

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года
Журнал выходит ежеквартально

Учредитель:
**Федеральный исследовательский центр
«Информатика и управление» Российской академии наук**

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета
академик РАН Г. И. Савин
академик РАН А. Л. Стемпковский
член-корреспондент РАН Ю. Б. Зубарев
профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)
профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)
профессор В. Ротарь (V. Rotar, San-Diego, USA)
профессор М. Финкельштейн (M. Finkelstein, Bloemfontein, South Africa)
профессор В. Хофкирхнер (W. Hofkirchner, Wien, Austria)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор
профессор, д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора
д.т.н. В. Н. Захаров
проф., д.ф.-м.н. А. И. Зейфман
проф., д.т.н. В. Д. Ильин
проф., д.т.н. К. К. Колин
проф., д.ф.-м.н. В. Ю. Королев
д.ф.-м.н. Р. В. Разумчик
д.ф.-м.н. В. И. Сеницын
проф., д.т.н. И. Н. Сеницын
проф., д.ф.-м.н. В. Г. Ушаков
д.ф.-м.н. А. К. Горшенин — отв. секретарь
к.ф.-м.н. С. А. Христочевский

Редакция

к.ф.-м.н. Е. Н. Арутюнов
С. Н. Стригина

© Федеральный исследовательский центр «Информатика
и управление» Российской академии наук, 2024

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science

Журнал входит в систему Российского индекса научного цитирования (РИНЦ)
Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory
и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНТИ
и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 34 № 2 Год 2024

СОДЕРЖАНИЕ

Условно-оптимальная фильтрация и управление стохастическими системами, не разрешенными относительно производных, со случайными параметрами И. Н. Сеницын	3
Моделирование нестационарного стохастического процесса посредством его канонического разложения на основе вейвлет-нейронной сети И. Н. Сеницын, В. И. Сеницын, Э. Р. Корепанов, Т. Д. Конашенкова	21
О моделировании совместного обслуживания одноадресного и многоадресного трафика с приоритизацией А. С. Румянцева, Ф. А. Москалева, А. К. Самуйлов, С. Я. Шоргин, Ю. В. Гайдамака	40
Статистический критерий стабильности системы массового обслуживания, основанный на последовательности времен пребывания М. П. Кривенко	55
Формализация синтеза самосинхронных счетчиков Ю. А. Степченков, Ю. Г. Дьяченко, Н. В. Морозов, Д. Ю. Степченков, Д. Ю. Дьяченко	66
Эволюция и перспективы концепции программного определения В. Б. Егоров	83
Моделирование динамических геообъектов на основе концепции фазовой вариантности Д. А. Никишин	95
Интеллектуальные системы поддержки принятия решений в медицине: понятие, проблемы, подходы к разработке С. Б. Румовская	107
Метод поиска оптимальных значений параметров алгоритма разрешения сущностей для конкретно-исторических данных И. М. Адамович, О. И. Волков	123
Об авторах	134
Правила подготовки рукописей статей	136
Requirements for manuscripts	140

УСЛОВНО-ОПТИМАЛЬНАЯ ФИЛЬТРАЦИЯ И УПРАВЛЕНИЕ СТОХАСТИЧЕСКИМИ СИСТЕМАМИ, НЕ РАЗРЕШЕННЫМИ ОТНОСИТЕЛЬНО ПРОИЗВОДНЫХ, СО СЛУЧАЙНЫМИ ПАРАМЕТРАМИ

*И. Н. Синицын*¹

Аннотация: Разработано методическое обеспечение синтеза фильтров и регуляторов для управления типовыми гауссовскими стохастическими системами (СтС) со случайными параметрами, описываемыми многокомпонентными интегральными каноническими представлениями (МК ИКП), не разрешенными относительно старших производных и допускающими приведение к дифференциальным. В основу синтеза положены методы теории условно-оптимальной фильтрации и управления В. С. Пугачёва. Особое внимание уделено синтезу управления на основе локальных критериев для нелинейностей, описывающих зависимости от старших производных, обладающих зонами неоднозначности и нечувствительности, гистерезиса и насыщения. Приводятся иллюстративные примеры. Предложены направления дальнейших исследований.

Ключевые слова: многокомпонентное интегральное каноническое представление (МК ИКП); случайные параметры; стохастическая система, не разрешенная относительно производных (СтС НРОП); условно-оптимальное управление (УОУ); условно-оптимальный фильтр (УОФ)

DOI: 10.14357/08696527240201

EDN: XLRCQE

1 Введение

В [1] рассмотрены вопросы синтеза нормальных субоптимальных фильтров (НСОФ) для дифференциальных СтС, не разрешенных относительно производных (НРОП). Представлены уравнения состояния и наблюдения нелинейных дифференциальных СтС НРОП. Синтез НСОФ выполнен при следующих условиях: (1) отсутствуют пуассоновские шумы в наблюдениях; (2) коэффициент при гауссовском шуме не зависит от состояния. Подробно рассмотрен синтез НСОФ при аддитивных шумах в уравнениях состояния и наблюдения. В [2] разработано методическое обеспечение для негладких правых частей уравнений СтС НРОП. Рассмотрены вопросы аналитического моделирования нормальных стохастических процессов (СтП) на основе нелинейных регрессионных моделей. Особое внимание уделено методам гауссовской фильтрации и экстраполяции. Изучены вопросы условно-оптимальной фильтрации и экстраполяции для СтС НРОП

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

с параметрическими шумами. В [3] разработано методическое и алгоритмическое обеспечение аналитического моделирования, оценивания и идентификации для существенно нестационарных процессов в СтС НРОП. Дан обзор профильных публикаций и изучены основные классы регрессионных уравнений СтС НРОП. Основные результаты:

- (1) для общего вида нелинейных СтС НРОП приведены оптимальные алгоритмы совместной фильтрации и распознавания;
- (2) для линейных гауссовских СтС НРОП получены простые алгоритмы;
- (3) для СтС НРОП, линейных относительно состояния X_t и нелинейных относительно наблюдений Y_t , получены соответствующие алгоритмы;
- (4) в случае (3) методом нормальной аппроксимации (МНА) получен простой алгоритм.

В статье [4] рассматривается задача разработки методического обеспечения и алгоритмов условно-оптимальной фильтрации процессов в СтС НРОП со случайными переменными параметрами, описываемых МК ИКП, и моделирования точности условно-оптимальных фильтров (УОФ). Приводятся сведения по теории МК ИКП и их преобразованиям. Отдельный раздел посвящен вопросам приведения дифференциальных СтС НРОП к дифференциальным. Приводятся основные результаты по теории УОФ, рассматривается УОФ для систем с мультипликативными шумами.

Теория условно-оптимальной фильтрации и управления непрерывными, дискретными и непрерывно-дискретными СтС восходит к работам В. С. Пугачёва, И. Е. Казакова, И. Н. Сеницына, В. И. Шина и Н. К. Моцука [5–10]. Современное изложение теории условно-оптимального управления (УОУ) дано в [11].

Рассмотрим обобщение [1–4] на случай приводимых непрерывных и дискретных СтС НРОП в части УОФ, а также теорию УОУ.

2 Многокомпонентные интегральные канонические представления случайных параметров

Как известно [4, 12], МК ИКП векторного СтП $X(t) = [X_1(t) \cdots X_n(t)]^T$ и матрицы ковариационных функций $K^x(t, t') = [K_{hl}^x(t, t')]$ имеет вид:

$$\left. \begin{aligned} X(t) &= m^x(t) + \sum_{\rho=1}^r \int_{\Lambda_\rho} V_\rho(\lambda) x_\rho(t, \lambda) d\lambda; \\ X_h(t) &= m_h^x(t) + \sum_{\rho=1}^r \int_{\Lambda_\rho} V_\rho(\lambda) x_{\rho h}(t, \lambda) d\lambda \quad (h = \overline{1, n}); \end{aligned} \right\} \quad (1)$$

$$K_{hl}(t, t') = \sum_{\rho=1}^r \int_{\Lambda_\rho} G_\rho(\lambda) x_{\rho h}(t, \lambda) \overline{x_{\rho l}(t', \lambda)} d\lambda \quad (h, l = \overline{1, n}), \quad (2)$$

где $V_1(\lambda), \dots, V_r(\lambda)$ — некоррелированные белые шумы, определяемые формулой

$$V_\rho(\lambda) = \sum_{h=1}^n \int_T \overline{a_{\rho h}(t, \lambda)} X_h^0(t) dt \quad (\lambda \in \Lambda_\rho; \quad \rho = \overline{1, n}).$$

Интенсивности белых шумов $V_\rho(\lambda)$ определяются формулой:

$$G_\rho(\lambda) = \sum_{h, l=1}^n \int_{\Lambda_\rho} \int_T \int_T \overline{a_{\rho h}(t, \lambda)} a_{\rho l}(t', \lambda') K_{hl}^x(t, t') dt dt' d\lambda' \quad (\lambda \in \Lambda_\rho; \quad \rho = \overline{1, r}).$$

Для вычисления координатных функций $x_{\rho h}(t', \lambda')$ и функций $a_{\rho h}(t, \lambda)$ используются формулы:

$$x_{\rho h}(t, \lambda) = \frac{1}{G_\rho(\lambda)} \sum_{l=1}^n \int_T a_{\rho l}(t', \lambda) K_{hl}^x(t, t') dt' \quad (\lambda \in \Lambda_\rho; \quad \rho = \overline{1, r}; \quad h = \overline{1, n}); \quad (3)$$

$$\sum_{h=1}^n \int_T \overline{a_{\mu h}(t, \lambda)} a_{\rho h}(t, \lambda') dt = \delta_{\rho\mu} \delta(\lambda - \lambda') \quad (\lambda \in \Lambda_\mu; \quad \lambda' \in \Lambda_\rho; \quad \rho, \mu = \overline{1, r}); \quad (4)$$

$$\sum_{\rho=1}^r \int_{\Lambda_\rho} \overline{a_{\rho l}(t', \lambda)} x_{\rho h}(t, \lambda) d\lambda = \delta_{hl} \delta(t - t') \quad (h, l = \overline{1, n}),$$

выражающие необходимые и достаточные условия представления векторного СтП посредством МК ИКП.

Теорема 2.1. В условиях (3) и (4) из МК ИКП (1) вытекает МК ИКП матрицы его ковариационных функций (2).

Теорема 2.2. Если известно МК ИКП векторного СтП (1), то вектор математического ожидания и матрица ковариационных функций линейного преобразования $Y_t(t) = A_t X_t$ допускают МК ИКП, определяемые формулами:

$$X_p(t) = m_p^x(t) + \sum_{\rho=1}^r \int_{\Lambda_\rho} V_\rho(\lambda) x_{\rho p}(t, \lambda) d\lambda \quad (p = \overline{1, n});$$

$$Y_p(s) = m_p^y(s) + \sum_{\rho=1}^r \int_{\Lambda_\rho} V_\rho(\lambda) y_{\rho p}(s, \lambda) d\lambda,$$

$$y_{\rho p}(s, \lambda) = \sum_{h=1}^n A_{\rho h} x_{\rho h}(t, \lambda) \quad (p = \overline{1, m}),$$

$$K^y(s, s') = [K_{pq}^y(s, s')],$$

$$K_{pq}^y(s, s') = \sum_{\rho=1}^r \int_{\Lambda_\rho} G_\rho(\lambda) y_{\rho p}(s, \lambda) \overline{y_{\rho q}(s', \lambda)} d\lambda \quad (p, q = \overline{1, m}).$$

В задачах нелинейной корреляционной теории невырожденные безынерционные скалярные и векторные существенно нелинейные преобразования $Y_t = \varphi_t(X_t)$ заменяют оптимальными в среднеквадратичном смысле линейными регрессионными преобразованиями [12]. Задача эквивалентной регрессионной линеаризации детерминированной векторной нелинейной функции $Y = \varphi(X)$ при использовании критерия минимума среднеквадратичной ошибки совпадает с классической задачей линейного регрессионного анализа. В этом случае оптимальная линейная среднеквадратичная регрессия вектора Y на вектор X определяется формулой:

$$m^y(X) = gX + a,$$

где

$$g = K^{yx}(K^x)^{-1};$$

$$a = m^y - gm^x.$$

Пусть $f(y, x)$ — совместная плотность случайных векторов Y и X ; m^x и K^x — математическое ожидание и ковариационная матрица вектора x , $\det |K^x| \neq 0$. Формула (5) при этом принимает вид:

$$g = K^{yx}(K^x)^{-1} = \int_{-\infty}^{\infty} [m^y(x) - m^y] (x - m^x)^T (K^x)^{-1} f_1(x) dx,$$

где $f_1(x)$ — плотность случайного вектора X . Эта формула вместе с приближенной формулой

$$m^y(X) \approx m^y + g(X - m^x)$$

дает статистическую линеаризацию регрессии $m^y(X)$ по Казакову.

Теорема 2.3. *Если существуют конечные моменты первого и второго порядка векторного СтП $X = X(t)$, то векторное нелинейное преобразование $Y = \varphi(X)$ допускает линейную среднеквадратичную регрессию $Y = Y(t)$ на $X = X(t)$, определяемую по Казакову формулами (5) и (6).*

Теорема 2.4. *Оптимальная среднеквадратичная линеаризация посредством МК ИКП (теорема 2.1) первого рода определяется следующими формулами:*

$$\varphi_t(X_t) \approx m_t^{(1)y}(X_t) = \varphi_{0t}^{(1)}(m_t^x, K_t^x) + g_t^{(1)}(m_t^x, K_t^x) \left[\sum_{\rho=1}^r \int_{\Lambda_\rho} V_\rho(\lambda) x_t(\lambda) d\lambda - m_t^x \right],$$

где

$$\varphi_{0t}^{(1)}(m_t^x, K_t^x) = M_N [\varphi(X_t)]; \quad g_t^{(1)}(m_t^x, K_t^x) = K_t^{yx} (K_t^x)^{-1}.$$

Здесь M_N — символ математического ожидания для нормального распределения.

Теорема 2.5. *Оптимальная среднеквадратичная линеаризация посредством МК ИКП (теорема 2.1) второго рода определяется следующими формулами:*

$$\begin{aligned} \varphi_t(X_t) \approx m_t^{(2)y}(X_t) &= \varphi_{0t}^{(2)}(m_t^x, K_{\rho t}^x) + \\ &+ \sum_{\rho=1}^r \int_{\Lambda_\rho} g_{\rho t}^{(2)}(m_t^x, K_{\rho t}^x) V_\rho(\lambda) x_{\rho t} d\lambda - g_t^{(2)}(m_t^x, K_t^x) m_t^x, \end{aligned}$$

где

$$\varphi_{0t}^{(2)}(m_t^x, K_{\rho t}^x) = M_N [\varphi(X_t)]; \quad g_t^{(2)}(m_t^x, K_{\rho t}^x) = K_{\rho t}^{yx} (K_{\rho t}^x)^{-1}.$$

3 Условно-оптимальная фильтрация

Следуя [4, 12], в качестве исходной СтС НРОП рассмотрим детерминированную векторную систему уравнений

$$\Phi = \Phi(t, X_t, \bar{X}_t, Y_t, \Theta_t, Q_t) = 0. \quad (7)$$

Здесь X_t — вектор состояния; $\bar{X}_t = [\dot{X}_t^\Gamma \dots X_t^{(l)\Gamma}]^\Gamma$ — расширенный вектор производных состояния, состоящий из l -го порядка производных по времени; Y_t — вектор наблюдений; Θ_t — вектор случайных параметров, описываемый МК ИКП (см. разд. 2); Q_t — вектор возмущений; Φ — нелинейная функция переменных $\bar{X}_t = [X_t^\Gamma \bar{X}_t^\Gamma Q_t^\Gamma]^\Gamma$ и Θ_t , допускающая при фиксированном векторе случайных параметров Θ_t среднеквадратичную регрессионную линеаризацию по \bar{X}_t и Q_t вида

$$\Phi \approx \Phi_0 + \sum_{j=1}^k k_{\bar{X},j}^\Phi \bar{X}_t^j + k_Q^\Phi Q_t^0.$$

Пусть вектор возмущений Q_t связан с гауссовским белым шумом V_0 линейным уравнением формирующего фильтра ($\Phi\Phi$):

$$\dot{Q}_t = a_t^Q Q_t + a_{0t}^Q + b_t^Q V_0, \quad (8)$$

где

$$M V_0 = 0; \quad M [V_0(\Theta_t) V_0(\Theta_t)^\Gamma] = \nu_0^{V_0} \delta(t - \tau).$$

Тогда при фиксированном Θ_t и при условиях $\det k_{\bar{X},l}^\Phi \neq 0$ и $\det k_Q^\Phi \neq 0$ дифференциальная СтС НРОП (7) приводится к дифференциальной СтС следующего вида (**теорема 3.1**):

$$\dot{\bar{X}}_{1t} = \bar{X}_{2t}, \dots, \dot{\bar{X}}_{(l-1)t} = \bar{X}_{lt}, \dot{\bar{X}}_{lt} = - \left(k_{\bar{X},l}^\Phi \right)^{-1} \bar{X}_t^{(l)} - \left(k_{\bar{X},l}^\Phi \right)^{-1} \left(k_Q^\Phi \right)^{-1} Q_t$$

и (8). Матрицы коэффициентов $k_{\bar{X},l}^\Phi$ и k_Q^Φ неявно зависят от первых двух вероятностных моментов $m_t^{\bar{X}}$ и $K_t^{\bar{X}}$.

Пусть объектовая СтС НРОП допускает приведение к дифференциальной, измерительная система вполне наблюдаема, наблюдения влияют на объект, а уравнение наблюдения разрешено относительно Y_t . Тогда в качестве исходных приведенных дифференциальных уравнений объекта и измерительной системы можно принять следующую:

$$\left. \begin{aligned} \dot{X}_t &= A^\pi \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, \Theta_t, t \right) = \\ &= a^\pi \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, \Theta_t, t \right) + b^\pi \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, \Theta_t, t \right) V_0(\Theta_t); \\ \dot{Y}_t &= B \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, t \right) = \\ &= a_1 \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, t \right) + b_1 \left(X_t, Y_t, m_t^{\bar{X}}, K_t^{\bar{X}}, t \right) V_0(\Theta_t). \end{aligned} \right\} \quad (9)$$

Здесь a^π , a_1 , b^π и b_1 — известные векторно-матричные функции отмеченных переменных; V_0 — векторный нормальный (гауссовский) белый шум интенсивности $\nu_0^Q = \nu_0^Q(\Theta_t)$.

Теория УОФ для СтС, приводимых к дифференциальным, со случайными параметрами подробно представлена в [1–4], для дискретных и непрерывно-дискретных систем — в [11, 12]. Рассмотрим теорию УОУ для СтС НРОП, приводимых к дифференциальным.

4 Условно-оптимальное управление

Как известно [13, 14], уравнения систем управления путем ввода при необходимости дополнительных переменных состояния можно представить в виде системы дифференциальных разностных и смешанных дифференциально-разностных уравнений первого порядка, разрешенных относительно производных. Добавив к уравнениям всех линейных частей системы зависимости между входными и выходными сигналами всех нелинейных звеньев, получим нелинейную систему дифференциально-разностных уравнений. Следуя [11–14], обозначим через $z(t) = [z_1(t) \cdots z_p(t)]^T$ расширенный вектор состояния системы, состоящий из непрерывной компоненты $z'(t)$ и дискретной компоненты $z''(t)$, $z(t) = = z'(t) + z''(t)$, а через $x(t) = [x_1(t) \cdots x_n(t)]^T$ — векторный входной сигнал.

Тогда уравнения процессов в такой системе будут иметь следующий вид:

$$z = z' + z''; \quad \dot{z}' = f(z, t); \quad z''_{k+1} = \varphi_k(z_k); \quad y = g(z, t). \quad (10)$$

Здесь дискретная компонента определяется соотношением:

$$z'' = z''(t) = \sum_{k=0}^{\infty} z''_k \mathbf{1}_{A_k}(t), \quad (11)$$

где $\mathbf{1}_{A_k}(t)$ — индикатор интервала $A_k = [t^{(k)}, t^{(k+1)})$; $z_k = [z_k'^T z_k''^T]^T = z(t^{(k)})$; $f(z, t)$ и $g(z, t)$ — некоторые известные функции. В качестве начальных условий принимаются соответствующие количества при $t = t_0$:

$$z_0 = z(t_0) = z'(t_0) + z''(t_0); \quad y(t_0) = g(z_0, t_0). \quad (12)$$

В случае стохастических возмущений как случайных функций, зависящих от времени и случайных параметров, уравнения (10)–(12) принимают вид:

$$\begin{aligned} Z &= Z' + Z''; \quad \dot{Z}' = f(Z, Q(t); \Theta_t, t); \quad Z''_{k+1} = \varphi_k(Z_k, Q_k, \Theta_t); \\ Y &= g(Z, Q(t), \Theta_t, t); \\ Z''(t) &= \sum_{k=0}^{\infty} Z''_k \mathbf{1}_{A_k}(t); \end{aligned}$$

$$Z_0 = Z(t_0) = Z'(t_0) + Z''(t_0); \quad Y_0 = g(Z_0, Q(t_0), \Theta_t, t_0).$$

Здесь $Q(t)$ — некоторая векторная случайная функция; Q_k ($k = 0, 1, 2, \dots$) — векторные случайные величины; $\Theta(t)$ — некоторые случайные параметры, допускающие ИКП (МК ИКП) согласно разд. 2.

Идея УОУ была впервые выдвинута В. С. Пугачёвым в докладе на рабочем семинаре ИФАК по дифференциальным играм (г. Сочи, апрель 1980 г.), а затем появились работы [5–10], в которых были даны алгоритмы нахождения УОУ дискретной системой. Однако в этих работах задачи управления и оценивания (получения необходимой для управления информации) разделены, что представляет собой отход от основной идеи УОУ. При этом приближенный подход на практике часто приводит к приемлемым результатам.

Идея УОУ заключается в том, что оптимальное управление следует искать не в классе всех возможных управлений, как это делается в теории оптимального управления [13, 14], а в ограниченном классе допустимых управлений, которые могут вычисляться в реальном времени в процессе работы системы управления. В качестве класса допустимых управлений, удовлетворяющих этому условию,

можно взять класс управлений, определяемых скалярным или векторным обыкновенным дифференциальным уравнением заданной структуры (случай непрерывного управления) или разностным уравнением заданной структуры (случай дискретного управления).

Рассмотрим приведенную СтС НРОП, описываемую стохастическими дифференциальными уравнениями Ито (9).

Класс допустимых управлений в случае непрерывного управления определим векторным дифференциальным уравнением

$$dU_t = [\alpha_t \xi(Y_t, U_t, \Theta_t, t) + \gamma_t] dt + \beta_t \eta(Y_t, U_t, \Theta_t, t) dY_t, \quad (13)$$

где структурные функции ξ и η заданы, а коэффициенты α_t и β_t (матрицы) и свободный член γ_t (вектор) подлежат оптимизации.

Класс допустимых управлений в случае дискретного управления целесообразно определить разностным уравнением

$$U_{n+1} = \delta_n \zeta_n(Y_n, U_n, \Theta_n) + \gamma_n, \quad (14)$$

где $Y_n = Y(t_n)$; $U_n = U(t_n)$, $\{t_n\}$ — последовательность моментов времени, в которые производятся наблюдение и управление; ζ_n — заданные функции; δ_n и γ_n — параметры, подлежащие оптимизации.

Для учета ограничения $U_t \in \mathcal{D}$ можно заменить уравнение (13) уравнением

$$dU_t = [\alpha_t \xi(Y_t, U_t, \Theta_t, t) + \gamma_t] dt + \beta_t \eta(Y_t, U_t, \Theta_t, t) dY_t - \\ - \max \left\{ 0, n(U_t)^T [\alpha_t \xi(Y_t, U_t, \Theta_t, t) + \gamma_t] dt + n(U_t)^T \beta_t \eta(Y_t, U_t, \Theta_t, t) dY_t \right\} \times \\ \times n(U_t) \mathbf{1}_{\partial \mathcal{D}}(U_t). \quad (15)$$

Здесь в дополнение к предыдущим обозначениям $n(U_t)$ — единичный вектор внешней нормали к границе $\partial \mathcal{D}$ области \mathcal{D} в точке U_t ; $\mathbf{1}_{\partial \mathcal{D}}(U_t)$ — индикатор множества ($\partial \mathcal{D} = 1$ при $U_t \in \partial \mathcal{D}$ и $\partial \mathcal{D} = 0$ при $U_t \notin \partial \mathcal{D}$). Дополнительный член в (15) по сравнению с (13) обеспечивает удерживание на границе $\partial \mathcal{D}$ области \mathcal{D} при достижении вектором U_t границы $\partial \mathcal{D}$ с положительной проекцией вектора $(\alpha_t \xi + \gamma_t) dt + \beta_t \eta dY_t$ на внешнюю нормаль к границе.

Аналогично в случае дискретного управления ограничение $U_n \in \mathcal{D}$ учитывается заменой уравнения (15) уравнением

$$U_{n+1} = \delta_n \zeta_n(Y_n, U_n, \Theta_n) + \gamma_n - \\ - \max \left\{ 0, n(U_n, \Theta_n)^T [\delta_n \zeta_n(Y_n, U_n, \Theta_n) - U_n] \right\} n(U_n, \Theta_n) \mathbf{1}_{\partial \mathcal{D}}(U_n, \Theta_n). \quad (16)$$

Что касается выбора функций ξ и η в (15) и ζ_n в (16), то их целесообразно определить путем изучения возможных структур системы управления и динамических характеристик ее возможных элементов.

Так как для вычисления управления U_t в системе управления в каждый данный момент времени невозможно использовать информацию, получаемую после этого момента, то критерий качества управления в каждый данный момент не может зависеть от значений вектора состояния системы X_t и управления U_t в последующие моменты времени. В этом заключается коренное отличие задач теории УОУ от задач классической теории оптимального управления [13, 14]. Поэтому за критерий качества управления в теории УОУ можно принять математическое ожидание некоторого функционала от $X_{t_0}^t = \{X(\tau) : \tau \in [t_0, t]\}$ и $U_{t_0}^t = \{U(\tau) : \tau \in [t_0, t]\}$:

$$\rho(\Theta_t) = M \ell(X_{t_0}^t, U_{t_0}^t, \Theta_t, t), \quad (17)$$

где ℓ — некоторая функция потерь, определяющая цену результатов управления при данных реализациях $x_{t_0}^t$ и $u_{t_0}^t$ стохастических процессов $X_{t_0}^T$ и $U_{t_0}^T$.

Задача нахождения УОУ состоит в нахождении функций времени α_t , β_t и γ_t в (15) или последовательностей $\{\delta_n\}$ и $\{\gamma_n\}$ в (16) из условия минимума критерия (17) в каждый момент времени. Эта задача, по существу, многокритериальная и вследствие этого однозначного решения иметь не может. Поэтому понятие УОУ необходимо уточнить. В связи с этим, следуя Пугачёву [5, 10], будем называть условно-оптимальным такое управление, которое получается в результате минимизации критерия (17) выбором значений α_t , β_t и γ_t , а (15) — в каждый момент t (или δ_n и γ_n в (16) в каждый момент t_n) при данных значениях α_t , β_t и γ_t (или δ_n и γ_n) во все предшествующие моменты времени.

Эффективное решение поставленной задачи нахождения УОУ легко получается в случае, когда функция потерь ℓ зависит только от значений X_t и U_t в один момент t . В этом случае для вычисления критерия (17) в каждый данный момент t необходимо знать совместное одномерное распределение СтП X_t и U_t . Это распределение легко находится применением любого из приближенных методов решения задачи анализа стохастических дифференциальных систем к системе (3), (4), (15) или (3), (4), (16). При применении любого из этих методов задача сводится к численному интегрированию обыкновенных дифференциальных уравнений для параметров распределения с минимизацией на каждом шаге критерия (17) путем выбора значений α_t , β_t и γ_t или δ_n и γ_n в данный момент. Эта задача совершенно аналогична задаче условно-оптимальной фильтрации (см. разд. 3).

Таким образом, приходим к следующим результатам.

Теорема 4.1. Пусть СтС НРОП при фиксированном векторе случайных параметров $\Theta_t = [\Theta_{1t} \cdots \Theta_{n\Theta_t}]^T$ допускает приведению к виду (3), (4). Тогда УОУ в случае непрерывного неограниченного управления определяется (13), а критерий качества (17) вычисляется на каждом шаге численного интегрирования для вероятностных параметров распределения составного СтП $[X_t^T Y_t^T U_t^T]^T$.

Теорема 4.2. В условиях теоремы 4.1 в случае ограничения на управление УОУ оно имеет вид (15).

Для дискретного УОУ вместо формулы (13) применяется формула (14), а вместо (15) — (16). Функционирование системы управления при применении УОУ сводится к интегрированию дифференциального уравнения (15) или к вычислению U_n по рекуррентной формуле (16).

Для получения безусловных вероятностных характеристик точности УОУ для нормальных (гауссовских) случайных параметров $\Theta_t = [\Theta_{1t} \dots \Theta_{n\Theta t}]^T$ в рамках корреляционной теории применяют методы чувствительности. Метод предполагает малость средних отклонений этих параметров по сравнению с их математическими ожиданиями. В его основе при $\Theta = \Theta_t$ лежат следующие формулы [15]:

$$\rho(\Theta) = \rho(m^\Theta) + \sum_{i=1}^{n^\Theta} R'_i \Theta_i^0 + \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} R''_{ij} \Theta_i^0 \Theta_j^0;$$

$$M_N \rho(\Theta) = \rho(m^\Theta) + \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} R''_{ij} R_{ij}^\Theta = R' \Theta^0 + \Theta^{0T} R'' \Theta^0; \quad (18)$$

$$\begin{aligned} M_N [\rho^2(\Theta)] &= \rho^2(m^\Theta) + \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} R'_i R'_j K_{ij}^\Theta + \\ &+ \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} \sum_{k=1}^{n^\Theta} \sum_{l=1}^{n^\Theta} R''_{ij} R''_{kl} [K_{ij}^\Theta K_{kl}^\Theta + K_{ik}^\Theta K_{jl}^\Theta + K_{il}^\Theta K_{jk}^\Theta] + \\ &+ 2\rho(m^\Theta) \sum_{i,j=1}^{n^\Theta} K''_{ij} K_{ij}^\Theta = \rho^2(m^\Theta) + R'^T K^\Theta R' + \\ &+ 2\rho(m^\Theta) \text{tr}(R'' K^\Theta) + (\text{tr} R'' K^\Theta)^2 + 2 \text{tr}(R'' K^\Theta)^2; \quad (19) \end{aligned}$$

$$\begin{aligned} D_\rho &= \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} R'_i R'_j K_{ij}^\Theta + \sum_{i=1}^{n^\Theta} \sum_{j=1}^{n^\Theta} \sum_{k=1}^{n^\Theta} \sum_{l=1}^{n^\Theta} R''_{ij} R''_{kl} [K_{ik}^\Theta K_{jl}^\Theta + K_{il}^\Theta K_{jk}^\Theta] = \\ &= R'^T K^\Theta R' + 2 \text{tr}(R'' K^\Theta)^2; \quad (20) \end{aligned}$$

$$\varepsilon_4 = \varepsilon_2^{1/4}; \quad \varepsilon_2 = M_N [\rho^2(\Theta)] - \rho^2(m^\Theta). \quad (21)$$

Здесь приняты обозначения: $\Theta = m^\Theta + \Theta^0$, $\Theta^0 = [\Theta_1^0 \dots \Theta_{n^\Theta}^0]^T$; $R' = \nabla \rho(\Theta) = (\partial \rho(\Theta) / \partial \Theta_i)_{\Theta=m^\Theta}$ и $R'' = [R''_{ij}] = 1/2 \nabla \nabla^T \rho(\Theta)$, $R''_{ij} = 1/2 (\partial^2 \rho(\Theta) / \partial \Theta_i \partial \Theta_j)_{\Theta=m^\Theta}$ — вектор и матрица первых и вторых функций

чувствительности; M_N — символ математического ожидания для нормального распределения Θ , tr — символ следа матрицы.

Теорема 4.3. В рамках корреляционной теории при условиях теорем 4.1 и 4.2 точность УОУ (13) или (15) определяется формулами (18)–(21).

5 Задачи условно-оптимального управления по локальным критериям

Следуя [8, 9], применительно к приведенной СтС НРОП рассмотрим задачи синтеза УОУ, когда управление ищется как функция текущих измерений (а не оценки) и времени.

Задача 5.1. Пусть вектор состояния приведенной системы X и вектор измерений Y описываются следующей гауссовской системой стохастических дифференциальных уравнений Ито

$$\dot{X}_t = a^{\text{п}}(X_t, U_t, t) + b^{\text{п}}(X_t, U_t, t) V_0; \quad \dot{Y} = a_1(Y_t, X_t, t) \quad (22)$$

с соответствующими начальными данными. Здесь $X_t \in R^{n^X}$; $Y_t \in R^{n^Y}$; $a^{\text{п}}$, $b^{\text{п}}$, a_1 , a_0 и b_0 — известные детерминированные функции (соответствующей размерности) своих аргументов; $U_t \in R^{n^U}$ — вектор управления. Предполагается также, что вектор V_0 не зависит от начального состояния $[X_0^T Y_0^T]^T$.

Требуется определить оптимальное управление U_t в классе допустимых управлений вида

$$U_t = \alpha_t \xi(Y_t, \dot{Y}_t, t) + \gamma_t, \quad (23)$$

доставляющих локальный минимум некоторому функционалу I в каждый текущий момент времени $t \geq t_0$:

$$I = M F(X_t, t) + M \int_{t_0}^t F_1(U_\tau, \tau) d\tau \rightarrow \min. \quad (24)$$

В (23) и (24) $\xi(Y, \dot{Y}, t)$ — заданная функция, определяющая класс (структуру) допустимых уравнений; $\alpha \in \mathcal{M}_{n^U \times s}$ и $\gamma \in R^{n^U}$ — неизвестные функции времени ($\mathcal{M}_{n^U \times s}$ — пространство вещественных $n^U \times s$ матриц), определяемые по критерию (24); $F : R^{n^X} \times [t_0, \infty) \rightarrow R$, $F_1 : R^{n^U} \times [t_0, \infty) \rightarrow R$ — заданные функционалы.

Для определения локально-оптимального управления при фиксированном Q продифференцируем по формуле Ито функционал (24) по времени и подставим в полученный результат (23):

$$\begin{aligned}
 J = \dot{I} &= \mathbb{M} \left[\dot{F}_t + F_x^T a^n + \frac{1}{2} \text{tr} (F_{xx} \sigma^n) \right] + \mathbb{M} F_1 = \\
 &= \mathbb{M} \left\{ \dot{F}_t + F_x^T a^n (X_t, \alpha_t \xi + \gamma_t, t) + \frac{1}{2} \text{tr} [F_{xx} \sigma^n (X_t, \alpha_t \xi + \gamma_t, t)] \right\} + \\
 &\quad + \mathbb{M} F_1 (\alpha_t \xi + \gamma_t, t) \quad (\sigma^n = b^n \nu_0 b^{nT}).
 \end{aligned}$$

Под УОУ будем понимать также управление U_t из класса (23), которое минимизирует J в каждый момент времени. Тогда задача определения неизвестных функций α_t и γ_t сводится к задаче минимизации $J(\alpha_t, \gamma_t)$ по α_t и γ_t при условии, что этот минимум существует. Искомые оптимальные коэффициенты α_t и γ_t определяются следующим образом:

$$(\alpha_t, \gamma_t) = \underset{\alpha, \gamma}{\text{argmin}} J(\alpha, \gamma). \quad (25)$$

Для существования минимума J достаточно, например, чтобы α и γ были ограниченными функциями для любого t .

Рассмотрим ряд случаев, в которых решение (25) упрощается.

Случай 1 (аддитивное управление). В этом случае система (22) описывается уравнением

$$\dot{X}_t = a^n(X_t, t) + B(t)U_t + b^n(X_t, t)V_0, \quad B \in \mathcal{M}_{n \times n \times U}.$$

Тогда решение (25) примет вид:

$$(\alpha_t, \gamma_t) = \underset{(\alpha, \gamma)}{\text{argmin}} \mathbb{M} \left[F_x^T B (\alpha_t \xi + \gamma_t) + F_1 (\alpha_t \xi + \gamma_t, t) \right].$$

Если в этом случае функционал $F_1 \equiv 0$, а α_t и γ_t лежат в некотором параллелепипеде

$$|\alpha_{ij}(t)| \leq \alpha_{ij}^*; \quad |\gamma_i(t)| \leq \gamma_i^*,$$

то оптимальные коэффициенты α_t и γ_t определяются по формулам:

$$\alpha_t = -\alpha^* \text{sgn} \left[B^T \mathbb{M} \left(F_x \xi^T \right) \right]; \quad \gamma_t = -\gamma^* \text{sgn} \left[B^T \mathbb{M} (F_x) \right]. \quad (26)$$

Как и в теории УОФ, реализация УОУ требует знания вероятностных моментов расширенного вектора $[X_t^T Y_t^T]^T$, которые определяются заранее в процессе проектирования, и они не зависят от текущих наблюдений. Поэтому оптимальные управления (26) легко могут быть реализованы в реальном масштабе времени.

Случай 2 (квадратичный критерий). При $F = X_t^T \Omega X_t$ и $F_1 = 0$ решение (25) имеет вид:

$$(\alpha_t, \gamma_t) = \underset{\alpha, \gamma}{\operatorname{argmin}} \left\{ 2M \left[X_t^T \Omega a^\Pi(X_t, \alpha_t \xi + \gamma_t, t) \right] + M \operatorname{tr} [\Omega \sigma^\Pi(X_t, \alpha_t \xi + \gamma_t, t)] \right\}.$$

Если в рамках этого случая приведенная система (22) линейная, т. е.

$$a^\Pi(X_t, U_t, t) = AX_t + BU_t; \quad b^\Pi(X_t, U_t, t) = b^\Pi(t),$$

управление U_t ищется в классе $U_t = \alpha_t Y_t + \gamma_t$ ($\xi(Y_t, t) \equiv Y_t$), а коэффициенты α_t и γ_t лежат в параллелепипеде, то оптимальные значения α_t и γ_t имеют вид:

$$\alpha_t = -\alpha^* \operatorname{sgn} \left[B^T \Omega M \left(X_t Y_t^T \right) \right]; \quad \gamma_t = -\gamma^* \operatorname{sgn} \left[B^T \Omega M (X_t) \right].$$

Задача 5.2. Пусть вектор состояния и вектор наблюдений описываются следующей системой стохастических дифференциальных уравнений Ито:

$$\begin{aligned} \dot{X}_t &= a^\Pi(X_t, t) + BU_t + b^\Pi(X_t, t) V_0, \quad X_t \in R^{n^X}, \quad V_0 \in R^{n^V}; \\ \dot{Y}_t &= a_1(X_t, Y_t, t) + b_1(X_t, Y_t, t) V_0, \quad Y_t \in R^{n^Y}, \end{aligned}$$

где $V_0 = V_0(t)$ — нормальный белый шум с интенсивностью $\nu_0 = \nu_0(t)$. Управление $U_t \in R^{n^U}$ ищется в виде

$$U_t = \alpha_t \xi(Y_t) + \beta \dot{Y}_t + \gamma_t$$

с ограничением только на свободный член γ_t , $|\gamma_i| \leq \gamma_i^*$.

Критерий оптимальности имеет вид:

$$I = M F(X_t, t) + M \int_{t_0}^t c^T(\tau) U_\tau d\tau \rightarrow \min, \quad c(\tau) \in R^{n^U}.$$

Отсюда получаем следующие уравнения для оптимальных α_t , β_t и γ_t :

$$\begin{aligned} B^T \left\{ M \left(F_x a_1^T \right) + M \left[F_{xx} (B \beta_t b_1 \nu_0 b_1^T + b^\Pi \nu_0 b_1^T) \right] \right\} + c M \left(a_1^T \right) &= 0; \\ \alpha_t = -\alpha^* \operatorname{sgn} M \left[\xi(F_x^T B + c) \right]^T, \quad \gamma_t = -\gamma^* \operatorname{sgn} \left[B^T (F_x) + c \right]. \end{aligned} \quad (27)$$

В частном случае, когда $b^\Pi = [b'(t) \ 0]$, $b_1 = [0 \ b'(t)]$, коэффициент α_t в (27) определяется в явном виде:

$$\alpha_t = - \left[B^T M (F_{xx}) B \right]^{-1} \left\{ c M (a_1^T) + M \left(F_x a_1^T \right) \right\} (b_1' \nu_0 b_1'^T)^{-1}. \quad (28)$$

Задача 5.3. Рассмотрим систему (7) при $\gamma_t = \gamma_0 \Theta_t$, $\nu_1 = \nu_{10}$ и $\nu_2 = \nu_{20}$:

$$\left. \begin{aligned} \Phi &= \Phi_1(\dot{X}_{1t}) + \gamma_0 \Theta_t X_{1t} + X_{2t} = 0; \\ \dot{X}_{2t} &= aX_{2t} + V_1; \quad \dot{Y}_t = b_1 X_{1t} + V_2. \end{aligned} \right\} \quad (29)$$

Здесь $[X_{1t} X_{2t}]^T = X_t$ — вектор состояния; V_1 и V_2 — независимые скалярные белые шумы с интенсивностями ν_1 и ν_2 ; a и b_1 — известные количества; $\varphi_1 = \varphi_1(\dot{X}_{1t})$ — функция, допускающая регрессионную линеаризацию

$$\Phi_1 \approx \Phi_{10} + k_1 \dot{X}_{1t}^0,$$

где $\Phi_{10} = \Phi_{10}(m_{1t}^{\dot{X}}, D_{1t}^{\dot{X}})$; $k_1 = k_1(m_{1t}^{\dot{X}}, D_{1t}^{\dot{X}})$. При $k_1 \neq 0$ уравнения (29) приводятся к виду:

$$\dot{X}_t = a_{1t}(\Theta_t)X_t + a_{0t} + V_1; \quad \dot{Y}_t = b_{1t}X_t + V_2,$$

где

$$\begin{aligned} a_{0t} &= \begin{bmatrix} -\Phi_{10} k_1^{-1} \\ 0 \end{bmatrix}; \quad a_{1t} = a_1(\Theta_t) = \begin{bmatrix} \gamma_0 \Theta_t & k_1^{-1} \\ b_1 & 0 \end{bmatrix}; \\ b_{1t} &= [b_1 \ 0], \quad \tilde{R} = \begin{bmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{bmatrix}. \end{aligned}$$

Уравнения для условных характеристик УОФ имеют следующий вид:

$$\dot{\hat{X}}_t = a_{1t}(\Theta_t)\hat{X}_t + a_{0t} + \tilde{\beta}_t(Z_t - b_{1t}\hat{X}_t), \quad \tilde{\beta}_t = \tilde{R}_t b_{1t}^T \nu_{20t}^{-1}; \quad (30)$$

$$\dot{\tilde{R}}_t = a_{1t}(\Theta_t)\tilde{R}_t + \tilde{R}_t a_{1t}^T(\Theta_t) - \tilde{R}_t b_{1t}^T \nu_{20t}^{-1} b_{1t} \tilde{R}_t + \nu_{10}. \quad (31)$$

Здесь

$$\begin{aligned} \Theta_t &= m_t^\Theta + \int_{\Lambda} V(\lambda) x^\Theta(\lambda, t) d\lambda; \\ K_t^\Theta(t_1, t_2) &= \int_{\Lambda} G(\lambda) x^\Theta(\lambda, t_1) x^\Theta(\lambda, t_2) d\lambda; \\ x(\lambda, t) &= q_1(t) q_2^{-1}(\lambda) \mathbf{1}(t - \lambda), \quad q_1(t) = e^{-\alpha t}, \quad q_2(t) = q_1(t) \int_{t_0}^t q_1^{-2}(\lambda) d\lambda; \\ K^\Theta(t_1, t_2) &= \begin{cases} q_1(t_2) q_2(t_1) & \text{при } t_1 < t_2; \\ q_1(t_1) q_2(t_2) & \text{при } t_1 > t_2. \end{cases} \end{aligned}$$

Выполняя статистическую линейризацию функции $\Theta_t X_{1t}$ в (29) по известной формуле

$$\Theta_t X_{1t} \approx m_t^\Theta m_{1t} + K_t^{\Theta X_1} + m_{1t} \Theta_t^0 + m_t^\Theta X_{1t}^0,$$

согласно (30) и (31), приходим к уравнениям для безусловных характеристических функций $\hat{\hat{X}}_t$ и $\hat{\hat{R}}_t$, определяющих точность УОФ. Формулы (27) и (28) задачи 5.2 определяют характеристики УОУ для произвольного локального критерия при фиксированном случайном параметре Θ_t . Точность УОУ по методу МК ИКП определяется на основе теорем 2.3 и 2.4.

6 Заключение

Разработано методическое обеспечение синтеза фильтров и регуляторов для управления типовыми гауссовскими СтС со случайными параметрами, описываемыми МК ИКП, не разрешенными относительно старших производных и допускающими приведение к дифференциальным. В основу синтеза положены обобщения методов теории условно-оптимальной фильтрации и управления В. С. Пугачёва. Дано решение трех типовых задач синтеза управления на основе локальных критериев для нелинейностей, описывающих зависимости от старших производных, обладающих зонами неоднозначности и нечувствительности, гистерезиса и насыщения [14, 16]. Приводится иллюстративный пример.

Применения: нелинейная корреляционная оценка потенциальной и инструментальной точности функционирования фильтров и регуляторов в сложных системах информатики и управления в условиях существенно нестационарных гауссовских, например ударных, воздействий.

Дальнейшее развитие теории УОФ и УОУ должно учитывать нелинейные эффекты, связанные с накоплением возмущений, дрейфов, а также с выбросами процессов. В этом случае потребуются учет негауссовости распределений.

Литература

1. Синицын И. Н. Аналитическое моделирование и оценивание нестационарных нормальных процессов в стохастических системах, не разрешенных относительно производных // Системы и средства информатики, 2022. Т. 32. № 2. С. 58–71. doi: 10.14357/08696527220206. EDN: YMGERJ.
2. Sinitsyn I. N. Analytical modeling and estimation of normal processes defined by stochastic equations with unsolved derivatives // Mathematics Statistics Research, 2021. Vol. 3. Iss. 1. Art. 139. 7 p. doi: 10.36266/JMSR/139.
3. Синицын И. Н. Совместная фильтрация и распознавание нормальных процессов в стохастических системах, не разрешенных относительно производных // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 85–93. doi: 10.14357/19922264220211. EDN: SMJCBV.

4. Сеницын И. Н. Субоптимальная фильтрация в стохастических системах, не разрешенных относительно производных, со случайными параметрами // Информатика и её применения, 2024 Т. 18. Вып. 1. С. 2–10. doi: 10.14357/19922264240101. EDN: KUWMKJ.
5. Pugachev V. S. Conditionally optimal control // 8th World Congress of IFAC Proceedings. — Pergamon Press, 1981. Vol. 1. P. 373–376.
6. Казаков И. Е. Синтез условно-оптимального управления по локальному критерию в нелинейной стохастической системе // Автоматика и телемеханика, 1987. № 12. С. 72–80.
7. Казаков И. Е. Условно-оптимальное управление в нелинейной дискретной стохастической системе // Техн. киберн., 1988. № 3. С. 165–169.
8. Сеницын И. Н., Мошук Н. К., Шин В. И. Общая теория условно-оптимальной фильтрации в стохастических дифференциальных системах // Системы и средства информатики, 1995. № 7. С. 75–85.
9. Мошук Н. К., Сеницын И. Н., Шин В. И. Проблемы условно-оптимальной фильтрации в стохастических дифференциальных системах в реальном масштабе времени // Тезисы докл. 1-го Междунар. симпозиума по проблемам модульных информационно-вычислительных систем и сетей. — Пушкин, 1996. С. 37.
10. Академик Пугачёв Владимир Семёнович: к столетию со дня рождения / Предисловие академика С. В. Емельянова; под ред. И. Н. Сеницына. — М.: ТОРУС ПРЕСС, 2011. 376 с.
11. Сеницын И. Н., Шаламов А. С. Лекции по теории систем интегрированной логистической поддержки. — 2-е изд. — М.: ТОРУС ПРЕСС, 2019. 1072 с. EDN: HVAZTX.
12. Сеницын И. Н. Канонические представления случайных функций. Теория и применения. — 2-е изд. — М.: ТОРУС ПРЕСС, 2023. 816 с. doi: 10.30826/94588-308-6. EDN: XHBITA.
13. Основы автоматического управления / Под ред. В. С. Пугачёва. — 3-е изд. — М.: Наука, 1974. 720 с.
14. Справочник по теории автоматического управления / Под ред. А. А. Красовского. — М.: Наука, 1987. 712 с.
15. Евланов Л. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 568 с.
16. Петров В. В., Гордеев А. А. Нелинейные сервомеханизмы. — М.: Машиностроение, 1979. 472 с.

Поступила в редакцию 13.11.23

CONDITIONALLY-OPTIMAL FILTRATION AND CONTROL FOR STOCHASTIC SYSTEMS WITH RANDOM PARAMETERS AND UNSOLVED DERIVATIVES

I. N. Sinitsyn

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

Abstract: Methodological support for synthesis of conditionally: optimal filtering and control for continuous, discrete, and continuous-discrete stochastic systems with unsolved derivatives (StS USD) and random parameters is developed. A survey of conditionally-optimal filtering (COF) and conditionally optimal control (COC) is presented. It is supposed that random parameters are described by multicomponent integral canonical expansions (MC ICE). Synthesis of COF and COC is based on Pugachev’s COC concepts. Special attention is paid to the three COC typical problems in the case of local criteria. Applications: nonlinear correlational estimation of potential and instrumental accuracy of COF and COC at nonstationary impact disturbances in StS USD. An illustrative example is given. Future research is connected with effects of disturbances accumulation, drifts, and excursions based on finite dimension distribution of stochastic processes.

Keywords: conditionally optimal control (COC); conditionally optimal filtering (COF); multicomponent integral canonical expansions (MC ICE); random parameters; stochastic systems with unsolved derivatives (StS USD)

DOI: 10.14357/08696527240201

EDN: XLRCQE

References

1. Sinitsyn, I. N. 2022. Analiticheskoe modelirovanie i otsenivanie nestatsionarnykh normal’nykh protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel’no proizvodnykh [Analytical modeling and estimation of nonstationary normal processors with unsolved derivatives]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):58–71. doi: 10.14357/08696527220206. EDN: YMGERJ.
2. Sinitsyn, I. N. 2021. Analytical modeling and estimation of normal processes defined by stochastic differential equations with unsolved derivatives. *J. Mathematics Statistics Research* 3(1):139. 7 p. doi: 10.36266/JMSR/139.
3. Sinitsyn, I. N. 2022. Sovmestnaya fil’tratsiya i raspoznavanie normal’nykh protsessov v stokhasticheskikh sistemakh, ne razreshennykh otnositel’no proizvodnykh [Joint filtration and recognition of normal processes in stochastic systems with unsolved derivatives]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):85–93. doi: 10.14357/19922264220211. EDN: SMJCBĀ.
4. Sinitsyn, I. N. 2024. Suboptimal’naya fil’tratsiya v stokhasticheskikh sistemakh, ne razreshennykh otnositel’no proizvodnykh, so sluchaynymi parametrami [Suboptimal filtering in stochastic systems with random parameters and unsolved deriva-

- tives]. *Informatika i ee Primeneniya — Inform. Appl.* 18(1):2–10. doi: 10.14357/19922264240101. EDN: KUWMKJ.
5. Pugachev, V. S. 1981. Conditionally optimal control. *8th World Congress of IFAC Proceedings*. Pergamon Press. 1:373–376.
 6. Kazakov, I. E. 1987. Synthesis of conditionally optimal control by local criterion in nonlinear stochastic system. *Automat. Rem. Contr.* 48:1591.
 7. Kazakov, I. E. 1988. Uslovno-optimal'noe upravlenie v nelineynoy diskretnoy stokhasticheskoy sisteme [Conditional optimal control in a nonlinear discrete stochastic system]. *Sov. J. Comput. Syst. Sc.* 3:165–169.
 8. Sinitsyn, I. N., N. K. Moshchuk, and V. I. Shin. 1995. Obschchaya teoriya uslovno-optimal'noy fil'tratsii v stokhasticheskikh differentsial'nykh sistemakh [General theory of conditional-optimal filtering in stochastic differential systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 7:75–85.
 9. Moshchuk, N. K., I. N. Sinitsyn, and V. I. Shin. 1996. Problemy uslovno-optimal'noy fil'tratsii v stokhasticheskikh differentsial'nykh sistemakh v real'nom mashtabe vremeni [Problems of conditionally optimal filtering in stochastic differential systems in real time]. *1st Symposium (International) on the Problems of Modular Information Computing Systems and Networks Proceedings*. Pushkin. 37.
 10. Sinitsyn, I. N., ed. 2011. *Akademik Pugachev Vladimir Semenovich: k stoletiyu so dnya rozhdeniya* [Academician Pugachev Vladimir Semenovich: To the centenary of his birth]. Moscow: TORUS PRESS. 376 p.
 11. Sinitsyn, I. N., and A. S. Shalamov. 2019. *Lektsii po teorii sistem integrirrovannoy logisticheskoy podderzhki* [Lectures on theory of integrated logistic support systems]. 2nd ed. Moscow: TORUS PRESS. 1072 p. EDN: HVAZTX.
 12. Sinitsyn, I. N. 2023. *Kanonicheskie predstavleniya sluchaynykh funktsiy. Teoriya i primeneniya* [Canonical expansions of random functions. Theory and applications]. 2nd ed. Moscow: TORUS PRESS. 816 p. doi: 10.30826/94588-308-6. EDN: XHBITA.
 13. Pugachev, V. S. ed. 1974. *Osnovy avtomaticheskogo upravleniya* [Foundations of automatic control]. 3rd ed. Moscow: Nauka. 720 p.
 14. Krasovskiy, A. A., ed. 1987. *Spravochnik po teorii avtomaticheskogo upravleniya* [Handbook for automatic control]. Moscow: Nauka. 712 p.
 15. Evlanov, L. G., and V. M. Konstantinov. 1976. *Sistemy so sluchaynymi parametrami* [Systems with random parameters]. Moscow: Nauka. 568 p.
 16. Petrov, V. V., and A. A. Gordeev. 1979. *Nelineynnye servomekhanizmy* [Nonlinear servomechanisms]. Moscow: Mashinostroenie. 472 p.

Received November 13, 2023

Contributor

Sinitsyn Igor N. (b. 1940)— Doctor of Science in technology, professor, Honored scientist of RF, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; sinitsin@dol.ru

МОДЕЛИРОВАНИЕ НЕСТАЦИОНАРНОГО СТОХАСТИЧЕСКОГО ПРОЦЕССА ПОСРЕДСТВОМ ЕГО КАНОНИЧЕСКОГО РАЗЛОЖЕНИЯ НА ОСНОВЕ ВЕЙВЛЕТ-НЕЙРОННОЙ СЕТИ*

И. Н. Синицын¹, В. И. Синицын², Э. Р. Корепанов³, Т. Д. Конашенкова⁴

Аннотация: Построено каноническое разложение (КР) скалярного стохастического процесса (СтП), заданного на конечном промежутке времени, с применением технологии вейвлет-нейронной сети (ВНС). Разработана архитектура трехслойной ВНС. Активационные функции скрытого слоя строятся на основе выбранного ортонормированного базиса вейвлетов с компактными носителями. Для обучения ВНС применяется метод наискорейшего спуска. Дан алгоритм построения КР ковариационной функции заданного СтП на основе ВНС. Каноническому разложению ковариационной функции соответствует КР СтП в виде линейной комбинации базисных вейвлетов с коэффициентами — случайными величинами (СВ) с нулевыми математическими ожиданиями и дисперсиями, определяемыми с помощью предлагаемого алгоритма. Дан иллюстративный пример. Проведено сравнение рекуррентного метода КР на основе вейвлетов Хаара с предлагаемым новым методом построения КР на основе ВНС.

Ключевые слова: вейвлет; вейвлет-нейронная сеть; искусственная нейронная сеть; каноническое разложение; ковариационная функция; моделирование; стохастический процесс

DOI: 10.14357/08696527240202

EDN: YFHFIFIN

1 Введение

Как известно [1–4], КР СтП представляет собой линейную комбинацию некоррелированных СВ с коэффициентами в виде детерминированных функций. Каноническое разложение СтП упрощает выполнение различных математических операций, особенно линейных, над случайными функциями. Это

*Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

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

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

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

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, tkonashenkova64@mail.ru

объясняется тем, что зависимость функции от случайного аргумента сводится к зависимости функции от детерминированных, так называемых координатных, функций, а следовательно, все операции над исследуемой случайной функцией сводятся к соответствующим операциям над координатными функциями. Каноническое разложение СтП удобно для решения задач как вероятностного (аналитического), так и статистического моделирования. Методы моделирования на основе КР СтП в рамках корреляционной теории оказываются универсальными. При этом, следуя прямой и обратной теоремам Пугачёва, обычно рассматривают моделирование КР СтП при известной ковариационной функции и моделирование КР ковариационной функции при известном КР СтП. Создание теории КР случайной функции связано с именами Лоэва, Колмогорова, Карунена и Пугачёва в 1940–1950 гг. Развитие теории канонических представлений случайных функций представлено в работах Пугачёва, Марченко, Сеницына [1–5].

В последние десятилетия для анализа нестационарных СтП в стохастических системах (СтС) успешно применяют вейвлеты [6–8]. В работе [9] дано развитие теории КР СтП на основе вейвлет-технологий. Для построения КР нестационарного СтП были выбраны вейвлеты с компактными носителями [10]. Вейвлет-каноническое разложение (ВлКР) СтП строится на основе коэффициентов разложения ковариационной функции СтП по ортонормированному двумерному базису вейвлетов с компактными носителями. Вычислительные эксперименты показали, что моделирование нестационарного СтП с применением ВлКР позволяет ограничиться небольшим числом членов КР [11]. В [12–16] метод ВлКР был успешно применен для синтеза оптимальных по различным байесовым критериям линейных и нелинейных СтС. Вычислительные эксперименты подтвердили высокую точность и простоту разработанных методов.

В настоящее время при решении задач имитационного моделирования широко используются подходы, основанные на искусственных нейронных сетях (ИНС) [17]. Искусственная нейронная сеть — это набор методов анализа данных, имеющих сетевую структуру и обладающих двумя характерными чертами. Во-первых, нейросетевой алгоритм всегда имеет распределенную сетевую структуру, состоящую из элементов, называемых «формальными нейронами». Во-вторых, нейросетевой алгоритм всегда представляет собой «обучаемый алгоритм». Все сетевые алгоритмы уникальны. Отличительными чертами каждого сетевого алгоритма выступают следующие элементы:

- число используемых формальных нейронов — элементов сетевой структуры;
- выделение в данной сетевой структуре упорядоченных образований, называемых слоями;
- число используемых слоев формальных нейронов;
- наличие или отсутствие дополнительных входов и выходов в отдельных нейронных слоях;
- наличие или отсутствие обратных связей и внутренних итераций;

- структура алгоритма обучения;
- использование алгоритма обучения как в виде отдельной стадии, так и в процессе функционирования.

Огромное внимание уделяется построению нейросетевых алгоритмов, использующих вейвлеты в качестве активационной функции скрытого слоя [18, 19]. При описании таких ИНС часто используют термин «вейвлон», который обозначает нейрон, активационной функцией которого служит вейвлет. Вейвлет-нейронная сеть — ИНС, скрытый слой которой состоит из вейвлонов, чьи сдвиги и сжатия задаются при инициализации и в процессе обучения не меняются. Вейвлет-нейронные сети считаются относительно изученными. В [19] показано, что использование многомерных вейвлет-функций позволяет провести более точную и тонкую настройку работы сети с отдельными входами или группами входов.

В настоящей статье авторы рассматривают вопрос построения КР СтП в заданном промежутке времени с применением ВНС, учитывая положительный опыт построения и применения ВлКР. В разд. 2 описано построение ВлКР СтП на основе рекуррентного алгоритма [3]. В разд. 3 предлагается новый способ построения КР СтП на основе трехслойной ВНС (далее — КРВНС): первый слой — входной; второй — скрытый; третий — выходной. В скрытом слое входные данные преобразуются в элементы пространства признаков, заданного на основе ортонормированного базиса вейвлетов с компактными носителями. В выходном слое формируется выходное значение сети, которое сравнивается с целевым значением. Расхождение между выходным и целевым значениями используется для настройки параметров ВНС. В разд. 4 приведен пример, иллюстрирующий высокую эффективность разработанного метода. Заключение содержит основные выводы и направления дальнейших исследований.

2 Построение вейвлет-канонического разложения стохастического процесса на основе рекуррентного алгоритма

Пусть скалярный СтП $X(t)$ для $t \in T$ задан нулевым математическим ожиданием и ковариационной функцией

$$K_X(t_1, t_2) = M \left[X(t_1) \overline{X(t_2)} \right] \quad (t_1 \in T, t_2 \in T), \quad K_X(t_1, t_2) \in \mathcal{L}^2(T \times T).$$

Построим КР заданного СтП $X(t)$ в виде

$$X(t) = \sum_{v=1}^L V_v x_v(t), \quad (1)$$

где V_v — некоррелированные СВ с нулевыми математическими ожиданиями и дисперсиями G_v ; $x_v(t)$ — координатные функции, удовлетворяющие соотношениям [2, 3]:

$$V_v = \int_T a_v(t)X(t) dt;$$

$$x_v(t) = \frac{1}{G_v} \int_T K_X(t_1, \tau)a_v(\tau) d\tau;$$

$$\int_T x_v(\tau)a_\mu(\tau) d\tau = \delta_{v\mu}.$$

Функции $a_v(t)$ находятся из условия некоррелированности СВ V_v , поэтому их можно определить бесчисленным множеством способов.

Тогда, согласно прямой теореме Пугачёва, КР ковариационной функции СтП $X(t)$ примет вид:

$$K_X(t_1, t_2) = \sum_{v=1}^L G_v x_v(t_1)x_v(t_2).$$

Таким образом, для построения КР СтП $X(t)$ достаточно найти функции $x_v(t)$ и $a_v(t)$ и дисперсии G_v СВ V_v ($v = \overline{1, L}$).

Для построения ВлКР заданного СтП $X(t)$ определим в пространстве $\mathcal{L}^2(T)$ ортонормированный базис вейвлетов с компактными носителями вида

$$\{\varphi_{00}(t), \psi_{jk}(t)\}, \quad (2)$$

где $\varphi_{00}(t) = \varphi(t)$ — масштабирующая функция; $\psi_{00}(t) = \psi(t)$ — материнский вейвлет, $\psi_{jk}(t) = \sqrt{2^j} \psi(2^j t - k)$ — вейвлеты уровня j для $j = \overline{1, J}$; $k = \overline{0, 2^j - 1}$, J — максимальный уровень вейвлет-разрешения. Тогда параметр $L = 2^{J+1}$ равен числу базисных функций. Для удобства представим вейвлет-базис (2) в виде

$$g_1(t) = \varphi_{00}(t); \quad g_2(t) = \psi_{00}(t); \quad g_v(t) = \psi_{jk}(t), \quad (3)$$

где $j = \overline{1, J}$, $k = \overline{0, 2^j - 1}$, $v = 2^j + k + 1$.

В пространстве $\mathcal{L}^2(T \times T)$ определим двумерный ортонормированный вейвлет-базис в виде тензорного произведения двух вейвлет-базисов вида (2) для случая, когда масштабирование по обоим переменным происходит одинаково. В этом случае базисные функции имеют вид:

$$\Phi^A(t_1, t_2) = \varphi_{00}(t_1)\varphi_{00}(t_2); \quad \Psi^H(t_1, t_2) = \varphi_{00}(t_1)\psi_{00}(t_2);$$

$$\Psi^B(t_1, t_2) = \psi_{00}(t_1)\varphi_{00}(t_2); \quad \Psi_{jkn}^D(t_1, t_2) = \psi_{jk}(t_1)\psi_{jn}(t_2),$$

где $j = \overline{1, J}$; $k, n = \overline{0, 2^j - 1}$. Тогда двумерное вейвлет-разложение заданной ковариационной функции примет вид:

$$K_X(t_1, t_2) = a^x \Phi^A(t_1, t_2) + h^x \Psi^H(t_1, t_2) + b^x \Psi^B(t_1, t_2) + \sum_{j=0}^J \sum_{k=0}^{2^j-1} \sum_{n=0}^{2^j-1} d_{jkn}^x \Psi_{jkn}^D(t_1, t_2),$$

где

$$a^x = \int_T \int_T K_X(t_1, t_2) \Phi^A(t_1, t_2) dt_1 dt_2; \quad h^x = \int_T \int_T K_X(t_1, t_2) \Psi^H(t_1, t_2) dt_1 dt_2;$$

$$b^x = \int_T \int_T K_X(t_1, t_2) \Psi^B(t_1, t_2) dt_1 dt_2; \quad d_{jkn}^x = \int_T \int_T K_X(t_1, t_2) \Psi_{jkn}^D(t_1, t_2) dt_1 dt_2.$$

Введем вспомогательные параметры $k_{v\mu}$ ($v, \mu = \overline{1, L}$):

$$k_{11} = a^x; \quad k_{12} = h^x; \quad k_{21} = b^x; \quad k_{22} = d_{00}^x;$$

$$k_{v\mu} = d_{jkn}^x \left(v = 2^j + k + 1; \mu = 2^j + n + 1; i = \overline{1, J}; n = \overline{0, 2^j - 1} \right),$$

остальные $k_{v\mu} = 0$.

Дисперсии G_v СВ V_v вычисляются по рекуррентным формулам:

$$c_{v1} = -\frac{k_{v1}}{G_1} \quad (v = \overline{2, L});$$

$$c_{v\mu} = -\frac{1}{G_\mu} \left(k_{v\mu} - \sum_{\lambda=1}^{\mu-1} G_\lambda c_{\mu\lambda} c_{v\lambda} \right) \quad (\mu = \overline{2, v-1}; v = \overline{3, L});$$

$$G_1 = k_{11}; \quad G_v = k_{vv} - \sum_{\lambda=1}^{v-1} G_\lambda |c_{v\lambda}|^2 \quad (v = \overline{2, L}).$$

Для определения координатных функций $x_v(t)$ ($v = \overline{1, L}$) сначала вычисляются вспомогательные функции

$$h_1^x(t) = a^x \varphi_{00}(t) + b^x \psi_{00}(t); \quad h_2^x(\bar{\tau}) = h^x \varphi_{00}(t) + d_{000}^x \psi_{00}(t);$$

$$h_v^x(t) = \sum_{k=0}^{2^j-1} d_{jkn}^x \psi_{jk}(t) \quad \left(v = \overline{3, L}; v = 2^j + n + 1; n = \overline{0, 2^j - 1}; j = \overline{1, J} \right),$$

затем вычисляются координатные функции $x_v(t)$ последовательно по формулам:

$$x_1(t) = \frac{1}{G_1} h_1^x(t); \quad x_v(t) = \frac{1}{G_v} \left(\sum_{\lambda=1}^{v-1} d_{v\lambda} h_\lambda^x(t) + h_v^x(t) \right);$$

$$d_{v\lambda} = c_{v\lambda} + \sum_{\mu=\lambda+1}^{v-1} c_{v\mu} d_{\mu\lambda} \quad (\lambda = \overline{1, v-2}); \quad d_{v, v-1} = c_{v, v-1}; \quad v = \overline{2, L}.$$

При этом функции $a_v(t)$ выражаются через базисные вейвлет-функции вида (3):

$$a_1(t) = g_1(t), \quad a_v(t) = \sum_{\lambda=1}^{v-1} d_{v\lambda} g_\lambda(t) + g_v(t) \quad (v = \overline{2, L}).$$

Разработанный алгоритм был реализован для вейвлетов Хаара в виде инструментального программного обеспечения «СТИТ-КРВЛ.1» для среды MATLAB.

3 Построение канонического разложения стохастического процесса на основе трехслойной вейвлет-нейронной сети

Обобщая результаты построения ВлКР СтП $X(t)$, изложенные в разд. 2, будем искать его КР (1) для координатных функций в виде:

$$x_v(t) = g_v(t) \quad (v = \overline{1, L}).$$

Тогда, согласно прямой теореме Пугачёва, КР ковариационной функции $K_X(t_1, t_2)$ примет вид:

$$K_X(t_1, t_2) = \sum_{v=1}^L D_v g_v(t_1) g_v(t_2). \quad (4)$$

Тем самым задача построения КР СтП $X(t)$ сведется к задаче аппроксимации ковариационной функции $K_X(t_1, t_2)$ линейной зависимостью нелинейных двумерных вейвлет-функций вида (4), которую можно вычислить с использованием многослойного персептрона [17]. После определения дисперсий D_v в (4), согласно обратной теореме Пугачёва, получим искомое КР СтП $X(t)$:

$$X(t) = \sum_{v=1}^L V_v g_v(t), \quad (5)$$

где V_v — некоррелированные СВ с нулевыми математическими ожиданиями и дисперсиями D_v .

Для нахождения неизвестных D_v в (4) задается множество маркированных примеров

$$T = \{(\tau^i, d^i)\}_{i=1}^N, \quad \tau^i = [t_1^i \ t_2^i]^T, \quad d^i = K_X(t_1^i, t_2^i).$$

Для обучения многослойного перцептрона с учителем применим алгоритм обратного распространения ошибки [17]. Он представляет собой градиентный метод и обладает двумя основными свойствами: (1) процедуру метода легко просчитать локально; (2) процедура метода реализует градиентный спуск в пространстве весов, при котором синаптические веса обновляются от примера к примеру.

В качестве активационных функций нейронов скрытого слоя рассмотрим функции, построенные на основе вейвлет-базиса (3):

$$\Psi_v(\tau) = \Psi_v(t_1, t_2) = g_v(t_1)g_v(t_2) \quad (v = \overline{1, L}).$$

Вектор признаков

$$\Psi(\tau) = [\Psi_0(\tau) \ \Psi_1(\tau) \ \dots \ \Psi_L(\tau)]^T$$

представляет собой «образ», индуцированный в пространстве признаков при предъявлении вектора входного сигнала $\tau = [t_1 \ t_2]^T$.

Пусть $w = [w_1 \ w_2 \ \dots \ w_L]^T$ — вектор весовых коэффициентов, связывающих пространство признаков с выходным пространством, $y = [y_1 \ y_2 \ \dots \ y_N]^T$ — действительный отклик системы, т. е. вектор оценки желаемого отклика системы $d = [d_1 \ d_2 \ \dots \ d_N]^T$. При этом вектор весовых коэффициентов $w = w(n)$ меняется при каждой итерации алгоритма n .

Функционирование ВНС можно описать следующим образом:

– входной сигнал

$$\tau^i = [t_1^i \ t_2^i]^T \quad (i = \overline{1, N});$$

– скрытый слой ВНС задает пространство признаков

$$\Psi_v(\tau^i) = \Psi_v(t_1^i, t_2^i) = g_v(t_1^i) g_v(t_2^i) \quad (v = \overline{1, L}; \ i = \overline{1, N});$$

– выходной слой ВНС вычисляет отклик системы

$$y^i(n) = \sum_{v=1}^L w_v(n) \Psi_v(\tau^i),$$

или в векторной форме

$$y^i(n) = \Psi^T(\tau^i) w(n) \quad (i = \overline{1, N}),$$

для итерации n алгоритма.

Выходной сигнал $y^i(n)$ для каждого i -го примера сравнивается с соответствующим желаемым откликом системы $d^i(n)$. Их разность определяет сигнал ошибки

$$e^i(n) = d^i - y^i(n) = d^i - \Psi^T(\tau^i)w(n).$$

Функция стоимости задается в виде среднеквадратичной ошибки (СКО; Mean Squared Error — MSE) для всего набора примеров:

$$E(w(n)) = \frac{1}{2N} \sum_{i=1}^N (e^i(n))^2. \quad (6)$$

Далее надо решить задачу безусловной оптимизации функции стоимости $E(w)$ по отношению к вектору весов w : $E(w) \rightarrow \min$. Необходимое условие оптимальности — равенство нулю вектора градиента скорости:

$$\nabla E(w^*) = 0,$$

где

$$\nabla E(w) = \left[\frac{\partial E}{\partial w_1} \quad \frac{\partial E}{\partial w_2} \quad \dots \quad \frac{\partial E}{\partial w_L} \right]^T,$$

$$\frac{\partial E}{\partial w_l} = \frac{1}{N} \sum_{i=1}^N \Psi_l(\tau^i) \left(d^i - \sum_{j=1}^L w_j(n) \Psi_j(\tau^i) \right) \quad (l = \overline{1, L}).$$

Для удобства обозначим градиент $\nabla E(w) = g(n)$.

Методы безусловной оптимизации основаны на алгоритме последовательного спуска. Исходное значение $w(0) = 0$. Далее генерируется последовательность весовых векторов $w(1), w(2), \dots$ такая, что значение функции стоимости уменьшается при каждой итерации алгоритма: $E(w(n+1)) < E(w(n))$. Применим метод наискорейшего спуска. Корректировка весов осуществляется по формуле:

$$w(n+1) = w(n) - \eta g(n), \quad (7)$$

где $\eta = const$ — параметр скорости обучения, $0 < \eta < 1$. Тогда с учетом (7) имеем

$$E(w(n+1)) = E(w(n)) - \eta \|g(n)\|^2.$$

В общем случае шагов итерации может быть бесконечно много. Из опыта известно, что для выхода из процедуры адаптации весов выполняется условие $|E(w(n))| \leq \leq E_a$ для $E_a = const$ — параметра, определяющего точность приближенных вычислений.

В результате получаем алгоритм КР ковариационной функции на основе обучения ВНС методом обратного распространения ошибки.

Алгоритм. Построение КР ковариационной функции скалярного СтП на основе ВНС (рис. 1).

1. Задание множества обучающих примеров: вектора входного сигнала $\tau^i = [t_1^i \ t_2^i]^T$ и желаемого отклика d^i ($i = \overline{1, N}$).
2. Задание параметра скорости обучения $0 < \eta < 1$.
3. Инициализация весов. При $n = 0$ $w(0) = 0$. Последующие шаги начинаются с $n = 1, 2, \dots$

4. Вычисление функций скрытого слоя:

$$\Psi_v(\tau^i) = \Psi_v(t_1^i, t_2^i) = g_v(t_1^i)g_v(t_2^i) \quad (v = \overline{1, L}, \quad i = \overline{1, N}).$$

5. Вычисление фактического отклика:

$$y^i(n) = \Psi^T(\tau^i)w(n) \quad (i = \overline{1, N}).$$

6. Вычисление ошибки $e^i(n) = d^i - y^i(n)$ для каждого примера и СКО для всего набора обучающих примеров (6). Начиная со второго шага, проверка условия окончания адаптации весов:

$$|E(w(n))| \leq E_a.$$

Если условие выполняется, выход из алгоритма, в противном случае переход к п. 7.

7. Адаптация вектора весов. Изменяем вектор весов:

$$w(n+1) = w(n) - \eta g(n).$$

8. Увеличиваем номер итерации n на 1 и возвращаемся к п. 5.

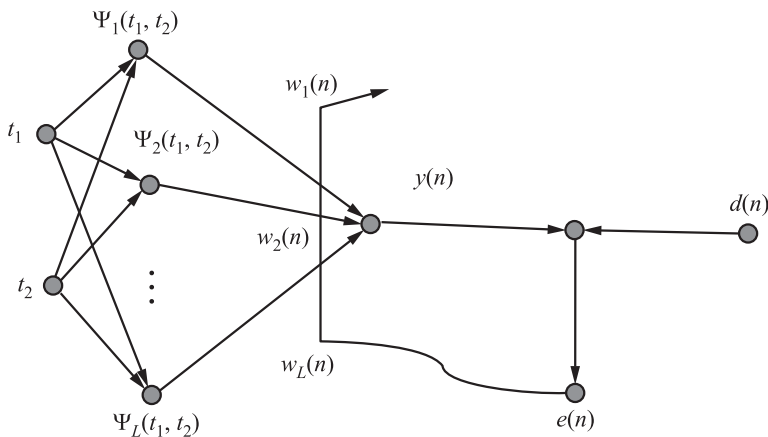


Рис. 1 Граф передачи сигнала в ВНС

После обучения ВНС проводится тестирование для проверки качества аппроксимации.

На рис. 1 представлен граф прохождения сигнала в ВНС в процессе обучения, основанного на коррекции СКО (6). Сигнал ошибки $e(n)$ инициализирует механизм управления, цель которого заключается в применении последовательности корректировок к синаптическим весам w . Эти изменения нацелены на пошаговое приближение выходного сигнала y к желаемому d . На рис. 1 видно, что обучение на основе коррекции ошибок — это пример замкнутой системы с обратной связью. Из теории управления известно, что устойчивость такой системы определяется параметром обратной связи, т. е. параметром скорости обучения η . Для обеспечения устойчивости и сходимости итеративного процесса обучения требуется тщательный подбор этого параметра. Пошаговая корректировка синаптических весов продолжается, как правило, до тех пор, пока система не достигнет устойчивого состояния, при котором синаптические веса практически стабилизируются. В этой точке процесс обучения останавливается [17]. При этих значениях синаптических весов достигается минимум функции стоимости (6).

Преимущества предлагаемого подхода нахождения минимума функции (6) по сравнению со стандартными методами, основанными на решении системы уравнений $\partial E/\partial w_l = 0$ ($l = \overline{1, L}$) и приводящими к сложным математическим процессам, заключаются в простоте вычисления градиента $\nabla E(w)$ и сходимости алгоритма при $0 < \eta < 1$.

4 Пример

Рассмотрим задачу моделирования СтП $X(t)$ на промежутке $[t_0, t_0 + T]$ с нулевым математическим ожиданием и нестационарной ковариационной функцией $K_X(t_1, t_2) = D \exp\{-\alpha|t_2 - t_1| + \beta(t_2 + t_1)\}$ с помощью ВНС. Каноническое разложение СтП $X(t)$ строим в виде (5), где СВ V_v имеют нулевые математические ожидания и дисперсии $D_v = w_v$. Значения синаптических весов w_v определяются из условия минимума СКО представления ковариационной функции $K_X(t_1, t_2)$ ее КР

$$\text{MSE} = E(w) = \frac{2}{N} \sum_{i=1}^N \left(K_X(t_1^i, t_2^i) - \sum_{v=1}^L w_v g_v(t_1^i) g_v(t_1^i) \right)^2 :$$

$$\min_w E(w) = \min_w \left\{ \frac{2}{N} \sum_{i=1}^N \left(K_X(t_1^i, t_2^i) - \sum_{v=1}^L w_v g_v(t_1^i) g_v(t_2^i) \right)^2 \right\}$$

для заданного набора входных данных

$$T = \left\{ \tau^i = [t_1^i \ t_2^i]^T, \ d^i = K_X(t_1^i, t_2^i); \ i = \overline{1, N} \right\}$$

при следующих исходных данных: $D = 1$; $\alpha = 1$; $\beta = 0,001$; $t_0 = 0$; $T = 8$.

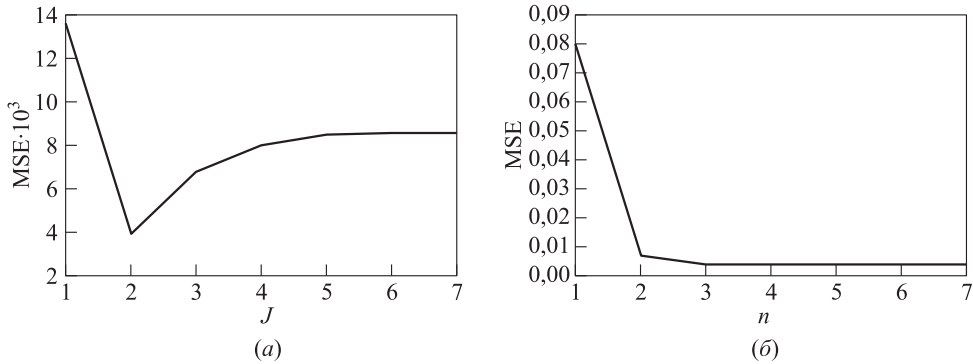


Рис. 2 Графики зависимости MSE аппроксимации $K_X(t_1, t_2)$ от уровня разрешения J базиса Хаара (а) и от числа n итераций алгоритма при $J = 2$ (б)

Для проведения вычислительных экспериментов разработано инструментальное программное обеспечение (ИПО) в среде MATLAB. Вначале экспериментально было определено, что разработанный алгоритм обеспечивает максимальную скорость обучения ВНС без потери устойчивости при параметре скорости обучения $\eta = 0,8$ и при числе тестовых примеров, равном числу базисных вейвлет-функций, $N = L$. Далее была выявлена зависимость MSE от максимального уровня разрешения J базиса Хаара, представленная на рис. 2, а. Минимальное значение, равное 0,0039, MSE принимает при $J = 2$, т. е. для 8 базисных функций Хаара. Тем самым было экспериментально определено оптимальное число нейронов скрытого слоя ВНС, равное 8. На рис. 2, б представлен график зависимости MSE от числа n итераций алгоритма для $\eta = 0,8$ и $N = L = 8$. MSE принимает минимальное значение, равное 0,0039, при $n = 4$.

На рис. 3, а представлены графики точных значений D_{true} дисперсии $D_X(t) = K_X(t, t) = D \exp\{2\beta t\}$ и ее оценки D_{prib} для выбранных значений параметров $\eta = 0,8$ и $N = L = 8$. На рис. 3, б изображен график модуля ошибки оценки дисперсии.

В итоге КРВНС СтП примет вид:

$$X(t) = \sum_{v=1}^8 V_v g_v(t).$$

С помощью замены переменных $\bar{t} = (t - t_0)/T$ выразим базисные функции $g_v(t)$ через вейвлеты Хаара $h_v(\bar{t})$, определенные на промежутке $[0, 1]$:

$$g_1(t) = h_1(\bar{t}) = \begin{cases} 1, & \text{если } \bar{t} \in [0, 1]; \\ 0, & \text{если } \bar{t} \notin [0, 1]; \end{cases}$$

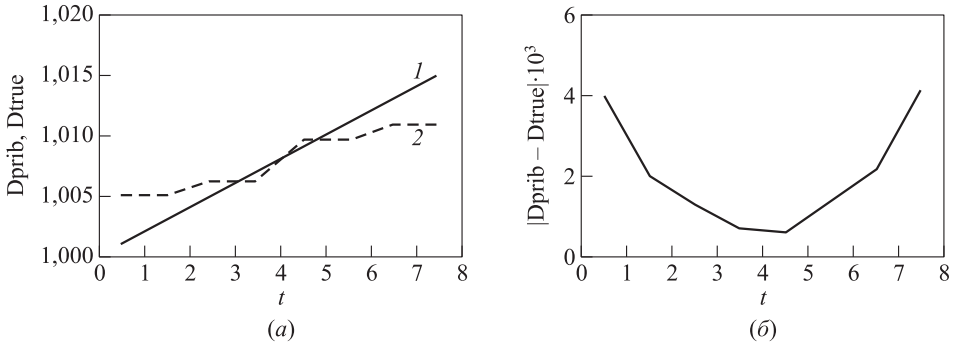


Рис. 3 Графики точных значений D_{true} (1) дисперсии СтП и ее оценки D_{prib} (2) (a) и модуля ошибки оценки дисперсии (б) на отрезке $[0, 8]$

$$g_2(t) = h_2(\bar{t}) = \begin{cases} 1, & \text{если } \bar{t} \in \left[0, \frac{1}{2}\right); \\ -1, & \text{если } \bar{t} \in \left[\frac{1}{2}, 1\right); \\ 0, & \text{если } \bar{t} \notin [0, 1]; \end{cases}$$

$$g_3(t) = h_3(\bar{t}) = \begin{cases} \sqrt{2}, & \text{если } \bar{t} \in \left[0, \frac{1}{4}\right); \\ -\sqrt{2}, & \text{если } \bar{t} \in \left[\frac{1}{4}, \frac{1}{2}\right); \\ 0, & \text{если } \bar{t} \notin \left[0, \frac{1}{2}\right); \end{cases}$$

$$g_4(t) = h_4(\bar{t}) = \begin{cases} \sqrt{2}, & \text{если } \bar{t} \in \left[\frac{1}{2}, \frac{3}{4}\right); \\ -\sqrt{2}, & \text{если } \bar{t} \in \left[\frac{3}{4}, 1\right); \\ 0, & \text{если } \bar{t} \notin \left[\frac{1}{2}, 1\right); \end{cases}$$

$$g_i(t) = h_i(\bar{t}) = \psi_{2k}(\bar{t}) = \begin{cases} 2, & \text{если } \bar{t} \in \left[\frac{k}{4}, \frac{k+0,5}{4}\right); \\ -2, & \text{если } \bar{t} \in \left[\frac{k+0,5}{4}, \frac{k+1}{4}\right); \\ 0, & \text{если } \bar{t} \notin \left[\frac{k}{4}, \frac{k+1}{4}\right) \quad (k = 0, 3; i = 4 + k + 1). \end{cases}$$

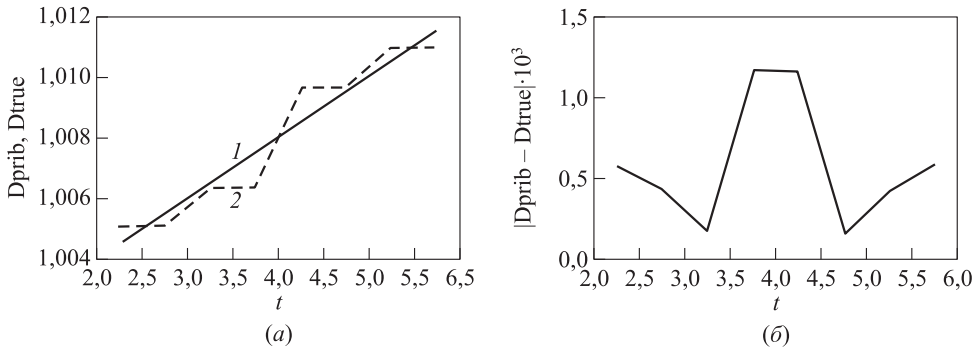


Рис. 4 Графики точных значений $D_{true}(t)$ дисперсии СтП $X(t)$ и ее оценки D_{prib} (2) (a) и модуля ошибки оценки дисперсии (б) при тестировании ВНС на отрезке $[2, 6]$

Случайные величины V_v ($v = \overline{1, 8}$) имеют дисперсии, равные синаптическим весам:

$$\begin{aligned} D[V_1] &= w_1 = 0,2437; & D[V_2] &= w_2 = 0,1878; \\ D[V_3] &= w_3 = 0,1285; & D[V_4] &= w_4 = 0,1295; \\ D[V_5] &= w_5 = 0,0792; & D[V_6] &= w_6 = 0,0795; \\ D[V_7] &= w_7 = 0,0798; & D[V_8] &= w_8 = 0,0801. \end{aligned}$$

Обученная ВНС была протестирована для различных значений $t \in [0, 8]$, отличных от точек из набора обучения ВНС. Вычислительные эксперименты показали, что во всех случаях $MSE = 0,0039$. На рис. 4, а представлены графики значений D_{true} дисперсии $D_X(t) = K_X(t, t) = D \exp\{2\beta t\}$ и ее оценки D_{prib} для $t \in [2, 6] \subset [0, 8]$. На рис. 4, б изображен график модуля ошибки оценки дисперсии.

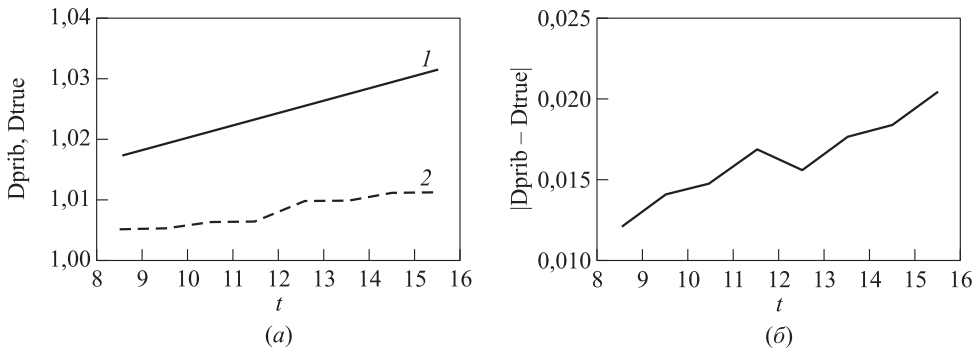


Рис. 5 Графики точных значений $D_{true}(t)$ дисперсии СтП $X(t)$ и ее оценки D_{prib} (2) (a) и модуля ошибки оценки дисперсии (б) при тестировании ВНС на отрезке $[8, 16]$

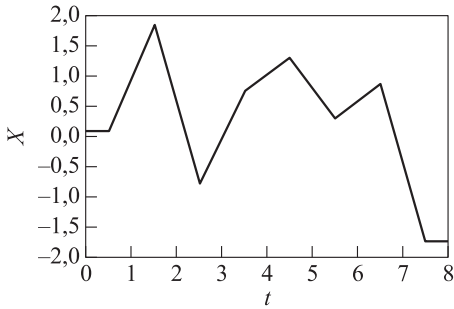


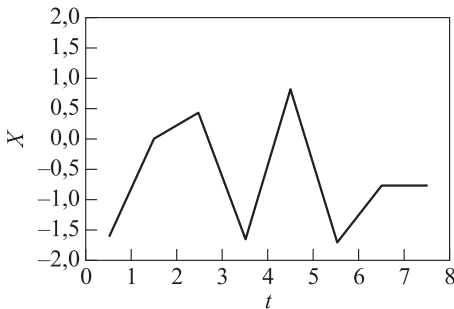
Рис. 6 График реализации СтП $X(t)$ на отрезке $[0, 8]$, построенный на основе ВНС

для базиса Хаара с максимальным уровнем разрешения $J = 2$. Среднеквадратичная ошибка представления ковариационной функции $K_X(t_1, t_2)$ ее КР равна 0,0033. На рис. 7 показаны результаты вычислительного эксперимен-

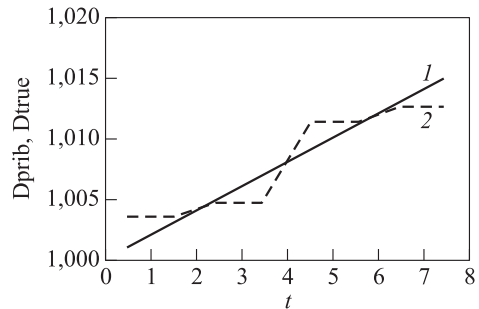
Кроме того, обученная ВНС была протестирована на интервале $[8, 16]$. В этом случае получено $MSE = 0,0040$. Результаты тестирования представлены на рис. 5.

На рис. 6 представлен результат моделирования СтП $X(t)$ на отрезке $[0, 8]$ с помощью разработанной ВНС.

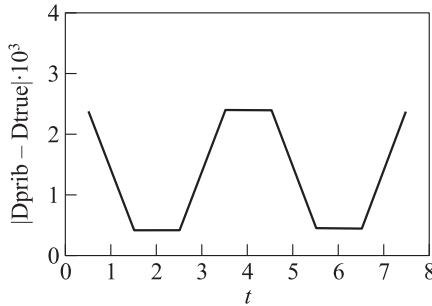
Для сравнения было построено ВлКР данного СтП $X(t)$ на отрезке $[0, 8]$ с применением ИПО «СТИГ-КРВЛ.1», реализующего рекуррентный алгоритм, изложенный в разд. 2,



(а)



(б)



(в)

Рис. 7 Графики реализации СтП $X(t)$ (а), точных значений $D_{true}(t)$ дисперсии СтП $X(t)$ и ее оценки D_{prib} (2) (б) и модуля ошибки оценки дисперсии (в) на отрезке $[0, 8]$, полученные с применением ИПО «СТИГ-КРВЛ.1»

та. Оба метода дают одинаковый порядок точности. При этом ВНС проще в математическом и алгоритмическом плане.

5 Заключение

Разработан новый метод построения КР скалярного СтП, заданного на конечном промежутке времени, на основе трехслойной ВНС. Скрытый слой задается на основе выбранного ортонормированного базиса вейвлетов с компактными носителями. Для обучения ВНС применяется метод наискорейшего спуска.

В результате вычислительных экспериментов было определено, что разработанный алгоритм обеспечивает максимальную скорость обучения ВНС без потери устойчивости при параметре скорости обучения, равном 0,8, и при числе тестовых примеров, равном числу базисных вейвлет-функций. Сравнительный анализ двух методов построения КР: ВлКР (рекуррентный) и КРВНС — показали одинаковую вычислительную точность. Преимущество КРВНС по сравнению с ВлКР заключается в представлении СтП в виде линейной комбинации базисных вейвлет-функций со случайными коэффициентами, что упрощает выполнение различных математических операций.

Полученные результаты допускают обобщение на случай векторного СтП, а также применение при разработке новых методов анализа и синтеза СтС различного назначения.

Литература

1. Пугачёв В. С. Теория случайных функций и ее применение к задачам автоматического управления. — М.: Физматгиз, 1962. 884 с.
2. Синицын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009. 768 с.
3. Синицын И. Н. Канонические представления случайных функций. Теория и применения. — 2-е изд. — М.: ТОРУС ПРЕСС, 2023. 816 с.
4. Sinitsyn I. N. Developing the theory of stochastic canonic expansions // Pattern Recognition Image Analysis, 2023. Vol. 33. Iss. 4. P. 862–887. doi: 10.1134/S1054661823040429.
5. Марченко Б. Г. Метод стохастических интегральных представлений и его приложения в радиотехнике. — Киев: Наукова думка, 1973. 191 с.
6. Meyer Y. Analysis at Urbana. Vol. 1. Analysis in function spaces. — Cambridge, U.K.: Cambridge University Press, 1989. 436 p.
7. Mallat S. G. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$ // T. Am. Math. Soc., 1989. Vol. 315. Iss. 1. P. 69–87. doi: 10.2307/2001373.
8. Rhif M., Ben Abbes A., Farah I. R., Martínez B., Sang Y. Wavelet transform application for/in non-stationary time-series analysis: A review // Appl. Sci., 2019. Vol. 9. Iss. 7. Art. 1345. 22 p. doi: 10.3390/app9071345.

9. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Конашенкова Т. Д. Дополнение «Синтез нестационарных систем методом вейвлет канонических разложений» // Сеницын И. Н. Канонические представления случайных функций. Теория и приложения. — 2-е изд. — М.: ТОРУС ПРЕСС, 2023. С. 752–807.
10. Добеши И. Десять лекций по вейвлетам. — М.–Ижевск: НИЦ «Регулярная и хаотическая динамика», 2004. 464 с. (*Daubechies I. Ten lectures on wavelets.* — Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992. 352 p.)
11. Сеницын И. Н., Сергеев И. В., Корепанов Э. Р., Конашенкова Т. Д. Инструментальное программное обеспечение анализа и синтеза стохастических систем высокой доступности (IV) // Системы высокой доступности, 2017. Т. 13. № 3. С. 55–69. EDN: ZSQHLR.
12. Sinitsyn I., Sinitsyn V., Korepanov E., Konashenkova T. Bayes synthesis of linear nonstationary stochastic systems by wavelet canonical expansions // Mathematics, 2022. Vol. 10. Iss. 9. Art. 1517. 14 p. doi: 10.3390/math10091517.
13. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Конашенкова Т. Д. Инструментальное программное обеспечение анализа и синтеза стохастических систем высокой доступности (XVII) // Системы высокой доступности, 2023. Т. 19. № 2. С. 5–24. doi: 10.18127/j20729472-202302-01. EDN: PCUHWT.
14. Sinitsyn I., Sinitsyn V., Korepanov E., Konashenkova T. Synthesis of nonlinear nonstationary stochastic systems by wavelet canonical expansions // Mathematics, 2023. Vol. 11. Iss. 9. Art. 2059. 18 p. doi: 10.3390/math11092059.
15. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Конашенкова Т. Д. Инструментальное программное обеспечение анализа и синтеза стохастических систем высокой доступности (XVIII) // Системы высокой доступности, 2023. Т. 19. № 3. С. 18–34. doi: 10.18127/j20729472-202303-02. EDN: SQPRDO.
16. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Конашенкова Т. Д. Синтез оптимальных по функциональным бейсовым критериям стохастических систем методом вейвлет-канонических разложений // Системы высокой доступности, 2023. Т. 19. № 4. С. 37–50. doi: 10.18127/j20729472-202304-03. EDN: WYQSQT.
17. Хайкин С. Нейронные сети: полный курс / Пер. с англ. — 2-е изд. — СПб.: Диалектика, 2020. 1104 с. (*Haykin S. Neural networks. A comprehensive foundation.* — 2nd ed. — Prentice-Hall of India Pvt. Ltd., 1999. 842 p.)
18. Терехов С. А. Вейвлеты и нейронные сети // Научная сессия МИФИ-2001: III Всеросс. научн.-техн. конф. «Нейроинформатика-2001»: лекции по нейроинформатике. — М.: МИФИ, 2001. С. 142–181.
19. Veitch D. Wavelet neural networks and their application in the study of dynamical system // Networks, 2005. Vol. 1. No. 8. P. 313–320.

Поступила в редакцию 15.03.24

NONSTATIONARY STOCHASTIC PROCESS MODELING BY CANONICAL EXPANSION AND WAVELET NEURAL NETWORK

I. N. Sinitsyn, V. I. Sinitsyn, E. R. Korepanov, and T. D. Konashenkova

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

Abstract: For scalar stochastic processes (StP) at finite time intervals and their canonical expansions (CE), a technology based on wavelet neural networks (WNN) is constructed. For WNN learning, the method of steepest descent was used. The three-layer WNN architecture is presented. The activation functions of the latent layer are based on chosen wavelet basis with general compact carrier. For StP covariance function, a special WNN algorithm of CE construction is developed. The covariance function CE corresponds to CE StP in the form of linear combination of wavelet basis with zero mathematical expectations and variances defined by the suggested algorithm. A numerical example illustrates CE WNN preference with wavelet CE.

Keywords: canonical expansion; covariance function; modeling; stochastic process; wavelet; wavelet neural network

DOI: 10.14357/08696527240202

EDN: YFHFIN

Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

References

1. Pugachev, V. S. 1962. *Teoriya sluchaynykh funktsiy i ee primeneniye k zadacham avtomaticheskogo upravleniya* [Theory of random functions and its application to automatic control problems]. Moscow: Fizmatgiz. 884 p.
2. 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.
3. Sinitsyn, I. N. 2023. *Kanonicheskie predstavleniya sluchaynykh funktsiy. Teoriya i primeneniya* [Canonical expansions of random functions. Theory and applications]. 2nd ed. Moscow: TORUS PRESS. 816 p.
4. Sinitsyn, I. N. 2023. Developing the theory of stochastic canonic expansions. *Pattern Recognition Image Analysis* 33(4):862–887. doi: 10.1134/S1054661823040429.
5. Marchenko, B. G. 1973. *Metod stokhasticheskikh integral'nykh predstavleniy i ego prilozheniya v radiotekhnike* [Method of stochastic integral representations and its applications in radio engineering]. Kiev: Naukova dumka. 191 p.

6. Meyer, Y. 1989. *Analysis at Urbana. Vol. 1. Analysis in function spaces*. Cambridge, U.K.: Cambridge University Press. 436 p.
7. Mallat, S. G. 1989. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *T. Am. Math. Soc.* 315(1):69–87. doi: 10.2307/2001373.
8. Rhif, M., A. Ben Abbes, I. R. Farah, B. Martínez, and Y. Sang. 2019. Wavelet transform application for/in non-stationary time-series analysis: A review. *Appl. Sci.* 9(7):1345. 22 p. doi: 10.3390/app9071345.
9. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, and T. D. Konashenkova. 2023. Dopolnenie “Sintez nestatsionarnykh sistem metodom veyvlet kanonicheskikh razlozheniy” [Appendix “Synthesis of nonstationary Systems by wavelet canonical expansions”]. Sinitsyn, I. N. *Kanonicheskie predstavleniya sluchaynykh funktsiy. Teoriya i primeneniya* [Canonical expansions of random functions. Theory and applications]. 2nd ed. Moscow: TORUS PRESS. 752–807.
10. Daubechies, I. 1992. *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics. 352 p.
11. Sinitsyn, I. N., I. V. Sergeev, E. R. Korepanov, and T. D. Konashenkova. 2017. Instrumental’noe programmnoe obespechenie analiza i sinteza stokhasticheskikh sistem vysokoy dostupnosti (IV) [Software tools for analysis and synthesis of stochastic systems with high availability (IV)]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 13(3):55–69. EDN: ZSQHLR.
12. Sinitsyn, I., V. Sinitsyn, E. Korepanov, and T. Konashenkova. 2022. Bayes synthesis of linear nonstationary stochastic systems by wavelet canonical expansions. *Mathematics* 10(9):1517. 14 p. doi: 10.3390/math10091517.
13. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, and T. D. Konashenkova. 2023. Instrumental’noe programmnoe obespechenie analiza i sinteza stokhasticheskikh sistem vysokoy dostupnosti (XVII) [Software tools for analysis and synthesis of stochastic systems with high availability (XVII)]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 19(2):5–24. doi: 10.18127/j20729472-202302-01. EDN: PCUHWI.
14. Sinitsyn, I., V. Sinitsyn, E. Korepanov, and T. Konashenkova. 2023. Synthesis of non-linear nonstationary stochastic systems by wavelet canonical expansions. *Mathematics* 11(9):2059. 18 p. doi: 10.3390/math11092059.
15. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, and T. D. Konashenkova. 2023. Instrumental’noe programmnoe obespechenie analiza i sinteza stokhasticheskikh sistem vysokoy dostupnosti (XVIII) [Software tools for analysis and synthesis of stochastic systems with high availability (XVIII)]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 19(3):18–34. doi: 10.18127/j20729472-202303-02. EDN: SQPRDO.
16. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, and T. D. Konashenkova. 2023. Sintez optimal’nykh po funktsional’nym beyesovym kriteriyam stokhasticheskikh sistem metodom veyvlet-kanonicheskikh razlozheniy [Methodological support of functional Bayes synthesis of stochastic systems based on wavelet canonical expansions]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 19(4):37–50. doi: 10.18127/j20729472-202304-03. EDN: WYQSQT.
17. Haykin, S. 1999. *Neural networks. A comprehensive foundation*. 2nd ed. Prentice-Hall of India Pvt. Ltd. 842 p.
18. Terekhov, S. A. 2001. Veyvlety i neyronnye seti [Wavelets and neural networks]. *Nauchnaya sessiya MIFI-2001: III Vseross. nauchn.-tekhn. konf. “Neyroinformatika-2001”: lektzii po neyroinformatike* [Scientific session MEFHI-2001: 3rd All-Russian

Scientific and Technical Conference “Neuroinformatics-2001”: Lectures on neuroinformatics]. Moscow: MIFI. 142–181.

19. Veitch, D. 2005. Wavelet neural networks and their application in the study of dynamical system. *Networks* 1(8):313–320.

Received March 15, 2024

Contributors

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; sinitsin@dol.ru

Sinitsyn Vladimir I. (b. 1968) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; vsinitsin@ipiran.ru

Korepanov Eduard R. (b. 1966) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; ekorepanov@ipiran.ru

Konashenkova Tatiana D. (b. 1964) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; tkonashenkova64@mail.ru

О МОДЕЛИРОВАНИИ СОВМЕСТНОГО ОБСЛУЖИВАНИЯ ОДНОАДРЕСНОГО И МНОГОАДРЕСНОГО ТРАФИКА С ПРИОРИТИЗАЦИЕЙ*

*А. С. Румянцева¹, Ф. А. Москалева², А. К. Самуйлов³, С. Я. Шоргин⁴,
Ю. В. Гайдамака⁵*

Аннотация: Увеличение числа пользователей и высокие требования к сети привели к необходимости внедрения технологии передачи данных на миллиметровых волнах (mmWave), чтобы повысить скорость передачи данных и минимизировать задержки. Переход на миллиметровый диапазон волн в пятом поколении (5G) сетей связи учитывает особенности реализации предоставления услуг, в числе которых чувствительность направленных антенных лучей к блокировке радиосигнала. Кроме этого, при эффективном планировании сети, предоставляющей одно- и многоадресные услуги, требуется достигать баланса между входящими в противоречие решениями для этих двух типов трафика. В статье с применением методов теории массового обслуживания разработана математическая модель для описания механизма приоритизации при обслуживании комбинации одно- и многоадресного трафика, для каждого из которых принимается во внимание возможность появления блокаторов прямой видимости (Line of Sight, LoS). Особенности распространения сигнала учитываются в модели через разбиение зоны покрытия на две области. Построенная математическая модель позволяет найти аналитические выражения для вероятностей потери одно- и многоадресных соединений, а также объема занятого ресурса как отдельными типами трафика в разных областях зоны покрытия, так и во всей системе в целом. На балансе между перечисленными основными показателями производительности основан метод распределения радиоресурса базовой станции (БС) сети, который позволяет следить за соблюдением требований к качеству для обоих типов услуг.

Ключевые слова: mmWave; одноадресные соединения; многоадресные соединения; LoS; блокировка LoS; математическая модель

DOI: 10.14357/08696527240203

EDN: QXWAUL

* Исследование выполнено за счет гранта Российского научного фонда № 22-79-10053 (<https://rscf.ru/project/22-79-10053>).

¹ Российский университет дружбы народов, 1132223493@rudn.ru

² Российский университет дружбы народов, moskaleva-fa@rudn.ru

³ Российский университет дружбы народов, samuylov-ak@rudn.ru

⁴ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

⁵ Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaidamaka-yuv@rudn.ru

1 Введение

В сетях 5G для поддержания высокоскоростной передачи данных с низкими задержками появилась необходимость привлечения незадействованного ранее радиоресурса, поскольку используемая полоса частот ниже 6 ГГц не справлялась с растущей нагрузкой от устройств mMTC (massive Machine-Type Communication) в условиях жестких требований URLLC (Ultra-Reliable Low Latency Communication) приложений [1]. По этой причине для передачи данных начали использоваться миллиметровые волны (с частотами от 24 до 100 ГГц) [2], что обеспечило дополнительную полосу пропускания, а значит, и большую скорость передачи данных. При этом пришлось учитывать особенности передачи данных на высоких частотах, одна из которых заключается в чувствительности сигнала к блокировке прямой видимости между приемо-передающими устройствами [3]. Блокатором сигнала может стать практически любой объект, поскольку длина миллиметровых волн значительно меньше размеров большинства предметов в окружающей среде. В результате нарушения прямой видимости происходит существенное падение скорости передачи, что может приводить даже к разрыву соединения. В рамках концепции нарезки радиоресурсов сети [4] предусмотрено перераспределение радиоресурса БС для выделения дополнительного ресурса с целью поддержания соединения на период блокировки сигнала [5–7], что должно обеспечивать качество предоставления услуги на требуемом уровне. При этом в сетях с одноадресной и многоадресной передачей приоритет следует отдавать поддержанию многоадресных соединений, обслуживающих большее число пользователей сети [8–11]. Авторы [9] используют математический аппарат для исследования влияния двух типов трафика друг на друга при возможности нарушения прямой видимости только для многоадресных соединений. Особенность [8, 11] заключается в учете при моделировании обслуживания двух типов трафика расстояния от БС до пользовательского устройства (ПУ). Авторы [10] при построении математической модели принимают во внимание возможности перекрытия прямой видимости для обоих типов трафика и учитывают особенности распространения сигнала, разбивая зону покрытия на несколько областей по отдаленности от БС, при этом изучение характеристик модели происходит с помощью имитационного моделирования.

В данной работе, служащей продолжением [12], построена математическая модель в виде системы массового обслуживания (СМО) с приоритетным обслуживанием для изучения характеристик системы через стационарные вероятности, полученные как решение системы уравнений равновесия. Модель учитывает особенности распространения сигнала с помощью разбиения зоны покрытия на две области по отдаленности пользователей от БС и возможность появления блокатора прямой видимости для обоих типов соединений. В работе исследуются следующие показатели эффективности: вероятности блокировки доступа и прерывания соединений, вероятности потери и среднее число занятого ресурса сети.

2 Системная модель

На рис. 1 изображена зона покрытия антенны БС в диапазоне миллиметровых волн в виде сектора в 120° , ограниченного окружностью, радиус которой определяется моделью распространения сигнала. Зона покрытия, в которой случайным образом распределены стационарные пользователи, генерирующие запросы на предоставление услуги посредством одноадресных и многоадресных соединений, и подвижные блокаторы, время от времени перекрывающие прямую видимость между БС и ПУ, условно разбита на две области по отдаленности от БС: ближнюю и дальнюю.

Требования к объему ресурса для поддержания одно- и многоадресного соединения от пользователей, находящихся в ближней области, ниже соответствующих требований в дальней области, что обусловлено простейшей моделью распространения сигнала Free Space Path Loss, FSPL [5].

Появление блокатора прямой видимости для ПУ в ближней области не приводит к мгновенному прерыванию соединения, поскольку сначала планировщиком БС делается попытка выделить дополнительный объем ресурса, необходимый для поддержания соединения. Многоадресные соединения имеют приоритет над одноадресными: если недостаточно свободного ресурса для поддержания многоадресного соединения в ближней области, прерываются одноадресные соединения, число которых определяется объемом необходимого ресурса. При

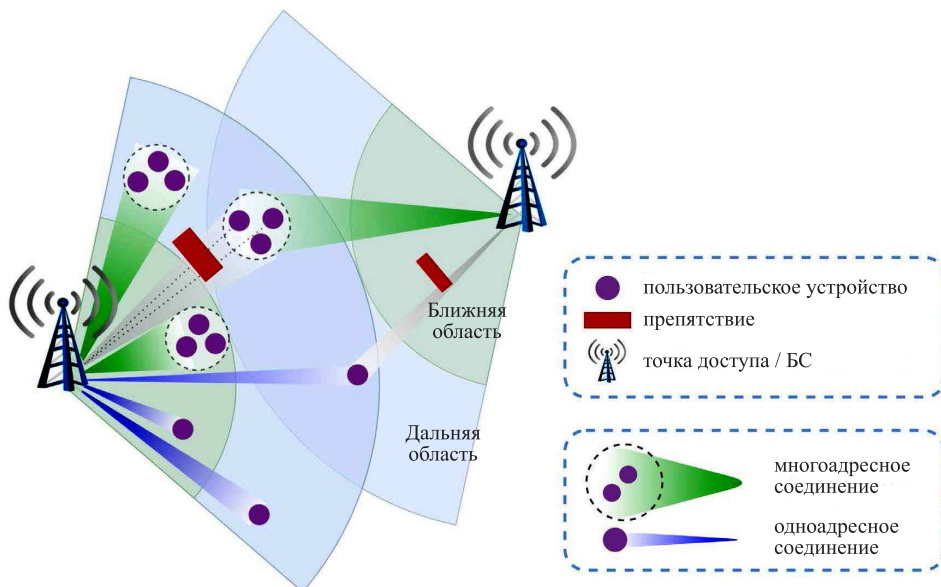


Рис. 1 Схема зоны покрытия БС

появлении блокиатора для ПУ в дальней области качество передачи данных в радиоканале резко падает, поэтому планировщик БС делает попытку хендовера — переключения обслуживания соединения на соседнюю БС, которая может взять на себя поддержание этого соединения в случае наличия свободного ресурса.

Потеря запроса на предоставление услуги возможна из-за блокировки доступа запроса в систему или из-за прерывания уже установленного соединения. Блокировка доступа — это потеря запроса в момент поступления в систему из-за недостатка на БС свободного ресурса для установления соединения. Блокировка доступа может произойти как для многоадресного запроса, так и для одноадресного, причем как в ближней области, так и в дальней. Прерывание соединения — это потеря принятого запроса во время его обслуживания. Прерывание соединений разных типов в ближней или дальней областях может быть вызвано разными причинами. Для одноадресных соединений прерывание может произойти в двух случаях: первое — если потребуется освободить ресурс для поддержания многоадресного соединения в ближней области, второе — если появится блокиатор прямой видимости для соответствующего ПУ, причем прерывание одноадресного соединения из ближней области в таком случае происходит при недостатке свободного ресурса для его поддержания на исследуемой БС, прерывание одноадресного соединения в дальней области — при недостатке свободного ресурса на соседней БС (неудачная попытка хендовера). Прерывание многоадресных соединений возможно только при появлении блокиатора прямой видимости для одного из ПУ. При этом прерваться может только многоадресное соединение в дальней области, поскольку при нехватке доступного ресурса многоадресное соединение в ближней области всегда поддерживается за счет прерывания одноадресных.

Схема функционирования системы для описанного алгоритма планировщика радиоресурсов БС показана на рис. 2.

В следующем разделе представлена математическая модель для анализа показателей эффективности системы.

3 Математическая модель

Для построения математической модели введем следующие упрощающие предположения.

1. Механизм приоритизации: в случае недостатка ресурса для поддержания многоадресного соединения в ближней области при появлении блокиатора прямой видимости происходит прерывание соответствующего числа одноадресных соединений в порядке убывания занятого ресурса на одно соединение. Как правило, больший объем требуется для соединений в условиях прямой видимости (LoS) в дальней области и для соединений в условиях отсутствия прямой видимости (nLoS, non Line-of-Sight) в ближней области, а меньший ресурс — для соединений в условии прямой видимости (LoS) в ближней области. Указанный механизм создает неявную приоритизацию

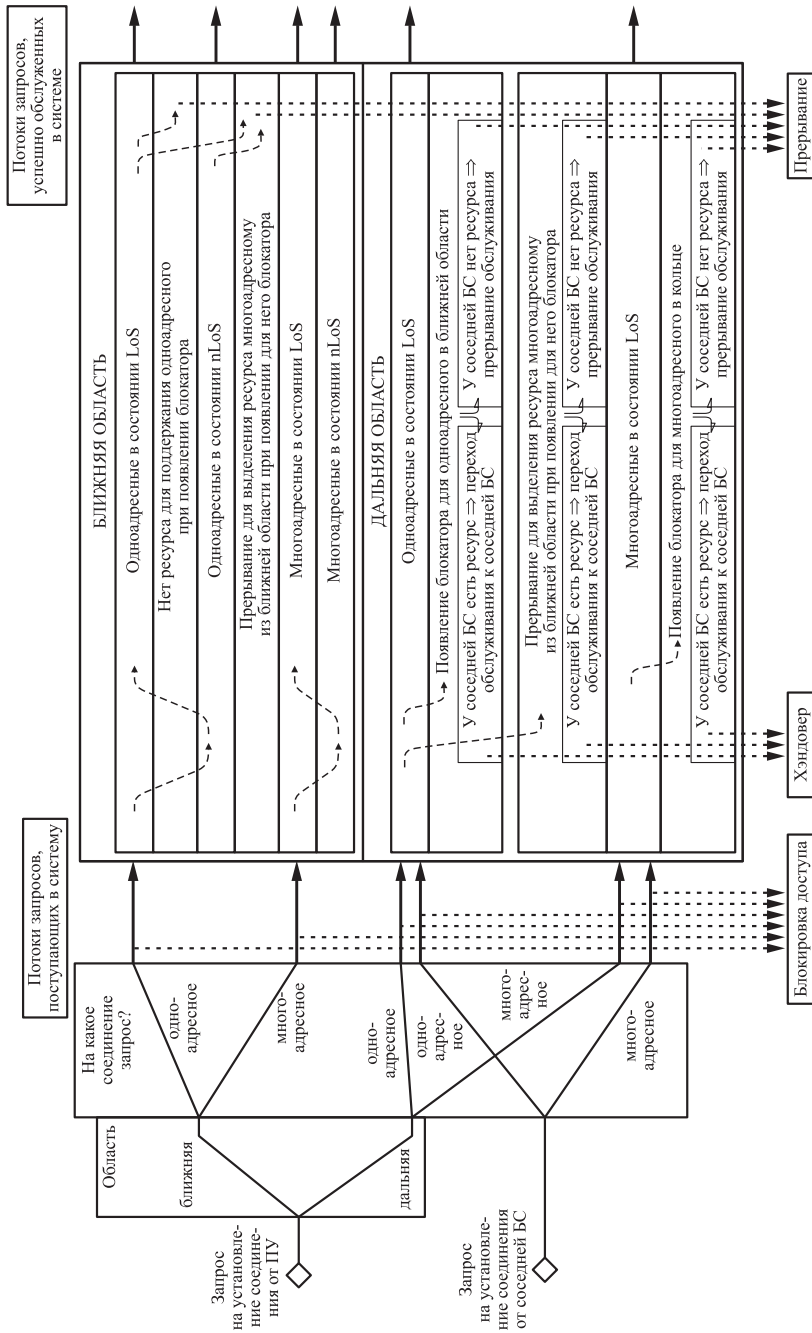


Рис. 2 Схема функционирования системы

обслуживания одноадресных заявок с небольшими требованиями к ресурсу в условиях прямой видимости (LoS).

2. В момент поступления запроса для установления соединения требуется ресурс, соответствующий условиям прямой видимости. Предположение действует как для основной БС, так и для соседних БС.
3. Длительность соединения, переключенного в результате хендовера с соседней БС, равна длительности соединения, установленного при поступлении запроса на исследуемую БС.

Процесс разделения ресурсов БС может быть смоделирован с помощью ресурсной СМО [13] с неограниченным числом приборов и ограниченным числом единиц ресурса, равным R . Интенсивности поступления заявок на одно- и многоадресные соединения соответственно в ближней и дальней областях равны λ_U^c , λ_M^c и λ_U^r , λ_M^r , причем

$$\lambda_U^r = \lambda_U^{AP} + \lambda_U^H; \quad \lambda_M^r = \lambda_M^{AP} + \lambda_M^H,$$

где $\lambda_{U/M}^{AP}$ — интенсивность поступления заявок на одноадресные/многоадресные соединения напрямую на исследуемую БС; $\lambda_{U/M}^H$ — интенсивность поступления заявок на одноадресные/многоадресные соединения в результате хендовера с соседних БС. Для описания процедуры обслуживания заявок используются параметры экспоненциального распределения длительности обслуживания заявок на одно- и многоадресные соединения соответственно в ближней и дальней областях μ_U^c , μ_M^c и μ_U^r , μ_M^r . Блокировка прямой видимости между устройствами определяется следующими параметрами:

- θ_{in}^c и θ_{in}^r (σ_{in}^c и σ_{in}^r) — параметры экспоненциального распределения случайной величины длительности интервала времени между соседними появлениями блокиатора LoS для многоадресных (одноадресных) соединений в ближней и дальней областях соответственно;
- θ_{out}^c и σ_{out}^c — параметры экспоненциального распределения случайной величины длительности интервала блокировки LoS для многоадресных и одноадресных соединений соответственно в ближней области.

Далее обозначим объемы требований к ресурсам для соединений двух типов в условиях прямой видимости (LoS) и при ее отсутствии (nLoS) в ближней и дальней областях:

- d_{LoS}^c и d_{nLoS}^c (b_{LoS}^c и b_{nLoS}^c) — требования к ресурсу одноадресных (многоадресных) сессий пользователей в ближней области в условиях прямой видимости (LoS) и при ее отсутствии (nLoS) соответственно, причем $d_{LoS}^c < d_{nLoS}^c$ и $b_{LoS}^c < b_{nLoS}^c$;
- d^r (b^r) — требования к ресурсу одноадресных (многоадресных) сессий пользователей в дальней области (в условиях прямой видимости, поскольку возможно только такое состояние ПУ).

Пусть

$$\Delta d^c = d_{nLoS}^c - d_{LoS}^c; \quad \Delta b^c = b_{nLoS}^c - b_{LoS}^c.$$

Будем рассматривать систему, в которой ресурс $R \geq b_{nLoS}^c + b^r$, что обеспечивает возможность включения многоадресных соединений в ближней и дальней областях.

Функционирование системы можно описать с помощью пятимерного случайного процесса (СП)

$$\mathbf{X}(t) = \{(m^c(t), u^c(t), u_{nLoS}^c(t), m^r(t), u^r(t)), t \geq 0\},$$

где $m^c(t) \in \{0, b_{LoS}^c, b_{nLoS}^c\}$ — объем ресурса, выделенный БС для запроса на установление многоадресного соединения в ближней области в момент t ; $u^c(t) \in \{0, 1, \dots, [R/d_{LoS}^c]\}$ — число всех одноадресных соединений в ближней области (всех возможных: и LoS, и nLoS) в момент t ; $u_{nLoS}^c(t) \in \{0, 1, \dots, [R/d_{nLoS}^c]\}$ — число одноадресных соединений в условиях nLoS в ближней области в момент t ; $m^r(t) \in \{0, b^r\}$ — объем ресурса, выделенный БС для заявок на установление многоадресного соединения в дальней области в момент t (возможны только условия LoS, поскольку соединений в условиях отсутствия прямой видимости (nLoS) в дальней области нет); $u^r(t) \in \{0, 1, \dots, [R/d^r]\}$ — число одноадресных соединений в дальней области в состоянии nLoS в момент t (возможны только условия LoS, поскольку соединений в условиях отсутствия прямой видимости (nLoS) в дальней области нет).

Тогда пространство состояний примет вид:

$$\begin{aligned} \mathbb{X} = & \left\{ (m^c, u^c, u_{nLoS}^c, m^r, u^r) : m^c \in \{0, b_{LoS}^c, b_{nLoS}^c\}, u^c \in \left\{ 0, 1, \dots, \left[\frac{R}{d_{LoS}^c} \right] \right\}, \right. \\ & u_{nLoS}^c \in \left\{ 0, 1, \dots, \left[\frac{R}{d_{nLoS}^c} \right] \right\}, m^r \in \{0, b^r\}, u^r \in \left\{ 0, 1, \dots, \left[\frac{R}{d^r} \right] \right\} : \\ & \left. u^c \geq u_{nLoS}^c, m^c + u^c d_{LoS}^c + u_{nLoS}^c \Delta d^c + m^r + u^r d^r \leq R \right\}. \end{aligned}$$

В стационарном режиме заслуживающие интереса метрики могут быть выражены через стационарные вероятности процесса $\mathbf{X}(t)$, представляющего собой решение системы уравнений равновесия. Стандартный подход позволяет получить метрики суммированием стационарных вероятностей по соответствующему подпространству состояний модели, как это сделано, например, в [13]. Для иллюстрации в следующем разделе приведен пример численного эксперимента по анализу вероятности потери одно- и многоадресных соединений в ближней и дальней областях и во всей системе в целом, а также коэффициента использования ресурса отдельными типами трафика и всей системой в целом.

4 Численный анализ

В данном разделе проводится численный анализ влияния системных параметров, представленных в таблице [8, 14], на исследуемые характеристики.

Увеличение интенсивности поступления одноадресных заявок в ближнюю область λ_U^c ведет к повышению вероятности потери заявок всех четырех потоков (рис. 3). Важно заметить, что потери заявок, интенсивность поступления которых не увеличивается (B_U^c , B_M^c и B_M^r), связаны преимущественно с блокировкой их доступа (B_U^{c-arr} , B_M^{c-arr} и B_M^{r-arr}). Отметим, что ресурс занят преимущественно одноадресными заявками из ближней области ($\overline{R_U^c}$), т. е. заявками, интенсивность поступления которых возрастает (рис. 4). При $\lambda_U^c > 40$ одноадресные заявки из ближней области начинают занимать почти весь ресурс сети, полностью вытесняя заявки других потоков. В условиях отсутствия заявок исключается возможность их прерывания, поэтому прерывание вносит незначительный вклад в потери много- и одноадресных заявок в ближней области.

Вероятность потери самих одноадресных заявок из ближней области (рис. 5, B_U^c) зависит в большей степени от вероятности прерывания (B_U^{c-pr}). Прерывание

Системные параметры

Обозначение	Описание	Значения по умолчанию
R	Емкость системы	1024 Мбит/с
d_{LoS}^c d_{nLoS}^c d^r	Требование к ресурсу для одноадресного соединения в состоянии LoS, nLoS в ближней и дальней областях	30 Мбит/с 45 Мбит/с 60 Мбит/с
b_{LoS}^c b_{nLoS}^c b^r	Требование к ресурсу для многоадресного соединения в состоянии LoS, nLoS в ближней и дальней областях	40 Мбит/с 60 Мбит/с 80 Мбит/с
μ_U^c μ_U^r μ_M^c μ_M^r	Интенсивность обслуживания одно- и многоадресных заявок в ближней и дальней областях	$2,0 \text{ с}^{-1}$ $1,5 \text{ с}^{-1}$ $1,5 \text{ с}^{-1}$ $1,0 \text{ с}^{-1}$
$\sigma_{in}^c = \sigma_{in}^r = \theta_{in}^c = \theta_{in}^r$	Интенсивность появления блокатора LoS для одно- и многоадресных соединений в ближней и дальней областях	4 с^{-1}
$\sigma_{out}^c = \theta_{out}^c$	Интенсивность ухода блокатора LoS для одно- и многоадресных соединений в ближней области	$1,25 \text{ с}^{-1}$
λ_U^c λ_M^c λ_U^r λ_M^r	Интенсивность поступления заявок на одно- и многоадресные соединения в ближней и дальней областях	$[1, \dots, 20]$ заявок/с 12 заявок/с 4 заявки/с 10 заявок/с

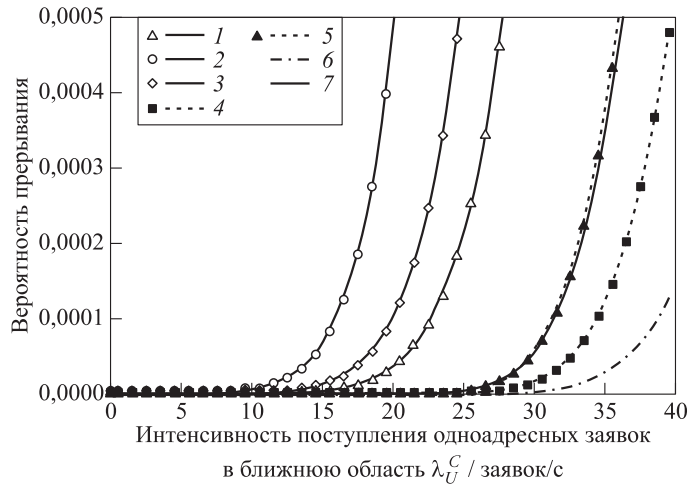


Рис. 3 Вероятности прерывания в зависимости от интенсивности поступления одноадресных заявок в ближнюю область: 1 — B_M^{r-pr} ; 2 — B_U^{r-pr} ; 3 — B_U^{c-pr} ; 4 — B_U^{r-prM} ; 5 — B_U^{c-prM} ; 6 — $B_U^{c-prM,LoS}$; 7 — $B_U^{c-prM,nLoS}$

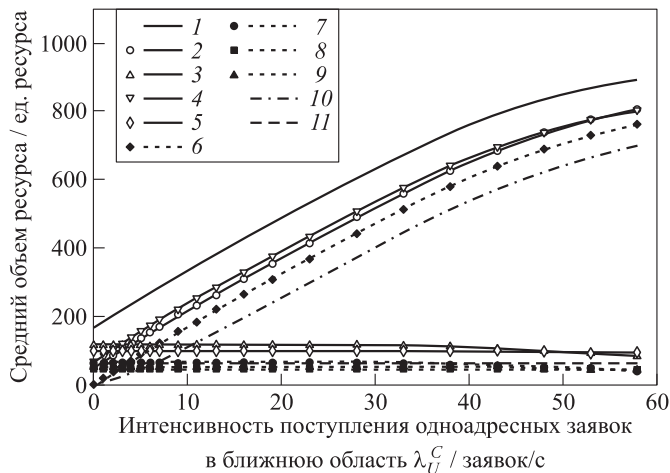


Рис. 4 Средний объем занимаемого ресурса в зависимости от интенсивности поступления одноадресных заявок в ближнюю область: 1 — \bar{R} ; 2 — \bar{R}^c ; 3 — \bar{R}^r ; 4 — \bar{R}_U ; 5 — \bar{R}_M ; 6 — \bar{R}_U^c ; 7 — \bar{R}_U^r ; 8 — \bar{R}_M^c ; 9 — \bar{R}_M^r ; 10 — $\bar{R}_{U^c,LoS}$; 11 — $\bar{R}_{U^c,nLoS}$

преимущественно происходит для одноадресных заявок в ближней области при появлении блокиатора (B_U^{c-prU}) в состоянии, когда свободного ресурса в системе недостаточно для поддержания из-за высокой загрузки системы.

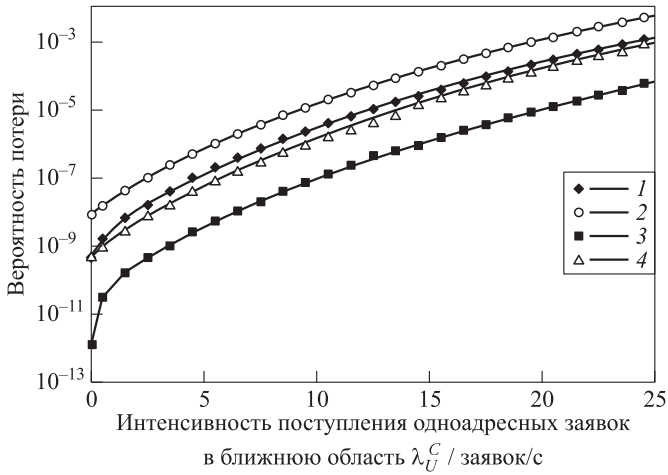


Рис. 5 Вероятности потери в зависимости от интенсивности поступления одноадресных заявок в ближнюю область: 1 — B_U^c ; 2 — B_U^r ; 3 — B_M^c ; 4 — B_M^r

Вероятности потери заявок иллюстрируют порядок прерывания заявок трех потоков при росте нагрузки одноадресными заявками в ближней области. Поскольку для рассмотренного набора исходных данных $d^r > d_{nLoS}^c > d_{LoS}^c$, то первым при росте нагрузки прерывается обслуживание одноадресных заявок в дальней области (B_U^{r-pr}). Вторыми кандидатами на прерывание становятся одноадресные заявки в ближней области в условиях nLoS (B_U^{c-pr}), а затем, когда одноадресных заявок из дальней и ближней областей в условиях nLoS станет недостаточно для поддержания обслуживания многоадресной заявки, подключатся прерывания одноадресных заявок из ближней области в состоянии LoS.

Заметим, что при исследованном наборе значений параметров системы (см. таблицу) и интенсивности поступления заявок на одноадресные соединения в ближнюю область менее 20 заявок/с значения вероятностей потери заявок, определяющие вероятности отказов в предоставлении услуги, удовлетворяют требованиям стандартов для услуг 4К видео по запросу (одноадресная передача) и прямой трансляция в формате 4К (мультивещание) [14], а именно: вероятности потери не более 10^{-3} . Это свидетельствует о том, что предложенный механизм приоритизации позволяет обеспечить требуемое качество предоставления услуги.

5 Заключение

Результаты численного анализа показали, что высокая нагрузка запросами на представление услуги по принципу многоадресной доставки данных ведет к невному резервированию ресурса многоадресными соединениями в системе,

сопровождающемуся ростом вероятности блокировки доступа запросов одноадресной доставки данных вследствие уменьшения свободного ресурса сети. При этом исследуемый механизм приоритизации, определяющий порядок выбора заявок для прерывания обслуживания, приводит к вытеснению менее приоритетных одноадресных заявок вне зависимости от условий прямой видимости между БС и ПУ. Однако в диапазоне нагрузок, полученном в результате исследования, значения вероятностей потери заявок, определяющие вероятности отказов в предоставлении услуги, удовлетворяют требованиям стандартов, что свидетельствует о требуемом качестве предоставления услуги.

Литература

1. *Bang J., Kim J.H.* Predicting power density of array antenna in mmWave applications with deep learning // *IEEE Access*, 2021. Vol. 9. P. 111030–111038. doi: 10.1109/ACCESS.2021.3102825.
2. Cisco Visual Networking Index: Global mobile data traffic forecast update. — San Jose, CA, USA: Cisco, 2019.
3. *Ruiz C. G., Pascual-Iserte A., Muñoz O.* Analysis of blocking in mmWave cellular systems: Application to relay positioning // *IEEE T. Commun.*, 2021. Vol. 69. Iss. 2. P. 1329–1342. doi: 10.1109/TCOMM.2020.3038177.
4. Telecommunication management; Study on management and orchestration of network slicing for next generation network. 3GPP TR 28.801. Version 15.1. 2018.
5. *Moltchanov D., Sopin E., Begishev V., Samuylov A., Koucheryavy Y., Samouylov K.* A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems // *IEEE Commun. Surv. Tut.*, 2022. Vol. 24. Iss. 2. P. 1072–1116. doi: 10.1109/COMST.2022.3156207.
6. *Власкина А. С., Бурцева С. А., Кочеткова И. А., Шоргин С. Я.* Управляемая система массового обслуживания с эластичным трафиком и сигналами для анализа нарезки ресурсов в сети радиодоступа // *Информатика и её применения*, 2022. Т. 16. Вып. 3. С. 90–96. doi: 10.14357/19922264220312. EDN: MFWTPO.
7. *Adou Y., Markova E., Gaidamaka Y.* Modeling and analyzing preemption-based service prioritization in 5G networks slicing framework // *Future Internet*, 2022. Vol. 14. Iss. 10. Art. 299. 16 p. doi: 10.3390/fi14100299.
8. *Samuylov A., Moltchanov D., Krupko A., Kovalchukov R., Moskaleva F., Gaidamaka Yu.* Performance analysis of mixture of unicast and multicast sessions in 5G NR systems // *10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. — Piscataway, NJ, USA: IEEE, 2018. P. 287–293. doi: 10.1109/ICUMT.2018.8631230.
9. *Москалева Ф. А., Бесчастный В. А., Самуйлов А. К., Молчанов Д. А., Гайдамака Ю. В.* Построение модели обслуживания двух типов трафика точкой доступа в диапазоне mmWave // *Сб. трудов XIII Всеросс. совещания по проблемам управления*. — М.: ИПУ им. В. А. Трапезникова РАН, 2019. С. 2985–2991. doi: 10.25728/vspu.2019.2985. EDN: DTBIMZ.
10. *Бесчастный В. А., Гайдамака Ю. В.* Модель обслуживания трафика одноадресных и многоадресных соединений высокочастотной сети 5G // *Современные ин-*

формационные технологии и ИТ-образование, 2019. Т. 15. № 2. С. 263–273. doi: 10.25559/SITITO.15.201902.263-273. EDN: BWKUBC.

11. *Samuylov A., Moltchanov D., Kovalchukov R., Pirmagomedov R., Gaidamaka Y., Andreev S., Koucheryavy Y., Samouylov K.* Characterizing resource allocation trade-offs in 5G NR serving multicast and unicast traffic // *IEEE T. Wirel. Commun.*, 2020. Vol. 19. Iss. 5. P. 3421–3434. doi: 10.1109/TWC.2020.2973375.
12. *Румянцева А. С., Москалева Ф. А., Гайдамака Ю. В.* Математическая модель обслуживания двух типов трафика точкой доступа в диапазоне mmWave // *Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем: Мат-лы Всеросс. конф. с международным участием.* — М.: РУДН, 2022. С. 55–60. EDN: VFTDZQ.
13. *Naumov V., Gaidamaka Yu., Yarkina N., Samouylov K.* Matrix and analytical methods for performance analysis of telecommunication systems. — Cham: Springer Nature, 2021. 308 p. doi: 10.1007/978-3-030-83132-5.
14. *Schoolar D., Lambert P., Nanbin W., Liang Z.* 5G service experience-based network planning criteria. — Ovum, 2019. <https://carrier.huawei.com/~media/cnbgv2/download/products/services/5g-planning-criteria-white-paper.pdf>.

Поступила в редакцию 20.02.24

ON MODELING SHARED SERVICE OF UNICAST AND MULTICAST TRAFFIC WITH PRIORITIZATION

*A. S. Rumyantseva¹, F. A. Moskaleva¹, A. K. Samouylov¹, S. Ya. Shorgin²,
and Yu. V. Gaidamaka^{1,2}*

¹RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

Abstract: The increase in the number of users and high network requirements have led to the need to introduce millimeter wave data transmission technology (mmWave) in order to increase data transfer speeds and minimize delays. The transition to the millimeter wave band in the fifth generation of communication networks allows the specifics of the implementation of the provision of services, including the sensitivity of directional antenna beams to radio signal blocking. In addition, when effectively planning a network providing unicast and multicast services, it is necessary to achieve a balance between opposite solutions for these two types of traffic. In this article, using the methods of the queuing theory, a mathematical model has been developed to describe the prioritization mechanism for servicing a combination of unicast and multicast traffic, for each of which the possibility of line-of-sight (LoS) blockers is regarded. The features of signal

propagation are taken into account in the model by dividing the coverage area into two areas. The constructed mathematical model allows one to find analytical expressions for the probabilities of loss of unicast and multicast connections as well as the amount of occupied resource both by individual types of traffic in different areas of the coverage area and in the entire system as a whole. The method of distribution of the radio resource of the network base station is based on the balance between the listed main performance indicators which allows one to monitor compliance with quality requirements for both types of services.

Keywords: mmWave; unicast connections; multicast connections; LoS; LoS blocking; mathematical model

DOI: 10.14357/08696527240203

EDN: QXWAUL

Acknowledgments

The reported study was funded by the Russian Science Foundation, project No. 22-79-10053 (<https://rscf.ru/project/22-79-10053>).

References

1. Bang, J., and J. H. Kim. 2021. Predicting power density of array antenna in mmWave applications with deep learning. *IEEE Access* 9:111030–111038. doi: 10.1109/ACCESS.2021.3102825.
2. Cisco Visual Networking Index: Global mobile data traffic forecast update. 2019. San Jose, USA.
3. Ruiz, C. G., A. Pascual-Iserte, and O. Muñoz. 2021. Analysis of blocking in mmWave cellular systems: Application to relay positioning. *IEEE T. Commun.* 69(2):1329–1342. doi: 10.1109/TCOMM.2020.3038177.
4. 3GPP TR 28.801. 2018. Telecommunication management; Study on management and orchestration of network slicing for next generation network. Version 15.1.0.
5. Moltchanov, D., E. Sopin, V. Begishev, A. Samuylov, Y. Koucheryavy, and K. Samouylov. 2022. A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems. *IEEE Commun. Surv. Tut.* 24(2):1072–1116. doi: 10.1109/COMST.2022.3156207.
6. Vlaskina, A. S., S. A. Burtseva, I. A. Kochetkova, and S. Ya. Shorgin. 2022. Up-ravlyaemaya sistema massovogo obsluzhivaniya s elastichnym trafikom i signalami dlya analiza narezki resursov v seti radiodostupa [Controllable queuing system with elastic traffic and signals for analyzing network slicing]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3):90–96. doi: 10.14357/19922264220312. EDN: MFWTPO.
7. Adou, Y., E. Markova, and Y. Gaidamaka. 2022. Modeling and analyzing preemption-based service prioritization in 5G networks slicing framework. *Future Internet* 14(10):299. 16 p. doi: 10.3390/fi14100299.
8. Samuylov, A., D. Moltchanov, A. Krupko, R. Kovalchukov, F. Moskaleva, and Yu. Gaidamaka. 2018. Performance analysis of mixture of unicast and multicast sessions in 5G NR systems. *10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 287–293. doi: 10.1109/ICUMT.2018.8631230.

9. Moskaleva, F. A., V. A. Beschastny, A. K. Samuilov, D. A. Molchanov, and Yu. V. Gaydamaka. 2019. Postroenie modeli obsluzhivaniya dvukh tipov trafika tochkoy dostupa v diapazone mmWave [Building a model for servicing two types of traffic with an access point in the mmWave range]. *Sbornik trudov XIII Vseross. soveshchaniya po problemam upravleniya* [13th All-Russian Meeting on Management Problems Proceedings]. Moscow: V. A. Trapeznikov Institute of Management Problems of the Russian Academy of Sciences. 2985–2991. doi: 10.25728/vspu.2019.2985. EDN: DTBIMZ.
10. Beschastnyy, V. A., and Yu. V. Gaydamaka. 2019. Model' obsluzhivaniya trafika odnoadresnykh i mnogoadresnykh soedineniy vysokochastotnoy seti 5G [Modeling mixture of unicast and multicast communications in 5G high frequency networks]. *Sovremennyye informatsionnye tekhnologii i IT-obrazovanie* [Modern Information Technology and IT-education] 15(2):263–273. doi: 10.25559/SITITO.15.201902.263-273. EDN: BWKUBC.
11. Samuylov, A., D. Moltchanov, R. Kovalchukov, R. Pirmagomedov, Y. Gaidamaka, S. Andreev, Y. Koucheryavy, and K. Samouylov. 2020. Characterizing resource allocation trade-offs in 5G NR serving multicast and unicast traffic. *IEEE T. Wirel. Commun.* 19(5):3421–3434. doi: 10.1109/TWC.2020.2973375.
12. Rumyantseva, A. S., F. A. Moskaleva, and Yu. V. Gaydamaka. 2022. Matematicheskaya model' obsluzhivaniya dvukh tipov trafika tochkoy dostupa v diapazone mmWave [Mathematical model of servicing two traffic types by mmWave access point]. *Informatsionno-telekommunikatsionnye tekhnologii i matematicheskoe modelirovanie vysokotekhnologichnykh sistem: Mat-ly Vseross. konf. s mezhdunarodnym uchastiem* [Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems: All-Russian Conference with International Participation Proceedings]. Moscow: RUDN. 55–60. EDN: VFTDZQ.
13. Naumov, V., Y. Gaidamaka, N. Yarkina, and K. Samouylov. 2021. *Matrix and analytical methods for performance analysis of telecommunication systems*. Cham: Springer Nature. 308 p. doi: 10.1007/978-3-030-83132-5.
14. Scholar, D., P. Lambert, W. Nanbin, and Z. Liang. 2019. 5G service experience-based network planning criteria. Ovum. Available at: <https://carrier.huawei.com/~media/cnbgv2/download/products/services/5g-planning-criteria-white-paper.pdf> (accessed April 2, 2024).

Received February 20, 2024

Contributors

Rumyantseva Aleksandra S. (b. 2001) — student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; 1132223493@rudn.ru

Moskaleva Faina A. (b. 1996) — PhD student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; moskaleva-fa@rudn.ru

Samouylov Andrey K. (b. 1988) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Department of Probability Theory and Cyber

Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; samuylov-ak@rudn.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gaidamaka-yuv@rudn.ru

СТАТИСТИЧЕСКИЙ КРИТЕРИЙ СТАБИЛЬНОСТИ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ, ОСНОВАННЫЙ НА ПОСЛЕДОВАТЕЛЬНОСТИ ВРЕМЕН ПРЕБЫВАНИЯ

*М. П. Кривенко*¹

Аннотация: Рассматривается задача статистической проверки стабильности систем массового обслуживания (СМО) на основе характеристик последовательности значений времени пребывания заявок в системе. Для ее решения предлагается использовать критерий единичного корня. Устанавливается связь между количественными и временными характеристиками системы в стиле формулы Литтла для нестационарного случая, что позволяет сопоставить gate-стабильность со стационарностью последовательности времен пребывания. В качестве критериев стационарности рассматриваются три базовые модели и соответствующие тесты Дики–Фуллера. Обсуждаются ограничения применяемых статистических таблиц. Проведенное имитационное исследование позволяет сделать выводы: все три теста решают поставленную задачу анализа стабильности с явной потерей качества принимаемых решений в случае модели с детерминированным линейным трендом, в области стабильности систем критерии ведут себя хуже, чем в области нестабильности; явно проявляются ожидаемые тенденции применения тестов; некоторые отклонения от ожидаемых результатов говорят о необходимости расширять спектр применяемых моделей и снимать ограничения используемых статистических таблиц.

Ключевые слова: система массового обслуживания; анализ временных рядов; тесты единичного корня; тесты Дики–Фуллера; имитационное моделирование

DOI: 10.14357/08696527240204

EDN: RVWXXQ

1 Введение

Функционирование СМО предполагает, что через ее вход и выход соответственно прибывают и убывают заявки. Для каждого неотрицательного времени t пусть $A(t)$ обозначает число прибывших заявок (входной процесс), $D(t)$ — покинувших систему (выходной процесс). Тогда общее число $N(t)$ заявок в системе в момент t равно

$$N(t) = N(0) + A(t) - D(t).$$

Данная общая модель входного–выходного процесса $N(t)$ подразумевает следующие свойства: $A(t)$ и $D(t)$ суть неубывающие непрерывные справа процессы, $N(t) \geq 0$ для всех $t \geq 0$.

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

Для времени прибытия и убытия k -й заявки, $k = 1, 2, \dots$, введем самостоятельные обозначения A_k и D_k . Тогда

$$A(t) = \# \{k : A_k \leq t\}; \quad D(t) = \# \{k : D_k \leq t\}.$$

Также определим: $N(t) = \# \{k : A_k \leq t < D_k\} = A(t) - D(t)$ — число заявок в системе; $V_k = D_k - A_k$ — время пребывания k -й заявки в системе.

Пусть $1\{E\}$ — индикатор события E . Тогда

$$N(t) = \sum_{k=1}^{\infty} 1\{A_k \leq t < D_k\};$$

$$V_k = \int_0^{\infty} 1\{A_k \leq t < D_k\} dt.$$

2 Связь определений стабильности

Ранее стабильность системы со ссылками на основополагающие работы была определена в [1] через поведение входного–выходного процесса $N(t)$ или, что эквивалентно, на основе входного и выходного процессов $A(t)$ и $D(t)$. Если требуется сделать это с помощью временного ряда $\{V_k, k = 1, 2, \dots\}$, то необходимо установить связь между количественными и временными характеристиками СМО, например в стиле формулировок закона Литтла, но при отказе от условия стационарности.

Заменим равенство Литтла на ограничения снизу и сверху для $\int_0^t L(s) ds$ на основе представления

$$\int_0^t N(s) ds = \int_0^t \sum_{k=1}^{\infty} 1\{A_k \leq s < D_k\} ds.$$

Сначала после отбрасывания части слагаемых имеем:

$$\int_0^t \sum_{k=1}^{\infty} 1\{A_k \leq s < D_k\} ds \geq \int_0^t \sum_{\{k:D_k < t\}} 1\{A_k \leq s < D_k\} ds =$$

$$= \sum_{\{k:D_k < t\}} \int_0^t 1\{A_k \leq s < D_k\} ds = \sum_{\{k:D_k < t\}} \int_0^{\infty} 1\{A_k \leq s < D_k\} ds = \sum_{\{k:D_k < t\}} V_k.$$

Далее за счет расширения области интегрирования получаем:

$$\begin{aligned} & \int_0^t \sum_{k=1}^{\infty} 1 \{A_k \leq s < D_k\} ds = \\ & = \int_0^t \left[\sum_{\{k:A_k < t\}} 1 \{A_k \leq s < D_k\} + \sum_{\{k:A_k \geq t\}} 1 \{A_k \leq s < D_k\} \right] ds = \\ & = \int_0^t \sum_{\{k:A_k < t\}} 1 \{A_k \leq s < D_k\} ds \leq \int_0^{\infty} \sum_{\{k:A_k < t\}} 1 \{A_k \leq s < D_k\} ds = \sum_{\{k:A_k < t\}} V_k. \end{aligned}$$

Таким образом приходим к базовым соотношениям:

$$\sum_{\{k:D_k < t\}} V_k \leq \int_0^t N(s) ds \leq \sum_{\{k:A_k < t\}} V_k.$$

После деления всех частей полученных неравенств на t и использования представлений вида

$$\frac{\sum_{\{k:\dots\}} V_k}{t} = \frac{\#\{k:\dots\}}{t} \frac{\sum_{\{k:\dots\}} V_k}{\#\{k:\dots\}}$$

получаем элементы формулы Литтла и важные для статистического контроля стабильности выводы: для стабильного случая время пребывания заявки в системе в среднем постоянно, а для нестабильного — растет по крайней мере линейно.

Обратим внимание, что левая и правая части полученных базовых соотношений становятся равными в периоды простоя, т. е. для таких моментов времени t , когда $N(t) = 0$. Для стабильных систем наличие таковых возможно, но, строго говоря, ниоткуда не следует. Поэтому при использовании равенств, например для обоснования формулы Литтла для стабильных систем, в [2] пришлось дополнительно вводить условие существования периодов простоя.

С точки зрения статистического контроля функционирования СМО важнее другое: для нестабильных систем $N(t)$ линейно растет при увеличивающемся t , поэтому время пребывания в системе в среднем также растет по крайней мере линейно, т. е. нестабильность на основе входного и выходного потоков порождает изменения характеристик последовательности времен пребывания в системе.

3 Тесты единичного корня анализа стационарности

Тематика «единичные корни» относится к использованию общей линейной модели (ОЛМ) при анализе временных рядов в первую очередь в области эконометрики, ее применение также отмечается [3] в отдельных отраслях геонауки,

в физиологии и медицинской диагностике, при социологических исследованиях, в ходе анализа веб-трафика. Вместе с тем публикации, посвященные тестированию единичного корня применительно к СМО, автору статьи найти не удалось и в первую очередь из-за малочисленности работ по статистическому контролю функционирования таких систем в целом.

Как оказывается, анализируемые последовательности времен пребывания обладают специфическими свойствами [1]: высокой степенью временной связанности и наличием или отсутствием стационарности. Эти свойства интересны как особенности временных рядов вообще и актуальны с точки зрения установления стабильности. Последнее приобретает первостепенное значение при реализации методологии Бокса–Дженкинса: если, например, в области эконометрики наличие тренда обычно становится мешающим фактором при решении задач прогнозирования, то при установлении стабильности СМО выявление роста среднего времени пребывания становится ключевым. При этом выбор адекватной модели должен не только приводить к непротиворечивым оценкам ее параметров и сводить к минимуму ошибочные решения в первую очередь при контроле стабильности, но и по возможности формировать основу решения иных задач анализа данных, как, например, прогноза.

Тестирование единичного корня предполагает, что исследуемый временной ряд $\{y_k, k = 1, 2, \dots\}$ может быть представлен как

$$y_k = d_k + s_k + \varepsilon_k,$$

где d_k — детерминированный компонент (постоянное смещение, функционально описанный тренд, сезонные изменения и т. п.); s_k — стохастический компонент (процессы авторегрессии — скользящего среднего общего и частного вида); ε_k — стационарный процесс ошибок. Задача состоит в том, чтобы определить, является s_k нестационарным или стационарным процессом.

Стационарность далее понимается в широком смысле, когда математическое ожидание случайного процесса постоянно, а корреляционная функция зависит только от разности аргументов. Интегрированный (разностно-стационарный) процесс $I(n)$ порядка n — это тот, для которого надо перейти к разностям n раз, чтобы он стал стационарным. Для обычным образом определенной модели авторегрессии — скользящего среднего и характеристического полинома $\varphi(z)$ ее авторегрессионной части $AR(p)$ — стационарность моделируемого процесса определяется свойствами корней $\varphi(z)$: они все должны быть вне единичного круга [4].

Процесс единичного корня определяется через $AR(1)$ при наличии единичного по модулю корня z_1 , при этом внимание сосредоточивается на $z_1 = 1$. Когда $z_1 = -1$, процесс будет осциллировать (знак меняется на противоположный при переходе от одного элемента временного ряда к другому), что не соответствует ни представлению о принципах функционирования СМО, ни имеющимся результатам наблюдений времен пребывания заданий. Для единичного корня временной

ряд принимает вид $y_k = y_{k-1} + \varepsilon_k$ и становится интегрированным процессом первого порядка.

Критерии единичного корня нацелены на нулевую гипотезу, состоящую в том, что наблюдаемый временной ряд — интегрированный процесс 1-го порядка $I(1)$. Базовая модель временного ряда представляет собой AR (1) с возможностью пополнения ее постоянной составляющей и линейным трендом и приводит к тестам Дики–Фуллера (Dickey–Fuller, DF). Расширенный вариант DF-тестов использует модель авторегрессии — скользящего среднего с компонентами разностей, после «удаления» которых во вспомогательной регрессии он может использовать критические значения DF-теста. Конкурирующая гипотеза ориентирована на предположение о стационарности.

Перечисленные критерии не исчерпывают все существующие варианты тестов единичного корня [5], их сравнительный анализ на основе трех основных аспектов — мощности теста, его обоснованности и простоты использования был проведен в [6]. Подобное имитационное комплексное исследование относится к редким и достаточно узко направленным (в частности, рассмотрение только модели AR (1) и только с положительным значением параметра), но оно может стать отправной точкой при выборе критерия стабильности СМО на основе единичного корня в пользу DF-тестов.

Базовый вариант DF-тестов основывается на представлении

$$y_k = \rho y_{k-1} + \varepsilon_k, \quad k = 1, 2, \dots, \quad y_0 = 0. \quad (1)$$

Нулевая гипотеза $H_0 : \rho = 1$, т. е. наблюдаемый процесс — $I(1)$ и поэтому нестационарный. Конкурирующая гипотеза $H_1 : |\rho| < 1$, т. е. речь идет о $I(0)$ и стационарности.

Модель (1) может быть расширена путем добавления константы

$$y_k = \mu + \rho y_{k-1} + \varepsilon_k. \quad (2)$$

или линейного тренда

$$y_k = \mu + \beta k + \rho y_{k-1} + \varepsilon_k. \quad (3)$$

Тестовая статистика имеет вид $(\hat{\rho} - 1)/SE(\hat{\rho})$, где $\hat{\rho}$ — оценка наименьших квадратов коэффициента ρ ; $SE(\hat{\rho})$ — ее стандартная ошибка для регрессионной модели; p -значения табулированы методом статистических испытаний для стандартных процентилей.

Принимая во внимание поведение $\{V_k\}$ для стабильной и нестабильной системы, рассмотрим подробнее модели (2) и (3).

Модель (2) при предположениях $\mu \neq 0$ и $\rho = 1$ можно переписать как

$$y_k = y_0 + \mu k + E_k, \quad (4)$$

где $E_k = \sum_{i=1}^k \varepsilon_i$. Тогда для нестабильной системы асимптотически получаем процесс с доминирующим линейным ростом, так как по вероятности $E_k = O(k^{1/2})$. Отсюда видно, что по крайней мере общий характер поведения среднего $\{V_k\}$ получает в этой модели свое отражение. В случае стабильной системы, для которой $\rho \in [0, 1)$, по аналогии с (4) получаем

$$y_k = \rho^k y_0 + \mu \sum_{i=0}^k \rho^i + \sum_{i=0}^k \rho^i \varepsilon_{k-i}.$$

Это приводит к постоянству среднего для $\{V_k\}$, которое за счет выбора y_0 и μ может стать нулевым, т. е. опять же получаем нужное поведение исследуемого временного ряда.

Естественной альтернативной моделью для $\{V_k\}$ может стать комбинация линейного тренда и стационарного временного ряда:

$$y_k = \beta_0 + \beta_1 k + x_k, \quad (5)$$

где $x_k = \rho x_{k-1} + \varepsilon_k$. После подстановки авторегрессии в (5) получаем представление вида (3), а именно:

$$y_k = \gamma_0 + \gamma_1 k + \rho y_{k-1} + \varepsilon_k,$$

где

$$\gamma_0 = \beta_0(1 - \rho) + \rho\beta_1; \quad \gamma_1 = \beta_1(1 - \rho).$$

Модель (5) включает детерминированный линейный временной тренд, не связанный со значением ρ . Если $\rho = 1$, то $\gamma_0 = \beta_1$, $\gamma_1 = 0$ и $y_k = \beta_1 + y_{k-1} + \varepsilon_k$, т. е. тренд (5) с наложенной на него авторегрессией фактически дает AR (1) с константой.

4 Таблицы тестов

Набор таблиц встречается в ряде изданий, но не всегда в одинаковом виде и с достаточными обоснованиями. Подробное описание построения таблиц с оценкой погрешности приведено в [7], вместе с отдельными уточнениями их можно найти в приложении А [8]. Более важным, чем незначительные отклонения в табличных данных, оказывается то, что обычно не приводятся условия, при которых они получены.

Так, для модели (2) предельное распределение статистики табулировано в предположении, что $\mu = 0$, а для (3) — $\beta = 0$ (значение μ в модели (3) на распределение не влияет). Если $\mu \neq 0$ для (2) или $\beta \neq 0$ для (3), то предельные распределения статистик критерия становятся нормальными с соответствующими моментами [7], что объясняет следующее: если приняты описанные

модели (2) и (3), а для проверки гипотезы $\rho = 1$ используются соответствующие статистики, то гипотеза будет принята с вероятностью, большей, чем та, которая соответствует нулевым значениям μ и β . Эту специфику использования таблиц необходимо принимать во внимание при верификации соответствующего программного обеспечения, а также в ходе сравнительного анализа мощности критериев значимости.

Следует обратить внимание еще на одну деталь применения критериев на практике. В приведенных рассуждениях начальное значение фиксировано и обычно $y_0 = 0$. Но если предельные распределения статистик критериев не зависят от y_0 , то для малых объемов данных это не так. Актуальным это может стать при разделении наблюдаемых процессов на части с последующим объединением результатов их статистического анализа (подобно тому, как вводятся в рассмотрение фазы и фрагменты в [1]). Для того чтобы снизить последствия указанной зависимости, достаточно просто все элементы временного ряда из фрагмента сдвинуть по значению на величину начального (первого) элемента, в результате чего оно станет нулевым. Не надо забывать, что цель состоит в контроле стационарности временного ряда $\{V_k\}$, а одинаковый сдвиг всех наблюдаемых значений в этом смысле не влияет на результаты.

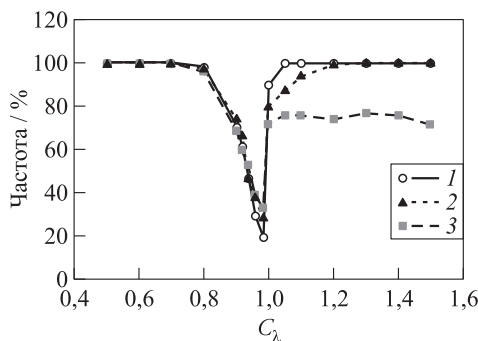
И, наконец, предельные распределения статистик справедливы для ошибок, которые представляют собой независимые и одинаково распределенные (необязательно нормально) случайные величины с нулевым средним и некоторой дисперсией [9].

5 Мощность тестов единичного корня

В условиях многовариантности тестов единичного корня и отсутствия практики их применения для статистического контроля функционирования СМО становится актуальным сравнение предлагаемых подходов и решений. При этом надо помнить, что при решении задачи выявления нестабильности попутно реализуется фундаментальный этап построения модели протекающих процессов.

Сосредоточив внимание на трех описанных моделях для $\{V_k\}$, продолжим рассмотрение задачи контроля стабильной работы двухпроцессорной системы обработки заданий со случайным выбором числа требуемых процессоров [1]. Функционирование подобной системы $M/M/2$ определяется параметрами: λ — интенсивность поступления заданий на обработку; p_1 — вероятность того, что для выполнения задания требуется один процессор; μ — среднее время обработки задания процессором/процессорами. Условие стабильности функционирования этой системы принимает вид $\lambda/\mu < 2/(2 - p_1^2)$, с помощью которого будет осуществляться выбор моделируемых систем: параметры p_1 и μ принимаются базовыми со значениями $p_1 = 0,5$ и $\mu = 1$, а λ выбирается как $C_\lambda \cdot 2\mu/(2 - p_1^2)$, где C_λ — индикатор стабильности работы системы (при $C_\lambda < 1$ система стабильна). Имитационное исследование проводилось для фрагментов ряда $\{V_k\}$, начинающихся с $k \approx 5000$ (чтобы исключить влияние начального этапа функционирования

65mm



Зависимость частоты правильных решений DF-тестов для СМО с различными значениями индикатора стабильности C_λ : 1 — модель (1); 2 — модель (2); 3 — модель (3)

СМО) и имеющих размер из набора $n = 100, 200, 400, 800$ (чтобы проследить влияние увеличения объема анализируемых данных), и включало 1000 экспериментов. Уровень значимости α для DF-тестов принимался равным 5% или 10% (чтобы выяснить степень важности использования определенного значения α при принятии решений). Иллюстрацией характера полученных результатов может служить рисунок, на котором для $n = 800$ и $\alpha = 10\%$ приведены графики частоты правильных решений DF-тестов для различных вариантов системы.

Из результатов моделирования можно сделать общие выводы:

- все три критерия решают поставленную задачу анализа стабильности с явной потерей качества принимаемых решений в случае модели (3);
- в области стабильности ($C_\lambda < 1$) критерии ведут себя хуже, чем в области нестабильности, рост ошибок наблюдается при приближении к границе $C_\lambda = 1$;
- явно проявляются ожидаемые тенденции применения тестов, а именно: рост правильности принимаемых решений с увеличением объема анализируемых данных; рост частоты принятия гипотезы о нестабильности с уменьшением уровня значимости (что приводит к росту ошибок при анализе стабильных систем);
- некоторые отклонения от ожидаемых результатов (например, конкретные значения α иногда не отвечают результатам моделирования) говорят о неполной адекватности принимаемых моделей данных и ограниченных возможностях таблиц DF-тестов.

6 Заключение

Установленная связь между количественными и временными характеристиками СМО для нестационарного случая позволила сопоставить gate-стабильность со стационарностью ряда времен пребывания заявок в системе. Это позволяет расширить набор статистических процедур контроля стабильности за счет критериев стационарности типа тестов Дики–Фуллера. Результаты экспериментов показали, что они остаются надежным вариантом для тестирования единичного корня, особенно в случае временных рядов с большим числом наблюдений. Также они остаются одними из наиболее часто используемых тестов, поскольку их решающее преимущество заключается в простоте конструкции и осуществимости.

Таким образом, проведенное исследование создает предпосылки для существенного расширения в будущем как спектра привлекаемых моделей, так и собственно методов анализа данных.

Литература

1. *Кривенко М. П.* Статистический критерий стабильности системы массового обслуживания, основанный на входном и выходном потоках // Информатика и её применения, 2024. Т. 18. Вып. 1. С. 54–60. doi: 10.14357/19922264240108. EDN: JNJJMU.
2. *El-Taha M., Stidham S., Jr.* Sample-path analysis of queueing systems. — New York, NY, USA: Springer, 1999. 302 p. doi: 10.1007/978-1-4615-5721-0.
3. *Herranz E.* Unit root tests // Wiley Interdisciplinary Reviews Computational Statistics, 2017. Vol. 9. Iss. 3. Art. e1396. 20 p. doi: 10.1002/wics.1396.
4. *Chan N. H.* Time series: Applications to finance. — New York, NY, USA: Wiley, 2002. 225 p.
5. *Kocenda E., Cerný A.* Elements of time series econometrics: An applied approach. — Prague, Czech Republic: Charles University, Karolinum Press, 2007. 228 p.
6. *Arltova M., Fedorova D.* Selection of unit root test on the basis of length of the time series and value of AR (1) parameter // Statistika Statistics Economy J., 2016. Vol. 96. No. 3. P. 47–64.
7. *Dickey D. A.* Estimation and hypothesis testing in nonstationary time series: PhD Diss. — Ames, IA, USA: Iowa State University Digital Repository, 1976. 128 p. doi: 10.31274/rtd-180813-2848.
8. *Fuller W. A.* Introduction to statistical time series. — 2nd ed. — Hoboken, NJ, USA: Wiley, 1996. 728 p. doi: 10.1002/9780470316917.
9. *Hasza D. P.* Estimation in nonstationary time series: PhD Diss. — Ames, IA, USA: Iowa State University Digital Repository, 1977. 130 p. doi: 10.31274/rtd-180813-5232.

Поступила в редакцию 15.03.24

STATISTICAL CRITERION FOR QUEUING SYSTEM STABILITY BASED ON SOJOURN TIMES

M. P. Krivenko

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

Abstract: The article discusses the problem of statistical test of queuing system stability based on the characteristics of the sojourn time series $\{V_k\}$. To solve it, it is proposed to use the tests for unit roots. A relationship is established between the quantitative and temporal characteristics of the system in the style of the Little’s formula for the nonstationary case which makes it possible to link the rate-stability with the stationarity $\{V_k\}$. Three basic models and the corresponding Dickey–Fuller tests are considered as the stationarity test. The limitations of the used statistical tables are discussed. The conducted simulation study allows one to draw the following conclusions: all three criteria solve the problem of stability analysis with a clear loss of the quality of decisions made in the case of a model with a deterministic linear trend, in the field of system stability, the test behaves worse than in the field of instability; the expected trends in the use of tests are clearly manifested; and some deviations from the expected results indicate the need to expand the range of the models used and to remove the limitations of the statistical tables used.

Keywords: queueing system; time series analysis; unit root tests; Dickey–Fuller tests; simulation

DOI: 10.14357/08696527240204

EDN: RVWXXQ

References

1. Krivenko, M. P. 2024. Statisticheskiy kriteriy stabil’nosti sistemy massovogo obsluzhivaniya, osnovanny na vkhodnom i vykhodnom potokakh [Statistical criterion for queuing system stability based on input and output flows]. *Informatika i ee Primeneniya — Inform. Appl.* 18(1):54–60. doi: 10.14357/19922264240108. EDN: JNJJMU.
2. El-Taha, M., and S. Stidham, Jr. 1999. *Sample-path analysis of queueing systems*. New York, NY: Springer. 302 p. doi: 10.1007/978-1-4615-5721-0.
3. Herranz, E. 2017. Unit root tests. *Wiley Interdisciplinary Reviews Computational Statistics* 9(3):e1396. 20 p. doi: 10.1002/wics.1396.
4. Chan, N. H. 2002. *Time series: Applications to finance*. New York, NY: Wiley. 225 p.
5. Kocenda, E., and A. Cerný. 2007. *Elements of time series econometrics: An applied approach*. Prague, Czech Republic: Charles University, Karolinum Press. 228 p.
6. Arltova, M., and D. Fedorova. 2016. Selection of unit root test on the basis of length of the time series and value of $AR(1)$ parameter. *Statistika Statistics Economy J.* 96(3):47–64.

7. Dickey, D. A. 1976. Estimation and hypothesis testing in nonstationary time series. Ames, IA: Iowa State University Digital Repository. PhD Diss. 128 p. doi: 10.31274/rtd-180813-2848.
8. Fuller, W. A. 1996. *Introduction to statistical time series*. 2nd ed. Hoboken, NJ: Wiley. 728 p. doi: 10.1002/9780470316917.
9. Hasza, D. P. 1977. Estimation in nonstationary time series. Ames, IA: Iowa State University Digital Repository. PhD Diss. 130 p. doi: 10.31274/rtd-180813-5232.

Received March 15, 2024

Contributor

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

ФОРМАЛИЗАЦИЯ СИНТЕЗА САМОСИНХРОННЫХ СЧЕТЧИКОВ

*Ю. А. Степченков¹, Ю. Г. Дьяченко², Н. В. Морозов³, Д. Ю. Степченков⁴,
Д. Ю. Дьяченко⁵*

Аннотация: Самосинхронные (СС) схемы обладают высокой надежностью. Они гарантируют обнаружение и локализацию любых константных неисправностей и демонстрируют высокий уровень сбоеустойчивости. Однако проектирование СС-схем более трудоемко в сравнении с синхронными аналогами из-за необходимости построения дополнительной индикаторной подсхемы и соблюдения принципов функционирования схем, относящихся к классу СС-схем. Автоматическое преобразование исходной схемы, описанной аппаратно как синхронная схема, в СС-реализацию обеспечивается с помощью процедуры формализованной десинхронизации. Но при синтезе последовательностных СС-устройств, в том числе СС-счетчиков, формальная десинхронизация приводит к чрезмерной аппаратной избыточности и, как следствие, к их низкому быстродействию. Статья обосновывает подход к синтезу СС-счетчиков, базирующийся на формализации эвристических методов их построения и обеспечивающий гарантированное получение действительно СС-реализации, функционирующей в полном соответствии с исходным описанием и обладающей близкими к оптимальным потребительскими характеристиками.

Ключевые слова: автоматизированный синтез; самосинхронная схема; счетчик; десинхронизация; предустановка; индикация

DOI: 10.14357/08696527240205

EDN: KDIEOJ

1 Введение

Успешное развитие современной экономики невозможно без широкого внедрения цифровой микроэлектроники во все области деятельности человека. Цифровизация жизнедеятельности человека порождает сильную зависимость экономики и общества от цифровой техники, от надежности, безотказности аппаратуры

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

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

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

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, stepchenkov@mail.ru

⁵Федеральный исследовательский центр «Информатика и управление» Российской академии наук, diaden87@gmail.com

на интегральных микросхемах. Современная микроэлектроника реализуется в основном на принципах синхронного управления процессами, базирующегося на глобальном «дереве» синхронизации [1]. С увеличением сложности микросхем и снижением норм топологического проектирования растет площадь кристаллов синхронных интегральных микросхем и, как следствие, объем аппаратных затрат и энергопотребления, приходящихся на тактовое «дерево». Разработчики синхронных схем вынуждены закладывать в проектируемые синхронные схемы «запас прочности» в расчете на наихудший случай и избыточность аппаратуры, обеспечивающий ее работоспособность в изменяющихся условиях эксплуатации и в процессе старения [2].

Альтернативой синхронным схемам служат асинхронные схемы, и в частности их подкласс — СС-схемы. В них нет глобальной синхронизации. Они функционируют на основе запрос-ответной дисциплины [3–5]. Самосинхронные схемы свободны от перечисленных недостатков синхронных аналогов, гарантируют надежную работу при любых предельных условиях эксплуатации и характеризуются повышенной устойчивостью к логическим сбоям [6, 7]. Тем не менее подавляющее большинство практических цифровых схем относятся к синхронному классу. Развитию СС-схем препятствует большая сложность и нестандартность процесса их ручной разработки и отсутствие программных средств, обеспечивающих автоматизацию всего маршрута проектирования СС-аппаратуры и при этом не требующих высокой квалификации от разработчика.

Современные системы автоматизированного проектирования (САПР) цифровых устройств, используемые в промышленных масштабах, обеспечивают эффективное проектирование схем синхронного типа. Они поддерживаются и развиваются зарубежными фирмами, например Cadence, Synopsys, Mentor Graphics и др. Использование синхронных САПР для разработки СС-схем приводит к неутешительным результатам: характеристики полученных СС-схем многократно уступают синхронным аналогам [8, 9]. Промышленных САПР СС-схем пока нет. Сейчас за рубежом есть несколько программных средств разработки асинхронных схем, например Balsa [10]. Они служат инструментом синтеза асинхронной схемы, используя в качестве исходного задания ее описание на поведенческом или алгоритмическом уровне. При этом они используют форму исходного описания синтезируемой схемы, специфичную именно для СС-схем: специальный язык, сигнальные графы и т. д. Такой подход затрудняет работу проектировщиков, но не гарантирует получение корректного с точки зрения СС-дисциплины результата [11].

Более простым и интуитивно понятным представляется подход к синтезу СС-схем, преобразующий исходное описание схемы, заданное в синхронном виде, в асинхронное с помощью метода десинхронизации [12–14]. С помощью анализа связности устройств в составе схемы, формальных правил замены синхронных элементов их СС-аналогами и замены глобальной синхронизации совокупностью локальных сигналов управления запрос-ответным взаимодействием этот метод позволяет достаточно просто получить асинхронную схему. Однако

синтезированные таким путем схемы используют для формирования сигналов управления элементы задержки вместо проверки и подтверждения завершения переключения схемы в текущую фазу работы (индицирования [15]) и из-за этого теряют ряд преимуществ полноценных СС-схем, в частности независимость работы от задержек элементов.

Отечественные программные средства проектирования СС-схем касаются лишь вопросов диагностирования [16] и анализа на самосинхронность [17]. Они обеспечивают проверку корректности СС-реализации синтезированной схемы, но не способны облегчить и ускорить ее разработку.

Кроме того, прямое формальное преобразование синхронного описания в СС-схему приводит к большой избыточности последовательностных СС-устройств, например двоичных счетчиков. Цель статьи — оценка возможности применения имеющихся программных средств синтеза цифровых устройств для синтеза двоичных СС-счетчиков и разработка формализованного метода синтеза СС-счетчиков, обеспечивающего получение корректных схем с приемлемыми характеристиками. Научная новизна описываемого подхода заключается в разработке алгоритма формализованного преобразования исходного Verilog-описания двоичного синхронного счетчика в СС-реализацию, функционирующую в полном соответствии с исходным описанием с учетом специфики работы СС-схем.

2 Автоматизация логического синтеза самосинхронных счетчиков

Формализованный синтез СС-счетчиков лежит в русле общей методологии автоматизации проектирования СС-схем [18]. Пусть исходное описание синтезируемой схемы представлено на языке Verilog на поведенческом уровне с использованием блоков `always`. Распознать синхронный счетчик в исходном Verilog-описании достаточно просто: в блоке `always`, описывающем счетчик, присутствует оператор инкремента или декремента векторной переменной по фронту (переднему или заднему) некоторого сигнала, чаще всего глобального тактового сигнала. Дополнительные операторы в теле блока `always`, описывающего счетчик, специфицируют его предустановку по внешнему сигналу разрешения или по заданному состоянию самого счетчика.

Рисунок 1 демонстрирует пример описания двоичного счетчика с асинхронным сбросом на языке Verilog. Состояние `State` инкрементируется при по-

```
always@(posedge T or posedge res) begin
  if(res) // Начальный сброс по условию res = 1
    State <= 4'h0;
  else // Инкрементация счетчика
    State <= State + 1;
end
```

Рис. 1 Verilog-описание двоичного счетчика

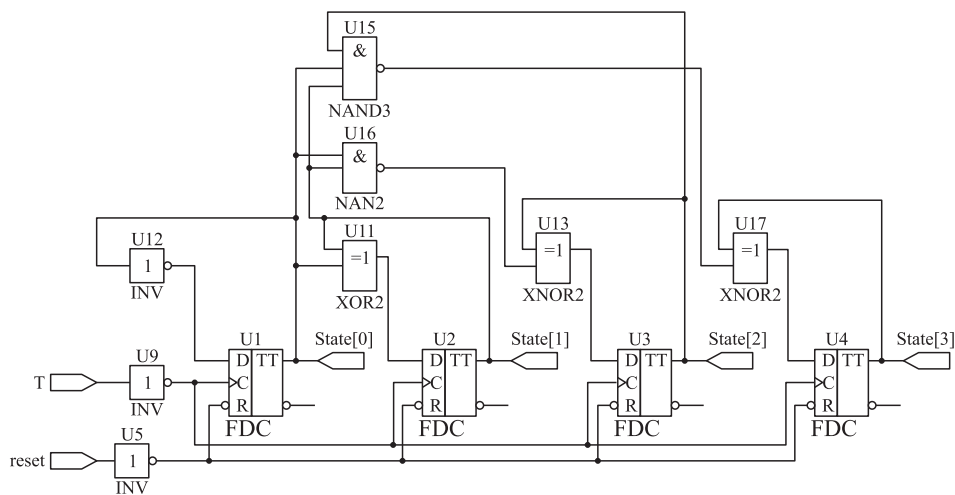


Рис. 2 Результат синтеза Yosys суммирующего счетчика с асинхронным сбросом

ступлении положительного фронта сигнала T и пассивном уровне входа сброса res, что соответствует суммирующему счетчику. Поскольку в теле блока always нет дополнительных условий форматирования состояния State (восьмеричного, десятичного кодирования), это двоичный счетчик.

Результат синтеза двоичного счетчика по его исходному поведенческому описанию зависит от уровня формализации алгоритма преобразования блока always в схему на элементах библиотеки стандартных ячеек. Например, подсистема Yosys [19] реализует любой двоичный счетчик в виде регистра хранения на двухтактных триггерах, состояние которого изменяется в соответствии с заданным типом счета (суммирующим, вычитающим, реверсивным) с помощью параллельной записи в разряды регистра информации, подготовленной комбинационной схемой. Рисунок 2 иллюстрирует результат синтеза подсистемой Yosys синхронного счетчика, заданного Verilog-кодом на рис. 1.

Такой подход к синтезу счетчиков обеспечивает построение в едином стиле всех разновидностей счетчиков, с любым типом предустановки и направления счета. Информационный вход i -го разряда регистра вычисляется как функция

$$X_i = \left(\bigwedge_{k=0}^{i-1} \text{State}[k] \right) \oplus \text{State}[i], \quad (1)$$

где символ « \wedge » означает логическое произведение. При $i = 0$ функция (1) упрощается до вида:

$$X_0 = \sim \text{State}[0].$$

Схема на рис. 2 логически обоснована, так как обеспечивает максимальное быстродействие в синхронной схемотехнике за счет избыточности аппаратных затрат, которая, однако, не всегда оказывается оправданной. Суммарная аппаратная сложность схемы счетчика на рис. 2 равна 180 транзисторам при изготовлении в технологии КМДП (комплементарный металл – диэлектрик – полупроводник). Превратить такую схему в СС-аналог можно путем прямого преобразования с помощью следующей процедуры, реализующей методы, изложенные в работах [18, 20]:

- дуализировать комбинационную часть, используя традиционное для комбинационных СС-схем парафазное со спейсером кодирование всех информационных сигналов;
- подставить вместо синхронных двухтактных триггеров, обозначенных именем FDC на рис. 2, подходящий двухтактный СС-триггер с входом управления и парафазным со спейсером информационным входом;
- ввести преобразователь бифазных выходов СС-триггеров в парафазные со спейсером сигналы, поскольку комбинационная часть СС-счетчика работает именно с такими сигналами;
- добавить индикаторную подсхему, учитывающую необходимость дополнительной индикации входа управления либо в каждом триггере, либо групповым методом из-за парафазного со спейсером типа информационного входа разрядов СС-счетчика.

В результате получается СС-схема, показанная на рис. 3 в базе библиотеки стандартных ячеек для проектирования СС-схем [21]. Схема использованного в ней СС-триггера R1RE24 представлена на рис. 4. Здесь R, S — информационный парафазный вход с нулевым спейсером; RT — вход асинхронного сброса; E — вход управления, аналог тактового сигнала в синхронных триггерах; Q, QB — бифазный информационный выход; I — индикаторный выход. Суммарная аппаратная сложность схемы на рис. 3 равна 376 КМДП-транзисторам.

В то же время максимальное быстродействие за счет избыточности, во-первых, далеко не всегда востребовано, а во-вторых, в СС-варианте оказывается проблематичным вследствие чрезмерного усложнения схемы. Для решения вопроса целесообразности использования параллельного СС-счетчика сравним его с последовательным вариантом СС-счетчика.

3 Сравнение вариантов самосинхронных счетчиков

Последовательный СС-счетчик также синтезируется на основе последовательного синхронного аналога. Рисунок 5 демонстрирует реализацию последовательного синхронного счетчика с асинхронным сбросом. Его аппаратная сложность — 116 КМДП-транзисторов.

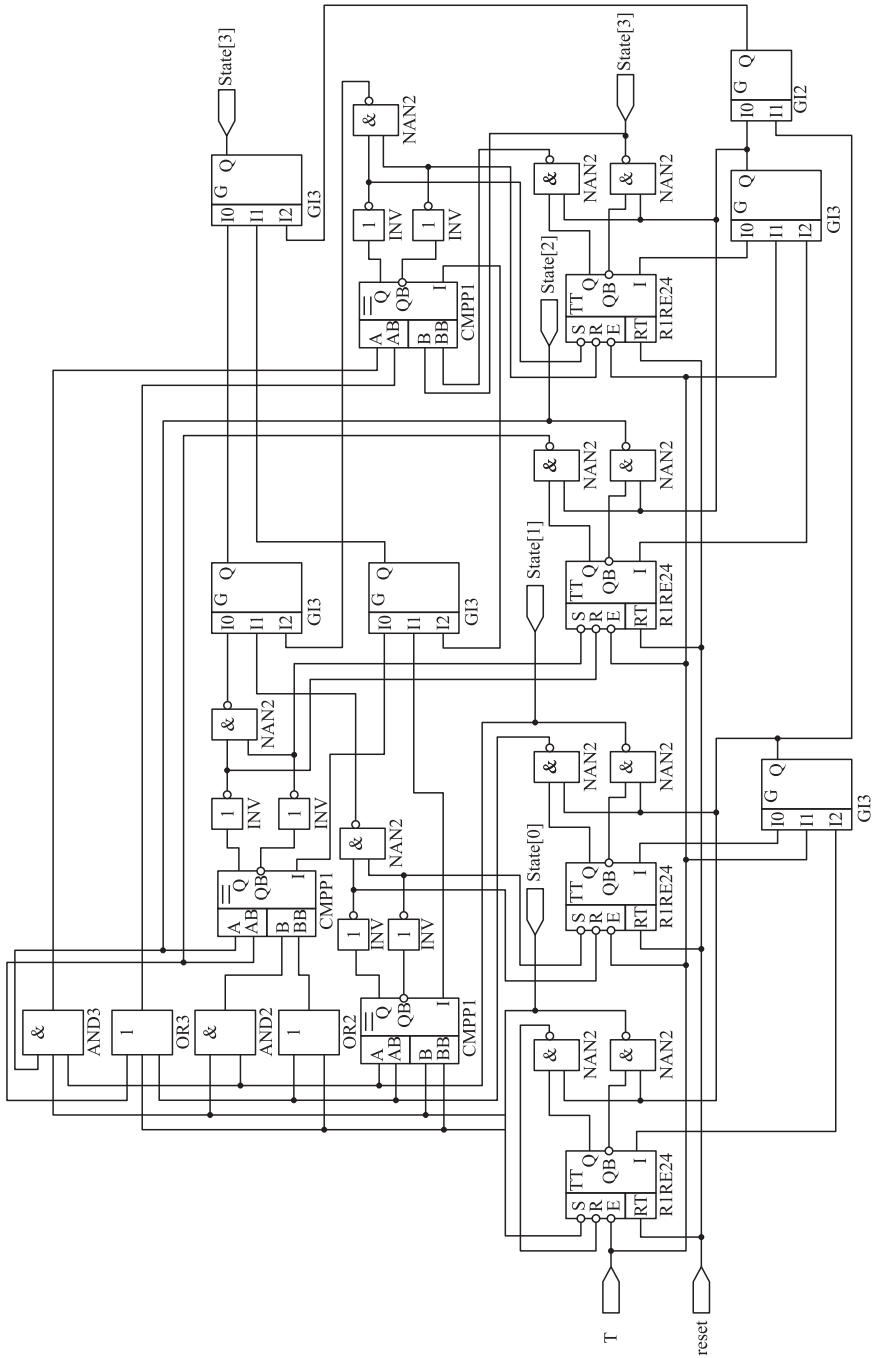


Рис. 3 Результат преобразования параллельного счетчика в СС-реализацию

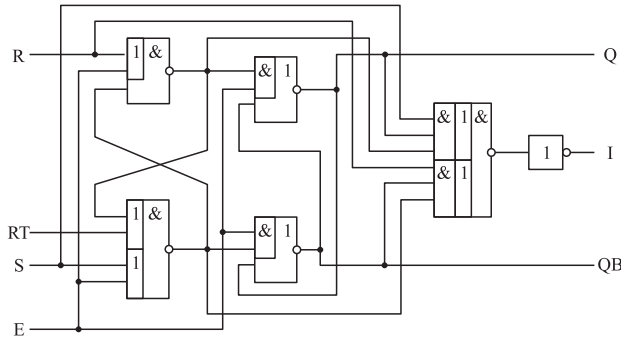


Рис. 4 Схема триггера R1RE24

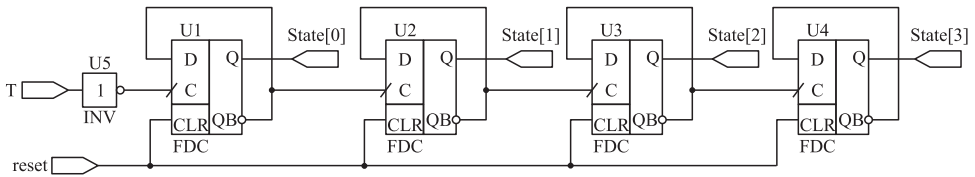


Рис. 5 Последовательный синхронный двоичный счетчик

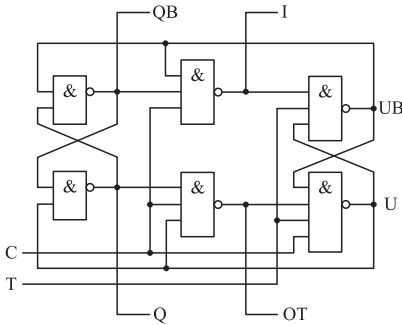


Рис. 6 Схема триггера C1C

Она превращается в СС-реализацию с асинхронным сбросом путем замены триггеров, обозначенных именем FDC на рис. 5, триггером C1C из библиотеки [21] и добавления индикаторной под-схемы. Схема триггера C1C изображена на рис. 6. Здесь T — счетный вход с нулевым спейсером; C — вход СС сброса; Q, QB — бифазный информационный выход; I — индикаторный выход; OT — счетный выход; U, UB — внутренний бифазный сигнал, выход бистабильной ячейки.

Результирующая схема четырехразрядного последовательного СС-счетчика показана на рис. 7. Ее аппаратная сложность — 154 КМДП-транзистора — почти в 2,5 раза ниже, чем в параллельном СС-счетчике на рис. 3.

Сравнение вариантов четырехразрядного СС-счетчика, представленных на рис. 3 и 7, показывает, что в СС-реализации, в отличие от синхронной схемотехники, параллельный счетчик не самый быстрый. Моделирование на транзисторном уровне (программа Spectre, Cadence) в 65-нанометровой КМДП-технологии,

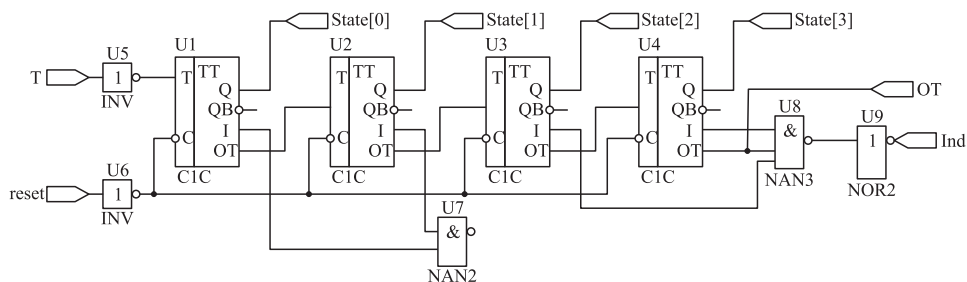


Рис. 7 Самосинхронная реализация последовательного счетчика с асинхронным сбросом

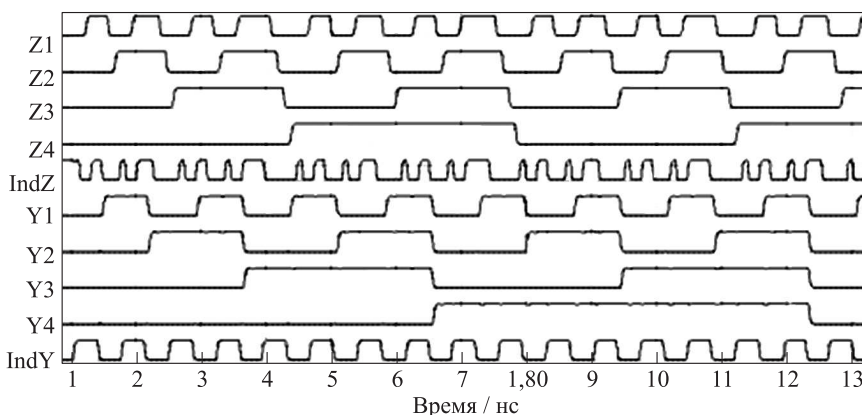


Рис. 8 Моделирование вариантов последовательного (выходы Z^* и IndZ) и параллельного (выходы Y^* и IndY) четырехразрядного СС-счетчика с асинхронным сбросом

результаты которого иллюстрирует рис. 8, выявило следующие различия в поведении вариантов СС-счетчика:

- (1) в параллельном варианте (см. рис. 3) времена пребывания СС-счетчика в рабочей и спейсерной фазах примерно равны и не зависят от текущего состояния счетчика, в то время как в последовательном счетчике (см. рис. 7) длительности фаз варьируются в зависимости от числа разрядов, переключающихся в процессе перехода из одного состояния счетчика в другое;
- (2) в автономном режиме, когда выход общего индикатора счетчика замкнут на его счетный вход, среднее быстродействие СС-счетчика (величина, обратная по отношению к времени прохода счетчика по всем состояниям) в последовательном варианте оказывается в 1,7 раза выше, чем в параллельном счетчике.

Моделирование двух аналогичных вариантов реализации восьмиразрядного СС-счетчика демонстрирует, что с увеличением разрядности преимущество

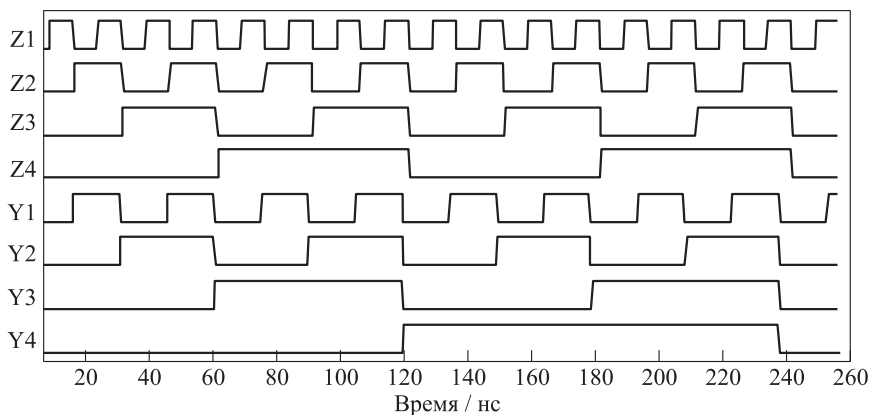


Рис. 9 Моделирование вариантов восьмиразрядного СС-счетчика с асинхронным сбросом

последовательного счетчика над параллельным растет. На рис. 9 показаны диаграммы переключения четырех старших разрядов восьмиразрядных последовательного и параллельного СС-счетчиков. Последовательный счетчик работает почти в 2 раза быстрее.

Причина преимущества последовательного счетчика в быстродействии заключается в более простой схеме и индикации в сравнении с параллельным СС-счетчиком. В параллельном варианте функция (1) информационного входа i -го разряда счетчика требует декомпозиции части $\bigwedge_{k=0}^{i-1} \text{State}[k]$ при $i > 4$, поэтому с увеличением разрядности комбинационная часть параллельного счетчика становится все более декомпозированной, увеличиваются сложность и задержка индикаторной подсхемы.

Дополнительное преимущество последовательного варианта СС-счетчика — его простая масштабируемость, что облегчает его параметризацию при автоматизированном синтезе. Поэтому для СС-реализации целесообразно использовать последовательный счетчик.

4 Формализация синтеза самосинхронных счетчиков

Самосинхронный счетчик целесообразно синтезировать на основе библиотеки параметризованных СС-счетчиков. Пример параметризованного Verilog-описания модуля СС-счетчика со сбросом представлен на рис. 10.

Параметр WIDTH задает разрядность счетчика и настраивается под требуемое значение в каждом конкретном случае в процессе синтеза, причем характер сброса — асинхронный или самосинхронный — не влияет на внутреннюю реализацию счетчика. Соответствующая специфика с успехом определяется окружением счетчика. Выходы модуля счетчика позволяют организовать СС-сброс в окру-

```

//--- Модуль прямого СС-счетчика со сбросом ---
module ST_Counter_Res (T, Res, OT, Ind, State, StateB);
parameter WIDTH = 4; // Разрядность счетчика, 4 бита по умолчанию
input T; // Счетный вход, активный передний фронт
input Res; // Вход начального сброса, активный уровень "0"
output OT; // Счетный выход
output Ind; // Индикаторный выход, спейсер=0
output [WIDTH-1:0] State, StateB; // Информационный выход в бифазном коде

wire [WIDTH-1:0] cind, tOT;
wire [WIDTH-1:1] bind;
genvar i;

//-- Первый разряд счетчика--
C1C Cnr_Inst (.T(T), .C(Res), .Q(State[0]), .QB(StateB[0]), .I(cind[0]), .OT(tOT[0]));

//-- Генерация остальных разрядов счетчика
generate
    for (i=1; i < WIDTH; i=i+1)
    begin: WIDTH_Bit_Counter
        // i-й разряд счетчика -
        C1C Cnr_Inst (.T(OT[i-1]), .C(Res), .Q(State[i]), .QB(StateB[i]), .I(bind[i]), .OT(tOT[i]));
        // Индикатор в рекурсивной форме -
        assign cind[i] = cind[i-1] & bind[i];
    end
endgenerate

//-- Полный индикатор счетного режима --
assign Ind = cind[WIDTH-1] & tOT[WIDTH-1];
assign OT = tOT[WIDTH-1];

endmodule

```

Рис. 10 Параметризованное Verilog-описание модуля СС-счетчика со сбросом

жении счетчика. Разрядность счетчика WIDTH извлекается из тела блока always или из объявления соответствующей переменной в исходном Verilog-описании. Индикаторная подсхема выполняет функцию логического «И» от совокупности выходов I всех разрядов и выхода OT последнего (старшего) разряда:

$$\text{Ind} = \text{OT}_N \wedge \left(\bigwedge_{k=1}^N I_k \right),$$

где N — разрядность счетчика. На рис. 10 она описана рекурсивно и при реализации допускает любую декомпозицию.

При автоматизированном синтезе СС-счетчика обязательно соблюдение полярности сигналов управления, в данном случае — активный передний фронт счетного входа TI и активный низкий уровень сигнала на входе начального сброса Res. При необходимости полярность входов иерархического модуля

```
//--- Подстановка 4-разрядного СС-счетчика с асинхронным сбросом ---  
assign TI = ~T;           // Активный передний фронт счетного входа  
assign Res = ~reset;     // Активный низкий уровень входа сброса  
ST_Counter_Res #(4) Inst_ST_Counter4 (T(TI), .Res(Res), .State(State),  
                                     .StateB(StateB), .Ind(Ind_ST_C4_Res), .OT(OT_ST_C4_Res));
```

Рис. 11 Подстановка модуля СС-счетчика со сбросом в Verilog-описание

согласуется с переменными в исходном Verilog-описании с помощью дополнительных инверторов. Кроме того, появляется инверсная компонента бифазного информационного выхода StateB. В результате блок always с асинхронным сбросом замещается фрагментом Verilog-кода, показанным на рис. 11.

Синтез СС-счетчика реализуется с помощью следующей формализованной процедуры на примере исходного описания на рис. 1:

- (1) **анализ заголовка блока always.** Цель — определить число и типы входов управления счетчиком, их активные уровни и фронты. В данном случае заголовок содержит в списке чувствительности два сигнала: T и res. У них обоих активным считается передний фронт — переключение сигнала из логического нуля в логическую единицу. Следовательно, можно предположить, что по крайней мере один из них служит сигналом синхронизации устройства, описываемого блоком always;
- (2) **специфицирование сигналов в блоке always.** В списке чувствительности счетчика указываются счетный сигнал (глобальный тактовый или производный от него), сигналы инициализации и их активные уровни или перепады. На рис. 1 таких сигналов два: T и res. Анализ тела блока always на рис. 1 показывает, что сигнал T в теле блока не используется, сигнал же res высоким уровнем разрешает обнуление некоторой переменной State, а низким уровнем разрешает инкремент этой же переменной. Значит, T — синхросигнал с активным передним фронтом, а res — сигнал сброса. Просмотр объявлений переменных в исходном Verilog-описании всей схемы позволяет определить разрядность переменной State. Поскольку она инкрементируется в блоке always, то это состояние счетчика, а T — его счетный вход с активным передним фронтом;
- (3) **подбор СС-аналога для синхронного устройства, описываемого данным блоком always.** Поскольку тело блока always описывает суммирующий счетчик, переключающийся по переднему фронту сигнала T и обнуляющийся по сигналу res = 1, в СС-варианте он реализуется на счетном триггере С1С из библиотеки [21], к входу сброса которого подключается инверсия сигнала res;
- (4) **синтез индикаторной подсхемы.** Обязательный компонент индикаторной подсхемы СС-счетчика — индикатор счетного режима. Его целесообразно включать непосредственно в модуль самого счетчика. Подтверждение

успешной СС-предустановки счетчика целесообразно формировать в окружении счетчика;

- (5) **оформление счетчика в виде отдельного модуля.** Блок `always` в исходном Verilog-описании синтезируемой схемы замещается иерархическим модулем СС-счетчика и атрибутами, специфицирующими его особенности функционирования. При последующем синтезе атрибуты помогут корректно реализовать и встроить СС-счетчик в общую схему.

5 Заключение

Известные логические синтезаторы синхронных схем производят двоичные счетчики с разными типами сброса и установки единообразно — с помощью регистра хранения с параллельной записью следующего состояния счетчика, которое вычисляется комбинационной схемой на основе текущего состояния и сигнала, определяющего направление счета. Все разряды такого счетчика обновляются одновременно, что соответствует параллельной архитектуре счетчика. При наличии сброса или установки счетчика входы предустановки формируются той же комбинационной схемой, что предвычисляет и новое состояние счетчика.

Однако последовательный СС-счетчик обладает в 2–2,5 раза меньшей сложностью и в 1,7–2 раза лучшим быстродействием в сравнении с параллельным СС-счетчиком, поэтому для синтеза СС-счетчика предпочтительней использовать его последовательную реализацию.

Самосинхронные счетчики целесообразно синтезировать с использованием параметризованных шаблонов модуля счетчика и его окружения, специфицирующих особенности функционирования СС-счетчика с различными вариантами предустановки и направления счета. При этом модуль счетчика индицирует только свою работу в счетном режиме. Самосинхронная предустановка и направление счета управляются и индицируются окружением СС-счетчика, которое организует необходимую процедуру СС-предустановки или изменения направления счета.

Литература

1. *Hennessy J. L., Patterson D. A.* Computer architecture: A quantitative approach. — 6th ed. — San Mateo, CA, USA: Morgan Kaufmann, 2019. 936 p.
2. *Sparsø J.* Introduction to asynchronous circuit design. — Copenhagen, Denmark: DTU Compute, Technical University of Denmark, 2020. 255 p. https://backend.orbit.dtu.dk/ws/files/215895041/JSPA_async_book_2020_PDF.pdf.
3. *Jiang W., Sha E. H.-M., Zhuge Q., Yang L., Chen X., Hu J.* On the design of time-constrained and buffer-optimal self-timed pipelines // IEEE T. Comput. Aid. D., 2019. Vol. 38. No. 8. P. 1515–1528. doi: 10.1109/TCAD.2018.2846642.
4. *Zakharov V., Stephenkov Y., Diachenko Y., Rogdestvenski Y.* Self-timed circuitry retrospective // Conference (International) on Engineering Technologies and Computer

- Science Proceedings. — Piscataway, NJ, USA: IEEE, 2020. P. 63–69. doi: 10.1109/EnT48576.2020.00018.
5. *Kushnerov A., Medina M., Yakovlev A.* Towards hazard-free multiplexer based implementation of self-timed circuits // 27th Symposium (International) on Asynchronous Circuits and Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2021. P. 17–24. doi: 10.1109/ASYNC48570.2021.00011.
 6. *Stepchenkov Y. A., Kamenskih A. N., Diachenko Y. G., Rogdestvenski Y. V., Diachenko D. Y.* Improvement of the natural self-timed circuit tolerance to short-term soft errors // Advances Science Technology Engineering Systems J., 2020. Vol. 5. No. 2. P. 44–56. doi: 10.25046/aj050206.
 7. *Соколов И. А., Степченко Ю. А., Рождественский Ю. В., Дьяченко Ю. Г.* Приближенная оценка эффективности синхронной и самосинхронной методологий в задачах проектирования сбоеустойчивых вычислительно-управляющих систем // Автоматика и телемеханика, 2022. № 2. С. 122–132. doi: 10.31857/S0005231022020088. EDN: PYRRXN.
 8. *Сурков А. В.* Использование Synopsys Design Compiler для синтеза самосинхронных схем // Программные продукты и системы, 2014. № 4. С. 24–30. doi: 10.15827/0236-235X.108.024-030.
 9. *Власов А. О., Сурков А. В.* Маршрут проектирования самосинхронных конвейерных схем с использованием возможностей САПР // Программные продукты и системы, 2015. № 4. С. 110–115. doi: 10.15827/0236-235x.112.110-115.
 10. *Edwards D., Bardsley A., Jani L., Plana L., Toms W.* Balsa: A tutorial guide. — Manchester, U.K.: The University of Manchester, 2006. 165 p. <https://apt.cs.manchester.ac.uk/ftp/pub/apt/balsa/3.5/BalsaManual3.5.pdf>.
 11. *Sparsø J., Furber S.* Principles of asynchronous circuit design: A systems perspective. — Dordrecht, Netherlands: Kluwer Academic Publs., 2001. 337 p. doi: 10.1007/978-1-4757-3385-3.
 12. *Andrikos N., Lavagno L., Pandini D., Sotiriou C. P.* A fully automated desynchronization flow for synchronous circuits // 44th ACM/IEEE Design Automation Conference Proceedings. — Piscataway, NJ, USA: IEEE, 2007. P. 982–985. doi: 10.1145/1278480.1278722.
 13. *Zhou R., Chong K.-S., Gwee B.-H., Chang J. S., Ho W.-G.* Synthesis of asynchronous QDI circuits using synchronous coding specifications // Symposium (International) on Circuits and Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2014. P. 153–156. doi: 10.1109/ISCAS.2014.6865088.
 14. *Степченко Ю. А., Хилько Д. В., Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю., Орлов Г. А.* Методика десинхронизации при синтезе самосинхронных схем // Системы и средства информатики, 2024. Т. 34. № 1. С. 33–43. doi: 10.14357/08696527240103. EDN: XGZCWU.
 15. *Плеханов Л. П.* Основы самосинхронных электронных схем. — М.: Бином; Лаборатория знаний, 2013. 208 с.
 16. *Гаврилов С. В., Денисов А. Н., Малашевич Н. И., Росляков А. С., Фёдоров Р. А.* Диагностирование самосинхронных функциональных ячеек средствами САПР «КОВЧЕГ» // Известия высших учебных заведений. Электроника, 2011. № 1(87). С. 40–45. EDN: NDDTDX.
 17. *Рождественский Ю. В., Морозов Н. В., Рождественскене А. В.* Подсистема событийного анализа самосинхронных схем АСПЕКТ // Проблемы разработки

- перспективных микро- и нанoeлектронных систем. — М.: ИПИМ РАН, 2010. С. 26–31.
18. *Зацаринный А. А., Степченко Ю. А., Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю.* Автоматизация синтеза самосинхронных схем // Системы высокой доступности, 2023. Т. 19. № 3. С. 48–56. doi: 10.18127/j20729472-202303-04. EDN: IAPVXR.
 19. Yosys open synthesis suite. <https://yosyshq.net/yosys>.
 20. *Степченко Ю. А., Степченко Д. Ю., Дьяченко Ю. Г., Морозов Н. В., Плеханов Л. П.* Замена синхронных триггеров самосинхронными аналогами в процессе десинхронизации схемы // Системы и средства информатики, 2023. Т. 33. № 4. С. 4–15. doi: 10.14357/08696527230401. EDN: VPLSHI.
 21. *Степченко Ю. А., Денисов А. Н., Дьяченко Ю. Г. и др.* Библиотека функциональных ячеек для проектирования самосинхронных полужаказных БМК микросхем серий 5503/5507. — М.: Техносфера, 2017. 367 с. <http://www.technosphera.ru/lib/book/497>.

Поступила в редакцию 15.03.24

SELF-TIMED COUNTER SYNTHESIS FORMALIZATION

*Yu. A. Stepchenkov, Yu. G. Diachenko, N. V. Morozov, D. Yu. Stepchenkov,
and D. Yu. Diachenko*

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

Abstract: Self-timed (ST) circuits have high reliability. They guarantee detection and localization of any persistent faults and demonstrate a high level of fault tolerance. However, designing ST circuits is more labor-intensive compared to synchronous circuits because one should construct an additional indication subcircuit and adhere to the principles of truly ST circuit implementation. Formalized desynchronization provides automatic conversion of the original synchronous circuit description into the self-timed one but when synthesizing sequential ST units, including ST counters, it leads to excessive hardware redundancy and, as a consequence, to their low performance. The article substantiates the approach to the ST counter synthesis based on the heuristic method formalization for their construction and ensuring the guaranteed resulted truly ST implementation that functions in full accordance with the original description and has close to optimal consumer characteristics

Keywords: automated synthesis; self-timed circuit; counter; desynchronization; preset; indication

DOI: 10.14357/08696527240205

EDN: KDIEOJ

References

1. Hennessy, J. L., and D. A. Patterson. 2019. *Computer architecture: A quantitative approach*. 6th ed. San Mateo, CA: Morgan Kaufmann. 936 p.
2. Sparsø, J. 2020. *Introduction to asynchronous circuit design*. Copenhagen, Denmark: DTU Compute, Technical University of Denmark. 255 p. Available at: https://backend.orbit.dtu.dk/ws/files/215895041/JSPA_async_book_2020_PDF.pdf (accessed May 1, 2024).
3. Jiang, W., E. H.-M. Sha, Q. Zhuge, L. Yang, X. Chen, and J. Hu. 2019. On the design of time-constrained and buffer-optimal self-timed pipelines. *IEEE T. Comput. Aid. D.* 38(8):1515–1528. doi: 10.1109/TCAD.2018.2846642.
4. Zakharov, V., Y. Stepchenkov, Y. Diachenko, and Y. Rogdestvenski. 2020. Self-timed circuitry retrospective. *Conference (International) on Engineering Technologies and Computer Science Proceedings*. Piscataway, NJ: IEEE. 63–69. doi: 10.1109/EnT48576.2020.00018.
5. Kushnerov, A., M. Medina, and A. Yakovlev. 2021. Towards hazard-free multiplexer based implementation of self-timed circuits. *27th Symposium (International) on Asynchronous Circuits and Systems Proceedings*. Piscataway, NJ: IEEE. 17–24. doi: 10.1109/ASYNC48570.2021.00011.
6. Stepchenkov, Y. A., A. N. Kamenskih, Y. G. Diachenko, Y. V. Rogdestvenski, and D. Y. Diachenko. 2020. Improvement of the natural self-timed circuit tolerance to short-term soft errors. *Advances Science Technology Engineering Systems J.* 5(2):44–56. doi: 10.25046/aj050206.
7. Sokolov, I. A., Yu. A. Stepchenkov, Yu. V. Rogdestvenski, and Yu. G. Diachenko. 2022. Approximate evaluation of the effectiveness of synchronous and self-timed methodologies in problems of designing failure-tolerant computing and control systems. *Automat. Rem. Contr.* 83(2):264–272. doi: 10.1134/S0005117922020084.
8. Surkov, A. V. 2014. Ispol'zovanie Synopsys Design Compiler dlya sinteza samosinkhronnykh skhem [Synthesis of burst-mode asynchronous schemes using Synopsys Design Compiler]. *Programmnyye produkty i sistemy [Software & Systems]* 4:24–30. doi: 10.15827/0236-235X.108.024-030.
9. Vlasov, A. O., and A. V. Surkov. 2015. Marshrut proektirovaniya samosinkhronnykh konveyernykh skhem s ispol'zovaniem vozmozhnostey SAPR [Automated design flow of self-timed pipelines using EDA tools]. *Programmnyye produkty i sistemy [Software & Systems]* 4:110–115. doi: 10.15827/0236-235x.112.110-115.
10. Edwards, D., A. Bardsley, L. Jani, L. Plana, and W. Toms. 2006. *Balsa: A tutorial guide*. Manchester, U.K.: The University of Manchester. 165 p. Available at: <https://apt.cs.manchester.ac.uk/ftp/pub/apt/balsa/3.5/BalsaManual3.5.pdf> (accessed April 10, 2024).
11. Sparsø, J., and S. Furber. 2001. *Principles of asynchronous circuit design: A systems perspective*. Dordrecht, Netherlands: Kluwer Academic Publ. 337 p. doi: 10.1007/978-1-4757-3385-3.
12. Andrikos, N., L. Lavagno, D. Pandini, and C. P. Sotiriou. 2007. A fully-automated desynchronization flow for synchronous circuits. *44th ACM/IEEE Design Automation Conference Proceedings*. Piscataway, NJ: IEEE. 982–985. doi: 10.1145/1278480.1278722.

13. Zhou, R., K.-S. Chong, B.-H. Gwee, J. S. Chang, and W.-G. Ho. 2014. Synthesis of asynchronous QDI circuits using synchronous coding specifications. *Symposium (International) on Circuits and Systems Proceedings*. Piscataway, NJ: IEEE. 153–156. doi: 10.1109/ISCAS.2014.6865088.
14. Stepchenkov, Yu. A., D. V. Khilko, Yu. G. Diachenko, N. V. Morozov, D. Yu. Stepchenkov, and G. A. Orlov. 2024. Metodika desinkhronizatsii pri sinteze samosinkhronnykh skhem [Desynchronization methodology at self-timed circuit synthesis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(1):33–43. doi: 10.14357/08696527240103. EDN: XGZCWU.
15. Plekhanov, L. P. 2013. *Osnovy samosinkhronnykh elektronnykh skhem* [Basics of self-timed electronic circuits]. Moscow: Binom. Laboratoriya znaniy. 208 p.
16. Gavrilov, S. V., A. N. Denisov, N. I. Malashevich, A. S. Roslyakov, and R. A. Fedorov. 2011. Diagnostirovanie samosinkhronnykh funktsional'nykh yacheek sredstvami SAPR “KOVCHEG” [Diagnostics of self-synchronous functional cells by means of CAD “Kovcheg”]. *Izvestiya vysshikh uchebnykh zavedeniy. Elektronika* [Proceedings of Universities. Electronics] 1(87):40–45. EDN: NDDTDX.
17. Rogdestvenski, Y. V., N. V. Morozov, and A. V. Rogdestvenskene. 2010. Podsystema sobytiynogo analiza samosinkhronnykh skhem ASPEKT [ASPECT: A suite of self-timed event-driven analysis]. *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem* [Problems of the perspective micro- and nanoelectronic systems development]. Moscow: IPPM RAN. 26–31.
18. Zatsarinnyy, A. A., Yu. A. Stepchenkov, Yu. G. Diachenko, N. V. Morozov, and D. Yu. Stepchenkov. 2023. Avtomatizatsiya sinteza samosinkhronnykh skhem [Self-timed circuits automated design]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 19(3):48–56. doi: 10.18127/j20729472-202303-04. EDN: IAPVXR.
19. Yosys open synthesis suite. Available at: <https://yosyshq.net/yosys> (accessed April 10, 2024).
20. Stepchenkov, Yu. A., D. Yu. Stepchenkov, Yu. G. Diachenko, N. V. Morozov, and L. P. Plekhanov. 2023. Zamena sinkhronnykh triggerov samosinkhronnymi analogami v protsesse desinkhronizatsii skhemy [Replacing synchronous triggers with self-timed counterparts during circuit desynchronization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):4–15. doi: 10.14357/08696527230401. EDN: VPLSHI.
21. Stepchenkov, Yu. A., A. N. Denisov, Yu. G. Diachenko, et al. 2017. *Biblioteka funktsional'nykh yacheek dlya proektirovaniya samosinkhronnykh poluzakaznykh BMK mikroskhem seriy 5503/5507* [Functional cell library for designing self-timed semi-custom integrated circuits on 5503/5507 gate arrays]. Moscow: Tekhnosfera. 367 p. Available at: <http://www.technosfera.ru/lib/book/497> (accessed April 10, 2024).

Received March 15, 2024

Contributors

Stepchenkov Yuri A. (b. 1951) — Candidate of Science (PhD) in technology, head of department, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; YStepchenkov@ipiran.ru

Diachenko Yuri G. (b. 1958) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; diaura@mail.ru

Morozov Nikolai V. (b. 1956) — senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; NMorozov@ipiran.ru

Stepchenkov Dmitri Yu. (b. 1973) — senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; stepchenkov@mail.ru

Diachenko Denis Yu. (b. 1987) — engineer-researcher, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; diaden87@gmail.com

ЭВОЛЮЦИЯ И ПЕРСПЕКТИВЫ КОНЦЕПЦИИ ПРОГРАММНОГО ОПРЕДЕЛЕНИЯ

*В. Б. Егоров*¹

Аннотация: Концепция программного определения родилась применительно к сетям передачи данных. Программное определение сетей (software-defined networking, SDN) изначально предполагало централизацию и автоматизацию администрирования сети с заменой маршрутизаторов на более дешевые коммутаторы. Концепция SDN претерпела существенные изменения при ее переносе из университетских лабораторий в дата-центры и провайдерские сети, в результате чего на практике SDN-сети стали представлять собой традиционные IP-сети с полнофункциональными маршрутизаторами, лишь дополненными возможностями SDN. Вслед за сетями программное определение освоило область хранения данных (software-defined storage, SDS), где приобрело новое важное свойство — сервисную ориентированность. При дальнейшем распространении программного определения на целые дата-центры (software-defined data centers, SDDC) и даже «облака» (software-defined clouds) это свойство стало доминирующим. Можно ожидать, что успехи SDDC сделают избыточными самостоятельные SDN и SDS, но широкому внедрению SDDC препятствует сложность создания для них специфического программного обеспечения (ПО) «оркестровки» (orchestration). Поэтому для ординарного владельца дата-центра едва ли не единственной доступной возможностью обретения SDDC оказывается покупка у крупного надежного поставщика комплексного решения «под ключ» с готовой «оркестровкой».

Ключевые слова: программно-определяемые сети; программно-определяемые хранилища; программно-определяемые дата-центры

DOI: 10.14357/08696527240206

EDN: KИJHNP

1 Введение

Концепция программного определения родилась в конце первого десятилетия XXI в. применительно к сетям передачи данных университетских лабораторий. Программное определение сетей предполагало централизацию и автоматизацию администрирования сети, что позволяло быстрее и проще реконфигурировать и перестраивать сети в условиях проведения разнообразных научных экспериментов. Обнадеживающие результаты, полученные на этом этапе, способствовали продвижению концепции SDN в сети дата-центров и провайдерские сети.

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

Хотя здесь достижения SDN оказались скромнее, в начале следующего десятилетия концепция программного определения освоила, и с гораздо большим успехом, новую область — хранение данных. Наконец, в последние годы все чаще речь заходит о программно-определяемых дата-центрах и даже программно-определяемых «облаках». В настоящей статье делается попытка объяснения такого распространения и универсальной применимости программного определения в различных информационно-коммуникационных технологиях и оценки его перспектив.

2 Программное определение сети

Концепция SDN переосмыслила такие фундаментальные сетевые понятия, как плоскости данных (data plane) и управления (control plane), которые ранее представляли собой некие абстракции, логически разделяющие две основные группы функций узлов сети — продвижение пакетов и маршрутизацию их потоков, но физически объединяемые в традиционных сетевых устройствах. SDN-подход «материализовал» обе плоскости, разделил их физически и «стянул» всю плоскость управления в логически единственную точку — программный контроллер сети (SDN-контроллер). Первоначально под SDN-контроллером понималось некое специфическое прикладное ПО, работающее на универсальном сервере или виртуальной машине (VM) и управляющее сетью от настройки отдельных ее узлов до маршрутизации и обеспечения соглашений по уровню обслуживания (service-level agreements, SLA). После концентрации всей плоскости управления в SDN-контроллере в плоскости данных концептуальной SDN-сети остались только простые коммутаторы, которые коммутировали кадры и пакеты на основании получаемых из контроллера готовых правил (политик) коммутации. Таким образом, важным преимуществом SDN-подхода наряду с программной автоматизацией плоскости управления предполагалось существенное упрощение аппаратуры плоскости данных за счет замены дорогих маршрутизаторов более дешевыми коммутаторами. Для унификации взаимодействия SDN-контроллера с такими коммутаторами был стандартизирован специальный протокол OpenFlow.

В лабораториях, где была придумана и опробована концепция SDN, и в дата-центрах управление плоскостью данных SDN-контроллером проще всего было организовывать «вне полосы» (out-of-band). При такой организации управления SDN-контроллер был связан со всеми коммутаторами выделенными каналами «точка-точка», на которых протокол OpenFlow мог функционировать на канальном уровне. При управлении «вне полосы» SDN-контроллер не был абонентом управляемой им сети и управляющее сетью прикладное ПО могло работать под любой универсальной операционной системой (ОС). Нацеленность этого прикладного ПО на управление физической аппаратурой плоскости данных ставило его в сильную зависимость от особенностей конкретной сети: ее топологии, технических характеристик узлов, пропускной способности отдель-

ных звеньев и, наконец, разнообразия условий SLA. Все это фактически делало SDN-контроллер уникальным для каждой конкретной сети и препятствовало его превращению в некий универсальный рыночный продукт.

Перенос исходной концепции SDN из лабораторий в дата-центры, провайдерские сети и далее в сферу «облачных» вычислений выявил у нее ряд слабых мест и внес в нее существенные коррективы.

Логически единый SDN-контроллер на практике приходилось как минимум дублировать в целях обеспечения доступности и безопасности сети, а в сложных и территориально распределенных сетях даже реализовывать множеством физических контроллеров, управляющих разными сегментами сети и взаимодействующих между собой вне рамок не только OpenFlow, но и самой концепции SDN. В такого рода сетях, в первую очередь провайдерских и «облачных», из-за практической невозможности реализации управления сетью «вне полосы» стал широко применяться способ управления «в полосе» (in-band), при котором фактически создавалась виртуальная OpenFlow-сеть, наложенная на физическую сеть плоскости данных. Переход от управления «вне полосы» к управлению «в полосе» имел два важных следствия.

Во-первых, при управлении «в полосе» SDN-сеть теряла возможность самоорганизации, в частности при начальном пуске (bootstrapping) и удаленном конфигурировании узлов, из-за возникающей зависимости наложенной OpenFlow-сети от физической сети плоскости данных [1]. В концептуальной SDN-сети при управлении «в полосе» неработоспособную по каким-либо причинам плоскость данных невозможно привести в работоспособное состояние, так как неработоспособной оказывается также и необходимая для этого OpenFlow-сеть. Поэтому в целях сохранения возможности самоорганизации при начальном пуске и в нештатных ситуациях на практике SDN-сети остались традиционными IP-сетями, лишь дополненными возможностями SDN. В узлах реальных SDN-сетей вместо концептуальных простых коммутаторов по-прежнему сохранились полнофункциональные маршрутизаторы, в которые лишь добавилась поддержка протокола OpenFlow.

Во-вторых, при управлении «в полосе» протокол OpenFlow был вынужден подняться с канального на сеансовый уровень над транспортными протоколами TCP (transmission control protocol) или TLS (transport layer security), а у SDN-контроллера, ставшего абонентом управляемой им сети, возникала необходимость в специализированной сетевой ОС, обеспечивающей поддержку виртуальной сети OpenFlow. В дальнейшем с такого рода ОС (SDN-ОС) произошла интересная метаморфоза. Ряд появившихся на рынке SDN-ОС, включая отечественные RUNOS [2] и B4N Controller [3], позиционировали себя, вероятно в рекламных целях, в качестве SDN-контроллеров, что создавало иллюзию возможности существования SDN-контроллера как универсального готового продукта. После узурпации SDN-ОС термина «SDN-контроллер» для прикладного ПО управления SDN-сетью выработался альтернативный термин «дирижер» (orchestrator), вероятно, как признание того, что ему по-прежнему вменялось согласование

(orchestration) работы всех компонентов сети для самых разных «партитур», т. е. физических конфигураций сети и вариаций SLA в ней. В отличие от относительно универсальных SDN-ОС, «дирижеры» по-прежнему плохо поддаются универсализации, что остается главным препятствием на пути массового внедрения SDN-сетей [4, 5].

3 Программное определение и виртуализация

Изначальная концепция SDN, будучи ориентированной на физическую аппаратуру плоскости данных, не была связана с виртуализацией. Виртуализация сетей и сетевых устройств развивалась самостоятельно.

Заметных успехов достигла виртуализация так называемых промежуточных устройств (middleboxes), к которым относят брандмауэры, средства балансировки нагрузок, системы обнаружения и защиты от вторжений и т. п. Попутно следует заметить, что к виртуализации сетевых устройств в литературе стал широко применяться неудачный термин «виртуализация сетевых функций» (network functions virtualization, NFV), что, вообще говоря, выглядит бессмыслицей [6]. Но независимо от терминологии суть виртуализации сетевых устройств сводилась к программной эмуляции на типовых серийных серверах той или иной сетевой аппаратуры. Подобно VM виртуальные сетевые устройства могли развертываться по мере надобности в нужных точках сети и ликвидироваться, когда нужда в них отпадала. Это подобие VM и, более того, возможность реализации в VM провоцировали отнесение управления такими виртуальными сетевыми объектами скорее к гипервизорам (мониторам VM), а не к SDN-контроллерам. Последние задействовались лишь косвенно для перенаправления физических потоков данных на соответствующие серверы.

Помимо промежуточных устройств сетевая виртуализация коснулась и таких концептуальных для SDN-сетей компонентов, как коммутаторы. Весьма популярным устройством стал виртуальный коммутатор vSwitch. Типичный vSwitch реализуется как часть гипервизора и им же управляется. Хотя многие реализации vSwitch включают поддержку протокола OpenFlow, виртуальные коммутаторы, как и виртуальные промежуточные устройства, в значительной степени остаются вне поля зрения SDN-контроллера. Все манипуляции с виртуальными сущностями, будь то наложенные сети, промежуточные устройства или коммутаторы, затрагивают SDN-контроллер только косвенно — необходимыми отражениями этих манипуляций на физическую инфраструктуру. Действия по управлению коммутаторами vSwitch могут производиться гипервизором без целеуказаний от SDN-контроллера и даже без его уведомления о таких действиях. Например, при миграции некой VM внутри сервера всей необходимой для перекоммутации связанных с ней сетевых трактов информацией обладает не SDN-контроллер, а гипервизор, которому «сподручней» осуществить требуемую перекоммутацию простой заменой своих адресных таблиц, при этом SDN-контроллер может такую реконфигурацию сети даже не заметить.

Хотя из-за нацеленности SDN на физическую инфраструктуру сети всякая сетевая виртуализация касается SDN-контроллера лишь опосредованно, адекватное отражение изменений в виртуальных средах на физическую сеть остается, тем не менее, задачей «дирижеров», которая еще более их усложняет и тем самым дополнительно тормозит внедрение SDN-сетей.

4 Программное определение систем хранения данных

На волне больших надежд, связанных с SDN, в начале второго десятилетия XXI в. идея программного определения проникла в системы хранения данных (СХД) и оформилась как программно-определяемое хранилище (software-defined storage, SDS). При этом SDS позаимствовало у SDN прежде не свойственную СХД терминологию, в том числе выделение плоскостей управления и данных, и, главное, централизацию плоскости управления в логически единственном контроллере (SDS-контроллере). Однако у SDS с самого начала обозначились четыре существенных отличия от SDN. Во-первых, концепция SDS не требовала введения специальных протоколов связи между плоскостями, как OpenFlow в случае сетей, поскольку оборудование хранения данных давно обзавелось своими интерфейсами и протоколами от простейших типа SCSI (small computer systems interface) до современных высокоскоростных вроде NVMe (nonvolatile memory express). Во-вторых, в отличие от SDN, где программное определение не было напрямую связано с виртуализацией сети и сетевых устройств, SDS принципиально реализуется на основе предварительно виртуализированного пула ресурсов хранения. Например, в дата-центре с гиперконвергентной инфраструктурой, где отсутствуют сетевое хранилище (network-attached storage) и сеть хранения данных (storage-area network), невозможно обойтись без преобразования множества распределенных по серверам локальных накопителей (direct-attached storage) в единый виртуальный пул ресурсов хранения [7]. В-третьих, если все сети по-своему уникальны и, соответственно, им требуются уникальные «дирижеры», то СХД выпускаются в виде типовых продуктов, для которых могут создаваться специализированные «дирижеры», в результате чего возникают законченные SDS-продукты [8]. Благодаря этому отличию SDS внедряются в дата-центры гораздо активнее SDN и, стартовав на несколько лет позже, уверенно обгоняют предтечу по числу реализаций. И наконец, в-четвертых, в SDS меняется сама цель программного определения. На этом отличии следует остановиться подробнее.

Первоначальной целью SDN было облегчение администрирования в сетях, в первую очередь экспериментальных лабораторных. Администратор SDN-сети получал возможность, не сходя с места за терминалом сетевого контроллера, управлять работой сети и отдельных ее узлов, оперативно перекоммутировать потоки данных для обхода поврежденных узлов или разгрузки перегруженных звеньев, экспериментировать с приоритетами потоков, а также собирать и анализировать разного рода статистику. SDN-контроллер практически не имел

касательства к приложениям, выполняемым на серверах-абонентах SDN-сети. При запуске некоего приложения на сервере или в ВМ гипервизор предоставлял ему сетевую поддержку из имеющихся в его распоряжении ресурсов и настраивал соответствующим образом свой vSwitch. Такого рода события внутри серверов могли оставаться вне поля зрения SDN-контроллера.

Взаимодействие приложения с СХД, в отличие от работы с сетью, не ограничивается получением или назначением ресурсов хранения при запуске приложения. В процессе работы приложение создает, модифицирует и уничтожает файлы, сохраняет и востребует сохраненные массивы данных. При этом могут подключаться дополнительные услуги сжатия, шифрования, дублирования или дедупликации данных, т. е. приложение в процессе работы может затребовать сервисы, которые в случае SDS обеспечиваются его контроллером. Если SDN-контроллер был задуман для обслуживания в первую очередь администратора сети, то SDS-контроллер «по роду службы» также вынужден обслуживать и имеющие доступ к СХД приложения. Таким образом, перенос программного определения с сетей на СХД добавил ему принципиальное свойство оперативного обслуживания запросов приложений, т. е. придал ему сервисную ориентированность.

5 Программное определение центров обработки данных

Следующий шаг в том же направлении, превращающий сервисную ориентированность из опции в доминанту программного определения, делает программно-определяемый дата-центр.

Программно-определяемому дата-центру дано множество определений, более или менее удачных. К сожалению, к числу удачных трудно отнести главное из них — определение разработчика стандартов для SDDC, целевой группы по распределенному управлению (distributed management task force, DMTF): «SDDC — это программная абстракция логических вычислений, сетевых, хранилищных и прочих ресурсов, представленных как ПО. Эти ресурсы динамически выявляются, разворачиваются и настраиваются на основе требований выполняемых работ (workloads). SDDC делает возможной «оркестровку» работ на основе политик, а также измерение потребляемых ресурсов и управление ими» [9]. В этом определении, если отвлечься от невнятной «программной абстракции, представленной как ПО», а также от той очевидности, что не SDDC делает возможной «оркестровку», а наоборот, «оркестровка» делает возможным SDDC, безусловно верным и важным остается то, что в SDDC ресурсы динамически разворачиваются на основе требований выполняемых работ, т. е. в SDDC сервисная ориентированность становится универсальной. Ее обеспечение в SDDC по-прежнему возлагается на «автоматизацию всех функций с помощью интеллектуального ПО, т. е. программный слой, посредством которого требования бизнес-логики транслируются на уровень управления инфраструктурой» [10].

Доминантность сервисной ориентированности подчеркивается и в целом ряде других определений SDDC:

- «В SDDC все элементы инфраструктуры, включая сеть, СХД, вычислительные ресурсы и безопасность, виртуализируются и предоставляются как услуга» (TechTarget [11]);
- «SDDC — это средство хранения данных, в котором все элементы инфраструктуры — сеть, СХД, центральный процессор и безопасность — виртуализированы и предоставляются как услуга. . .» (Forrester Research [9]).
- «SDDC — это выражение, используемое для обозначения дата-центра, где вся инфраструктура виртуализирована и предоставляется как услуга» (VMware [9]).

Если сервисная ориентированность SDDC становится общепринятой, то с пониманием средств ее достижения все не так просто. Wikipedia (в статье Software-defined data center) архитектурными компонентами SDDC называет компьютерную виртуализацию, SDN и SDS. Но с таким утверждением, при всей его кажущейся очевидности, можно согласиться лишь с определенными оговорками. Объединяя в дата-центре готовые SDN и SDS, получаем два (а на самом деле с учетом монитора ВМ даже три) независимых «дирижера», для организации взаимодействия между которыми понадобится «дирижер дирижеров». Но если до сих пор не преодолены трудности с «дирижерами» SDN, то вряд ли следует ожидать больших успехов с некими иерархическими «дирижерами». «Дирижер» SDDC все-таки видится скорее «одноэтажным», непосредственно управляющим всеми инфраструктурными компонентами дата-центра и обеспечивающим их бесконфликтную совместную работу. В этом плане адекватнее смотрится более общее определение SDDC как совокупности виртуализации всех инфраструктурных компонентов дата-центра с автоматизацией управления и «оркестровкой» [12]. При этом за рамками определения остается оговорка, что создание SDDC с комплексной «оркестровкой» и, соответственно, чрезвычайно сложным «дирижером» по силам лишь очень крупным специализированным компаниям, что подтверждается перечнем основных провайдеров SDDC [12]:

- VMware SDDC;
- VMware Cloud on Amazon Web Services;
- Azure VMware Solution;
- Google Cloud VMware Engine;
- VMware SDDC on IBM Cloud;
- Oracle Cloud VMware Solution;
- HPE Software-Defined Infrastructure (в партнерстве с VMware).

В этом перечне бросается в глаза упоминание в каждой строке VMware, т. е. на самом деле все перечисленные провайдеры распространяют на своих аппаратных

платформах ПО этой компании. Можно по-разному относиться к монополии VMware на ПО для SDDC, но в ней следует признать по крайней мере тот положительный аспект, что она обеспечивает высокий уровень совместимости программных продуктов в разных дата-центрах и даже в «облачных» сферах.

При комплексной поставке оборудования дата-центра совместно с соответствующим ПО, включая «дирижера», «пропадают» проблемы, вызванные совместимостью отдельных компонентов, а также снижается совокупная стоимость владения и риски при эксплуатации инфраструктуры, так как администрировать стек продуктов одного производителя значительно проще, чем «зоопарк» систем от разных вендоров» [13]. Поэтому для среднестатистического владельца дата-центра единственной реальной возможностью обретения SDDC остается, по-видимому, покупка «готового под ключ» комплексного продукта [14, 15], желательно у одного надежного вендора (т. е. фактически из приведенного выше перечня).

6 Заключение

Программно-определяемый подход в целом находит все более широкое применение в различных информационно-коммуникационных технологиях. Такая универсальная применимость объясняется его сущностью — автоматизацией процессов управления сложными объектами, такими как сети, СХД и целые дата-центры. Автоматизация труда администраторов такого рода объектов из весьма желательной возможности превращается в насущную необходимость для крупных дата-центров, а тем более при выходе в «облака» и глобальные сети.

После относительной неудачи в SDN концепция программного определения реабилитировалась в SDS и SDDC. В значительной степени ее успех в области СХД и дата-центров определило дополнение автоматизации администрирования сервисной ориентированностью, которая стала характерной чертой SDS и превратилась в настоятельную необходимость в SDDC.

Сегодня перспективы SDN выглядят весьма сомнительными. Будучи самой старой из технологий программного определения, она остается среди них «наименее зрелой» [16] и уже отстает от SDS по числу воплощений. Но и будущее SDS как самостоятельной технологии тоже под вопросом. Всякая автоматизация в конечном счете стремится к комплексным и всеобъемлющим решениям. Соответственно, магистральным путем развития и внедрения концепции программного определения видится ее комплексная реализация в SDDC. И тут необходимо иметь в виду, что вопреки распространенному мнению SDDC не столько включают в себя SDN и SDS, сколько заменяют их. Поэтому успехи SDDC, вероятно, сделают избыточными «самостоятельные» SDN и SDS.

В настоящий момент готовый SDDC предлагают несколько крупных фирм, но практически во всех случаях ПО программного определения в них предполагается от компании VMware, в чем есть и свой минус: безальтернативность и отсутствие конкуренции, — и свой плюс: сквозная уходящая в «облака»

совместимость «дирижеров». Для поддержания такой же совместимости в условиях конкуренции необходима разработка большого числа стандартов. Этим, вроде бы, занимается DMTF, но пока стандартами де-факто остаются решения компании VMware, на которые, вероятно, будут вынужденно ориентироваться и стандарты от DMTF.

Литература

1. *Егоров В. Б.* К вопросу о соотношении программно определяемых и традиционных IP-сетей // Системы и средства информатики, 2022. Т. 32. № 1. С. 73–82. doi: 10.14357/08696527220107. EDN: OGWOXL.
2. Контроллер RunOS // RunSDN, 2017. <https://runsdn.ru/index.php/products/runos>.
3. Компания Программируемые Сети (Brain4Net) выпустила первый в России SDN Controller операторского класса // Brain4Net, 2020. <https://brain4net.ru/brain4net-releases-carrier-grade-sdn-controller/>.
4. *Егоров В. Б.* Некоторые вопросы практической реализации концепции SDN // Системы и средства информатики, 2016. Т. 26. № 1. С. 109–120. doi: 10.14357/08696527160108. EDN: VZBFNT.
5. *Ганьжа Д.* Компонуемая инфраструктура в контексте SDDC // Ж. сетевых решений/LAN, 2018. № 04. <https://www.osp.ru/lan/2018/04/13054573>.
6. *Егоров В. Б.* К вопросу о совмещении программного определения сетей с виртуализацией сетевых функций // Системы и средства информатики, 2022. Т. 32. № 2. С. 36–46. doi: 10.14357/08696527220204. EDN: MTZDDZ.
7. *Егоров В. Б.* Программное определение сети в конвергентной и гиперконвергентной инфраструктурах // Системы и средства информатики, 2023. Т. 33. № 1. С. 105–113. doi: 10.14357/08696527230110. EDN: CAIIMP.
8. *Егоров В. Б.* Некоторые вопросы программного определения хранилища данных // Системы и средства информатики, 2021. Т. 31. № 2. С. 70–79. doi: 10.14357/08696527210207. EDN: UWZPYU.
9. Software Defined Data Center (SDDC) Definition. A White Paper from the OSD-DC incubator // DMTF, 2015. <https://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0501.1.0.0.pdf>.
10. *Дубова Н.* Программный дуализм // Открытые системы. СУБД, 2014. № 6. С. 14–16. EDN: SJKFNJ. <https://www.osp.ru/os/2014/06/13042309>.
11. *Kranz G.* SDDC (software-defined data center) // TechTarget, 2017. <https://www.techtarget.com/searchdatacenter/definition/SDDC-software-defined-data-center>.
12. *Crockett E.* Top software-defined data center providers (SDDC) // Datamation, 2022. <https://www.datamation.com/data-center/software-defined-data-center-providers/>.
13. Как построить полноценное гибридное облако и снизить стоимость владения ИТ-инфраструктурой // Snews-SoftLine, 2021. https://softline.cnews.ru/articles/2021-07-14_virtualnyj_tsod_j_gibridnoe_oblako.
14. *Егоров В. Б.* Некоторые вопросы программного определения центров обработки данных // Системы и средства информатики, 2020. Т. 30. № 2. С. 103–112. doi: 10.14357/08696527200210. EDN: FXXGJY.

15. Egorov V. B. Некоторые вопросы деагрегации и компонентности инфраструктуры центра обработки данных // Системы и средства информатики, 2023. Т. 33. № 2. С. 101–110. doi: 10.14357/08696527230210. EDN: OEXVDB.
16. Knorr E. What the software-defined data center really means // InfoWorld, 2012. <https://www.infoworld.com/article/2618097/what-the-software-defined-data-center-really-means.html>.

Поступила в редакцию 29.11.23

EVOLUTION OF THE “SOFTWARE-DEFINED” CONCEPT

V. B. Egorov

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

Abstract: The “software-defined” concept was born in relation to networks. Software-defined networking (SDN) assumed centralized automation of network administration and replacement of expensive routers with cheaper switches. The SDN has undergone significant changes when it was extended from university labs to data centers and provider networks. As a result, in practice, SDN networks represent traditional IP networks with full-featured routers, only supplemented with SDN functionality. Following networks, the “software-defined” concept has penetrated the data storage area where the software-defined storage (SDS) has acquired a new important feature — service orientation. With the further extension of the “software-defined” concept to entire software-defined data centers (SDDC) and software-defined clouds, this feature became dominant. It can be expected that the success of SDDCs would make redundant independent SDNs and SDSs. However, the widespread implementation of SDDCs is hindered by difficulties of creating specific orchestration software. Therefore, for an ordinary data center owner, the only opportunity to acquire an SDDC is the purchase of an integrated turnkey solution with ready orchestration software from a reliable supplier.

Keywords: software-defined data center (SDDC); software-defined networking (SDN); software-defined storage (SDS)

DOI: 10.14357/08696527240206

EDN: KHJHNP

References

1. Egorov, V. B. 2022. К вопросу о соотношении программно определяемых и традиционных IP-сетей [Interrelation between the software-defined and conventional IP-networks]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):73–82. doi: 10.14357/08696527220107. EDN: OGWOXL.
2. Kontroller RunOS [RunOS controller]. 2017. *Run SDN*. Available at: <https://runsdn.ru/index.php/products/runos> (accessed April 11, 2024).

3. Kompaniya Programmiruemye Seti (Brain4Net) vypustila pervyy v Rossii SDN Controller operatorskogo klassa [The Programmable Networks company (Brain4Net) has released the first carrier-class SDN Controller in Russia]. 2020. *Brain4Net*. Available at: <https://brain4net.ru/brain4net-releases-carrier-grade-sdn-controller/> (accessed April 11, 2024).
4. Egorov, V. B. 2016. Nekotorye voprosy prakticheskoy realizatsii kontseptsii SDN [Some issues of the SDN concept practical implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(1):109–120. doi: 10.14357/08696527160108. EDN: VZBFNT.
5. Ganzha, D. 2018. Komponuemaya infrastruktura v kontekste SDDC [Composable infrastructure in the context of SDDC]. *Zh. setevykh resheniy/LAN [LAN/Network Solutions Magazine]* 4:13054573. Available at: <https://www.osp.ru/lan/2018/04/13054573> (accessed April 11, 2024).
6. Egorov, V. B. 2022. K voprosu o sovmeshchenii programmnoy opredeleniya setey s virtualizatsiyey setevykh funktsiy [On combining the software-defined networking with the network functions virtualization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):36–46. doi: 10.14357/08696527220204. EDN: MTZDDZ.
7. Egorov, V. B. 2023. Programmnoe opredelenie seti v konvergentnoy i giperkonvergentnoy infrastrukturakh [The software-defined networking in converged and hyperconverged infrastructures]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(1):105–113. doi: 10.14357/08696527230110. EDN: CAIIMP.
8. Egorov, V. B. 2021. Nekotorye voprosy programmnoy opredeleniya khranilishcha dannykh [Some issues of the software-defined storage]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):70–79. doi: 10.14357/08696527210207. EDN: UWZPYU.
9. Software defined data center (SDDC) definition. A White Paper from the OSDDC incubator. 2015. *DMTF*. Available at: <https://www.dmtf.org/sites/default/files/standards/documents/DSP-ISO501.1.0.0.pdf> (accessed April 11, 2024).
10. Dubova, N. 2014. Programmnyy dualizm [Software dualism]. *Otkrytye sistemy. SUBD [Open Systems. DBMS]* 6:14–16. EDN: SJKFNJ. Available at: <https://www.osp.ru/os/2014/06/13042309> (accessed April 11, 2024).
11. Kranz, G. 2017. SDDC (software-defined data center). *TechTarget*. Available at: <https://www.techtarget.com/searchdatacenter/definition/SDDC-software-defined-data-center> (accessed April 11, 2024).
12. Crockett, E. 2022. Top software-defined data center providers (SDDC). *Datamation*. Available at: <https://www.datamation.com/data-center/software-defined-data-center-providers/> (accessed April 11, 2024).
13. Kak postroit’ polnotsennoe gibridnoe oblako i snizit’ stoimost’ vladeniya IT-infrastrukturoy [How to build a full-fledged hybrid cloud and reduce the ownership cost of the IT-infrastructure]. 2021. *Cnews-SoftLine*. Available at: https://softline.cnews.ru/articles/2021-07-14_virtualnyj_tsod_i_gibridnoe_oblako (accessed April 11, 2024).
14. Egorov, V. B. 2020. Nekotorye voprosy programmnoy opredeleniya tsentrov obrabotki dannykh [Some issues of software-defined data centers]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(2):103–112. doi: 10.14357/08696527200210. EDN: FXXGJY.

15. Egorov, V. B. 2023. Nekotorye voprosy dezagregatsii i komponuemosti infrastruktury tsentra obrabotki dannykh [Some issues of disaggregation and composability of the data center infrastructure]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):101–110. doi: 10.14357/08696527230210. EDN: OEXVDB.
16. Knorr, E. 2012. What the software-defined data center really means. *InfoWorld*. Available at: <https://www.infoworld.com/article/2618097/what-the-software-defined-data-center-really-means.html> (accessed April 11, 2024).

Received November 29, 2023

Contributor

Egorov Vladimir B. (b. 1948)—Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; VEgorov@ipiran.ru

МОДЕЛИРОВАНИЕ ДИНАМИЧЕСКИХ ГЕООбЪЕКТОВ НА ОСНОВЕ КОНЦЕПЦИИ ФАЗОВОЙ ВАРИАНТНОСТИ

*Д. А. Никишин*¹

Аннотация: Предложена структура образца базы геоданных (БГД) с поддержкой фазовой вариантности (ФВ) геоделей. Данная структура представляет собой развитие разработанной ранее мультимасштабной БГД с поддержкой представления геообъектов в различных прикладных контекстах. Рассмотрена сущность ФВ, ее принципиальные отличия от полноценного темпорального моделирования (ПТМ). Сделан вывод о том, что ФВ может найти широкое применение в специализированных геоинформационных системах (ГИС), требующих при решении своих задач учета состояния объектов на конкретный момент времени. База геоданных с использованием ФВ может использоваться в качестве оперативной БГД, поскольку ее функционирование может быть отделено от процесса непрерывного уточнения знаний о местности, реализуемого ведением основной темпоральной БГД.

Ключевые слова: геоданные; базы геоданных; концептуальная схема базы геоданных; темпоральное моделирование геообъектов; фазовая вариантность геоделей

DOI: 10.14357/08696527240207

EDN: GСJQET

1 Введение

В работе [1] была представлена концептуальная схема БГД с множественным представлением геообъектов в контексте разных уровней обобщения геоданных. В публикации [2] эта схема была расширена поддержкой вариантности для отдельных прикладных задач. Настоящая работа продолжает развитие данного направления исследований и посвящена реализации представления геообъектов в разновременных состояниях.

При проектировании БГД для ГИС исторически сложился подход к статическому моделированию ситуации и периодической актуализации этой модели с выведением ее предыдущего («исторического») состояния за рамки оперативного доступа (в архив). Моделирование динамики ситуации используется достаточно редко [3], однако регистрация изменений семантических и геометрических свойств геообъектов остается актуальным направлением развития функциональности БГД, поскольку позволяет восстанавливать состояние геообъектов на определенный момент времени, а также исследовать явления и процессы, протекающие на местности [4–7].

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

Термин «темпоральность» (*англ.* temporality) применительно к БГД был предложен в [7] в качестве обобщения пространственного и семантического компонентов при моделировании геообъектов, изменяющихся во времени. Это обусловлено тем, что часто применяемый термин «пространственно-временное» моделирование представляется лишь частично отражающим суть моделирования динамики геообъектов.

Вместе с тем реализация темпорального аспекта, с одной стороны, целесообразна лишь для ограниченного круга постоянно изменяющихся геообъектов (в том числе движущихся), а с другой, обеспечение темпоральности сопряжено с некоторыми трудностями [7–9], связанными как со специфической организацией такой БГД, так и с необходимостью регулярной актуализации состояния местности — сложной и затратной по времени и ресурсам задачей¹. В качестве альтернативы для разрешения этих трудностей автором предлагается использование фазовой вариантности.

Далее в разд. 2 рассматривается сущность и примеры использования ФВ. В разд. 3 показаны принципиальные отличия ФВ и ПТМ. В разд. 4 представлены проектные решения и структуры данных, предназначенные для построения образца БГД, поддерживающей генерализационную, прикладную и фазовую вариантность моделей геообъектов.

2 Сущность фазовой вариантности геоданных

Под «фазовой» вариантностью подразумевается особое проявление вариантности, когда отдельный геообъект может быть представлен рядом моделей, соответствующих его состоянию в различные моменты времени («фазам» существования геообъекта). Но в отличие от ПТМ речь идет лишь о частной задаче моделирования — условием здесь служит возможность заранее смоделировать состояние объекта в результате этого события. В противном случае, если событие не предусматривалось или его последствия невозможно смоделировать априори, потребуется актуализация модели геообъекта путем полевого обследования, что выходит за рамки ФВ и будет, по сути, проявлением ПТМ.

Циклические изменения могут быть либо установлены из опыта по многократным предшествующим наблюдениям геообъекта, например сезонные изменения по многолетним наблюдениям, либо задаваться predetermined планом его изменения, например графиком работы объекта или его жизненным циклом. Также может быть предусмотрено принудительное «переключение» фазы геообъекта по сигналу о возникновении некоторого ожидаемого (вероятного) события, например чрезвычайной ситуации типа прорыва плотины. Итак, условием использования ФВ принимается установленный цикл, планируемые или ожидаемые изменения геообъекта.

¹Использование дистанционных методов, даже при условии решения проблем автоматизации дешифрирования и измерения изображений, требует дополнительного наземного обследования местности.

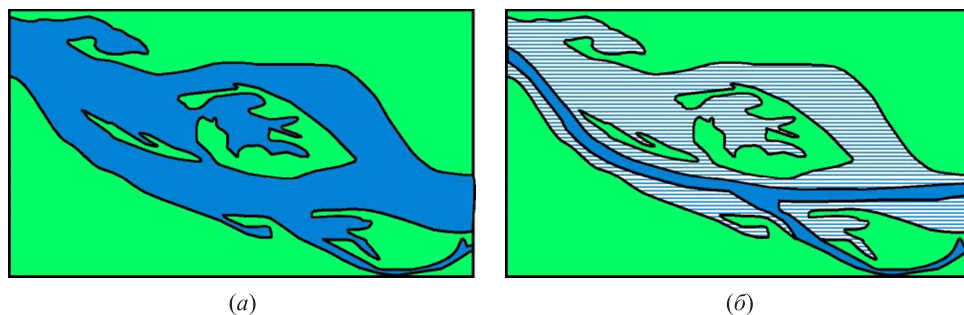


Рис. 1 Иллюстрация ФВ на примере сезонного цикла реки: (а) фаза половодья; (б) фаза межени (штриховкой показаны произошедшие изменения)

Характерным примером применения ФВ для сезонного цикла может служить водный режим реки, включающий фазы ее разлива (половодья), межени, становления, начала/окончания навигации и т. п. Так, во время половодья русло реки обычно расширяется, при этом возможно появление второстепенных русел (и соответственно островов между ними); в межень происходят обратные изменения (рис. 1). Учет этого обстоятельства может быть актуален при решении задач судоходства, движения вдоль берега, переправы через водную преграду, моделирования возможного подтопления местности и т. п. События становления/вскрытия реки — это пример, когда состояние геобъекта может влиять на взаимосвязанные объекты, например наличие/отсутствие зимнего пути («зимника»).

Актуальным примером цикла суточных колебаний служит приливно-отливные явления, обуславливающие изменение береговой линии. Это также может влиять на связанные объекты, например возможность движения по полосе осушки, обнажение подводных препятствий и т. п.

Примерами сезонных фаз также служат:

- периоды доступности/недоступности горного перевала для движения людей и транспорта и тому подобные случаи, что может определять возможность задействия соответствующего ребра графа транспортной сети при решении навигационных задач [10];
- вегетационные фазы растительности [11], влияющие на укрывающую способность местности, а также фазы пожароопасности, наличия клещей и тому подобных опасностей, что влияет на условия расположения людей и техники в массивах растительности.

Примерами плановых или ожидаемых изменений могут служить различные производственные циклы:

- цикл наполнения/срабатывания водохранилища — в целом этот цикл определяется природным циклом наполнения водохранилища в период половодья

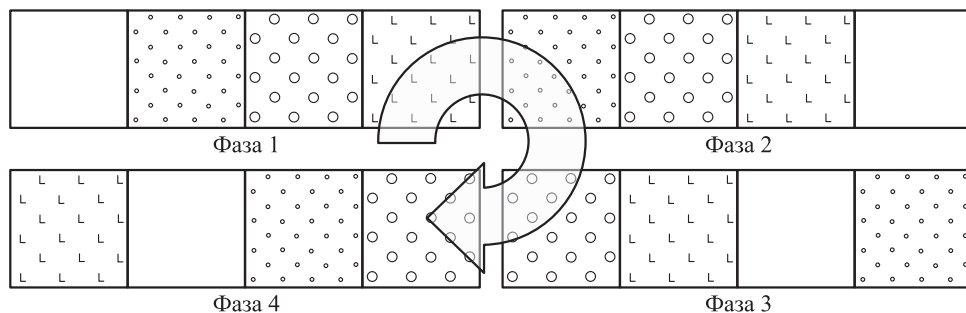


Рис. 2 Пример лесохозяйственного цикла: фаза 1 — пашня; фаза 2 — молодая поросль; фаза 3 — кондиционный лес; фаза 4 — вырубка

и срабатывания в межлетний период, но в данном случае также возможны принудительные действия, определяемые производственными, чрезвычайными и другими факторами;

- цикл угодий лесного хозяйства: пашня, молодая лесопосадка, высокоствольный лес, вырубка (рис. 2); цикл сельскохозяйственных угодий;
- этапы жизненного цикла зданий и сооружений (строительство/реконструкция, эксплуатация, консервация/заброшенность, руины) в рамках модели города; цикла накопления и вывоза сырья с открытых складов (площадок хранения), например угля, щебня, глины и т. п. в рамках производственной модели.

Как видно из приведенных примеров, ФВ (как и темпоральное моделирование вообще) применима (целесообразна) лишь для отдельных классов геообъектов, подмножеств объектов определенного класса объектов или отдельных объектов. Прочие, более «статичные» объекты могут актуализироваться с большой периодичностью (например, раз в 5–10 лет) или при возникновении определенных, разовых событий, сопровождаемых изменениями этих объектов.

В контексте топографической или морской карты ФВ применима лишь к ограниченному перечню объектов, хотя и позволяет более точно и качественно решать некоторые виды задач. Однако использование ФВ может найти более широкое применение в специализированных ГИС, требующих учета состояния своих объектов на конкретный момент времени. В частности, применением модели ФВ может стать реализация интерактивных моделей объектов и инфраструктур, в том числе для актуальной на сегодня парадигмы «цифрового двойника»¹.

¹Цифровой двойник — это цифровая (виртуальная) модель любого объекта, системы и т. п. Она точно воспроизводит форму и действия оригинала и синхронизирована с ним.

3 Сравнение фазовой вариантности и полноценного темпорального моделирования

Обе парадигмы (как ПТМ, так и ФВ) направлены на отражение процессов, происходящих на местности. Они позволяют восстанавливать ретроспективные состояния, а также прогнозировать будущее развитие ситуации на местности.

Однако, как говорилось выше, ФВ-модель отражает либо запланированные изменения, либо уже установленную ранее зависимость циклического изменения объекта во времени. Но для выявления такой зависимости необходимо иметь некий накопленный «темпоральный опыт» (многолетние и тому подобные наблюдения). Соответственно, создание ФВ-модели в общем случае предполагает наличие некой базовой темпоральной БГД, которая служит основой для разработки и уточнения циклов фазовой БГД.

Вместе с тем ФВ-модель по сравнению с ПТМ представляет собой более простой и компактный инструмент: с одной стороны, он позволяет отражать динамику геобъектов и отчасти прогнозировать ее; с другой стороны, ее функционирование отделено от процесса непрерывного уточнения знаний о местности. Так, ПТМ предполагает регулярную актуализацию информации о состоянии местности или отдельных геобъектов. Соответственно, ПТМ уместно лишь при необходимости продолжающихся исследований происходящих на местности процессов и соответствующих возможностях регулярного сбора информации, ее сохранения и последующего анализа. В противном случае возникают производительные осложнения, которые затрудняют применение БГД с поддержкой ПТМ (БГД-ПТМ) для решения повседневных, оперативных задач. Использование БГД с поддержкой ФВ моделей (БГД-ФВ) в качестве надстройки над БГД-ПТМ позволит оптимизировать решение этих оперативных задач. При этом БГД-ФВ может функционировать независимо от ведения БГД-ПТМ. Знания, заложенные в модели БГД-ФВ, могут лишь при необходимости актуализироваться по мере совершенствования знаний о геобъектах и происходящих с их участием процессах в базовой БГД-ПТМ.

4 Структура базы геоданных с поддержкой фазовой вариантности

По аналогии с ПТМ, ФВ-модель может реализовываться как временной ряд статических моделей-«кадров» с интерполяцией состояния между ними. Впоследствии может быть осуществлена аппроксимация этого временного ряда параметрической моделью [8], что даст возможность непосредственно интерполировать ситуацию на произвольный момент времени. В данной работе далее представлен только первый вариант.

Как и для ПТМ, в случае ФВ в состав каждой отдельной модели нужно ввести свойство «период актуальности модели», соответствующий периоду нахождения объекта в состоянии, соответствующем этой модели. Для определения моментов смены фаз должны быть выработаны критерии изменения состояния

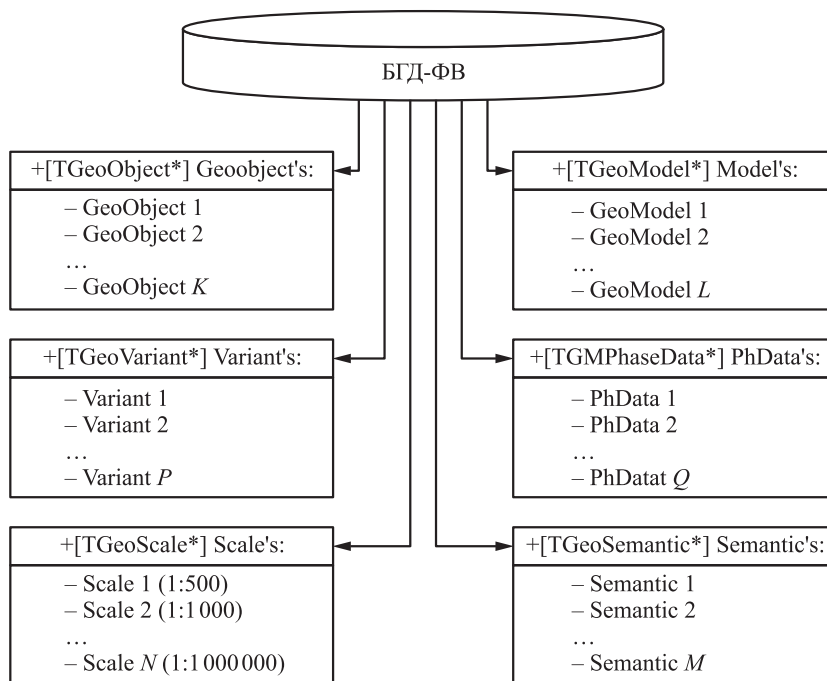


Рис. 3 Структура БГД-ФВ

объекта, обуславливающие переход от текущей к другой модели. Такими критериями могут быть определенные существенные изменения контура и других, непространственных свойств геообъекта. Основой для выделения фаз могут стать экстремальные состояния геообъекта; также возможны несколько промежуточных вариантов между этими состояниями.

На рис. 3 и 4 представлена¹ разработанная автором структура БГД-ФВ. Данная структура представляет собой развитие более ранней разработки мультимасштабной БГД с поддержкой представления геообъектов в различных прикладных контекстах [1, 2]. Исходным прототипом служит рассмотренная в [12] концептуальная схема БГД с поддержкой вариантного и темпорального аспектов геоданных.

На рис. 3 представлена структура БГД-ФВ. Она включает в себя 6 основных таблиц, 5 из которых были рассмотрены в [1, 2] и по своей структуре соответствуют основным концептам «геообъект» (TGeoObject), «геомодель» (TGeoModel), «семантическое свойство» (TGeoSemantic), «масштабный уровень» (TGeoScale) и «прикладной контекст» (TGeoVariant). Новая здесь только таблица, соответ-

¹В целях более четкого выделения структур данных и сокращения объема на рис. 3 и 4 некоторые элементы структур опущены или оформлены в упрощенном виде, с применением псевдокода.

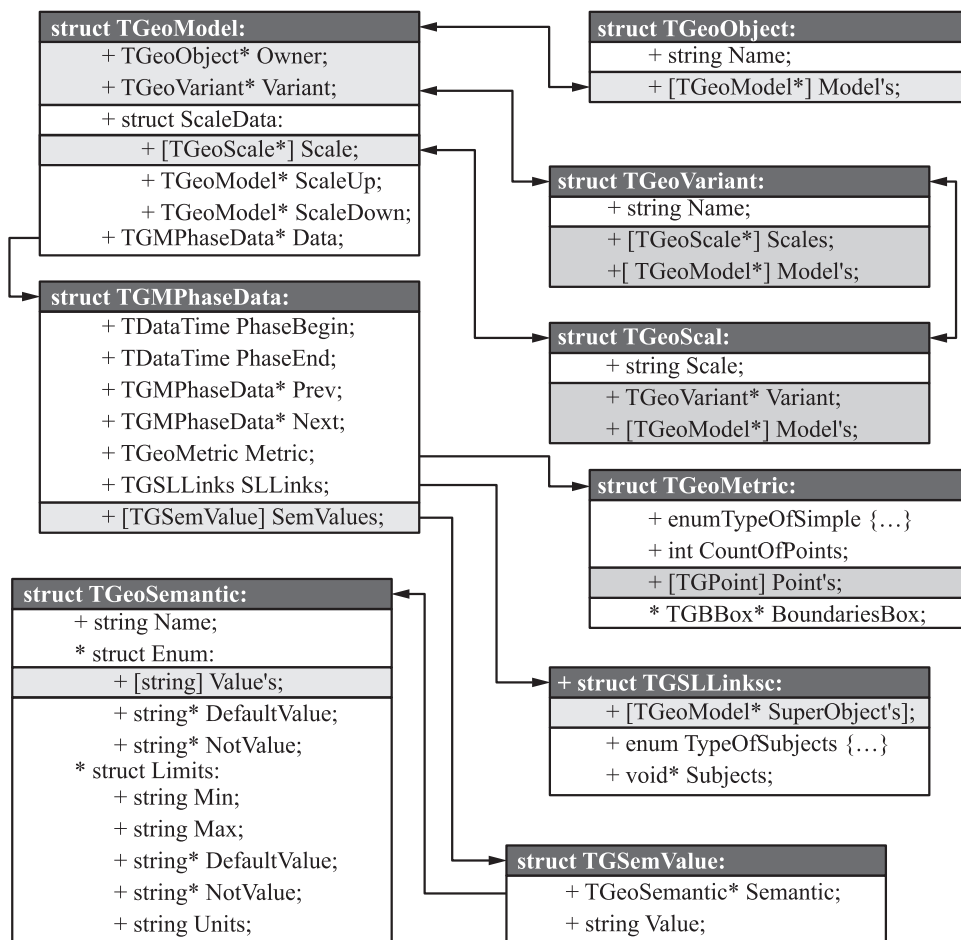


Рис. 4 Основные концепты данных для БГД-ФВ

ствующая концепту «фаза геобъекта» (TGMPhaseData). Эти таблицы содержат основные данные БГД и служат для обращения к соответствующим информационным объектам.

На рис. 4 представлены основные концепты БГД-ФВ и связей между ними. Эти концепты рассматриваются как структуры данных соответствующих информационных объектов БГД.

Большинство представленных на рис. 4 концептов были рассмотрены ранее в [1, 2], поэтому ниже остановимся только на изменениях.

1. В используемой ранее структуре данных модели геобъекта TGeoModel была выделена структура TGMPhaseData, которая унаследовала от TGeoModel

элементы `Metric`, `SLLinks` и `SemValues`, содержащие основные данные гео-модели. Это выделение сделано для обеспечения возможности приурочивать к одной структуре `TGeoModel` несколько вариантов метрики и семантики с соответствующими логическими связями.

Помимо этого структура `TGMPhaseData` была снабжена полями `PhaseBegin` и `PhaseEnd`, служащими для задания моментов начала и окончания периода актуальности соответствующих данных модели. Для обеспечения преемственности между смежными фазами также добавлены указатели `Prev` и `Next` на соответствующие объекты данных.

2. В самом концепте `TGeoModel` сохранены лишь данные привязки к соответствующему масштабному уровню, контексту использования, а также к гео-объекту-владельцу данной модели. Вместо остальных элементов в структуру `TGeoModel` введен указатель на один объект типа `TGMPhaseData`, что соответствует модели обычного, «стабильного» геообъекта. В случае наличия нескольких фазовых вариантов этому указателю соответствует одна из моделей (например, соответствующая текущему моменту времени). Если решаемая задача предполагает задание другого момента времени, то выполняется поиск соответствующей этому моменту фазового варианта при помощи указателей на смежные модели `Prev` и `Next`. Такое решение позволило избежать усложнения структуры данных для геообъектов, к которым нецелесообразно применение ФВ.

Практическая реализация БГД-ФВ предусматривается в виде создания специального программного обеспечения, возможно, с использованием в качестве основы стандартной реляционной системы управления базами данных.

Идея ФВ геообъектов, по мнению автора, обладает элементом новизны. А попытка реализации структуры БГД-ФВ геообъектов позволяет говорить о ее существенной новизне. При этом, несмотря на преемственность, структура БГД-ФВ потребовала существенной переработки структуры БГД, разработанной автором ранее и представленной в [2], что сделало эти структуры не полностью совместимыми.

Важной особенностью БГД-ФВ стала не просто реализация разновременных вариантов моделей геообъектов, а возможность объединения в одной БГД согласованных между собой множественных представлений геообъектов в аспектах генерализации, контекста использования и этапа жизненного цикла геообъекта. Это позволяет циклу ФВ опираться на представления геообъекта, имеющие место на смежных уровнях генерализации, в соответствующем контексте использования.

5 Заключение

Моделирование динамики изменений на местности и выявление протекающих на местности процессов остается актуальным направлением развития ГИС.

Вместе с тем реализация ПТМ сопряжена с рядом трудностей. В качестве альтернативного решения автором предлагается использование ФВ, предназначенной для моделирования циклических процессов или процессов, последствия которых могут быть predetermined заранее. Условием использования ФВ служит наличие установленного цикла либо априорное знание планируемых или ожидаемых изменений геообъекта.

Для выявления циклической зависимости изменений объекта во времени необходимо иметь накопленный «темпоральный опыт» (многолетние и тому подобные наблюдения), что в общем случае предполагает наличие соответствующей темпоральной БГД, которая будет служить основой для разработки и уточнения циклов фазовой БГД.

Вместе с тем ФВ-модель по сравнению с ПТМ представляет собой более простой и компактный инструмент: с одной стороны, она позволяет отражать динамику геообъектов и отчасти прогнозировать ее; с другой стороны, ее функционирование отделено от процесса ведения БГД-ПТМ, особенности реализации которой затрудняют решение оперативных задач. Использование БГД-ФВ в качестве надстройки над БГД-ПТМ позволит оптимизировать решение этих задач. При этом БГД-ФВ может функционировать независимо от процесса непрерывного уточнения знаний о местности, реализуемого ведением БГД-ПТМ. Знания, заложенные в модели БГД-ФВ, могут лишь при необходимости актуализироваться по мере совершенствования знаний о геообъектах и происходящих с их участием процессах в базовой БГД-ПТМ.

Фазовая вариантность может найти более широкое применение в специализированных ГИС, требующих при решении своих задач учета состояния объектов на конкретный момент времени. Возможным применением ФВ может стать реализация интерактивных моделей объектов и инфраструктур.

Литература

1. *Никишин Д. А.* Вариант концептуальной схемы базы геоданных с поддержкой обратимой генерализационной связанности моделей геообъектов // Системы и средства информатики, 2021. Т. 31. № 2. С. 119–128. doi: 10.14357/08696527210211. EDN: KСJYА.
2. *Никишин Д. А.* Подход к разработке концептуальной схемы базы геоданных с поддержкой многовариантного представления геообъектов // Системы и средства информатики, 2022. Т. 32. № 2. С. 128–136. doi: 10.14357/08696527220212. EDN: IXRLAA.
3. *Laurini R.* Some philosophical issues regarding geometric modeling for geographic information and knowledge systems // The philosophy of GIS / Ed. T. Tambassi. — Springer geography ser. — Cham: Springer, 2019. P. 25–50. doi: 10.1007/978-3-030-16829-2_2.
4. *Das M., Ghosh S. K.* Modeling spatio-temporal change pattern using mathematical morphology // 3rd IKDD Conference on Data Science Proceedings. — New York, NY, USA: ACM, 2016. Art. 4. 10 p. doi: 10.1145/2888451.2888458.

5. *Atluri G., Karpatne A., Kumar V.* Spatio-temporal data mining: A survey of problems and methods // *ACM Comput. Surv.*, 2018. Vol. 51. Iss. 4. Art. 83. 41 p. doi: 10.1145/3161602.
6. *Wang Qing, Lv Weifeng, Du Bowen.* Spatio-temporal anomaly detection in traffic data // 2nd Symposium (International) on Computer Science and Intelligent Control Proceedings. — New York, NY, USA: ACM, 2018. Art. 46. 5 p. doi: 10.1145/3284557.3284725.
7. *Никишин Д. А.* Структура и особенности генерализации в контексте функционирования темпоральной базы геоданных // *Системы и средства информатики*, 2020. Т. 30. № 1. С. 147–159. doi: 10.14357/08696527200113. EDN: SYGNFF.
8. *Siabato W., Claramunt C., Harri S., Manso-Callejo M. A.* A survey of modelling trends in temporal GIS // *ACM Comput. Surv.*, 2018. Vol. 51. Iss. 2. Art. 30. 41 p. doi: 10.1145/3141772.
9. *Никишин Д. А.* Обзор подходов к пространственно-временному моделированию и выявление основных тенденций развития T-GIS // *Системы и средства информатики*, 2022. Т. 32. № 3. С. 50–62. doi: 10.14357/08696527220305. EDN: TOWSDH.
10. *Jilani M., Corcoran P., Bertolotto M.* Automated highway tag assessment of OpenStreetMap road networks // 22nd ACM SIGSPATIAL Conference (International) on Advances in Geographic Information Systems Proceedings. — New York, NY, USA: ACM, 2014. P. 449–452. doi: 10.1145/2666310.2666476.
11. *Wang H., Chen X., Polys N., Sforza P.* A Web3D forest geo-visualization and user interface evaluation // 22nd Conference (International) on 3D Web Technology Proceedings. — New York, NY, USA: ACM, 2017. Art. 9. 9 p. doi: 10.1145/3055624.3075956.
12. *Дулин С. К., Никишин Д. А.* Подходы к интеграции прикладных концептуальных схем в составе унифицированной геоонтологии // *Системы и средства информатики*, 2020. Т. 30. № 2. С. 68–77. doi: 10.14357/08696527200207. EDN: VKGXOP.

Поступила в редакцию 05.03.24

AN APPROACH TO MODELING DYNAMIC GEOOBJECTS BASED ON THE CONCEPT OF PHASE VARIATION

D. A. Nikishin

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

Abstract: A structure and basic design solutions are proposed for constructing a sample geodatabase with support for phase variation of geomodels. This structure is a development of the previously developed multiscale geodatabase with support for representing geobjects in various application contexts. The

essence of phase variation and its fundamental differences from full-fledged temporal modeling are considered. It is concluded that the proposed approach can be widely used in specialized geographic information systems that require taking into account the state of their objects at a specific point in time as an operational geodatabase, since its operation can be separated from the process of continuous refinement of knowledge about the area implemented by maintaining the main temporal geodatabase.

Keywords: geodata; geodata databases; conceptual scheme of the geodata database; temporal modeling of geobjects; phase variation of geomodels

DOI: 10.14357/08696527240207

EDN: GCJQET

References

1. Nikishin, D. A. 2021. Variant kontseptual'noy skhemy bazy geodannykh s podderzhkoy obratimoy generalizatsionnoy svyazannosti modeley geoob"ektov [A variant of the conceptual schema of the geodata database with support for reversible generalization connectivity of geo object models]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):119–128. doi: 10.14357/08696527210211. EDN: KCJJYA.
2. Nikishin, D. A. 2022. Podkhod k razrabotke kontseptual'noy skhemy bazy geodannykh s podderzhkoy mnogovariantnogo predstavleniya geoob"ektov [An approach to the development of a conceptual scheme of a geodata database with support for a multivariate representation of geobjects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):128–136. doi: 10.14357/08696527220212. EDN: IXRLAA.
3. Laurini, R. 2019. Some philosophical issues regarding geometric modeling for geographic information and knowledge systems. *The philosophy of GIS*. Ed. T. Tambassi. Springer geography ser. Cham: Springer. 25–50. doi: 10.1007/978-3-030-16829-2_2.
4. Das, M., and S. K. Ghosh. 2016. Modeling spatio-temporal change pattern using mathematical morphology. *3rd IKDD Conference on Data Science Proceedings*. New York, NY: ACM. Art. No. 4. 10 p. doi: 10.1145/2888451.2888458.
5. Atluri, G., A. Karpatne, and V. Kumar. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Comput. Surv.* 51(4):83. 41 p. doi: 10.1145/3161602.
6. Wang, Qing, Weifeng Lv, and Bowen Du. 2018. Spatio-temporal anomaly detection in traffic data. *2nd Symposium (International) on Computer Science and Intelligent Control Proceedings*. New York, NY: ACM. Art. No. 46. 5 p. doi: 10.1145/3284557.3284725.
7. Nikishin, D. A. 2020. Struktura i osobennosti generalizatsii v kontekste funktsionirovaniya temporal'noy bazy geodannykh [Structure and features of generalization in the context of functioning of a temporal geodatabase]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1):147–159. doi: 10.14357/08696527200113. EDN: SYGNFF.
8. Siabato, W., C. Claramunt, S. Ilarri, and M. A. Manso-Callejo. 2018. A survey of modelling trends in temporal GIS. *ACM Comput. Surv.* 51(2):30. 41 p. doi: 10.1145/3141772.
9. Nikishin, D. A. 2022. Obzor podkhodov k prostranstvenno-vremennomu modelirovaniyu i vyyavlenie osnovnykh tendentsiy razvitiya T-GIS [Overview of approaches

- to space–time modeling and the main trends in the development of T-GIS]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):50–62. doi: 10.14357/08696527220305. EDN: TOWSDH.
10. Jilani, M., P. Corcoran, and M. Bertolotto. 2014. Automated highway tag assessment of OpenStreetMap road networks. *22nd ACM SIGSPATIAL Conference (International) on Advances in Geographic Information Systems Proceedings*. New York, NY: ACM. 449–452. doi: 10.1145/2666310.2666476.
 11. Wang, H., X. Chen, N. Polys, and P. Sforza. 2017. A Web3D forest geo-visualization and user interface evaluation. *22nd Conference (International) on 3D Web Technology Proceedings*. New York, NY: ACM. Art. No. 9. 9 p. doi: 10.1145/3055624.3075956.
 12. Dulin, S. K., and D. A. Nikishin. 2020. Podkhody k integratsii prikladnykh konseptual'nykh skhem v sostave unifikirovannoy geoontologii [Approaches to the integration of the application conceptual schemas in the unified geontology]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(2):68–77. doi: 10.14357/08696527200207. EDN: VKGXOP.

Received March 5, 2024

Contributor

Nikishin Dmitry A. (b. 1976)— Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; dmnikishin@mail.ru

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В МЕДИЦИНЕ: ПОНЯТИЕ, ПРОБЛЕМЫ, ПОДХОДЫ К РАЗРАБОТКЕ*

*С. Б. Румовская*¹

Аннотация: В неоднородной проблемной среде диагностики и прогностики в медицине, технике и т. д. представления об объекте и процессе как гетерогенной целостной системе не завершены в теории и не стали картиной мира узкопрофессиональных специалистов. Принятие решений в условиях ограниченности ресурсов, неоднозначности оценки ситуации, большого объема и разнообразия перерабатываемой информации об объекте сопровождается серьезными ошибками и рисками, в том числе и в медицине. Необходимы технологии виртуализации принятия решений. В работе рассмотрено понятие интеллектуальных систем поддержки принятия решений (ИСППР) в медицине, особенности и проблемы предметной области, а также подходы к разработке.

Ключевые слова: интеллектуальная система поддержки принятия решений; медицина и здравоохранение; проблемы разработки; подходы к разработке

DOI: 10.14357/08696527240208

EDN: IJDVVV

1 Введение

Принятие решений в условиях ограниченности времени и ресурсов, дефицита специалистов, большого объема и разнообразия перерабатываемой, зачастую атипичной и многозначной, информации об объекте в медицине сопровождается серьезными ошибками и рисками: по разным оценкам, примерно 50% россиян хотя бы раз сталкивались с ошибками врачей [1], и вследствие этого более чем у 70 тыс. человек в год возникали тяжелые осложнения [2]. Положение можно нивелировать путем повышения уровня поддержки персонализированной медицины информационными технологиями, в частности ИСППР. В работе рассмотрены: понятие ИСППР в медицине, особенности и проблемы предметной области (горячие точки, барьеры разработки и внедрения ИСППР), а также подходы к разработке.

2 Интеллектуальные системы поддержки принятия решений в медицине. Понятие, особенности и проблемы

Системы поддержки принятия решений (СППР) — компьютерная система, использующая базы знаний (БЗ), базы данных, базы моделей и представля-

* Исследование выполнено за счет гранта Российского научного фонда № 24-21-00392.

¹ Федеральний исследовательский центр «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

ющая выводы, оценки возможных альтернативных вариантов, рекомендации решения проблемы [3]. Интеллектуальные СППР сочетают строгие модели и методы поиска решения с нестрогими, эвристическими, базирующимися на знаниях специалистов-экспертов, моделях человеческих рассуждений, неклассических логиках и накопленном опыте [4].

Интеллектуальные СППР в медицине [5, 6] нацелены на снижение ошибок и повышение качества оказываемой медицинской помощи. Это системы, имитирующие рассуждения врача (построенные с применением методов математической статистики, нейронных сетей, математической логики) или моделирующие их — экспертные системы (ЭС) с БЗ. Кобринский [7] также выделяет системы на прецедентах (CBR, case-based reasoning); гибридные интеллектуальные системы (ГиИС) интегрируют различные технологии искусственного интеллекта (ИИ), модели физиологических систем организма и т. д.; системы с «доской объявлений», частично моделирующие работу консилиума.

Разработка и внедрение ИСППР в медицине осложняется особенностями слабоструктурированной предметной области [8]: необходимо учитывать и совокупность взаимосвязанных понятий и признаков, и интуитивную и визуальную составляющие клинического мышления опытных врачей; болезни не имеют постоянного симптомокомплекса; по ходу болезни диагностическая ценность признаков меняется; диагноз часто ставится на оттенках симптомов и признаков; субъективность оценок; неполнота признаков на ранних стадиях и при атипичных формах; нетрадиционное течение заболеваний и многоликость. К настоящему времени выделяют ряд проблем (горячих точек) в разработке медицинских интеллектуальных систем (МИС), в том числе ИСППР [8–14]. Рассмотрим их кратко далее.

Проблема получения полноценных и достоверных знаний. На характер экспертных знаний влияет отношение к научным школам, лево- и правополушарность экспертов, их рефлексивность. Необходима интеграция знаний экспертов. Дополнительные барьеры здесь: колоссальный объем накопленных знаний; искажение первичных медицинских знаний в ходе клинических исследований (данные с отрицательными результатами обычно не публикуются); отсутствие/недостоверность знаний — методы диагностики и лечения, базирующиеся на клинических рекомендациях, не всегда эффективны.

Когнитивные образы — непосредственная операция над образами для формирования диагностических гипотез пока недостижима.

Когнитивный барьер — связан с недостаточным обоснованием и пониманием метода получения рекомендуемого решения, в частности в технологии машинного обучения (МО). Исследования в данном направлении активно ведутся последние 10 лет, и есть ряд работ, например [15–17], освещающих понятие объяснимого ИИ, его методы повышения объяснимости и их применение в медицинских ИСППР (например, *post hoc* методы: LIME — Local Interpretable Model-agnostic Explanations; SHAP — SHapley Additive exPlanations и др.) и т. д.

Проблема полиморбидности — наличие нескольких заболеваний, одновременно не связанных и связанных единым доказанным патогенетическим и генетическим механизмом. Среди последних выделяют конкурирующие (каждое может привести к смерти), сочетанные (в совокупности могут привести к смерти) и фоновые (влияют на патогенез основной патологии, а также становятся предикторами тяжести ее течения и осложнений, угрожающих жизни). Однако преимущественно их особенности в ИСППР или не учитываются совсем, или учитываются фрагментарно.

Ассистирование на всех этапах лечебно-диагностического процесса и персонализация пользователя — необходимы системы, включающие взаимосвязанные БЗ, учитывающие механизм развития заболевания, одновременные факты о больном и характер лечебно-профилактических мероприятий. Также надо учитывать специализацию и уровень врача.

«Погружение» интеллектуальных систем поддержки принятия решений в информационные системы электронных медицинских карт. На данном пути возникает барьер недостаточной формализации и стандартизации данных — каждая система имеет собственную организацию данных, интероперабельности нет, а множественность оборудования влечет разнородность визуальных изображений. Остается проблема отсутствия доступных для исследователей банков обезличенных клинических данных.

Барьер на пути к широкому распространению медицинских интеллектуальных систем — нет требований к полноте и прозрачности документации МИС и к контролю в процессе эксплуатации. Первая в мире серия национальных стандартов по системам ИИ в медицине опубликована в [14]. Перечень разработанных в мире к настоящему времени документов по доверенному ИИ в медицине и здравоохранении, а также принципы, требования и механизмы регулирования разработки, применения и надзора за использованием технологий ИИ в российском здравоохранении представлены в [13]. Применение ИСППР повысит эффективность работы врачей, а также продолжительность и качество жизни граждан. Анализ публикаций в научной литературе по теме ИСППР в медицине, представленный далее, показал, что разработки и исследования в этой области активно ведутся во всем мире.

3 Подходы к разработке интеллектуальных систем поддержки принятия решений в медицине

3.1 Машинное обучение

Хасанов с соавт. [18] предложили рекуррентную искусственную нейронную сеть (ИНС) для прогнозирования динамики развития рожистой патологии: 27 нейронов на входе, 6 — в скрытом слое, 1 — на выходе.

Kui с соавт. [19] предложена предиктивная модель раннего выявления пациентов с высоким риском развития тяжелого острого панкреатита (ТОП) в течение

нескольких часов после госпитализации на базе XGBoost (eXtreme Gradient Boosting) и SHAP (для определения дескрипторов и их вклада в предсказание). В [20] также предложена модель на базе XGBoost для выявления случаев ТОП, а в [21] для раннего выявления ТОП исследовались модели логистической регрессии (ЛР) и пять алгоритмов AutoML (Automated Machine Learning) — лучшие результаты дал градиентный бустинг (GBM, gradient boosting machine). В [22] модель GBM прогнозирует сложность эндоскопической резекции гастроинтестинальной стромальной опухоли (размер опухоли и опыт эндоскопистов влияли на эффективность модели).

В [23] разработана модель вероятности развития парафарингеального абсцесса у пациентов с острыми воспалительными заболеваниями глотки. При анализе взаимосвязи между качественным признаком, выступающим в роли зависимого результирующего показателя, и подмножеством количественных признаков использовалась ЛР с пошаговыми алгоритмами включения и исключения предикторов. Выделенные предикторы ранжировали по степени силы связи с зависимой переменной методом сортировки предикторов по модулю стандартизованных коэффициентов регрессии.

В [24] описан индекс-тест — распознавание и анализ рентгенограмм (РГ) на основе трех сверточных нейронных сетей (СНС) типа U-NET. Сформированы две тестовые выборки: 140 РГ с соотношением норма/патология 50/50; и 150 — с 95/5. Индекс-тест применим для формирования групп риска по туберкулезу.

Cho с соавт. [25] использовали модель XGBoost для классификации поражений коронарных артерий как имеющих фракционный резерв кровотока $\leq 0,8$ по 12 физиологическим особенностям.

Тао с соавт. [26] предложили Bottleneck Residual U-Net для точной сегментации коронарной артерии на ангиограмме.

В [27] предложена модель многослойного перцептрона с одним входным, тремя скрытыми и одним выходным слоем для расчета риска сердечно-сосудистых заболеваний (ССЗ). Искусственная нейронная сеть повысила точность моделирования на 8,84% по сравнению с фрамингемской шкалой, которая способна наиболее точно определять здоровых пациентов, но не истинно больных.

В [28] предложена модель прогнозирования тяжелых миелотоксических (ТМ) осложнений химиотерапии (ХТ). База данных исследования — 47 552 курса ХТ (включают клинические и лабораторные показатели до и после ХТ). Регрессионная функция строилась на основе одной наиболее значимой переменной, затем по мере добавления оставшихся переменных оценивалось воздействие каждой. Построенные функции затем модифицировались вручную путем замены переменных на коллинеарные и инженерии признаков.

В [29] предлагается модель для выявления однонуклеотидных полиморфизмов, связанных с диабетом 2-го типа, и для прогнозирования риска его развития на базе случайного леса (RF, random forest) для поиска наиболее важных атрибу-

тов, а также с помощью метода опорных векторов и ЛР. Алгоритм RF превзошел другие методы. Атрибуты, полученные от RF, использовались для прогнозирования методом k-ближайших соседей.

Есть ряд разносторонних обзоров, рассматривающих применение методов МО для интеллектуальных систем поддержки принятия врачебных решений (ИСППВР) в разных областях медицины (см., например, [30, 31]).

3.2 Экспертные системы

В работе [32] описывается метод разработки СППР для назначения комплексного лечения. Он основывается на создании задачеориентированной оболочки, ключевой компонент которой — обобщенная онтология, позволяющая формировать знания о различных видах лечения (медикаментозном, восстановительном, хирургическом) вне зависимости от раздела медицины. Онтология реализована на облачной платформе IASaaS (Intelligent Applications, Control, and Platform as a Service) [33].

В [34] предложен прототип ЭС для диагностики подагры, в котором структура БЗ разработана с использованием онтологического подхода и элементов нечеткой логики, семантических сетей и фреймов (у объектов БЗ один из атрибутов — степень уверенности экспертов в однозначности определения заболевания симптомами, синдромами или их комбинациями).

Есть ряд нечетких ЭС (см., например, [35]). В [36] предложен метод оценки необходимости ортодонтического лечения пациентов с постоянным прикусом на базе байесовской сети: узел–переменная, а направленные дуги — причинно-следственные связи между переменными.

В [37] темпоральные CBR (Case-Based Reasoning) системы консультируют по поводу болюсного введения инсулина при сахарном диабете 1-го типа, используя временные последовательности для учета предшествующих событий при рекомендации болюсных доз инсулина, что позволяет изучать и улучшать прогнозы болюса.

В [38] описана концепция построения ИСППР на базе CBR (Case-Based Reasoning) в психиатрии, рекомендующих схемы лечения пациента и оценивающих принятое врачом решение с целью снижения риска повторной госпитализации. Модифицированный CBR-цикл реализует предварительную проверку пациента на предмет наличия обслуживания его случая ранее (учет временных характеристик ремиссии). Если найден, то сначала пересматривается прежнее решение, затем текущий случай направляется по классическому CBR-циклу.

В [39] для диагностики рака поджелудочной железы и хронического панкреатита применен метод автоматизированной поддержки исследований с пополняемыми базами фактов на основе обнаружения эмпирических закономерностей посредством ДСМ-рассуждений, порождающих гипотезы о предсказаниях.

3.3 Гибридные интеллектуальные системы

В [40] предлагается система для прогнозирования зависимости «структура–свойства» лекарственных соединений. Многоагентный и онтологический подходы позволили структурировать входные и выходные данные, оптимально распределить вычислительные ресурсы и координировать работу системы. Для прогнозирования свойств лекарственных соединений использовался алгоритм моделирования иммунной сети (МИС). Для выбора информативных дескрипторов использовался алгоритм серого волка (АСВ).

В [41] описано моделирование ГиИС коллективной выработки решений консиллиумами. Инструментальная среда «Виртуальный консиллиум» (ВК) включает две ИНС, девять нечетких систем (НС), две ЭС и девять генетических алгоритмов. Среда ВК разрабатывалась на примере диагностики артериальной гипертензии как модульная, гибкая, с библиотекой всех моделей.

В [42] предложена диагностическая система, комбинирующая однослойную ИНС и НС, для автоматического выявления астмы и хронической обструктивной болезни легких (ХОБЛ) в клинических условиях. Сначала запускается алгоритм предварительной классификации на базе анкеты с целью определения вероятности наличия ХОБЛ или астмы. Если вероятность $> 50\%$, то делается спирометрия, результаты которой подаются на вход ИНС-НС, которая ставит диагноз или запрашивает дополнительные тесты, после чего снова запускается.

Результаты анализа подходов из разд. 3 сведены в таблицу.

Интеллектуальные системы поддержки принятия врачебных решений (см. таблицу) — преимущественно системы, использующие методы одного какого-то класса. Интеллектуальные нейронные сети и ансамблевые методы превалируют последние несколько лет и хорошо себя зарекомендовали для решения различных задач в медицине, однако их решения непрозрачны (несмотря на начавшееся распространение методов объяснимого ИИ), а также не позволяют учитывать все типы переменных, которыми оперируют врачи в рассуждениях, в частности нечеткость. Имеет место акцент на развитие разработок ИСППВР как ГиИС.

4 Заключение

Диагноз и прогноз — показатели понимания и объяснения состояния больного. Представление о динамике и направлении развития заболевания формируется у организма как единого целого в рамках дифференциальной прогностики. В условиях гетерогенной, неспецифической, непредсказуемой клинической ситуации требуется коллективная выработка диагностических и прогностических решений с позиции сотрудничества, относительности и дополненности знаний, поэтому требуется синтез и/или реконфигурирование интегрированной прозрачной модели рассуждений коллектива специалистов для ИСППВР.

Анализ отечественных и зарубежных подходов к разработке ИСППР в медицине

Метод	Тип	Работы/назначение	Результаты
ИНС	Рекуррентные	[18] Прогнозирование развития рожистой	Ошибка прогноза — 0,21%
	МСП	[27] Расчет риска ССЗ	AUC — 0,84; точность — 78,8%; Se — 79%
DL	СНС	[24] Оценка диагностической точности системы скрининга туберкулеза легких	Для выборок 1 и 2 соответственно: Se — 87,2% и 75,0%; Sp — 60,0% и 53,5%
		[26] Анализ ангиограмм	AUC — 0,991; Se — 88%; Sp — 97,9%
Ансамблевые методы	XGBoost, SHAP	[19] Раннее выявление пациентов с высоким риском развития ТОП	AUC — 0,81; точность — 89,1%
		[20] Выявление случаев ТОП	AUC — 0,921
	AutoML (GBM), SHAP, LIME	[25] Прогнозирование фракционного резерва кровотока по ангиографии	AUC — 0,87; Se — 84%; Sp — 80%
		[21] Раннее выявление ТОП	AUC — 0,945 в тестовом наборе; Se — 58,3%
Регрессия логистическая	RF и k-NN	[22] Прогнозирование сложности эндоскопической резекции	Точность в проверочной и тестовой когортах — 93,5% и 91,1% соответственно
		[29] Прогнозирование риска диабета 2-го типа	Точность — 85,3%
	[23] Прогноз парафарингеального абсцесса	Se — 98,2%, Sp — 96,2%	
		[28] Оценка риска развития осложненной химиотерапии	Для всех видов миелотоксичности: Se — от 70,0% до 90,1%, Sp — от 60,2% до 79,9%

Окончание таблицы на с. 114

Анализ отечественных и зарубежных подходов к разработке ИСПП в медицине (окончание)

Метод	Тип	Работы/назначение	Результаты
ЭС	Онтологический подход	[32] Метод построения ИСПП по назначению персонализированного лечения	Снижает затраты на разработку и сопровождение ИСПП
	Нечеткие	[34] Диагностика подагры	AUC — 0,954; Se — 91,3%; Sp — 85,4%
	Вероятностные	[35] Диагностика и лечение рака груди	Точность системы — 90%
	СВР	[36] Оценка необходимости ортодонтического лечения	Высокая степень согласия между ИСПП и опытными ортодонтами: $\kappa > 0,894$
	ДСМ-метод	[37] Персонализация расчета болюса	Снижен индекс риска повышения глюкозы
		[38] Концепция ИСПП в психиатрии	Снизит риск повторной госпитализации
ГиИС		[39] Дифференциальная диагностика РПЖ	В 6 из 9 наблюдений вывод верный
		[40] Прогнозирование свойств лекарственных соединений	Точность — 93,5%
		[41] Виртуальный консилум	Точность — 92%
		[42] Дифференциальная диагностика астмы	Se — 96,45%; Sp — 98,71%

Обозначения: AUC — Area Under Curve; DL — deep learning; GBM — gradient boosting machine; k-NN — k-ближайших соседей; Se — чувствительность; Sp — специфичность; GA — генетический алгоритм; МСП — многослойный перцептрон; РПЖ — рак поджелудочной железы.

Литература

1. С врачебной ошибкой в России сталкивается каждый пятый пациент. <https://www.medikforum.ru/medicine/77161-s-vrachebnoy-oshibkoy-v-rossii-stalkivaetsya-kazhdyu-pyatyu-pacient.html>.
2. Минздрав назвал ежегодное количество осложнений от врачебных ошибок. <https://www.rbc.ru/rbcfreenews/5e3e9fc09a79473ae7c214a6?from=copy>.
3. *Фурман Н. Л.* Разработка систем информационно-аналитической поддержки принятия решений // Технологии информатизации и управления. — Минск: БГУ, 2011. Вып. 2. С. 80–85.
4. *Вагин В. Н., Еремеев А. П.* Некоторые базовые принципы построения интеллектуальных систем поддержки принятия решений реального времени // Известия РАН. Теория и системы управления, 2001. Вып. 6. С. 114–123.
5. *Гусев А. В., Зарубина Т. В.* Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации // Врач и информационные технологии, 2017. Вып. 2. С. 60–72. EDN: YSZMEL.
6. *Реброва О. Ю.* Жизненный цикл систем поддержки принятия врачебных решений как медицинских технологий // Врач и информационные технологии, 2020. Вып. 1. С. 27–37. doi: 10.37690/1811-0193-2020-1-27-37. EDN: CVANJA.
7. *Кобринский Б. А.* Системы поддержки принятия врачебных решений: история и современные решения // Методология и технология непрерывного профессионального образования, 2020. № 4(4). С. 22–38. doi: 10.24075/МТСРЕ.2020.022. EDN: LWVTZN.
8. *Кобринский Б. А.* Искусственный интеллект в медицине: состояние и горячие точки // XIX Национальная конф. по искусственному интеллекту с международным участием: Тр. конф. / Под ред. В. В. Борисова, Б. А. Кобринского. — Ростов-на-Дону; Таганрог: ЮФУ, 2021. С. 13–29.
9. *Малых В. Л.* Системы поддержки принятия решений в медицине // Программные системы: теория и приложения, 2019. Т. 10. № 2. С. 155–184. doi: 10.25209/2079-3316-2019-10-2-155-184. EDN: SECLOM.
10. *Ларина В. Н., Самородская И. В., Глибка К. В. и др.* Комплексная патология: терминология, оценка тяжести состояния пациентов и возможный подход к их ведению // Клиническая геронтология, 2019. Т. 25. № 1-2. С. 49–57. doi: 10.26347/1607-2499201901-02049-057. EDN: YXZNZR.
11. *Шадеркин И. А.* Слабые стороны искусственного интеллекта в медицине // Российский ж. телемедицины и электронного здравоохранения, 2021. Т. 7. Вып. 2. С. 50–52. doi: 10.29188/2712-9217-2021-7-2-50-52. EDN: JNVMQF.
12. *Гольдина Т. А., Бурмистров В. А., Ефименко И. В., Хорошевский В. Ф.* Искусственный интеллект в здравоохранении: Real World Data и Patient Voice — готовы ли мы к новым реалиям? // Медицинские технологии. Оценка и выбор, 2021. Т. 43. Вып. 2. С. 22–31. doi: 10.17116/medtech20214302122. EDN: GDCZFH.
13. *Гусев А. В., Астапенко Е. М., Иванов И. В. и др.* Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения // Вестник Росздравнадзора, 2022. Вып. 2. С. 25–33. EDN: BSMOFV.
14. *Шарова Д. Е., Гарбук С. В., Васильев Ю. А.* Системы искусственного интеллекта в клинической медицине. Первая в мире серия национальных стандартов // Стан-

- дарты и качество, 2023. Вып. 1. С. 46–51. doi: 10.35400/0038-9692-2023-1-304-22. EDN: SNMGQA.
15. Explainable artificial intelligence in medical decision support systems / Eds. A. L. Imoize, J. Hemanth, D.-T. Do, S.N. Sur. — The Institution of Engineering and Technology, 2022. 545 p. doi: 10.1049/PBHE050E.
 16. Explainable AI in healthcare: Unboxing machine learning for biomedicine / Eds. M. S. Raval, M. Roy, T. Kaya, R. Kapdi. — New York, NY, USA: Chapman and Hall/CRC, 2023. 328 p. doi: 10.1201/9781003333425.
 17. Medical data analysis and processing using explainable artificial intelligence / Eds. O. P. Jena, M. Panda, U. Kose. — Boca Raton, CA, USA: CRC Press, 2024. 269 p. doi: 10.1201/9781003257721.
 18. *Хасанов А. Г., Шайбаков Д. Г., Жернаков С. В. и др.* Нейронные сети для прогнозирования динамики развития заболеваний // Креативная хирургия и онкология, 2020. Т. 10. № 3. С. 198–204. doi: 10.24060/2076-3093-2020-10-3-198-204. EDN: DZNXJA.
 19. *Kui B., Pinter J., Molontay R., et al.* EASY-APP: An artificial intelligence model and application for early and easy prediction of severity in acute pancreatitis // Clinical Translational Medicine, 2022. Vol. 12. Art. e842. 13 p. doi: 10.1002/ctm2.842.
 20. *Thapa R., Iqbal Z., Garikipati A., et al.* Early prediction of severe acute pancreatitis using machine learning // Pancreatology, 2022. Vol. 22. Iss. 1. P. 43–50. doi: 10.1016/j.pan.2021.10.003.
 21. *Yin M., Zhang R., Zhou Z., et al.* Automated machine learning for the early prediction of the severity of acute pancreatitis in hospitals // Front. Cell. Infect. Mi., 2022. Vol. 12. Art. 886935. 11 p. doi: 10.3389/fcimb.2022.886935.
 22. *Liu L., Zhang R., Shi D., Li R., Wang Q., Feng Y., Lu F., Zong Y., Xu X.* Automated machine learning to predict the difficulty for endoscopic resection of gastric gastrointestinal stromal tumor // Frontiers Oncology, 2023. Vol. 13. Art. 1190987. 10 p. doi: 10.3389/fonc.2023.1190987.
 23. *Ястремский А. П., Захаров С. Д., Кузнецова Н. Е. и др.* Прогнозирование развития парафарингеального абсцесса у лиц, проживающих в условиях сибирского региона, с помощью логистической регрессии // Университетская медицина Урала, 2023. Т. 9. № 2. С. 24–28. doi: 10.36361/24129445_2023_9_2_24. EDN: SVVKYR.
 24. *Морозов С. П., Владимирский А. В., Ледихова Н. В. и др.* Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта // Туберкулез и болезни легких, 2018. Т. 96. Вып. 8. С. 42–49. doi: 10.21292/2075-1230-2018-96-8-42-49. EDN: XYNQEP.
 25. *Cho H., Lee J. G., Kang S. J., et al.* Angiography-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesion // J. Am. Heart Assoc., 2019. Vol. 8. Iss. 4. Art. e011685. 18 p. doi: 10.1161/JAHA.118.011685.
 26. *Tao X., Zhou X., Xu X., Xiong D. A.* Lightweight network for accurate coronary artery segmentation using X-ray angiograms // Frontiers Public Health, 2022. Vol. 10. Art. 892418. 16 p. doi: 10.3389/fpubh.2022.892418.
 27. *Гусев А. В., Гаврилов Д. В., Корсаков И. Н. и др.* Перспективы использования методов машинного обучения для предсказания сердечно-сосудистых заболеваний // Врач и информационные технологии, 2019. Вып. 3. С. 41–47. EDN: KKDZXQ.

28. Пензин О. В. Прогностическая модель для оценки риска развития миелотоксических осложнений химиотерапевтического лечения: Дис. . . . канд. мед. наук. — М., 2020. 168 с. EDN: AUFLJ.
29. López B., Torrent-Fontbona F., Viñas R., Fernández-Real J. M. Single nucleotide polymorphism relevance learning with Random Forests for type 2 diabetes risk prediction // *Artif. Intell. Med.*, 2018. Vol. 85. P. 43–49. doi: 10.1016/j.artmed.2017.09.005.
30. Moazemi S., Vahdati S., Li J., et al. Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review // *Front. Med.* — *Prс.*, 2023. Vol. 10. Art. 1109411. 18 p. doi: 10.3389/fmed.2023.1109411.
31. Khalfallah H. B., Jelassi M., Demongeot J., Saoud N. B. B. Decision support systems in healthcare: Systematic review, meta-analysis and prediction, with example of COVID-19 // *AIMS Bioengineering*, 2023. Vol. 10. Iss. 1. P. 27–52. doi: 10.3934/bioeng.2023004.
32. Ковалёв П. И., Грибова В. В., Окунь Д. Б. Онтология представления знаний о назначении персонализированного лечения // *Онтология проектирования*, 2023. Т. 13. № 2. С. 192–203. doi: 10.18287/2223-9537-2023-13-2-192-203. EDN: ZI-JWLZ.
33. Грибова В. В., Клещев А. С., Москаленко Ф. М. и др. Облачная платформа IASaaS для разработки оболочек интеллектуальных сервисов: состояние и перспективы развития // *Программные продукты и системы*, 2018. Т. 31. Вып. 3. С. 527–536. doi: 10.15827/0236-235X.123.527-536.
34. Осмоловский И. С., Зарубина Т. В. Разработка и апробация прототипа экспертной системы для диагностики подагры // *Социальные аспекты здоровья населения*, 2023. Т. 69. № 4. Ст. 15. 24 с. doi: 10.21045/2071-5021-2023-69-4-15. EDN: OFVRSI.
35. Okikiola F. M., Aigbokhan E. E., Mustapha A. M., et al. Design and implementation of a fuzzy expert system for diagnosing breast cancer // *J. Advances Mathematics Computer Science*, 2019. Vol. 32. Iss. 1. P. 1–14. doi: 10.9734/jamcs/2019/v32i130137.
36. Thanathornwong B. Bayesian-based decision support system for assessing the needs for orthodontic treatment // *Healthcare Informatics Research*, 2018. Vol. 24. Iss. 1. P. 22–28. doi: 10.4258/hir.2018.24.1.22.
37. Brown D., Aldea A., Martin C., Bayley I. Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support // *Artif. Intell. Med.*, 2018. Vol. 85. P. 28–42. doi: 10.1016/j.artmed.2017.09.007.
38. Озеров В. А., Руданов Ю. Г. Система поддержки принятия врачебных решений // *Системный анализ в проектировании и управлении: Сб. науч. тр. XXIV Междунар. научн. и учебно-практич. конф.: в 3 ч.* — СПб.: Политех-Пресс, 2020. Ч. 3. С. 260–272.
39. Шестерникова О. П., Финн В. К., Винокурова Л. В. и др. Интеллектуальная система для диагностики заболеваний поджелудочной железы // *Научно-техническая информация. Сер. 2. Информационные процессы и системы*, 2019. Вып. 10. С. 41–48.
40. Samigulina G., Samigulina Z. Ontological model of multi-agent smart-system for predicting drug properties based on modified algorithms of artificial immune systems // *Theor. Biol. Med. Model.*, 2020. Vol. 17. Iss. 1. Art. 12. 22 p. doi: 10.1186/s12976-020-00130-x.

41. *Rumovskaya S. B., Kolesnikov A. V., Kirikov I. A.* Modeling of collective decisions by the virtual council // Scientific Technical Information Processing, 2019. Vol. 46. Iss. 5. P. 356–365. doi: 10.3103/S0147688219050083.
42. *Badnjevic A., Gurbeta L., Custovic E.* An expert diagnostic system to automatically identify asthma and chronic obstructive pulmonary disease in clinical settings // Sci. Rep. — U.K., 2018. Vol. 8. Iss. 1. Art. 11645. 9 p. doi: 10.1038/s41598-018-30116-2.

Поступила в редакцию 29.02.24

INTELLIGENT DECISION SUPPORT SYSTEMS IN MEDICINE: CONCEPT, PROBLEMS, AND APPROACHES TO THE DEVELOPMENT

S. B. Rumovskaya

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

Abstract: Ideas about an object and a process as a heterogeneous integral system are not completed in a theory and have not become a world picture of highly professional specialists in the heterogeneous problematic environment of diagnostics and prognosis in medicine, technology, etc. Decision-making under conditions of limited resources, ambiguity in assessing the situation, and a large volume and variety of processed information about an object is accompanied by serious errors and risks. Decision-making virtualization technologies are needed. The paper discusses the concept of intelligent decision support systems in medicine, features and problems of the subject area, and approaches to the development.

Keywords: intelligent decision support system; medicine and healthcare; development problems; development approaches

DOI: 10.14357/08696527240208

EDN: IJDVVV

Acknowledgments

The research was supported by the Russian Science Foundation (project No. 24-21-00392).

References

1. S vrachebnoy oshibkoy v Rossii stalkivaetsya kazhdyy pyatyy patsient [Every fifth patient in Russia faces a medical error]. Available at: <https://www.medikforum.ru/medicine/77161-s-vrachebnoy-oshibkoy-v-rossii-stalkivaetsya-kazhdyy-pyatyy-pacient.html> (accessed April 17, 2024).

2. Minzdrav nazval ezhegodnoe kolichestvo oslozhneniy ot vrachebnykh oshibok [The Ministry of Health named the annual number of complications from medical errors]. Available at: <https://www.rbc.ru/rbcfreenews/5e3e9fc09a79473ae7c214a6?from=copy> (accessed April 17, 2024).
3. Furman, N. L. 2011. Razrabotka sistem informatsionno-analiticheskoy podderzhki prinyatiya resheniy [Development of systems for information and analytical support for decision making]. *Tekhnologii informatizatsii i upravleniya* [Technologies of Information and Control]. Minsk: BGU. 2:80–85.
4. Vagin, V. N., and A. P. Eremeev. 2001. Some basic principles of design of intelligent systems for supporting real-time decision making. *J. Comput. Sys. Sc. Int.* 40(6):953–961. EDN: LGMXGB.
5. Gusev, A. V., and T. V. Zarubina. 2017. Podderzhka prinyatiya vrachebnykh resheniy v meditsinskikh informatsionnykh sistemakh meditsinskoy organizatsii [Clinical decisions support in medical information systems of a medical organization]. *Vrach i informatsionnye tekhnologii* [Medical Doctor and IT] 2:60–72. EDN: YSZMEL.
6. Rebrova, O. Yu. 2020. Zhiznennyy tsikl sistem podderzhki prinyatiya vrachebnykh resheniy kak meditsinskikh tekhnologiy [Life cycle of decision support systems as medical technologies]. *Vrach i informatsionnye tekhnologii* [Medical Doctor and IT] 1:27–37. doi: 10.37690/1811-0193-2020-1-27-37. EDN: CVANJA.
7. Kobrinskiy, B. A. 2020. Sistemy podderzhki prinyatiya vrachebnykh resheniy: istoriya i sovremennyye resheniya [Clinical decision support systems in advanced qualification: History and modern trends]. *Metodologiya i tekhnologiya nepreryvnogo professional'nogo obrazovaniya* [Methodology and Technology of Continuous Professional Education] 4(4):22–38. doi: 10.24075/MTCPE.2020.022. EDN: LWVTZN.
8. Kobrinskiy, B. A. 2021. Iskusstvennyy intellekt v meditsine: sostoyanie i goryachie tochki [Artificial intelligence in medicine: Status and hot spots]. *XIX Natsional'naya konf. po iskusstvennomu intellektu s mezhdunarodnym uchastiem. Tr. konf.* [19th National Conference on Artificial Intelligence with International Participation Collection of Works]. Eds. V. V. Borisov and B. A. Kobrinskiy. Rostov-on-Don; Taganrog: YuFU. 13–29.
9. Malykh, V. L. 2019. Sistemy podderzhki prinyatiya resheniy v meditsine [Decision support systems in medicine]. *Programmnyye sistemy: teoriya i prilozheniya* [Program Systems: Theory and Applications] 10(2):155–184. doi: 10.25209/2079-3316-2019-10-2-155-184. EDN: SECLOM.
10. Larina, V. N., I. V. Samorodskaya, K. V. Glibko, et al. 2019. Kompleksnaya patologiya: terminologiya, otsenka tyazhesti sostoyaniya patsientov i vozmozhnyy podkhod k ikh vedeniyu [Complex pathology: Terminology, assessment of the severity of patient's condition and possible approach to the management of patients]. *Klinicheskaya gerontologiya* [Clinical Gerontology] 25(1-2):49–57. doi: 10.26347/1607-2499201901-02049-057. EDN: YXZNR.
11. Shaderkin, I. A. 2021. Slabye storony iskusstvennogo intellekta v meditsine [Weaknesses of artificial intelligence in medicine]. *Rossiyskiy zh. teleditsiny i elektronogo zdavookhraneniya* [Russian J. Telemedicine and e-Health] 7(2):50–52. doi: 10.29188/2712-9217-2021-7-2-50-52. EDN: JNVMQF.
12. Goldina, T. A., V. A. Burmistrov, I. V. Efimenko, and V. F. Khoroshevskiy. 2021. Iskusstvennyy intellekt v zdavookhraneni: Real World Data i Patient Voice — gotovy li my k novym realiyam? [Artificial intelligence in healthcare: Real World

- Data and Patient Voice — are we ready for new realities?]. *Meditsinskie tekhnologii. Otsenka i vybor* [Medical Technologies. Assessment and Choice] 43(2):22–31. doi: 10.17116/medtech20214302122. EDN: GDCZFH.
13. Gusev, A. V., E. M. Astapenko, I. V. Ivanov, *et al.* 2022. Printsipy formirovaniya doveriya k sistemam iskusstvennogo intellekta dlya sfery zdravookhraneniya [Principles for building confidence in artificial intelligence systems for healthcare]. *Vestnik Roszdravnadzora* 2:25–33. EDN: BSMOFV.
 14. Sharova, D. E., S. V. Garbuk, and Yu. A. Vasilyev. 2023. Sistemy iskusstvennogo intellekta v klinicheskoy meditsine. Pervaya v mire seriya natsional'nykh standartov [Artificial intelligence systems in clinical medicine. The world's first series of national standards]. *Standarty i kachestvo* [Standards and Quality] 1:46–51. doi: 10.35400/0038-9692-2023-1-304-22. EDN: SNMGQA.
 15. Imoize, A. L., J. Hemanth, D.-T. Do, and S. N. Sur, eds. 2022. *Explainable artificial intelligence in medical decision support systems*. The Institution of Engineering and Technology. 545 p. doi: 10.1049/PBHE050E.
 16. Raval, M. S., M. Roy, T. Kaya, and R. Kapdi, eds. 2023. *Explainable AI in healthcare: Unboxing machine learning for biomedicine*. New York, NY: Chapman and Hall/CRC. 328 p. doi: 10.1201/9781003333425.
 17. Jena, O. P., M. Panda, and U. Kose, eds. 2024. *Medical data analysis and processing using explainable artificial intelligence*. Boca Raton, CA: CRC Press. 269 p. doi: 10.1201/9781003257721.
 18. Khasanov, A. G., D. G. Shaybakov, S. V. Zhernakov, *et al.* 2020. Neyronnye seti dlya prognozirovaniya dinamiki razvitiya zabolevaniy [Neural networks in forecasting disease dynamics]. *Kreativnaya khirurgiya i onkologiya* [Creative Surgery and Oncology] 10(3):198–204. doi: 10.24060/2076-3093-2020-10-3-198-204. EDN: DZNXJA.
 19. Kui, B., J. Pinter, R. Molontay, *et al.* EASY-APP: An artificial intelligence model and application for early and easy prediction of severity in acute pancreatitis. *Clinical Translational Medicine* 12(6):e842. 13 p. doi: 10.1002/ctm2.842.
 20. Thapa, R., Z. Iqbal, A. Garikipati, *et al.* 2022. Early prediction of severe acute pancreatitis using machine learning. *Pancreatology* 22(1):43–50. doi: 10.1016/j.pan.2021.10.003.
 21. Yin, M., R. Zhang, Z. Zhou, *et al.* 2022. Automated machine learning for the early prediction of the severity of acute pancreatitis in hospitals. *Front. Cell. Infect. Mi.* 12:886935. 11 p. doi: 10.3389/fcimb.2022.886935.
 22. Liu, L., R. Zhang, D. Shi, R. Li, Q. Wang, Y. Feng, F. Lu, Y. Zong, and X. Xu. 2023. Automated machine learning to predict the difficulty for endoscopic resection of gastric gastrointestinal stromal tumor. *Frontiers Oncology* 13:1190987. 10 p. doi: 10.3389/fonc.2023.1190987.
 23. Yastremskiy, A. P., S. D. Zakharov, N. E. Kuznetsova, *et al.* 2023. Prognozirovanie razvitiya parafaringeal'nogo abstsessa u lits, prozhivayushchikh v usloviyakh sibirskogo regiona, s pomoshch'yu logisticheskoy regressii [Predicting the development of parapharyngeal abscess in people living in the siberian region using logistic regression]. *Universitetskaya meditsina Urala* [University Medicine of the Urals] 9(2):24–28. doi: 10.36361/24129445_2023_9_2_24. EDN: SVVKYR.
 24. Morozov, S. P., A. V. Vladzimirskiy, N. V. Ledikhova, *et al.* 2018. Otsenka diagnosticheskoy tochnosti sistemy skrininga tuberkuleza legkikh na osnove iskusstvennogo intellekta [Evaluation of diagnostic accuracy of the system for pulmonary tuberculosis

- screening based on artificial neural networks]. *Tuberkulez i bolezni legkikh* [Tuberculosis and Lung Diseases] 96(8):42–49. doi: 10.21292/2075-1230-2018-96-8-42-49. EDN: XYNQEP.
25. Cho, H., J. G. Lee, S. J. Kang, *et al.* 2019. Angiography-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesion. *J. Am. Heart Assoc.* 8(4):e011685. 18 p. doi: 10.1161/JAHA.118.011685.
 26. Tao, X., X. Zhou, X. Xu, and D. A. Xiong. 2022. Lightweight network for accurate coronary artery segmentation using X-ray angiograms. *Frontiers Public Health* 10:892418. 16 p. doi: 10.3389/fpubh.2022.892418.
 27. Gusev, A. V., D. V. Gavrilov, I. N. Korsakov, *et al.* 2019. Perspektivy ispol'zovaniya metodov mashinnogo obucheniya dlya predskazaniya serdechno-sosudistyykh zabolevaniy [Prospects for the use of machine learning methods for predicting cardiovascular disease]. *Vrach i informatsionnye tekhnologii* [Medical Doctor and IT] 3:41–47. EDN: KKDZXQ.
 28. Penzin, O. V. 2020. Prognosticheskaya model' dlya otsenki riska razvitiya mielotoksicheskikh oslozhneniy khimioterapevticheskogo lecheniya [Predictive model for assessing the risk of developing myelotoxic complications of chemotherapy treatment]. Moscow. PhD Thesis. 168 p. EDN: AUFLLJ.
 29. López, B., F. Torrent-Fontbona, R. Viñas, and J. M. Fernández-Real. 2018. Single nucleotide polymorphism relevance learning with random forests for type 2 diabetes risk prediction. *Artif. Intell. Med.* 85:43–49. doi: 10.1016/j.artmed.2017.09.005.
 30. Moazemi, S., S. Vahdati, J. Li, *et al.* 2023. Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review. *Front. Med. — Prc.* 10:1109411. 18 p. doi: 10.3389/fmed.2023.1109411.
 31. Khalfallah, H. B., M. Jelassi, J. Demongeot, and N. B. B. Saoud. 2023. Decision support systems in healthcare: Systematic review, meta-analysis and prediction, with example of COVID-19. *AIMS Bioengineering* 10(1):27–52. doi: 10.3934/bioeng.2023004.
 32. Kovalev, R. I., V. V. Gribova, and D. B. Okun'. 2023. Ontologiya predstavleniya znanii o naznachanii personifitsirovannogo lecheniya [Ontology of knowledge representation about the prescribed personalized treatment]. *Ontologiya proektirovaniya* [Ontology of Designing] 13(2):192–203. doi: 10.18287/2223-9537-2023-13-2-192-203. EDN: ZIJWLZ.
 33. Gribova, V. V., A. S. Kleshchev, F. M. Moskalenko, *et al.* 2018. Oblachnaya platforma IACPaaS dlya razrabotki oblochek intellektual'nykh servisov: sostoyanie i perspektivy razvitiya [IACPaaS cloud platform for the development of intelligent service shells: Current state and future evolution]. *Programmnye produkty i sistemy* [Software & Systems] 31(3):527–536. doi: 10.15827/0236-235X.123.527-536.
 34. Osmolovskiy, I. S., and T. V. Zarubina. 2023. Razrabotka i aprobatsiya prototipa ekspertnoy sistemy dlya diagnostiki podagry [Developing and testing a prototype expert system for gout diagnosis]. *Sotsial'nye aspekty zdorov'ya naseleniya* [Social Aspects of Population Health] 69(4):15. 24 p. doi: 10.21045/2071-5021-2023-69-4-15. EDN: OFVRSI.
 35. Okikiola, F. M., E. E. Aigbokhan, A. M. Mustapha, *et al.* 2019. Design and implementation of a fuzzy expert system for diagnosing breast cancer. *J. Advances Mathematics Computer Science* 32(1):1–14. doi: 10.9734/jamcs/2019/v32i130137.

36. Thanathornwong, B. 2018. Bayesian-based decision support system for assessing the needs for orthodontic treatment. *Healthcare Informatics Research* 24(1):22–28. doi: 10.4258/hir.2018.24.1.22.
37. Brown, D., A. Aldea, C. Martin, and I. Bayley. 2018. Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support. *Artif. Intell. Med.* 85:28–42. doi: 10.1016/j.artmed.2017.09.007.
38. Ozerov, V. A., and Yu. G. Rudanov. 2020. Sistema podderzhki prinyatiya vrachebnykh resheniy [Medical decision support system]. *Sistemnyy analiz v proektirovanii i upravlenii: Sb. nauch. tr. XXIV Mezhdunar. nauchnoy i uchebno-prakticheskoy konf.* [24th Scientific and Educational-Practical Conference (International) Proceedings]. Saint Petersburg: Politekh-Press. 3:260–272.
39. Shesternikova, O. P., V. K. Finn, L. V. Vinokurova, et al. 2019. An intelligent system for diagnostics of pancreatic diseases. *Automatic Documentation Mathematical Linguistics* 53(5):288–294. doi: 10.3103/S000510551905008X. EDN: OWVGBW.
40. Samigulina, G., and Z. Samigulina. 2020. Ontological model of multi-agent smart-system for predicting drug properties based on modified algorithms of artificial immune systems. *Theor. Biol. Med. Model.* 17(1):12. 22 p. doi: 10.1186/s12976-020-00130-x.
41. Rumovskaya, S. B., A. V. Kolesnikov, and I. A. Kirikov. 2019. Modeling of collective decisions by the virtual council. *Scientific Technical Information Processing* 46(5):356–365. doi: 10.3103/S0147688219050083.
42. Badnjevic, A., L. Gurbeta, and E. Custovic. 2018. An expert diagnostic system to automatically identify asthma and chronic obstructive pulmonary disease in clinical settings. *Sci. Rep. — U.K.* 8(1):11645. 9 p. doi: 10.1038/s41598-018-30116-2.

Received February 29, 2024

Contributor

Rumovskaya Sophiya B. (b. 1985) — Candidate of Sciences (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; sophiyabr@gmail.com

МЕТОД ПОИСКА ОПТИМАЛЬНЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ АЛГОРИТМА РАЗРЕШЕНИЯ СУЩНОСТЕЙ ДЛЯ КОНКРЕТНО-ИСТОРИЧЕСКИХ ДАННЫХ

И. М. Адамович¹, О. И. Волков²

Аннотация: Статья посвящена использованию метода коллективного разрешения сущностей, основанного на новом алгоритме реляционной кластеризации, представляющем собой модификацию жадного алгоритма агломеративной кластеризации, в конкретно-историческом исследовании при обработке номинативных источников. Предложен метод поиска оптимальных значений параметров алгоритма коллективного разрешения сущностей для задач, связанных с конкретно-историческими исследованиями. Метод основан на анализе специфики конкретно-исторических данных, их сравнении с тестовыми данными, для которых имеются оценки эффективности алгоритма, и процедуре нахождения оптимальных параметров процессов по схеме Гаусса–Зайделя, заключающемся в последовательном поиске оптимума функции поочередно по каждой переменной. Применение предложенного метода делает возможным использование рассматриваемого алгоритма разрешения сущностей в реальных конкретно-исторических исследованиях в задачах автоматизированного связывания записей в номинативных источниках.

Ключевые слова: конкретно-историческое исследование; распределенная технология; разрешение сущностей; параметры алгоритма; мера реляционного сходства

DOI: 10.14357/08696527240209

EDN: IJWCDL

1 Введение

В [1] рассматривалось применение метода коллективного разрешения сущностей, основанного на новом алгоритме реляционной кластеризации [2] в области программно-информационной поддержки конкретно-исторических исследований (ПКИИ). Этот алгоритм представляет собой модификацию жадного алгоритма агломеративной кластеризации, в котором исходное распределение кластеров обеспечивается специальной процедурой начальной загрузки.

Цель применения данного алгоритма состоит в автоматизированном поиске записей, которые ссылаются на одну и ту же сущность в разных номинативных (содержащих персональную информацию, включая имя персоны) источниках исторических данных. Данная задача считается одной из наиболее важных

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Adam@amsd.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Volkov@amsd.com

и сложных, особенно при обработке номинативных источников семейной структуры.

Анализ показал, что на данных с высоким уровнем неоднозначности ссылок, что соответствует специфике конкретно-исторических данных, рассматриваемый алгоритм более эффективен, чем традиционный способ разрешения ссылок, основанный на сравнении атрибутов. Из этого следует принципиальная возможность интеграции нового подхода в технологию ПККИ [3, 4], основанную на принципах краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий).

Для конкретизации параметров алгоритма в соответствии с особенностями предметной области необходимо определить метод поиска их оптимальных для конкретно-исторических данных значений. В статье предложен такой метод, основанный на анализе специфики конкретно-исторических данных, их сравнении с тестовыми данными, для которых имеются оценки эффективности алгоритма, и схеме, основанной на последовательном покоординатном поиске максимума.

2 Настраиваемые параметры алгоритма

Обозначения в соответствии с [1] следующие:

- $R = \{r_i\}$ — набор ссылок, соответствующих набору сущностей $E = \{e_i\}$;
- $r.E$ — сущность, соответствующая ссылке r ;
- $r.A_1, r.A_2, \dots, r.A_k$ — набор атрибутов ссылки r ;
- $H = \{h_i\}$ — набор гиперребер (групп ссылок);
- $h.R$ — набор ссылок, которые соединяет гиперребро h ;
- $r.H$ — набор гиперребер, в котором участвует ссылка r .

Алгоритм осуществляет разбиение ссылок на кластеры по признаку соответствия этих ссылок одной и той же сущности. Обозначения, связанные с кластерами, следующие:

- $C = \{c_i\}$ — текущий набор кластеров ссылок;
- $r.C$ — метка кластера для ссылки r ;
- $c.H = \bigcup_{r \in R \wedge r.C=c} \{h | h \in H \wedge r \in h.R\}$ — набор гиперребер кластера c ;
- $\text{Nbr}(c) = \bigcup_{h \in c.H, r \in h.R} \{c_j | c_j = r.C\}$ — окрестность кластера c ;
- $\text{Nbr}^2(c) = \bigcup_{c' \in \text{Nbr}(c)} \text{Nbr}(c')$ — окрестность второго порядка для кластера c ;
- $\text{Amb}(r.A_1) = |\sigma_{R.A_1=r.A_1}(R)|/|R|$ — оценка неоднозначности атрибута $R.A_1$, где $|\sigma_{R.A_1=r.A_1}(R)|$ обозначает число ссылок со значением $r.A_1$ для A_1 .

Если кластеры рассматривать как мультимножества, то арифметическая сумма кластеров, связанных с кластером c через набор его гиперребер, дает определенное окрестности $\text{Nbr}_B(c)$, учитывающее множественность вхождения различных соседних кластеров.

Мера сходства sim двух кластеров c_i и c_j в [2] определяется следующим образом:

$$\text{sim}(c_i, c_j) = (1 - \alpha) \text{sim}_A(c_i, c_j) + \alpha \text{sim}_R(c_i, c_j), \quad 0 \leq \alpha \leq 1, \quad (1)$$

где $\text{sim}_A()$ — мера сходства на основе атрибутов, способ вычисления которой подробно описан; $\text{sim}_R()$ — мера реляционного сходства.

В [2] для рассматриваемого алгоритма были обоснованы как наиболее подходящие, адаптированы и модифицированы некоторые меры реляционного сходства, и выбор одной из этих мер, подробно описанных в [1], наиболее подходящей для конкретно-исторических данных, относится к настраиваемым параметрам алгоритма. Также требуется определения весовой коэффициент α .

3 Меры реляционного сходства кластеров

Для задач разрешения сущностей есть множество методов оценки сходства: эквивалентность булевых предикатов, вычисление функции сходства значений, вычисление функции сходства множеств, вычисление функции сходства векторов, сходство на основе выравнивания, сходство фонетических данных, сходство, основанное на переводе, сходство, основанное на знаниях о предметной области, и др. [5]. Как показано в [2], рассматриваемый алгоритм использует меры сходства множеств. Этот класс мер также весьма широк [6], и некоторые из них успешно применяются для алгоритмической кластеризации [7], лежащей в основе рассматриваемого алгоритма. В [2] некоторые из них были обоснованы как наиболее подходящие для него, адаптированы и модифицированы в соответствии с его особенностями. В [1] эти меры, кратко перечисленные ниже, описаны подробно. Мера определяется через понятие окрестности Nbr . Для каждой меры определяется также вариант, обозначаемый индексом B , с опорой на оценку Nbr_B вместо Nbr .

1. Оценка общих соседей

$$\text{CommonNbr}(c_i, c_j) = \frac{1}{K} |\text{Nbr}(c_i) \cap \text{Nbr}(c_j)|,$$

где K — нормализующая константа.

2. Коэффициент Жаккарда

$$\text{Jaccard}(c_i, c_j) = \frac{|\text{Nbr}(c_i) \cap \text{Nbr}(c_j)|}{|\text{Nbr}(c_i) \cup \text{Nbr}(c_j)|}.$$

3. Оценка Адамик–Адар

$$\text{AdarNbr}(c_i, c_j) = \frac{\sum_{c \in \text{Nbr}(c_i) \cap \text{Nbr}(c_j)} u(c)}{\sum_{c \in \text{Nbr}(c_i) \cup \text{Nbr}(c_j)} u(c)},$$

где уникальность кластера

$$u(c) = \frac{1}{\log(|\text{Nbr}(c)|)}.$$

4. Оценка неоднозначности имен Адар AdarName определяется аналогично оценке Адамик–Адар при

$$u(c) = \frac{1}{\text{Avg}_{r \in C}(\text{Amb}(r.\text{Name}))}.$$

5. Оценка через окрестности более высокого порядка

$$\text{Path3Sim}(c_i, c_j) = \frac{1}{3} [\text{Jaccard}(\text{Nbr}(c_i), \text{Nbr}(c_j)) + \text{Jaccard}(\text{Nbr}^2(c_i), \text{Nbr}(c_j)) + \text{Jaccard}(\text{Nbr}(c_i), \text{Nbr}^2(c_j))].$$

4 Результаты тестирования алгоритма

В [2] приведены результаты тестирования алгоритма разрешения реляционных сущностей на трех наборах реальных данных с различным уровнем неоднозначности имен. Параметры наборов приведены в табл. 1.

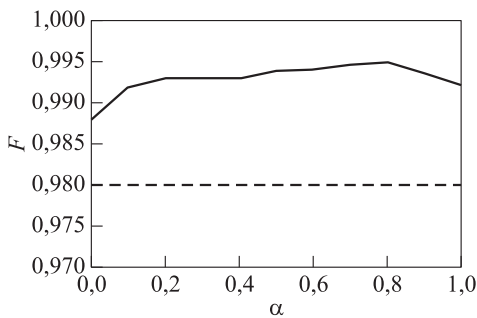
Таблица 2 показывает эффективность алгоритма, выраженную в значении меры F1 (среднее гармоническое точности и полноты) для различных мер реляционного сходства кластеров sim_R , описанных выше, на трех наборах тестовых данных, приведенных в табл. 1.

Таблица 1 Параметры тестовых наборов данных

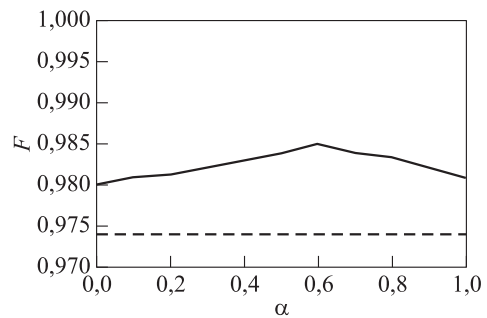
Наименование	Число документов	Максимальное число ссылок в документе	Доля неоднозначных имен	Среднее число сущностей на одно неоднозначное имя
CiteSeer	1504	10	0,2%	2,33
arXiv	29 555	10	4,2%	2,41
BioBase	156 156	100	84%	32,00

Таблица 2 Эффективность алгоритма для различных мер реляционного сходства кластеров sim_R

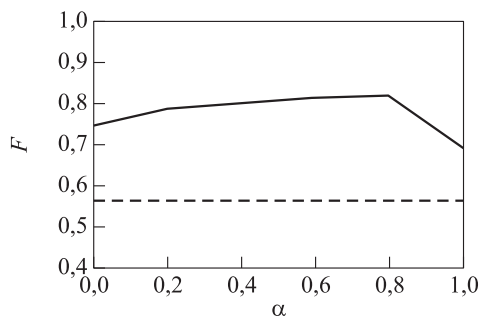
sim_R	CiteSeer	arXiv	BioBase
Common	0,994	0,984	0,814
Common _B	0,994	0,984	0,816
Jaccard	0,994	0,985	0,818
Jaccard _B	0,995	0,985	0,818
AdarNbr	0,994	0,984	0,815
AdarNbr _B	0,994	0,984	0,816
AdarName	0,994	0,985	0,819
AdarName _B	0,995	0,984	0,817
Path3Sim	0,994	0,984	0,812



(а)



(б)



(в)

Зависимость эффективности (значения меры F1) алгоритма (сплошная линия) в сравнении с подходом, основанным на сравнении атрибутов (штриховая линия), от параметра α на различных тестовых наборах данных: (а) CiteSeer; (б) arXiv; (в) BioBase

Из табл. 2 видно, что для наборов $agXiv$ со средним уровнем неоднозначности имен и BioBase с высоким уровнем неоднозначности имен наибольшую эффективность обеспечивает мера реляционного сходства $AdarName$. Для набора CiteSeer с низким уровнем неоднозначности имен эта мера лишь незначительно уступает в эффективности мерам $Jaccard_B$ и $AdarName_B$.

На рисунке приведены графики эффективности (значения меры $F1$) алгоритма для оптимальной в соответствии с табл. 2 меры реляционного сходства кластеров в зависимости от параметра α формулы (1) в сравнении с эффективностью традиционного атрибутивного подхода на тех же наборах тестовых данных. Примечательно, что алгоритм остается более эффективным даже при значении $\alpha = 0$, когда мера сходства кластеров sim не учитывает реляционное сходство. Это объясняется использованием процедурой начальной загрузки как атрибутов, так и реляционных связей.

5 Конкретно-исторические данные

В [8, 9] перечислены основные номинативные источники, представляющие наибольший интерес в конкретно-историческом исследовании. К персональным источникам относятся прежде всего метрические книги, а также метрические выписи, брачные обыски, венечные памяти и т. п. К источникам семейной структуры относятся прежде всего исповедные росписи и посемейные списки, а также ревизские сказки, писцовые, переписные, ландратские книги и т. п.

Направление исследования может быть любым, но, учитывая тот факт, что большинство населения в Российской Империи составляли крестьяне (77,1% в 1897 г.), а также то, что большинство крестьян центра России были официально наделены фамилиями, зафиксированными в документах только после отмены крепостного права в 1861 г., оценку специфики конкретно-исторических данных следует провести по крестьянам центра России без учета информации о фамилиях. Анализ переписных книг по «Танбовского уезда вотчины Солотчинского монастыря села Преображенского Пичаево тож», а также метрических книг Тульской губернии Елифанского уезда деревень Елизаветовка и Мельгуново без учета информации о фамилиях показал следующие результаты:

- максимальное число ссылок в документе — 9;
- доля неоднозначных имен — 5,5363%;
- среднее число сущностей на одно неоднозначное имя — 2,1875.

Эти показатели наиболее близки к показателям тестового набора данных $agXiv$, что позволяет сделать вывод о допустимости оценки эффективности применения алгоритма для целей большинства конкретно-исторических исследований именно на этом наборе. При этом для менее неоднозначных данных, содержащих фамилии, в оценке эффективности можно ориентироваться на набор CiteSeer. Безусловно, в исследованиях могут встречаться данные с большим

уровнем неоднозначности имен, например при исследовании групп с иным национальным составом. В этом случае оценки должны смещаться от оценок на наборе arXiv в сторону оценок на наборе BioBase, но незначительно, поскольку уровень неоднозначности имен 84% существенно превышает достижимый на конкретно-исторических данных. Для принятия решения по выбору значений настраиваемых параметров алгоритма следует учесть, что оценки эффективности на данных с низким уровнем неоднозначности имен менее значимы, поскольку объем данных, обрабатываемый в рамках одной задачи разрешения сущностей при проведении конкретно-исторического исследования, как правило, не превышает нескольких сот записей. При наличии фамилий или иных факторов, снижающих уровень неоднозначности, число неоднозначных имен будет крайне невелико, и каждый такой случай может быть обработан и разрешен исследователем вручную.

6 Метод поиска оптимальных значений параметров

Для рассматриваемого алгоритма при фиксированных значениях параметров известна его эффективность F_i для тестовых наборов данных D_i с различным значением неоднозначности p_i . Оценка его эффективности на реальных конкретно-исторических данных строится на предположении, что реальные данные можно представить как смесь данных с различной степенью неоднозначности p_i в некоторых пропорциях a_i , которые можно оценить заранее, как это сделано выше. Тогда эффективность F можно оценить как взвешенную сумму $\sum a_i F_i$. Опираясь на эти предположения, в качестве метода поиска оптимальных значений параметров α и sim_R можно использовать процедуру нахождения оптимальных параметров процессов по схеме Гаусса–Зайделя (метод покоординатного спуска), заключающуюся в последовательном поиске оптимума некоторой функции поочередно по каждой переменной [10]. В качестве оптимизируемой функции $F(\alpha, \text{sim}_R)$ рассмотрим взвешенную сумму мер F1, посчитанных на различных тестовых наборах данных:

$$F(\alpha, \text{sim}_R) = a_1 F1_{\text{CiteSeer}}(\alpha, \text{sim}_R) + a_2 F1_{\text{arXiv}}(\alpha, \text{sim}_R) + a_3 F1_{\text{BioBase}}(\alpha, \text{sim}_R). \quad (2)$$

Как следует из приведенного выше анализа соответствия наборов тестовых данных специфике конкретно-исторических данных и их значимости в конкретно-историческом исследовании, $a_1 \approx a_3$; $a_1, a_3 \ll a_2$.

При определении очередности переменных следует учесть дискретный характер переменной sim_R и непрерывный характер переменной α . Наиболее определяющая — дискретная переменная, поэтому поиск максимума для оптимизируемой функции следует начать с меры реляционного сходства.

Максимум эффективности на наборе тестовых данных arXiv достигается при использовании в качестве меры реляционного сходства кластеров коэффициента Жаккарда и оценки неоднозначности имен Адар. При снижении степени не-

однозначности эти оценки остаются равноэффективными, а при увеличении степени неоднозначности — оценка неоднозначности имен Адар демонстрирует незначительное преимущество перед коэффициентом Жаккарда. Из табл. 2 видно, что отклонение максимальной эффективности от среднего значения для BioBase втрое превышает отклонения для arXiv и CiteSeer. Поскольку весовой коэффициент a_2 в формуле (2) существенно превышает коэффициенты a_1 и a_3 , из этого следует, что в качестве меры реляционного сходства кластеров для конкретно-исторических данных наиболее подходящей следует считать оценку AdarName, но коэффициент Jaccard также может быть использован.

Далее оптимизация проводится по переменной α . Максимум эффективности алгоритма на наборе тестовых данных arXiv, как это видно на рисунке, достигается при значении 0,6 параметра α формулы (1). При этом максимум эффективности на наборах CiteSeer и BioBase достигается при значении параметра $\alpha = 0,8$ (см. рисунок). Из рисунка следует, что отклонение максимальной эффективности от среднего значения для BioBase значительно превышает отклонения для arXiv и CiteSeer. Поэтому, несмотря на малость коэффициента a_3 в формуле (2), максимум для набора BioBase не может не учитываться. Поскольку на наборе BioBase наблюдается значительный прирост эффективности при $\alpha = 0,8$ по отношению к уровню при $\alpha = 0,6$, а для наборов arXiv и CiteSeer разница в эффективности для $\alpha = 0,6$ и $\alpha = 0,8$ незначительна, следует принять значение параметра α формулы (1) при обработке конкретно-исторических данных, равным 0,8.

Учитывая дискретность одной из переменных и незначительность изменения эффективности для arXiv и CiteSeer по второй переменной, а также недостижимость степени неопределенности BioBase на конкретно-исторических данных, параметры можно принять в высокой степени сепарабельными, что означает достаточность одного цикла перебора переменных по схеме Гаусса–Зайделя для определения оптимума [10].

7 Выводы

В статье предложен метод оптимизации параметров алгоритма коллективного разрешения сущностей для задач, связанных с конкретно-историческими исследованиями. Проведен анализ специфики конкретно-исторических данных, а также сравнение этих данных с тестовыми данными алгоритма, что позволило применить предложенный метод и найти оптимальные значения параметров. Это делает возможным использование рассматриваемого алгоритма разрешения сущностей в реальных конкретно-исторических исследованиях в задачах автоматизированного связывания записей в номинативных источниках.

Литература

1. Адамович И. М., Волков О. И. Коллективное разрешение сущностей в технологии поддержки конкретно-исторических исследований // Системы и средства информ-

- матики, 2024. Т. 34. № 1. С. 128–138. doi: 10.14357/08696527240111. EDN: GQPGXQ.
2. *Bhattacharya I., Getoor L.* Collective entity resolution in relational data // ACM T. Knowl. Discov. D., 2007. Vol. 1. No. 1. Art. 5. 36 p. doi: 10.1145/1217299.1217304.
 3. *Адамович И. М., Волков О. И.* Технология распределенного автоматизированного анализа исторических текстов // Системы и средства информатики, 2016. Т. 26. № 3. С. 148–161. doi: 10.14357/08696527160311. EDN: WWSZIJ.
 4. *Адамович И. М., Волков О. И.* Единая технология поддержки конкретно-исторических исследований // Системы и средства информатики, 2019. Т. 29. № 1. С. 194–205. doi: 10.14357/08696527190116. EDN: MZLQGZ.
 5. *Вовченко А. Е., Калининченко Л. А., Ковалёв Д. Ю.* Методы разрешения сущностей и слияния данных в ETL-процессе и их реализация в среде Hadoop // Информатика и её применения, 2014. Т. 8. Вып. 4. С. 94–109. doi: 10.14357/19922264140412. EDN: PJYYLB.
 6. *Strehl A., Ghosh J., Mooney R.* Impact of similarity measures on web-page clustering // Workshop of Artificial Intelligence for Web Search. — AAAI Press, 2000. P. 58–64.
 7. *Collins-Elliot S.A.* Agglomerative clustering using cosine and Jaccard distances: A computational approach to Roman vessel taxonomy // Archeologia Calcolatori, 2016. Vol. 27. P. 7–25.
 8. *Антонов Д. Н.* Восстановление истории семей: метод, источники, анализ: Дис. . . . канд. ист. наук. — М., 2000. 290 с. EDN: QDBKMR.
 9. *Торвальдсен Г.* Номинативные источники в контексте всемирной истории переписей: Россия и Запад // Известия Уральского федерального университета. Сер. 2. Гуманитарные науки, 2016. Т. 18. № 3(154). С. 9–28. doi: 10.15826/izv2.2016.18.3.041. EDN: WYDBXL.
 10. *Хнаев О. А., Пчелинцев И. А.* Параметрическая оптимизация систем. Методы решения экстремальных задач // Модели, системы, сети в экономике, технике, природе и обществе, 2012. № 2(3). С. 146–152. EDN: RPUFKH.

Поступила в редакцию 15.03.24

METHOD FOR SEARCHING FOR OPTIMAL PARAMETER VALUES OF THE ENTITY RESOLUTION ALGORITHM FOR CONCRETE HISTORICAL DATA

I. M. Adamovich and O. I. Volkov

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

Abstract: The article is devoted to the use of the collective entity resolution method based on a new relational clustering algorithm, which is a modification of the greedy agglomerative clustering algorithm, in concrete historical investigation when processing nominative sources. The article proposes the method for

searching for optimal values of parameters of the collective entity resolution algorithm for tasks related to concrete historical investigation. The method is based on the analysis of the specifics of concrete historical data, their comparison with test data for which there are estimates of the effectiveness of the algorithm, and the procedure for finding the optimal process parameters according to the Gauss–Seidel scheme that consists in sequentially searching for the function optimum alternately for each variable. The application of the proposed method makes it possible to use the considered entity resolution algorithm in real concrete historical research in the tasks of automated record linkage in nominative sources.

Keywords: concrete historical investigation; distributed technology; entity resolution; algorithm parameters; relational similarity measure

DOI: 10.14357/08696527240209

EDN: IJWC DL

References

1. Adamovich, I. M., and O. I. Volkov. 2024. Kollektivnoe razreshenie sushchnostey v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [Collective entity resolution in technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(1):128–138. doi: 10.14357/08696527240111. EDN: GQPGXQ.
2. Bhattacharya, I., and L. Getoor. 2007. Collective entity resolution in relational data. *ACM T. Knowl. Discov. D.* 1(1):5. 36 p. doi: 10.1145/1217299.1217304.
3. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelenogo avtomatizirovannogo analiza istoricheskikh tekstov [The distributed automated technology of historical texts analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):148–161. doi: 10.14357/08696527160311. EDN: WWSZIJ.
4. Adamovich, I. M., and O. I. Volkov. 2019. Edinaya tekhnologiya podderzhki konkretno-istoricheskikh issledovaniy [Unified technology of concrete historical research support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):194–205. doi: 10.14357/08696527190116. EDN: MZLQZ.
5. Vovchenko, A. E., L. A. Kalinichenko, and D. Yu. Kovalev. 2014. Metody razresheniya sushchnostey i sliyaniya dannykh v ETL-protsesse i ikh realizatsiya v srede Hadoop [Methods of entity resolution and data fusion in the ETL-process and their implementation in the Hadoop environment]. *Informatika i ee primeneniya — Inform. Appl.* 8(4):94–109. doi: 10.14357/19922264140412. EDN: PJYYLB.
6. Strehl, A., J. Ghosh, and R. Mooney. 2000. Impact of similarity measures on web-page clustering. *Workshop of Artificial Intelligence for Web Search*. AAAI Press. 58–64.
7. Collins-Elliott, S. A. 2016. Agglomerative clustering using cosine and Jaccard distances: A computational approach to Roman vessel taxonomy. *Archeologia Calcolatori* 27:7–25.
8. Antonov, D. N. 2000. Vosstanovlenie istorii semey: metod, istochniki, analiz [Restoring family history: Method, sources, and analysis]. Moscow. PhD Diss. 290 p. EDN: QDBKMR.
9. Thorvaldsen, G. 2016. Nominativnye istochniki v kontekste vseмирnoy istorii perepisey: Rossiya i Zapad [Nominative data and global census history: Russia and the West]. *Izvestiya Ural'skogo federal'nogo universiteta. Ser. 2. Gumanitarnye nauki*

[Izvestia. Ural Federal University J. Ser. 2. Humanities and Arts] 18(3):9–28. doi: 10.15826/izv2.2016.18.3.041. EDN: WYDBXL.

10. Khnaev, O. A., and I. A. Pchelintsev. 2012. Parametricheskaya optimizatsiya sistem. Metody resheniya ekstremal'nykh zadach [Parametric optimization of systems. Methods for solving extremal problems]. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* [Models, Systems, Networks in Economics, Engineering, Nature and Society] 2(3):146–152. EDN: RPUFKH.

Received March 15, 2024

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; volkov@amsd.com

ОБ АВТОРАХ

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Волков Олег Игоревич (р. 1964) — ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дьяченко Денис Юрьевич (р. 1987) — инженер-исследователь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Егоров Владимир Борисович (р. 1948) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Конашенкова Татьяна Дмитриевна (р. 1964) — кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Морозов Николай Викторович (р. 1956) — старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Москалева Фаина Александровна (р. 1996) — аспирант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов

Никишин Дмитрий Александрович (р. 1976) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Румовская София Борисовна (р. 1985) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Румянцева Александра Сергеевна (р. 2001) — студентка кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов

Самуйлов Андрей Константинович (р. 1988) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов

Синицын Владимир Игоревич (р. 1968) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор кафедры «Мехатроника и теоретическая механика» Московского авиационного института (национального исследовательского университета)

Степченко Дмитрий Юрьевич (р. 1973) — старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Степченко Юрий Афанасьевич (р. 1951) — кандидат технических наук, заведующий отделом, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются 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 представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и 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 оформления представлены в интернете:

http://www.ipiran.ru/journal/template_iiep_ssi_2024.zip

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 города и страны и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2019_29_03_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 (ссылка 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”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

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 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 resursosbergayushchie tekhnologii nedropol’zovaniya i povyseniya 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.

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, NY: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatzionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichstva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publ. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
- указывать в поле subject (тема) название журнала и фамилию автора;
 - указывать в тексте письма название статьи, авторов и журнал, в который направляется статья;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

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: ssi@frccsc.ru (Стригина Светлана Николаевна)

<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

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 “Systems and Means of Informatics” 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).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.

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 authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. 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_jiep_ssi_2024.zip

8. 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/collected/2019_29_03_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2019_29_03_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 in 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).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalina, 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 resursosberegayushchie tekhnologii nedropol'zovaniya i povyseniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovaryia 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 Publ. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
 - the journal title and author's name in the "Subject" field;
 - the article title, authors' names, and the journal title, whereto the paper is being submitted, in the text of the e-mail;
 - 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. "System and Means of Informatics" 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: ssi@frcsc.ru (to Svetlana Strigina)

http://www.ipiran.ru/english/journal_systems.asp