T. Y., Millar C., Zaltsman A. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organisation and memory // *Technol. Forecast. Soc.*, 2020. Vol. 158. Art. ID: 120134.
18. Орлов Ю. Н., Федоров С. Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера–Планка // *Труды МФТИ*, 2016. Т. 8. № 2(30). С. 126–133.
19. Fuentes M. Non-linear diffusion and power law properties of heterogeneous systems: Application to financial time series // *Entropy*, 2018. Vol. 20. Iss. 9. Art. No. 649.

Поступила в редакцию 18.06.2019

MODELING OF THE STOCHASTIC DYNAMICS OF CHANGES IN NODE STATES AND PERCOLATION TRANSITIONS IN SOCIAL NETWORKS WITH SELF-ORGANIZATION AND MEMORY

D. O. Zhukov¹, T. Yu. Khvatova², and A. D. Zaltzman¹

¹Russian Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation

²Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya Str., St. Petersburg 195251, Russian Federation

Abstract: This paper explores the use of theoretical informatics applied for analyzing and modeling the processes in sociotechnical systems (social networks). A stochastic model of users' (network nodes) dynamic changes of states (opinions or moods) and the percolation threshold in a social network with random connections among nodes was developed. This model demonstrates the opportunity for jump-like transitions in states (opinions, moods) of the nodes in a social network over a short period of time without external influence. While developing the model, the probabilistic schemes of state-to-state transitions of nodes (users having certain opinions and views) were considered; a second-order nonlinear differential equation was derived; the boundary for calculating the probability density function for a system being in a certain state depending on the time interval was formulated. The differential equation of the model contains a member representing the opportunity for self-organization; it also considers the presence of memory. The results of analysis of the stochastic model support those previously obtained by the authors when investigating social network processes using the percolation theory. This theory was used for defining the time of reaching the threshold values for the share of social network nodes when certain opinions or preferences can spread freely within the whole social network.

Keywords: stochastic dynamics; states of social network nodes; system self-organization; processes involving memory; percolation in social networks

DOI: 10.14357/19922264210114

References

- Gasser, L. 1991. Social conceptions of knowledge and action: DAI foundations and open system semantics. *Artif. Intell.* 47(1-3):107–138.
- Jennings, N. R., P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge. 2001. Automated negotiation: Prospects, methods and challenges. *Group Decis. Negot.* 10(2):199–215.
- Plikynas, D., A. Raudys, and S. Raudys. 2015. Agent-based modelling of excitation propagation in social media groups. *J. Exp. Theor. Artif. In.* 27(4):373–388.
- Hay, J., and D. Flynn. 2014. The effect of network structure on individual behavior. *Complex Systems* 23(4):295–311.
- Hay, J., and D. Flynn. 2016. How external environment and internal structure change the behavior of discrete systems. *Complex Systems* 25(1):39–49.
- Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing. 2008. Mixed membership stochastic block-models. *J. Mach. Learn. Res.* 9:1981–2014.
- Box, G. E. P., and G. M. Jenkins. 1970. *Time series analysis: Forecasting and control*. Holden-day. 553 p.
- Gnedenko, B. V. 1961. *Kurs teorii veroyatnostey* [Probability theory]. Moscow: Fizmatlit. 406 p.
- Grimmett, G. R. 1989. *Percolation*. New York, NY: Springer-Verlag. 296 p.
- Lesko, S. A., and D. O. Zhukov. 2015. Percolation models of information dissemination in social networks. *IEEE Conference (International) on Smart City/SocialCom/SustainCom Together with DataCom Proceedings*. IEEE. 213–216.
- Block, M., T. Khvatova, D. Zhukov, and S. Lesko. 2015. Studying the structural topology of the knowledge sharing network. *11th European Conference on Management, Leadership and Governance Proceedings*. Lisbon, Portugal: Academic Conferences and Publishing International Ltd. 20–27.
- Khvatova, T. Yu., A. D. Zaltzman, and D. O. Zhukov. 2017. Information processes in social networks: Percolation and stochastic dynamics. *CEUR Workshop Procee.* 2064:277–288.
- Zhukov, D., T. Khvatova, S. Lesko, and A. Zaltzman. 2018. Managing social networks: Applying Percolation theory methodology to understand individuals' attitudes and moods. *Technol. Forecast. Soc.* 129:297–307.
- Lesko, S. A., A. S. Alyoshkin, and V. V. Filatov. 2019. Stokhasticheskie i perkolyatsionnye modeli dinamiki blokirovki vychislitel'nykh setey pri rasprostranении epidemii evolyutsioniruyushchikh komp'yuternykh virusov [Stochastic and percolating models of blocking computer networks dynamics during distribution of epidemics of evolutionary computer viruses]. *Rossiyskiy tekhnologicheskii zh.* [Russian Technological J.] 7(3):7–27.
- Khvatova, T., M. Block, D. Zhukov, and S. Lesko. 2016. How to measure trust: The percolation model applied to intraorganisational knowledge sharing networks. *J. Knowl. Manag.* 20 (5):918–935.

16. Zhukov, D. O., T. Yu. Khvatova, S. A. Les'ko, and A. D. Zal'tsman. 2018. Vliyaniye plotnosti svyazey na klas-terizatsiyu i porog perkolyatsii pri rasprostraneni infor-matsii v sotsial'nykh setyakh [The influence of the connec-tions density on clusterization and percolation threshold during information distribution in social networks]. *Infor-matika i ee Primeneniya — Inform. Appl.* 12(2):90–97.
17. Zhukov, D. O., T. Y. Khvatova, C. Millar, and A. Zaltc-man. 2020. Modelling the stochastic dynamics of tran-sitions between states in social systems incorporating self-organisation and memory. *Technol. Forecast. Soc.* 158:120134.
18. Orlov, Yu. N., and S. L. Fedorov. 2016. Generatsiya ne-statsionarnykh traektoriy vremennogo ryada na osnove uravneniya Fokkera—Planka [Generating nonstationary trajectories of a time series based on Fokker—Plank equation]. *Trudy MFTI [MIPT Proceedings J.]* 8(2):126–133.
19. Fuentes, M. 2018. Non-linear diffusion and power law properties of heterogeneous systems: Application to fi-nancial time series. *Entropy* 20(9):649. 8 p.

Received June 18, 2019

Contributors

Zhukov Dmitry O. (b. 1965) — Doctor of Science in technology, professor, Head of Department, Russian Techno-logical University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; zhukov_do@mirea.ru

Khvatova Tatiana Yu. (b. 1971) — Doctor of Science in economics, professor, Peter the Great St. Petersburg Poly-technic University, 29 Polytechnicheskaya Str., St. Petersburg 195251, Russian Federation; khvatova.ty@spbstu.ru

Zaltcman Anastasia D. (b. 1989) — lecturer, Russian Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; ad.zaltcman@gmail.com

О СИСТЕМНОЙ ИЕРАРХИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

С. Н. Гринченко¹

Аннотация: Искусственный интеллект (ИИ) рассмотрен с позиций информатико-кибернетического моделирования (ИКМ) процесса развития самоуправляющейся иерархической системы Человечества как природный феномен, теснейшим образом сопряженный с понятиями «когнитивные функции человека» и «интеллектуальная деятельность человека». Опираясь на информационно-коммуникационно-инфраструктурную составляющую определения ИИ и на ИКМ, понятие «человеко-аппаратурной интеллектуальной единицы» естественным образом обобщено на *все* уровни/ярусы системы Человечества, расположенные в ее иерархии выше и ниже относительно уровня/яруса «человек/личность». Как следствие, феномен «личностного естественно-искусственного интеллекта» дополняется феноменом «иерархического ИИ». Его формирование стало возможным начиная примерно с 1946 г. — с возникновением базисной информационной технологии (БИТ) компьютеров — и приняло взрывной характер примерно с 1979 г. — с возникновением БИТ телекоммуникаций/сетей. Приводятся типичные размеры ареалов уровней/ярусов в иерархии ИИ системы Человечества (указанные даты и размеры — результат модельного расчета).

Ключевые слова: искусственный интеллект; информационные технологии; информатико-кибернетическая модель; самоуправляющаяся иерархическая система Человечества; человеко-аппаратурная интеллектуальная единица

DOI: 10.14357/19922264210115

Что такое *искусственный интеллект*? Будем исходить из следующей формулировки: это «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе то, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений» [1].

Из определения ИИ следует, что это понятие теснейшим образом сопрягается с понятиями «когнитивные функции человека» и «интеллектуальная деятельность человека». Конкретизируем этот факт.

Тезис 1 (А. А. Зиновьева): «Человеческий интеллект есть не просто интеллект отдельно взятых людей. Это — интеллект человечества, аккумулирующийся в языковой практике человечества, включая практику научного познания... Разгадку тайны человеческого сознания надо искать не в головах отдельно взятых людей, а в интеллектуальной

деятельности всего человечества, осуществляемой в языке» [2].

Тезис 2 (авторский) состоит в рассмотрении Человечества — в контексте его ИКМ — не как суммы собственно людей как таковых, но как совокупности «человеко-аппаратурных интеллектуальных единиц», характерных для соответствующей «продвинутой» фазы его личностно-производственно-социальной метаэволюции². При этом «аппаратурная составляющая интеллектуальной единицы в общем случае содержит комбинацию «софтвера» (software), «хардвера» (hardware) и «брэйнваре» (brainware), иначе говоря: *памяти (баз данных) + аппаратной основы + алгоритмов преобразования информации*; аналогичным образом в ее «человеческой» составляющей (по крайней мере, для человека, находящегося в сознании) целесообразно выделить подобную же комбинацию, а именно: *продуктивную личностную память + работоспособный мозг + приемлемый уровень мыслительной деятельности*» [3, с. 92; 4, 5].

Тезис 3 (авторский) заключается в формализации понятия «интеллект» — с информатико-кибернетических позиций — как проявления деятельности имманентно присущего системам живой и «человеко-искусственной» природы механизма

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

² Метаэволюция — процесс последовательного наращивания числа уровней/ярусов иерархической системы в ходе ее формирования как таковой.

иерархической адаптивной поисковой оптимизации (целевых критериев энергетического характера), определяющего их приспособительное поведение [6, 7]. Введение такого модельного взгляда при рассмотрении основных структурных элементов окружающего мира эквивалентно переходу от их трактовки как *пассивных* и «*косных*», безропотно допускающих любые воздействия на себя извне — к представлению о них как об *активных*, перманентно стремящихся к достижению *собственных целей* (т. е. к энергетически оптимальным их состояниям), парирующих существенную часть подобных воздействий.

Представление об *аппаратурной составляющей* «человеко-аппаратурной интеллектуальной единицы» в иерархосетевой системе Человечества вполне соответствует вышеприведенному понятию ИИ в локальном его смысле, соотносящемся с отдельным *человеком/личностью*. При этом, опираясь на информационно-коммуникационно-инфраструктурную составляющую определения ИИ и на ИКМ, понятие «человеко-аппаратурной интеллектуальной единицы» естественным образом обобщается на *все* уровни/ярусы самоуправляющейся иерархосетевой системы Человечества, расположенные в ее иерархии выше и ниже относительно уровня/яруса «человек/личность». Как следствие, феномен «личностного естественно-искусственного интеллекта» (или «человеко-аппаратурной интеллектуальной единицы») дополняется феноменом «иерархического искусственного интеллекта» (или иерархии «социум-аппаратурных интеллектуальных общностей»).

«Человеко-аппаратурная интеллектуальная единица» — элементарное звено самоуправляющейся иерархосетевой системы Человечества — в своем достаточно полном виде, т. е. в форме триады «софтвр» (память) — «хардвр» (носители) — «брэйнвр» (алгоритмы), возникает начиная с момента создания:

- БИТ компьютеров (начиная примерно с 1946 г.),
- БИТ телекоммуникаций/сетей (начиная примерно с 1979 г.),
- перспективной нано-БИТ (начиная примерно с 1981 г.) и т. д.¹

Их предтечей были БИТ, не обладающие третьей составляющей вышеприведенной интеллектуальной триады — автоматическими *алгоритмами* преобразования информации, но в какой-то степени реализовывающие первую и вторую ее составля-

ющие — *память* на основе внешних (по отношению к человеку) ее пассивных *носителей*, ретроспективно:

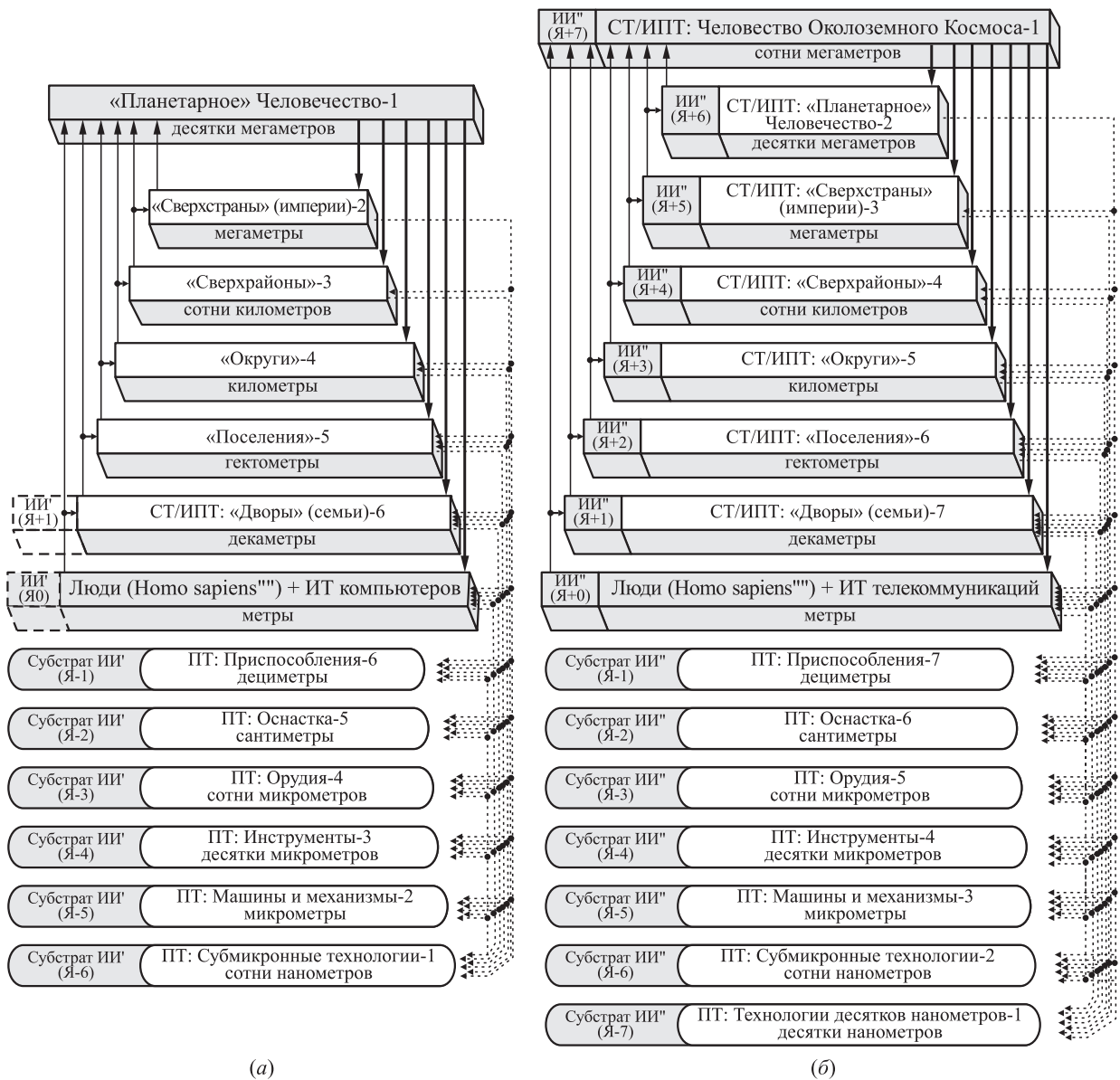
- БИТ тиражирования текстов/книгопечатания (начиная примерно с 1446 г. — «носитель» как книга/журнал/газета/листочка/инструктивный материал и т. п.);
- БИТ письменности/чтения (возникшая примерно 8,1 тыс. лет назад — «носитель» как рукопись).

Наконец, еще более ранняя в истории Человечества БИТ речи/языка (возникшая примерно 123 тыс. лет назад) в некоторой степени реализовывала лишь первую составляющую указанной интеллектуальной комбинации — память (чисто биологическую, управляемую сознанием человека, с носителем — корой головного мозга).

Согласно данной связке ИКМ и ИИ, инфраструктура ИИ пространственно структурируется на следующие типичные иерархические компоненты (рис. 1):

- (Я0) искусственная составляющая «личностного естественно-искусственного интеллекта» (в пределах пространства, соразмерного человеку/личности/индивиду — т. е. личного психологического пространства человека с характерным размером $\sim 4,2$ м — радиусом круга вокруг него);
- (Я+1) ИИ, ориентированный на решение задач уровня семьи — двора — малого коллектива (с характерным размером социума/инфраструктуры до ~ 64 м — радиусом круга той же площади);
- (Я+2) ИИ, ориентированный на решение задач уровня поселения — крупного коллектива (с характерным размером социума/инфраструктуры до ~ 1 км);
- (Я+3) ИИ, ориентированный на решение задач уровня округа (с характерным размером социума/инфраструктуры до ~ 15 км);
- (Я+4) ИИ, ориентированный на решение задач уровня «сверхрайона» (с характерным размером социума/инфраструктуры до ~ 222 км);
- (Я+5) ИИ, ориентированный на решение задач уровня «сверхстраны» (с характерным размером социума и инфраструктуры до ~ 3370 км);

¹Временные и пространственные количественные параметры ИКМ базируются на геометрической прогрессии со знаменателем $e^e = 15,15426\dots$, выявленной А. В. Жирмунским и В. И. Кузьминым при исследовании критических уровней в развитии биосистем [8].



(a)

(б)

Рис. 1 Подсистемы иерархосетевой системы Человечества на этапе развития: (а) компьютерном — примерно с 1946 г.; (б) телекоммуникационном — примерно с 1979 г.; СТ — социальные технологии; ИПТ — инфраструктурные производственные технологии; ПТ — производственные технологии; Я — ярусы. Восходящие стрелки (имеющие структуру «многие — к одному») отражают первую из трех основных составляющих контура поисковой оптимизации — *поисковую активность* представителей соответствующих ярусов в иерархии. Нисходящие сплошные (имеющие структуру «один — ко многим») стрелки отражают вторую составляющую — *целевые* критерии поисковой оптимизации энергетики системы Человечества. Нисходящие пунктирные («один — ко многим») стрелки отражают третью составляющую — *оптимизационную системную память* личностно-производственно-социального (результат адаптивных влияний представителей вышележащих иерархических ярусов на структуру вложенных в них нижележащих)

(Я+6) ИИ, ориентированный на решение задач уровня Планетарного Человечества (с характерным размером социума/инфраструктуры до ~ 51 тыс. км);

(Я+7) ИИ, ориентированный на решение задач уровня Человечества Околосземного Космоса (с характерным размером социума/инфраструкту-

ры в объеме, эквивалентном шару радиусом ~ 770 тыс. км), и т. д.

Производственные технологии (ПТ), которые должны обеспечивать реализацию субстрата ИИ указанных уровней/ярусов в иерархии системы Человечества, соответственно:

- (Я–1) ПТ с точностью до ~ 28 см;
- (Я–2) ПТ с точностью до ~ 1,8 см;
- (Я–3) ПТ с точностью до ~ 1,2 мм;
- (Я–4) ПТ с точностью до ~ 80 мкм;
- (Я–5) ПТ с точностью до ~ 5 мкм;
- (Я–6) ПТ с точностью до ~ 0,35 мкм;
- (Я–7) ПТ с точностью до ~ 23 нм и т. д.

На первом этапе формирования иерархического ИИ (рис. 1, а) возможны лишь начальные попытки этого — на базе возникших тогда прообразов локальных компьютерных сетей (ярус Я+1). Полностью этот процесс обеспечила БИТ телекоммуникаций (рис. 1, б).

Из определения «человеко-аппаратурной интеллектуальной единицы» следует, что «искусственный интеллект» — не автономная сущность, а неотъемлемый элемент триады [9] «естественный интеллект — интерфейс — искусственный интеллект», предназначение которой — выполнение творческих функций. Этим она отличается от интеллектуальных роботов, функция которых существенно шире и совпадает с основной функцией человека — формировать вокруг себя *вторую природу*. Именно в активности деятельности интеллектуальных роботов — промежуточного результата технологического развития Человечества — и заключается экзистенциальная — гуманитарная! — опасность для последнего.

Выводы

1. Высказывание С. Лема о том, что «как существуют двигатели разной мощности, так же могут быть разной силы искусственные интеллекты» [10], с позиций информатико-кибернетического подхода выглядит вполне оправданным.
2. Человек с его *естественным интеллектом* — как центральный субъект в системной пространственно-временной иерархии личностно-производственно-социальной природы (Человечества) и как элемент в системной иерархии живой природы — есть природный феномен. В ходе своего эволюционного саморазвития он естественным образом формирует окружающие его сущности, к каковым, несомненно, должен быть отнесен и «искусственный интеллект» как «феномен формирующейся второй природы».

3. Позиционирование ИТ локальных компьютеров и ИТ телекоммуникаций в качестве неотъемлемых составляющих совокупности монотонно усложняющихся в ходе цивилизационного развития информационных технологий позволяет рассматривать их появление и функционирование в широком контексте единой исторической ретроспективы и перспективы, давая возможность делать не только теоретические, но и практические выводы.

Литература

1. О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации от 10.10.2019 № 490. <http://publication.pravo.gov.ru/Document/View/0001201910110003>.
2. *Зиновьев А. А.* Логический интеллект. — М.: МосГУ, 2006. 282 с.
3. *Гринченко С. Н.* Метаэволюция (систем неживой, живой и социально-технологической природы). — М.: ИПИ РАН, 2007. 456 с.
4. *Гринченко С. Н.* «Человеко-аппаратурная интеллектуальная единица» как элемент информационного общества: кибернетический взгляд // Большая Евразия: Развитие, безопасность, сотрудничество. — М.: ИНИОН РАН, 2018. Вып. 2. Ч. 2. С. 685–688.
5. *Гринченко С. Н.* О генезисе информационного общества: информатико-кибернетическое модельное представление // Информатика и её применения, 2019. Т. 13. Вып. 2. С. 100–108.
6. *Гринченко С. Н.* Интеллект и «поисково-оптимизационная» картина мира // Открытое образование, 2005. № 2(49). С. 39–42.
7. *Гринченко С. Н.* О пространственном структурировании феномена «искусственный интеллект» // ИТНОУ: Информационные технологии в науке, образовании и управлении, 2019. № 4(14). С. 10–16.
8. *Жирмунский А. В., Кузьмин В. И.* Критические уровни в процессах развития биологических систем. — М.: Наука, 1982. 179 с.
9. *Гринченко С. Н.* Биполярность и триадичность: кибернетический взгляд на проблему // *Biocosmology — neo-Aristotelism*, 2016. Vol. 6. No. 1. P. 166–175.
10. *Лем С.* Artificial servility // Мегабитовая бомба / Пер. с польск. — 2002. (*Lem S.* Artificial servility // *Bomba megabitowa*. — Krakow, Poland: Wydawnictwo Literackie, 1999. 228 p.)

Поступила в редакцию 10.02.2020

ON THE SYSTEM HIERARCHY OF ARTIFICIAL INTELLIGENCE

S. N. Grinchenko

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

Abstract: Artificial intelligence (AI) is considered from the point of view of informatics-cybernetic modeling (ICM) of the development process of the self-controlling hierarchical-network system of Humankind as a natural phenomenon, closely associated with the concepts of “cognitive functions of man” and “intellectual activity of man.” Based on the information-communication-infrastructure component of the AI definition and on ICM, the concept of “human-hardware intellectual unit” is naturally generalized to all levels/tiers of the Humanity system located in its hierarchy above and below relative to the level/tier of “personality.” As a result, the phenomenon of “personal natural-artificial intelligence” is supplemented by the phenomenon of “hierarchical AI.” Its formation became possible starting from ~ 1946 — with the advent of basic information technology (BIT) of computers and took on an explosive character from ~ 1979 — with the advent of telecommunication/network BIT. Typical sizes of the ranges of levels/tiers in the hierarchy of the AI of the Humankind system are given (indicated dates and sizes are the result of a model calculation).

Keywords: artificial intelligence; information technology; informatics-cybernetic model; self-controlling hierarchical-network system of Humankind; human-hardware intellectual unit

DOI: 10.14357/19922264210115

References

1. O razvitii iskusstvennogo intellekta v Rossiyskoy Federatsii: Ukaz Prezidenta ot 10.10.2019 No. 490 [About strategy of scientific and technological development of the Russian Federation. Presidential Decree No. 490 dated 10.10.2019]. Available at: <http://static.kremlin.ru/media/events/files/ru/AH4x6HgKWANwVtMOFPDhcbRpvD1HCCsv.pdf> (accessed December 8, 2020).
2. Zinov'ev, A. A. 2006. *Logicheskiy intellekt* [The logical intellect]. Moscow: MosGU. 282 p.
3. Grinchenko, S. N. 2007. *Metaevolyutsiya (sistem nezhiyov, zhivoy i sotsial'no-tekhnologicheskoy prirody)* [Metaevolution (of inanimate, animate, and sociotechnological nature systems)]. Moscow: IPIRAN. 456 p.
4. Grinchenko, S. N. 2019. “Cheloveko-apparaturaya intellektual'naya edinita” kak element informatsionnogo obshchestva: kiberneticheskiy vzglyad [«Human-apparatus intellectual unit» as an element of the information society: Cybernetic view]. *Bol'shaya Evraziya: Razvitiye, bezopasnost', sotrudnichestvo* [Greater Eurasia: Development, security, cooperation] 2(2):685–688.
5. Grinchenko, S. N. 2019. O genezise informatsionnogo obshchestva: informatiko-kiberneticheskoe model'noe predstavlenie [On the genesis of the information society: Informatics-cybernetic model representation]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):100–108.
6. Grinchenko, S. N. 2005. Intellekt i “poiskovo-optimizatsionnaya” kartina mira [Intelligence and the “search-optimization” picture of the World]. *Otkrytoe obrazovanie* [Open Education] 2:39–42.
7. Grinchenko, S. N. 2019. O prostranstvennom strukturirovaniy fenomena “iskusstvennyy intellekt” [On the spatial structuring of the phenomenon of “artificial intelligence”]. *ITNOU: Informatsionnye tekhnologii v nauke, obrazovanii i upravlenii* [ITNOU: Information Technologies in Science, Education and Control] 14(4):10–16.
8. Zhirmunskiy, A. V., and V. I. Kuz'min. 1982. *Kriticheskie urovni v protsessakh razvitiya biologicheskikh sistem* [Critical levels in the development of biological systems]. Moscow: Nauka. 179 p.
9. Grinchenko, S. N. 2016. Bipolyarnost' i triadichnost': kiberneticheskiy vzglyad na problemu [Bipolarity and triad: A cybernetic view of the problem]. *Biokosmologiya — neo-Aristotelizm* 6(1):166–175.
10. Lem, S. 1999. *Artificial servility. Bomba megabitowa*. Krakow, Poland: Wydawnictwo Literackie. 228 p.

Received February 10, 2020

Contributor

Grinchenko Sergey N. (b. 1946) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sgrinchenko@ipiran.ru

О ТОЧНОСТИ НОРМАЛЬНОЙ АППРОКСИМАЦИИ ПРИ ОТСУТСТВИИ НОРМАЛЬНОЙ СХОДИМОСТИ*

В. Ю. Королев¹, А. В. Дорофеева²

Аннотация: При решении прикладных задач в самых разных областях принято использовать нормальное распределение в качестве модели статистических закономерностей в наблюдаемых данных с аддитивной структурой. В качестве критерия степени адекватности такой модели можно использовать оценки скорости сходимости в центральной предельной теореме (ЦПТ) теории вероятностей, устанавливающей, что при определенных условиях (например, при условии Линдеберга) суммарное воздействие большого числа случайных факторов проявляется в виде случайной величины с нормальным распределением. Классические оценки скорости сходимости в ЦПТ типа неравенства Берри–Эссеена доказаны при условии конечности третьих моментов слагаемых. Известны также оценки скорости сходимости при существовании моментов порядка $2 + \delta$ с $0 < \delta < 1$. Если существуют моменты лишь второго порядка, то сходимости в ЦПТ может быть как угодно медленной. Если же у слагаемых моменты второго порядка не существуют, то сходимости распределений сумм независимых случайных величин к нормальному закону не имеет места. Условия, гарантирующие справедливость ЦПТ, практически невозможно достоверно проверить при ограниченном объеме наблюдаемой выборки. Поэтому вопрос о том, какой может быть реальная точность нормальной аппроксимации, когда она теоретически не применима, но используется в практических вычислениях, представляет большой интерес. Более того, в некоторых ситуациях при имитационном моделировании, когда распределения слагаемых принадлежат области притяжения устойчивого закона с характеристическим показателем, меньшим двух, при увеличении числа слагаемых сначала наблюдается убывание расстояния между распределением нормированной суммы и нормальным законом и лишь при довольно большом числе слагаемых это расстояние начинает увеличиваться. В данной заметке предпринята попытка дать ответ на сформулированный выше вопрос и привести некоторые теоретические объяснения указанному эффекту.

Ключевые слова: центральная предельная теорема; точность нормальной аппроксимации; тяжелые хвосты; равномерное расстояние

DOI: 10.14357/19922264210116

1 Введение

При решении прикладных задач в самых разных областях принято использовать нормальное распределение в качестве модели статистических закономерностей в наблюдаемых данных с аддитивной структурой. В качестве критерия степени адекватности такой модели можно использовать оценки скорости сходимости в ЦПТ теории вероятностей, устанавливающей, что при определенных условиях (например, при условии Линдеберга) суммарное воздействие большого числа случайных факторов проявляется в виде случайной величины с нормальным распределением.

Классические оценки скорости сходимости в ЦПТ типа неравенства Берри–Эссеена доказаны

при условии конечности третьих моментов слагаемых.

Известны также оценки скорости сходимости при существовании моментов порядка $2 + \delta$ с $0 < \delta < 1$ (см. подробный обзор в [1]). Если существуют моменты лишь второго порядка, то сходимости в ЦПТ может быть как угодно медленной [2, 3]. Если же у слагаемых моменты второго порядка не существуют, то сходимости распределений сумм независимых случайных величин к нормальному закону не имеет места.

Условия, гарантирующие справедливость ЦПТ, практически невозможно достоверно проверить при ограниченном объеме наблюдаемой выборки. В частности, гистограмма, построенная по выборке из имеющего очень тяжелые хвосты распределения

* Работа выполнена при поддержке РФФИ (проект 18-07-01405).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vkorolev@cs.msu.ru

² Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, alex.dorofeyeva@gmail.com

Коши (у которого отсутствует даже математическое ожидание), при умеренном объеме выборки может быть визуально практически неотличимой от нормального распределения. Поэтому вопрос о том, какой может быть реальная точность нормальной аппроксимации, когда она теоретически не применима, но используется в практических вычислениях, представляет большой интерес. Более того, в некоторых ситуациях при имитационном моделировании, когда распределения слагаемых принадлежат области притяжения устойчивого закона с характеристическим показателем, меньшим двух, при увеличении числа слагаемых сначала наблюдается убывание расстояния между распределением нормированной суммы и нормальным законом и лишь при довольно большом числе слагаемых это расстояние начинает увеличиваться.

В данной заметке предпринята попытка дать ответ на сформулированный выше вопрос и привести некоторые теоретические объяснения указанному эффекту.

2 Обозначения и вспомогательные результаты

Пусть $n \in \mathbb{N}$, ξ_1, \dots, ξ_n — независимые необязательно одинаково распределенные случайные величины, заданные на вероятностном пространстве $(\Omega, \mathfrak{A}, \mathbb{P})$. Обозначим $F_j(x) = \mathbb{P}(\xi_j < x)$, $x \in \mathbb{R}$, $j \in \mathbb{N}$. Без существенного ограничения общности для удобства будем считать, что все функции распределения $F_j(x)$ непрерывны.

Обозначим $S_n = \xi_1 + \dots + \xi_n$. Индикатор множества (события) A обозначим $\mathbb{I}(A)$. Пусть $u > 0$. Очевидно,

$$\xi_j = \xi_j \mathbb{I}(|\xi_j| \leq u) + \xi_j \mathbb{I}(|\xi_j| > u).$$

Тогда

$$\begin{aligned} S_n &= \sum_{j=1}^n \xi_j \mathbb{I}(|\xi_j| \leq u) + \sum_{j=1}^n \xi_j \mathbb{I}(|\xi_j| > u) \equiv \\ &\equiv S_n^{(\leq u)} + S_n^{(> u)}. \end{aligned}$$

Если условиться считать равенство единице индикатора $\mathbb{I}(|\xi_j| \leq u)$ «успехом», а противоположное событие — «неудачей», то число $N_n(u)$ ненулевых слагаемых в сумме $S_n^{(\leq u)}$ будет случайной величиной, имеющей пуассон-биномиальное распределение с параметрами n и $p_j = p_j(u) = \mathbb{P}(|\xi_j| \leq u) = F_j(u) - F_j(-u)$, $j = 1, \dots, n$. Заметим, что при неограниченном увеличении u параметры p_j стремятся к единице.

Лемма 1. Пусть $A, B \in \mathfrak{A}$. Тогда

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) - \mathbb{P}(\overline{B}).$$

Доказательство элементарно.

Равномерное расстояние между функциями распределения F_ξ и F_η случайных величин ξ и η будем обозначать $\rho(F_\xi, F_\eta)$:

$$\rho(F_\xi, F_\eta) = \sup_x |F_\xi(x) - F_\eta(x)|.$$

Нормальную функцию распределения со средним $a \in \mathbb{R}$ и дисперсией $\sigma^2 > 0$ обозначим $\Phi_{a,\sigma}$:

$$\begin{aligned} \Phi_{a,\sigma}(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(z-a)^2}{2\sigma^2}\right\} dz = \\ &= \Phi_{0,1}\left(\frac{x-a}{\sigma}\right) = \Phi_{0,\sigma}(x-a), \quad x \in \mathbb{R}. \end{aligned}$$

Лемма 2. Для любых $a \in \mathbb{R}$, $\sigma > 0$, $b \in \mathbb{R}$

$$\rho(\Phi_{a+b,\sigma}, \Phi_{a,\sigma}) = 2\Phi_{0,\sigma}\left(\frac{|b|}{2}\right) - 1.$$

Доказательство. Заметим, что если $H(x)$ и $G(x)$ — дифференцируемые функции распределения, то $\rho(H, G)$ реализуется (т.е. точная верхняя грань $\sup_x |H(x) - G(x)|$ по x достигается) в одной из точек x , где $F'(x) = G'(x)$. Действительно,

$$\begin{aligned} \rho(H, G) &= \sup_x |H(x) - G(x)| = \\ &= \max\left\{\max_x [H(x) - G(x)], \max_x [G(x) - H(x)]\right\} \end{aligned}$$

и экстремум каждого из выражений в фигурных скобках достигается в такой точке, где производная соответствующего выражения равна нулю, что равносильно равенству производных функций распределения H и G , т.е. равенству соответствующих плотностей. В рассматриваемом случае нормальных плотностей последнее условие эквивалентно тому, что

$$\begin{aligned} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-a-b}{\sigma}\right)^2\right\} &= \\ = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2\right\}, \end{aligned}$$

или $(x - (a+b))^2 = (x-a)^2$. Решая данное уравнение, получаем $x - a = b/2$, откуда с учетом соотношения $\Phi_{0,\sigma}(-|b|) = 1 - \Phi_{0,\sigma}(|b|)$ вытекает требуемое утверждение.

Используя формулу Лагранжа, из леммы 2 легко получить известное неравенство:

$$\rho(\Phi_{a+b,\sigma}, \Phi_{a,\sigma}) \leq \frac{|b|}{\sigma\sqrt{2\pi}}$$

(см., например, неравенство (3.4) в книге [4]).

Лемма 3. Пусть $n \in \mathbb{N}$, ξ_1, \dots, ξ_n — случайные величины, a_1, \dots, a_n — положительные числа, такие что $a_1 + \dots + a_n = 1$. Тогда для любого $x > 0$

$$\mathbb{P}\left(\left|\sum_{j=1}^n \xi_j\right| \geq x\right) \leq \sum_{j=1}^n \mathbb{P}(|\xi_j| \geq a_j x).$$

Если дополнительно случайные величины ξ_1, \dots, ξ_n одинаково распределены, то

$$\mathbb{P}\left(\left|\sum_{j=1}^n \xi_j\right| \geq x\right) \leq n\mathbb{P}\left(|\xi_1| \geq \frac{x}{n}\right).$$

Доказательство. Заметим, что

$$\mathbb{P}\left(\left|\sum_{j=1}^n \xi_j\right| \geq x\right) \leq \mathbb{P}\left(\sum_{j=1}^n |\xi_j| \geq x\right).$$

Из геометрических соображений вытекает, что

$$\begin{aligned} \left\{\omega : \sum_{j=1}^n |\xi_j(\omega)| \geq x\right\} &\subseteq \{\omega : |\xi_1(\omega)| \geq a_1 x\} \cup \\ &\cup \left\{\omega : \sum_{j=2}^n |\xi_j(\omega)| \geq (1 - a_1)x\right\} \subseteq \\ &\subseteq \{\omega : |\xi_1(\omega)| \geq a_1 x\} \cup \{\omega : |\xi_2(\omega)| \geq a_2 x\} \cup \\ &\cup \left\{\omega : \sum_{j=3}^n |\xi_j(\omega)| \geq (1 - a_1 - a_2)x\right\} \subseteq \dots \\ &\dots \subseteq \bigcup_{j=1}^n \{\omega : |\xi_j(\omega)| \geq a_j x\}. \end{aligned}$$

Поэтому

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{j=1}^n \xi_j\right| \geq x\right) &\leq \mathbb{P}\left(\sum_{j=1}^n |\xi_j| \geq x\right) \leq \\ &\leq \mathbb{P}\left(\bigcup_{j=1}^n \{\omega : |\xi_j(\omega)| \geq a_j x\}\right) \leq \sum_{j=1}^n \mathbb{P}(|\xi_j| \geq a_j x). \end{aligned}$$

Лемма доказана.

3 Основные результаты

Рассмотрим оценку равномерного расстояния между распределением суммы $S_n = S_n^{(\leq u)} + S_n^{(>u)}$ и нормальным законом с соответствующими математическим ожиданием $a \in \mathbb{R}$ и дисперсией $\sigma > 0$, конкретный выбор которых прокомментируем ниже.

Теорема 1. Пусть $\epsilon > 0$, $u > 0$ — произвольны. Тогда для любых $a \in \mathbb{R}$, $\sigma > 0$

$$\begin{aligned} \rho(F_{S_n}, \Phi_{a,\sigma}) &\leq \rho\left(F_{S_n^{(\leq u)}}, \Phi_{a,\sigma}\right) + \\ &+ \sum_{j=1}^n [F_j(-u) + 1 - F_j(u)]. \end{aligned} \quad (1)$$

Доказательство. Пусть $\epsilon > 0$ произвольно. На основании леммы 1 имеем:

$$\begin{aligned} \mathbb{P}(S_n < x) &= \mathbb{P}\left(S_n < x; \left|S_n^{(>u)}\right| \leq \epsilon\right) + \\ &+ \mathbb{P}\left(S_n < x; \left|S_n^{(>u)}\right| > \epsilon\right) \geq \\ &\geq \mathbb{P}\left(S_n^{(\leq u)} < x - S_n^{(>u)}; \left|S_n^{(>u)}\right| \leq \epsilon\right) \geq \\ &\geq \mathbb{P}\left(S_n^{(\leq u)} < x - \epsilon; \left|S_n^{(>u)}\right| \leq \epsilon\right) \geq \\ &\geq \mathbb{P}\left(S_n^{(\leq u)} < x - \epsilon\right) - \mathbb{P}\left(\left|S_n^{(>u)}\right| \geq \epsilon\right). \end{aligned} \quad (2)$$

С другой стороны, очевидно:

$$\begin{aligned} \mathbb{P}(S_n < x) &= \mathbb{P}\left(S_n^{(\leq u)} < x - S_n^{(>u)}; \left|S_n^{(>u)}\right| \leq \epsilon\right) + \\ &+ \mathbb{P}\left(S_n < x; \left|S_n^{(>u)}\right| > \epsilon\right) \leq \\ &\leq \mathbb{P}\left(S_n^{(\leq u)} < x + \epsilon; \left|S_n^{(>u)}\right| \leq \epsilon\right) + \\ &+ \mathbb{P}\left(S_n < x; \left|S_n^{(>u)}\right| > \epsilon\right) \leq \\ &\leq \mathbb{P}\left(S_n^{(\leq u)} < x + \epsilon\right) + \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right). \end{aligned} \quad (3)$$

Легко видеть, что

$$|\mathbb{P}(S_n < x) - \Phi_{a,\sigma}(x)| = \max\{\mathbb{P}(S_n < x) - \Phi_{a,\sigma}(x), \Phi_{a,\sigma}(x) - \mathbb{P}(S_n < x)\}. \quad (4)$$

Используя (3) и лемму 2, получим:

$$\begin{aligned} \mathbb{P}(S_n < x) - \Phi_{a,\sigma}(x) &\leq \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right) + \\ &+ \left[\mathbb{P}\left(S_n^{(\leq u)} < x + \epsilon\right) - \Phi_{a,\sigma}(x + \epsilon)\right] + \\ &+ [\Phi_{a,\sigma}(x + \epsilon) - \Phi_{a,\sigma}(x)] \leq \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right) + \\ &+ \rho\left(F_{S_n^{(\leq u)}}, \Phi_{a,\sigma}\right) + \left[2\Phi_{0,\sigma}\left(\frac{\epsilon}{2}\right) - 1\right]. \end{aligned} \quad (5)$$

Используя (2) и лемму 2, получим

$$\begin{aligned} & \Phi_{a,\sigma}(x) - \mathbb{P}(S_n < x) \leq \Phi_{a,\sigma}(x) - \\ & - \mathbb{P}(S_n^{(\leq u)} < x - \epsilon) + \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right) = \\ & = [\Phi_{a,\sigma}(x) - \Phi_{a,\sigma}(x - \epsilon)] - \left[\mathbb{P}\left(S_n^{(\leq u)} < x - \epsilon\right) - \right. \\ & \quad \left. - \Phi_{a,\sigma}(x - \epsilon)\right] + \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right) \leq \\ & \leq \left[2\Phi_{0,\sigma}\left(\frac{\epsilon}{2}\right) - 1\right] + \rho(F_{S_n^{(\leq u)}}, \Phi_{a,\sigma}) + \\ & \quad + \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right). \quad (6) \end{aligned}$$

Подставив оценки (5) и (6) в (4), получим

$$\begin{aligned} \rho(F_{S_n}, \Phi_{a,\sigma}) & \leq \rho\left(F_{S_n^{(\leq u)}}, \Phi_{a,\sigma}\right) + \\ & + \left[2\Phi_{0,\sigma}\left(\frac{\epsilon}{2}\right) - 1\right] + \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right). \quad (7) \end{aligned}$$

Рассмотрим третье слагаемое в правой части (7). На основании леммы 3 по формуле полной вероятности, принимая во внимание тот факт, что $\epsilon/n > 0$ и $|\xi_j(\omega)|\mathbb{I}(|\xi_j(\omega)| > u) = 0$ для тех ω , для которых $|\xi_j(\omega)| \leq u, j = 1, \dots, n$, имеем:

$$\begin{aligned} \mathbb{P}\left(\left|S_n^{(>u)}\right| > \epsilon\right) & = \mathbb{P}\left(\left|\sum_{j=1}^n \xi_j \mathbb{I}(|\xi_j| > u)\right| > \epsilon\right) \leq \\ & \leq \sum_{j=1}^n \mathbb{P}\left(|\xi_j| \mathbb{I}(|\xi_j| > u) > \frac{\epsilon}{n}\right) = \\ & = \sum_{j=1}^n \left[\mathbb{P}\left(|\xi_j| \mathbb{I}(|\xi_j| > u) > \frac{\epsilon}{n} \mid |\xi_j| > u\right) \mathbb{P}(|\xi_j| > u) + \right. \\ & \quad \left. + \mathbb{P}\left(|\xi_j| \mathbb{I}(|\xi_j| > u) > \frac{\epsilon}{n} \mid |\xi_j| \leq u\right) \mathbb{P}(|\xi_j| \leq u)\right] = \\ & = \sum_{j=1}^n [1 - p_j(u)] \mathbb{P}\left(|\xi_j| \mathbb{I}(|\xi_j| > u) > \frac{\epsilon}{n} \mid |\xi_j| > u\right) \leq \\ & \leq \sum_{j=1}^n [F_j(-u) + 1 - F_j(u)]. \quad (8) \end{aligned}$$

Подставив (8) в (7) и устремив ϵ к нулю, получим требуемое. Теорема доказана.

На практике в качестве параметров a и σ можно брать, например,

$$\begin{aligned} a & = a(u) = \mathbb{E}S_n^{(\leq u)} = \mathbb{E} \sum_{j=1}^n \xi_j \mathbb{I}(|\xi_j| \leq u) = \\ & = \sum_{j=1}^n \mathbb{E}[\xi_j \mathbb{I}(|\xi_j| \leq u)] = \sum_{j=1}^n p_j(u) \int_{-u}^u x dF_j(x); \end{aligned}$$

$$\begin{aligned} \sigma^2 & = \sigma^2(u) = \mathbb{D}S_n^{(\leq u)} = \mathbb{D} \sum_{j=1}^n \xi_j \mathbb{I}(|\xi_j| \leq u) = \\ & = \sum_{j=1}^n \mathbb{D}[\xi_j \mathbb{I}(|\xi_j| \leq u)] = \\ & = \sum_{j=1}^n \left[p_j(u) \int_{-u}^u x^2 dF_j(x) - \left(p_j(u) \int_{-u}^u x dF_j(x) \right)^2 \right]. \end{aligned}$$

При фиксированном $u > 0$ первое слагаемое в правой части (1) убывает с увеличением n , тогда как второе возрастает. При этом существует $n_0 \geq 1$ такое, что при $1 \leq n \leq n_0$ вся правая часть (1) убывает, а при $n \geq n_0$ возрастает. В случае одинаково распределенных слагаемых второе слагаемое возрастает как kn , где $k > 0$. При этом за счет выбора очень большого u можно добиться произвольной малости коэффициента k и, как следствие, очень медленного роста второго слагаемого. Поэтому параметр n_0 может принимать довольно большие значения.

Задачу определения указанного n_0 конкретизируем для частного случая. Предположим, что

$$\rho\left(F_{S_n^{(\leq u)}}, \Phi_{a,\sigma}\right) \leq Cn^{-\gamma}, \quad (9)$$

где $\gamma > 0$, а коэффициент $C = C(u)$ определяется свойствами распределений слагаемых в сумме $S_n^{(\leq u)}$. Некоторые критерии справедливости (9) приведены, например, в [5]. Предположим также, что

$$\sum_{j=1}^n [F_j(-u) + 1 - F_j(u)] \leq kn, \quad (10)$$

где $k = k(u) > 0$. Несложно убедиться, что минимум функции

$$g(x) = \frac{C}{x^\gamma} + kx$$

достигается в точке

$$x_0 = \left(\frac{C\gamma}{k}\right)^{1/(1+\gamma)},$$

причем

$$\min_{x \geq 0} g(x) = g(x_0) = (\gamma + 1) \left(\frac{Ck\gamma}{\gamma^\gamma}\right)^{1/(\gamma+1)}. \quad (11)$$

Таким образом, в качестве n_0 выступает либо $[x_0]$, либо $[x_0] + 1$. При этом правую часть (11) можно рассматривать как приближенное значение наилучшей верхней оценки точности нормальной аппроксимации при справедливости условий (9) и (10).

Литература

1. Шевцова И. Г. Точность нормальной аппроксимации: методы оценивания и новые результаты. — М.: Аргатак-Медиа, 2016. 380 с.
2. Мацкявичюс В. К. О нижней оценке скорости сходимости в центральной предельной теореме // Теория вероятностей и ее применения, 1983. Т. 28. Вып. 3. С. 565–569.
3. Korolev V. Yu., Dorofeeva A. V. Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions // *Lith. Math. J.*, 2017. Vol. 57. No. 1. P. 38–58.
4. Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972. 414 с.
5. Ибрагимов И. А. О точности аппроксимации функций распределения сумм независимых случайных величин нормальным распределением // Теория вероятностей и ее применения, 1966. Т. 11. Вып. 4. С. 632–655.

Поступила в редакцию 13.10.2020

ON THE ACCURACY OF THE NORMAL APPROXIMATION UNDER THE VIOLATION OF THE NORMAL CONVERGENCE

V. Yu. Korolev^{1,2} and A. V. Dorofeeva¹

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

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

Abstract: When solving applied problems in various fields, it is conventional to use the normal approximation to the distribution of data with additive structure. As a criterion of the adequacy of such a model, it is possible to use bounds for the convergence rate in the central limit theorem (CLT) of the probability theory stating that under certain conditions (say, under the Lindeberg condition), the total effect of very many random factors acts as a random variable with the normal distribution. The classical bounds for the convergence rate in the CLT such as the Berry–Esseen inequality are proved under the condition that the third moments of the summands exist. Also, bounds are known that require the existence of the moments of orders $2 + \delta$ with $0 < \delta < 1$. If only the moments of the second order exist, then the convergence in the CLT can be arbitrarily slow. But if the moments of the summands of the second order do not exist, then the convergence of the distributions of sums of independent random variables to the normal law does not take place. It is practically impossible to reliably check the conditions of the central limit theorem with the limited size of the available sample. Therefore, the question of what is the real accuracy of the normal approximation if it is theoretically impossible is of great interest. Moreover, in some situations, in computer simulation of sums of random variables whose distributions belong to the domain of attraction of the stable distribution with the characteristic exponent less than two, as the number of summands grows, first, the distance between the distribution of the normalized sum and the normal law decreases and starts to increase only when the number of summands becomes sufficiently large. In this paper, an attempt is undertaken to give some theoretical explanation of this effect and to give an answer to the question posed above.

Keywords: central limit theorem; accuracy of normal approximation; heavy tails; uniform distance

DOI: 10.14357/19922264210116

Acknowledgments

Research supported by the Russian Foundation for Basic Research, project 18-07-01405.

References

1. Shevtsova, I. G. 2016. *Tochnost' normal'noy approksimatsii: metody otsenivaniya i novye resul'taty* [The accuracy of the normal approximation: Methods of estimation and new results]. Moscow: Argatak-Media. 380 p.
2. Mackevicius, V. K. 1984. A lower bound for the convergence rate in the central limit theorem. *Theor. Probab. Appl.* 28(3):596–601.
3. Korolev, V. Yu., and A. V. Dorofeeva. 2017. Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions. *Lith. Math. J.* 57(1):38–58.

4. Petrov, V. V. 1972. *Summy nezavisimykh sluchaynykh velichin* [Sums of independent random variables]. Moscow: Nauka. 416 p.
5. Ibragimov, I. A. 1966. On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables. *Theor. Probab. Appl.* 11(4):559–579.

Received October 13, 2020

Contributors

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of Department, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vkorolev@cs.msu.ru

Dorofeeva Alexandra V. (b. 1991) — graduate PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; alex.dorofeyeva@gmail.com

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Арутюнов Евгений Наполеонович (р. 1952) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бахтеев Олег Юрьевич (р. 1993) — ассистент кафедры интеллектуальных систем Московского физико-технического института; руководитель отдела исследований Компании Антиплагиат

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Брюхов Дмитрий Олегович (р. 1971) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Вышинский Леонид Леонидович (р. 1941) — кандидат физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Александр Анатольевич (р. 1994) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гребенькова Ольга Сергеевна (р. 1999) — студентка Московского физико-технического института

Гринченко Сергей Николаевич (р. 1946) — доктор технических наук, профессор, главный научный сотрудник Института проблем информатики Феде-

рального исследовательского центра «Информатика и управление» Российской академии наук

Дорофеева Александра Владимировна (р. 1991) — выпускница аспирантуры кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Жуков Дмитрий Олегович (р. 1965) — доктор технических наук, профессор, заведующий кафедрой Московского технологического университета МИРЭА

Зальцман Анастасия Дмитриевна (р. 1989) — преподаватель Московского технологического университета МИРЭА

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ковалёв Дмитрий Юрьевич (р. 1988) — научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ковалёв Сергей Протасович (р. 1972) — доктор физико-математических наук, ведущий научный сотрудник Института проблем управления им. В. А. Трапезникова Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кружков Михаил Григорьевич (р. 1975) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Кузнецова Рита Валерьевна (р. 1990) — аспирант Московского физико-технического института

Липатьев Александр Андреевич (р. 1988) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Недоливко Юлия Николаевна (р. 1998) — студент факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Павлов Юрий Леонидович (р. 1949) — доктор физико-математических наук, профессор, главный научный сотрудник Института прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН»

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

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный со-

трудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Московского физико-технического института

Ступников Сергей Александрович (р. 1978) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Флёров Юрий Арсениевич (р. 1942) — доктор физико-математических наук, профессор, член-корреспондент РАН, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хватова Татьяна Юрьевна (р. 1971) — доктор экономических наук, профессор Санкт-Петербургского политехнического университета Петра Великого

Чехович Юрий Викторович (р. 1976) — кандидат физико-математических наук, заведующий отделом Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шанин Иван Андреевич (р. 1991) — младший 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 ЛАТЭК (.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 Publs. 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 Publs. 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: ieep@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>